# MOLECULAR RECOGNITION OF SUBSTRATES BY PROTEIN FARNESYLTRANSFERASE AND GERANYLGERANYLTRANSFERASE-I

by

Corissa L. Lamphear

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

Doctoral Committee:

Professor Carol A. Fierke, Chair
Professor Robert S. Fuller
Professor Ayyalusamy Ramamoorthy
Associate Professor Anne B. Vojtek
Assistant Professor Patrick O'Brien

## DEDICATION

To the Lord, my strength.

"Humble yourselves, therefore, under God's mighty hand, that he may lift you up in due time. Cast all your anxiety on him because he cares for you."
1 Peter 5 : 6-7

## ACKNOWLEDGEMENTS

There are many people I'd like to thank that helped me through this journey of graduate school. First of all, I'd like to thank my advisor and mentor, Dr. Carol Fierke. I have appreciated all of the kind support and advice, both relating to science and life in general. I have learned so much under her guidance. I'd also like to thank my committee, Drs. Fuller, Raamamoorthy, Vojtek, and O'Brien, as well as my former committee member Dr. Palfey. I have really benefited from the helpful advice and comments about my work the past years. I truly appreciate your help.

I would also like to thank my collaborators. Dr. Shueler-Furman and Nir London from the Hebrew University of Jerusalem were a pleasure to work with on the computational studies with FTase. We had many interesting and stimulating conversations. I'd also like to thank Dr. Animesh Aditya and Dr. Richard Gibbs from Purdue University for collaborating on some of the peptide library studies with GGTase-I and providing the dns-GCxxx peptides. Additionally, I'm grateful for the help with the mass spectrometry on the farnesylated peptides from Dr. Philip Andrews and Dr. Eric Simon here at the University of Michigan.

Another professor that encouraged me early on in my career was my undergraduate advisor at Hope College, Dr. Leah Chase. Without her, I would not have become excited about science. She inspired me to be confident in my skills and encouraged me even when I did not believe in myself.

My labmates, both past and present, have been wonderful colleagues. I would like to thank Dr. James Hougland for mentoring me as a rotation student and beyond with the FTase project. I had many helpful discussions regarding my work with Drs. Terry Watt, Nathan Zahler, and John Hsieh so I thank them for taking the time to chat with me. Dr. Xiaomu Guan was a great labmate to share "bay awesome" with and discuss lipidation and various techniques. I especially thank Elaina Zverina for our brainstorms and collaboration working on the prenyltransferases. I thank Noah Wolfson for celebrating every "Friday" with me and for helping me fix broken instruments. I'm truly indebted to Andrea Stoddard for all of her advice for troubleshooting molecular biology problems, and I've enjoyed our friendship over copious amounts of No Thai and sushi. I thank all of the rotation students and undergraduates that have worked with me: Teng Xue, Jenna Hendershot, Carol Ann Pitcairn, Elia Wright, and Megan Novak. I truly enjoyed working with each of them, and I'm glad to continue the tradition of "bay awesome" with Elia and collaborate with her on future aspects of the *in vivo* project. I thank the rest of the Fierke lab for all of the helpful discussions, advice, and support.

A big thank you goes to all of my friends in the biological chemistry and chemistry departments. I don't know what I would have done without their support and friendship. I also thank all of the people I worked with in the Chemistry Professional Development Organization (CPDO), and I especially thank Dr. Shannon Watt for her mentoring and friendship, and her confidence in me. I want to thank my friends and my church family for welcoming me into their community, and for their friendship, patience, and support.

Lastly, I want to thank my family. I thank my in-laws Craig and Jayne Gilmer, for welcoming me into their family and for their support these past years. I'm grateful to my sister Brenna Lamphear, for just being there for me, no matter what. I know I can always call my best friend. I thank my parents Bruce and Linda Lamphear for their unconditional love and encouragement from the beginning. They have been a wonderful example to me, and they are great at reminding me to keep things in perspective. Finally, I thank my husband, Matthew Gilmer. Matt, I would not have made it through graduate school without your love, encouragement, and steady presence. Thank you for convincing me to get our dog, Macy; although she is a giant pain, her antics have been a constant source of laugher that you knew I needed. You have made these past five and a half years a bit easier for me, and for that I'm truly grateful. I love you and look forward to sharing many more adventures with you in life.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## ABBREVIATIONS

| | |
|---|---|
| $^{3}$[H]-FPP | Tritium labeled farnesyl pyrophosphate |
| $^{3}$[H]-GGPP | Tritium labeled geranylgeranyl pyrophosphate |
| AGOH | Anilinogeraniol |
| AGPP | 8-anilinogeranyl diphosphate |
| APT | Protein acylthioesterases |
| BGPP | Biotin-geranyl diphosphate analog |
| $Ca_1a_2X$ | C-terminal motif for prenylation |
| Dns | Dansyl |
| DTNB | 5,5′-Dithiobis(2-nitrobenzoic acid) |
| EDTA | Ethylenediaminetetraacetic acid |
| EGFP | Enhanced green fluorescent protein |
| FOH | Farnesol |
| FPP | Farnesyldiphosphate |
| FRET | Fluorescence resonance energy transfer |
| FTase | Farnesyltransfersase |
| FTI | FTase inhibitor |
| GGOH | Geranylgeraniol |
| GGPP | Geranylgeranyldiphosphate |
| GGTase-I | Geranylgeranyltransferase-I |
| GGTase-II | Geranylgeranyltransferase-I or RabGGTase |
| GGTI | GGTase inhibitor |
| GPI | glycosylphosphatidylinositol |
| HEPES | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| HEPPSO | N-(2-Hydroxyethyl)piperazine-N'-(2-hydroxypropanesulfonic acid) sodium salt |
| HGPS | Hutchinson-Gilford progeria syndrome |
| HMG-CoA reductase | 3-hydroxy-3-methylglutaryl-CoA-reductase |

| | |
|---|---|
| HTZ buffer | 50 mM HEPES pH 7.8, 10 μM ZnCl2 and 2 mM TCEP |
| ICMT | Isoprenylcysteine methyl transferase |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| MTO | Multiple turnover |
| $N_3$-FOH | Azido-farnesol |
| $N_3$-GGOH | Azido-geranylgeraniol |
| NMT | myristoyl-CoA:protein N-myristoyltransferase |
| NON | Non-substrate |
| PATs | Protein acyl transferases |
| PCR | Polymerase chain reaction |
| PMSF | Phenylmethanesulfonyl fluoride |
| PTM | Post-translational modification |
| RCE1 | Ras-converting enzyme 1 |
| REP | Rab escort protein |
| Shh | Sonic hedegehog protein |
| STO | Single turnover |
| TAME | $N_\alpha$-*p*-Tosyl-L-arginine methyl ester hydrochloride |
| TAMRA-alkyne | tetramethylrhodamine-alkyne |
| TCEP | Tris(2-carboxyethyl)phosphine hydrochloride |
| ZMPSTE24 | Zinc metalloprotease Ste24 |

**ABSTRACT**

**MOLECULAR RECOGNITION OF SUBSTRATES BY PROTEIN FARNESYLTRANSFERASE AND GERANYLGERANYLTRANSFERASE-I**

**by**

**Corissa L. Lamphear**

**Chair: Carol A. Fierke**

Prenylation is an important post-translational modification that targets proteins to the cellular membrane. Farnesyltransferase (FTase) catalyzes the attachment of the 15-carbon farnesyl moiety from farnesyldiphosphate to a cysteine near the C-terminus of a protein, while geranylgeranyltransferase-I (GGTase-I) catalyzes the analogous attachment of the 20-carbon geranylgeranyl group from geranylgeranyldiphosphate. Substrates of the prenyltransferases are involved in a myriad of signaling pathways and processes within the cell, therefore inhibitors targeting FTase and GGTase-I are being developed as therapeutics for treatment of diseases such as cancer, parasitic infection, and progeria. FTase and GGTase-I were proposed to recognize a $Ca_1a_2X$ motif, where C is the cysteine where the prenyl group is attached, $a_1$ and $a_2$ are aliphatic amino acids, and X confers specificity between FTase and GGTase-I with X being methionine, serine, glutamine, and alanine for FTase and leucine or phenylalanine for GGTase-I. Recent work indicates that the $Ca_1a_2X$ paradigm should be expanded; therefore, further studies are needed to define the prenylated proteome, to understand normal cellular processes,

and to determine the targets of prenyltransferase inhibitors. In this study, we probed the molecular recognition of GGTase-I by testing a 400 peptide library for activity with GGTase-I. The enzyme modifies two classes of substrates: multiple turnover substrates (MTO) and single turnover-only (STO) which undergo chemistry but not product release. Statistical analysis was used to determine that MTO substrates typically follow the $Ca_1a_2X$ definition, but the STO sequences are more diverse, further indicating GGTase-I recognizes a broader range of substrates. Additionally, with collaborators at the Hebrew University of Jerusalem, a computational program that predicts FTase substrates was developed, FlexPepBind. This novel method successfully predicted new peptide substrates with FTase and identified a new class of substrates containing a positively charged X residue. Lastly, to examine prenylation *in vivo*, we created a library of GFP-$Ca_1a_2X$ fusion proteins and measured protein localization using fluorescence microscopy. The identity of the C-terminal sequence caused the proteins to localize to different cellular compartments presumably due to modification status. Together, these studies provide insight into the *in vivo* specificity of prenyltransferases and the involvement of prenylation in various cellular processes.

# CHAPTER I

# INTRODUCTION[1]

## Prenylation Background

*General information: lipidation*

In the eukaryotic cell, an assortment of proteins resides in or at the membrane in order to perform necessary biological functions. Typical membrane proteins include those with transmembrane domains that span the lipid bilayer one or multiple times. However, another mode of targeting proteins to the cellular membranes is by the attachment of a lipid group post-translationally or co-translationally in a process called lipidation which increases the affinity of the protein to the membrane. Lipidation can be broadly classified into two groups: modifications that target the protein to the outer membrane or those that target a protein to inner membrane leaflet.

Two examples of lipid modifications that attach proteins to the outer membrane are the GPI anchor and the cholesterol modification. The glycosylphosphatidylinositol (GPI) moiety is a complex structure containing a phospholipid, sugars and ethanolamine and is attached to the C-terminal amino acid of a protein (*1, 2*) which targets the protein to the outer leaflet (*3-5*). This modification is thought to play a part in a variety of signal

[1] A portion of Chapter 1 is taken 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.

transduction pathways, partitioning to lipid rafts, cell communication, and prion disease pathogenesis (*1, 2, 6*). Cholesterylation also occurs at the C-terminus, forming acholesteryl ester. One example of a cholesterylated protein is the Sonic hedgehog (Shh) protein which is involved in organ development (*7*). The Shh protein has both cholesterol and palmitoyl modifications (*8, 9*) that are required for proper secretion and signaling (*10*).

Modifications that target proteins to the inner membrane leaflet include myristoylation, palmitoylation, and prenylation (Figure 1.1). Myristoylation is catalyzed by myristoyl-CoA:protein N-myristoyltransferase (NMT) (*11*) and typically involves the attachment of 14-carbon saturated fatty acid to an N-terminal glycine on a protein via an amide bond (*12, 13*). Although this modification was originally thought to solely occur co-translationally, recent studies have indicated that this modification can also occur post-translationally in apoptotic cells (*14, 15*). Myristoylated proteins are involved in a variety of signaling pathways; modified families include Gα proteins, the non-receptor protein kinases, and calcium binding proteins (*16*). Palmitoylation is the second type of acylation, which in most cases is the attachment of the 16-carbon palmitate at a cysteine forming a thioester bond (*17, 18*). Palmitoylation is unique from the other types of lipidation since it is readily reversible (*19, 20*). The attachment of palmitate is catalyzed by protein acyl transferases (PATs) (*21-23*) and removal is catalyzed by protein acylthioesterases (APTs) (*24*) but both processes can occur non-enzymatically as well. Palmitoylation does not occur at a specific consensus sequence, but the modification is found in four patterns including alone, proximal to protein transmembrane domains, at the C-terminus with a prenylation modification, and finally, near a myristoyl

modification (*22*). Examples of palmitoylated proteins include N- and H-Ras (*25*) (further discussed in a later section) and the Huntington protein, which is involved in Huntington's disease (*26*). Lastly, prenylation comes in two forms: the 15-carbon prenylation and the 20-carbon geranylgeranylation. Protein farnesyltransferase (FTase) catalyzes the transfer of the 15-carbon farnesyl moiety from farnesyldiphosphate (FPP) to a cysteine residue near the C-terminus of the target protein; protein geranylgeranyltransferase-I (GGTase-I) and geranylgeranyltransferase-II (GGTase-II or RabGGTase) catalyze the analogous attachment of a 20-carbon geranylgeranyl group from geranylgeranyldiphosphate (GGPP) to cysteine(s) near the C-terminus of the substrate protein (Figure 1.2) (*27, 28*). These hydrophobic modifications help to localize proteins to cellular membranes to carry out their function as well as to facilitate protein-protein interactions (*29-31*). GGTase-II (also called RabGGTase) modifies the Rab family of proteins at cysteines near the C-terminus in diverse motifs such as CC or CXC (*32-34*), and requires the Rab escort protein (REP) as an accessory protein to bind to substrates for activity (*35*). FTase and GGTase-I are termed the "CaaX" prenyltransferases, have similar modes of recognition, and are the focus of this work.

*FTase and GGTase-I properties*

FTase and GGTase-I are proposed to recognize the "$Ca_1a_2X$" motif on substrate proteins (*28, 31, 36*). This $Ca_1a_2X$ motif is canonically defined as: "C" is a cysteine four amino acids from the C-terminus where the prenyl group is attached forming a thioether bond; $a_1$ and $a_2$ are small aliphatic amino acids; and X is proposed to confer specificity of substrates for modification by FTase or GGTase-I, with X being methionine, serine,

glutamine, and alanine for FTase and leucine or phenylalanine for GGTase-I (*37-40*). Although many substrates are described by this paradigm, recent studies have indicated that the $Ca_1a_2X$ model should be revised, as many substrates fall outside the traditional $Ca_1a_2X$ definition (*41*). Additionally, there is evidence for a large pool of dual substrates for FTase and GGTase-I, with these proteins potentially modified by both enzymes (*42*). Currently, although many prenyltransferase substrates have been identified, the full extent of prenylation within the cell is still unclear (*41, 43-46*). Understanding how FTase and GGTase-I recognize substrates would aid in defining the prenylated proteome and in understanding the biological signaling pathways involving prenylated proteins.

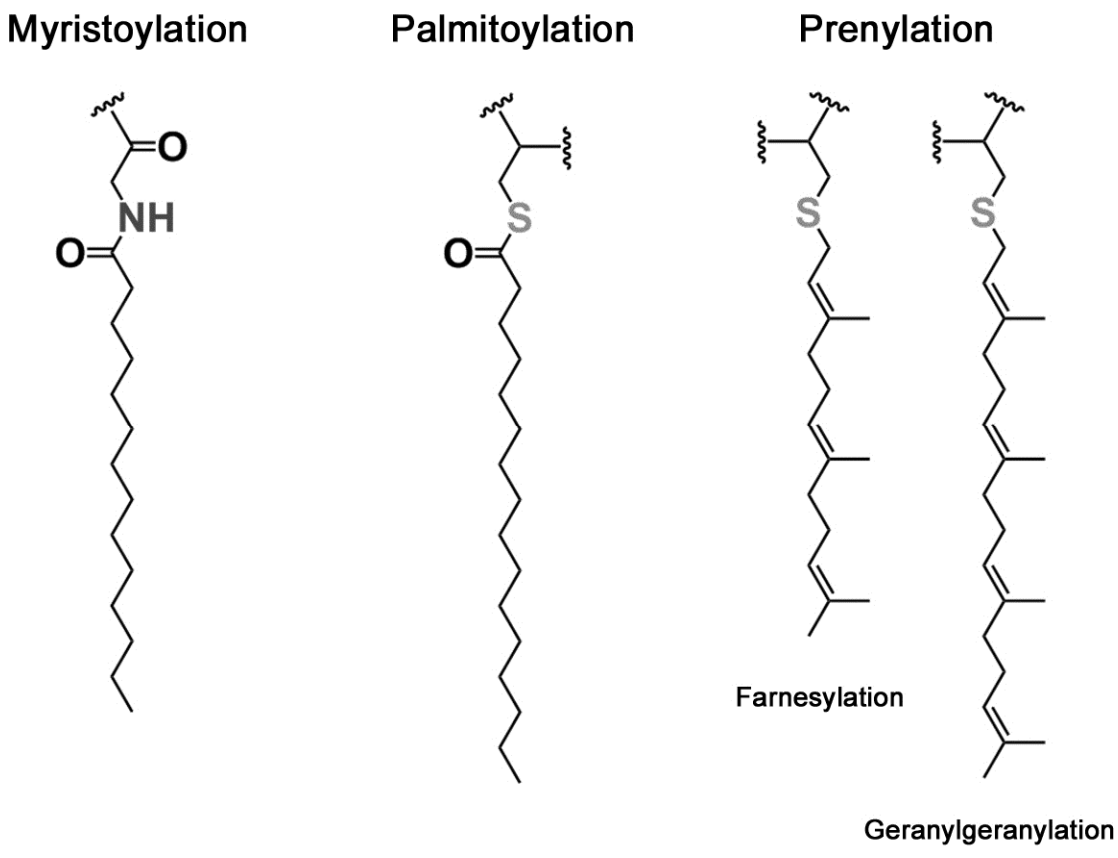


**Figure 1.1. Types of lipidation.** The myristoyl, palmitoyl, and prenyl groups can target proteins to the inner leaflet of the membrane.

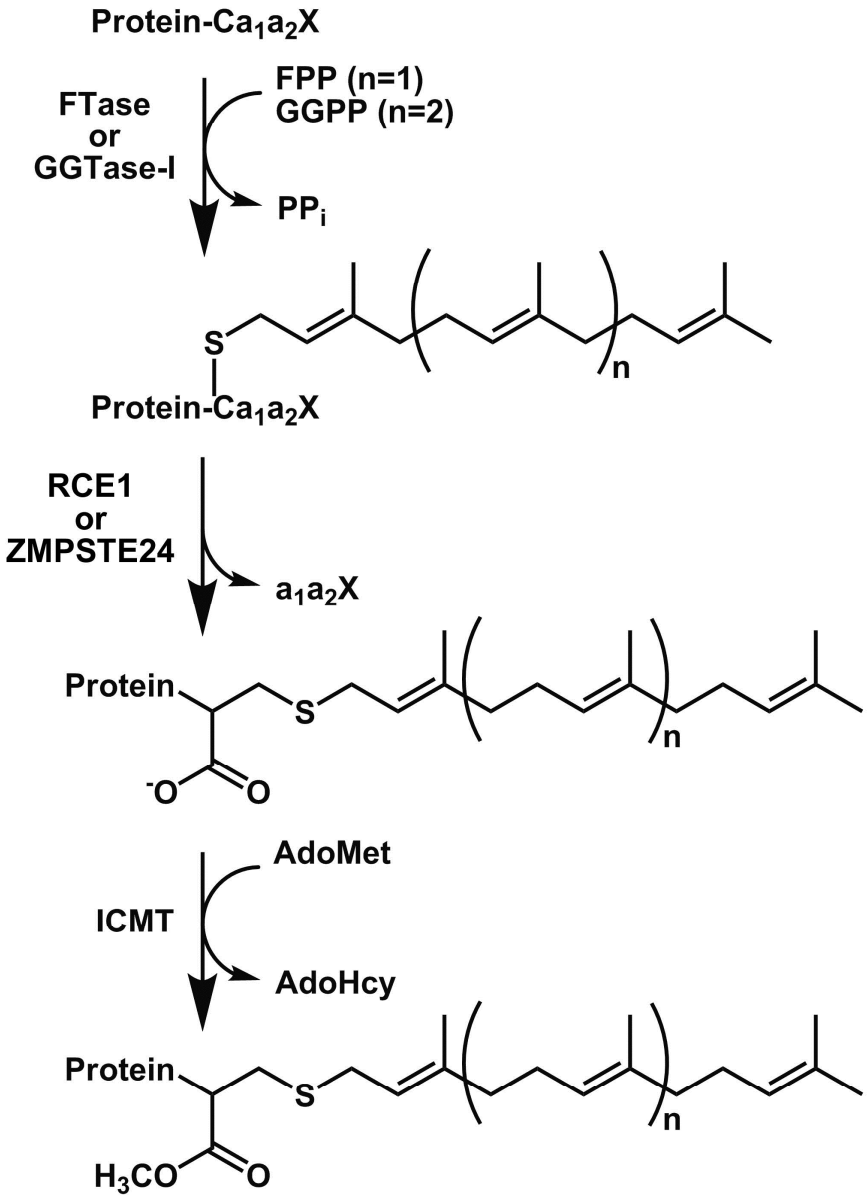


**Figure 1.2. Enzymes in the prenylation pathway.** FTase and GGTase-I catalyze the attachment of a farnesyl or geranylgeranyl group to the sulfur atom of a cysteine in the proposed C-terminal $Ca_1a_2X$ sequence of a substrate protein using farnesyldiphosphate (FPP) or geranylgeranyldiphosphate (GGPP) as the lipid donor (*27, 28*). Next the last three amino acids of the protein can be proteolyzed by RCE1 or ZMPSTE24, and the carboxy terminus can be methylated by ICMT.

FTase and GGTase-I are heterodimeric zinc-metalloenzymes, containing identical α subunits and distinct but homologous β subunits (*47*), with active sites in both enzymes composed predominantly of β subunit residues near the α/β interface (Figure 1.8 A and B (*48*)). The kinetic mechanism of FTase (and by similarity GGTase-I) is thought to be functionally ordered, with FPP binding first before the protein or peptide substrate followed by a conformational change in the prenyl diphosphate substrate that positions the $C_1$ of FPP or GGPP closer to the sulfur of the peptide (*49-52*). An active site $Zn^{2+}$ coordinates the sulfur of the cysteine, lowering the $pK_a$ and creating a reactive thiolate which performs a nucleophilic attack on the alpha carbon of FPP to form a thioether linkage between cysteine and the lipid (*53, 54*). Finally, diphosphate is rapidly released. In FTase, a bound $Mg^{2+}$ coordinates and stabilizes the developing negative charges on the pyrophosphate leaving group (*55, 56*), while Lys311β accomplishes this task in GGTase-I (*57*). There has been much debate in the literature about the nature of the transition state in the chemical step as various studies have suggested both $S_N1$-like (associative) and $S_N2$-like (dissociative) behavior (*58-60*). The latest work using $^3H$ α-secondary kinetic isotope effect measurements suggested an associative mechanism with dissociative character for the FTase reaction (*61*). New work using quantum mechanical molecular mechanical studies (QM/MM) with FTase has suggested that the transition state structure is actually peptide dependent. For instance, one peptide may exhibit an $S_N1$-like transition state, while the transition state for a different peptide shows $S_N2$-like behavior (Yue Yang, Bing Wang, Melek N. Ucisik, Guanglei Cui, Carol A. Fierke, and Kenneth M. Merz, Jr, unpublished data).

*History: the discovery of prenylation*

C-terminal prenylation was first discovered on the Rhodotorucine A mating factor in yeast in 1979 (*62*) and discovered in mammalian cells in 1984 during studies of the cholesterol biosynthesis pathway. Researchers were studying the effect of inhibitors of the enzyme 3-hydroxy-3-methylglutaryl-CoA-reductase (HMG-CoA reductase), which blocks the synthesis of mevalonate, and thus, isoprenoids (Figure 1.3). When cells were treated with radiolabeled mevalonate after treatment with lovastatin, the radiolabel was found to be incorporated onto cellular proteins (later found to be due to prenylation) and soon it was demonstrated that all mammalian cells displayed this behavior (*63, 64*). Then, in *Saccharomyces cerevisiae,* a gene that affected the labeling on the RAS proteins and the RAM a-mating factor was identified, and the only similarity between the two substrates was in the C-terminal CaaX sequence (*65*). Therefore, a connection was forged between the CaaX sequence and the modification, which is prenylation. These studies opened doors as more and more classes of modified proteins were found, and it was ascertained that the molecules responsible for the labeling were the isoprenoids.

*In vivo prenylation pathway: further modifications*

Additionally, *in vivo*, substrates can undergo further modification after prenylation (Figure 1.2). The last three amino acids of prenylated substrates can be proteolyzed by zinc metalloprotease Ste24 (ZMPSTE24) or Ras-converting enzyme 1 (RCE1) at the endoplasmic reticulum, followed by methylation of the carboxy terminus, catalyzed by isoprenylcysteine methyl transferase (ICMT) (*66-68*). These additional modifications can aid in membrane localization, but it is not yet known whether these

modifications occur on every prenylated substrate. Specificity studies of RCE1 and ZMPSTE24 are also limited, but it is known these enzymes require a prenylcysteine for recognition and that some substrates may overlap between the two proteases (*68-75*). Besides these modifications, some proteins (like the Ras proteins) are palmitoylated or contain an upstream polybasic region of the CaaX sequence that aid in membrane association (Figure 1.4) (*76-80*). Therefore, it will be useful to understand which modifications or combination of modifications occurs on various prenylation pathway substrates.

*Prenylation and disease implications*

Many proteins are modified by the prenyltransferases, including the small Ras and Rho GTPase superfamilies (*31, 81*) and the nuclear lamins (*82*), and often the prenyl modification is essential for the biological function (*31, 83*). Prenyltransferases are being investigated as targets for inhibitors to treat a variety of diseases, including cancer (*84*), Hutchinson-Gilford progeria syndrome (*85*), and parasitic diseases such as malaria (*86*). Prenyltransferase inhibitors were initially developed to target Ras protein signaling pathways implicated in cancer, but it was later determined that inhibitor efficacy is the result of modulating prenylation of non-Ras proteins (*87*). Studies of the pleitropic effects of statin treatment, which blocks the biosynthesis of FPP and GGPP and therefore affects prenylation, suggest the involvement of prenylated proteins in diseases such as leukemia (*88*), asthma (*89*), and cardiovascular disease (*90*). Therefore, understanding how prenyltransferases recognize their substrates, which substrates are prenylated *in vivo,* and, finally, what substrates are responsible for inhibitor efficacy are important outstanding

questions. More details about the modifications and biology of Ras family and the lamins are discussed in the next section.

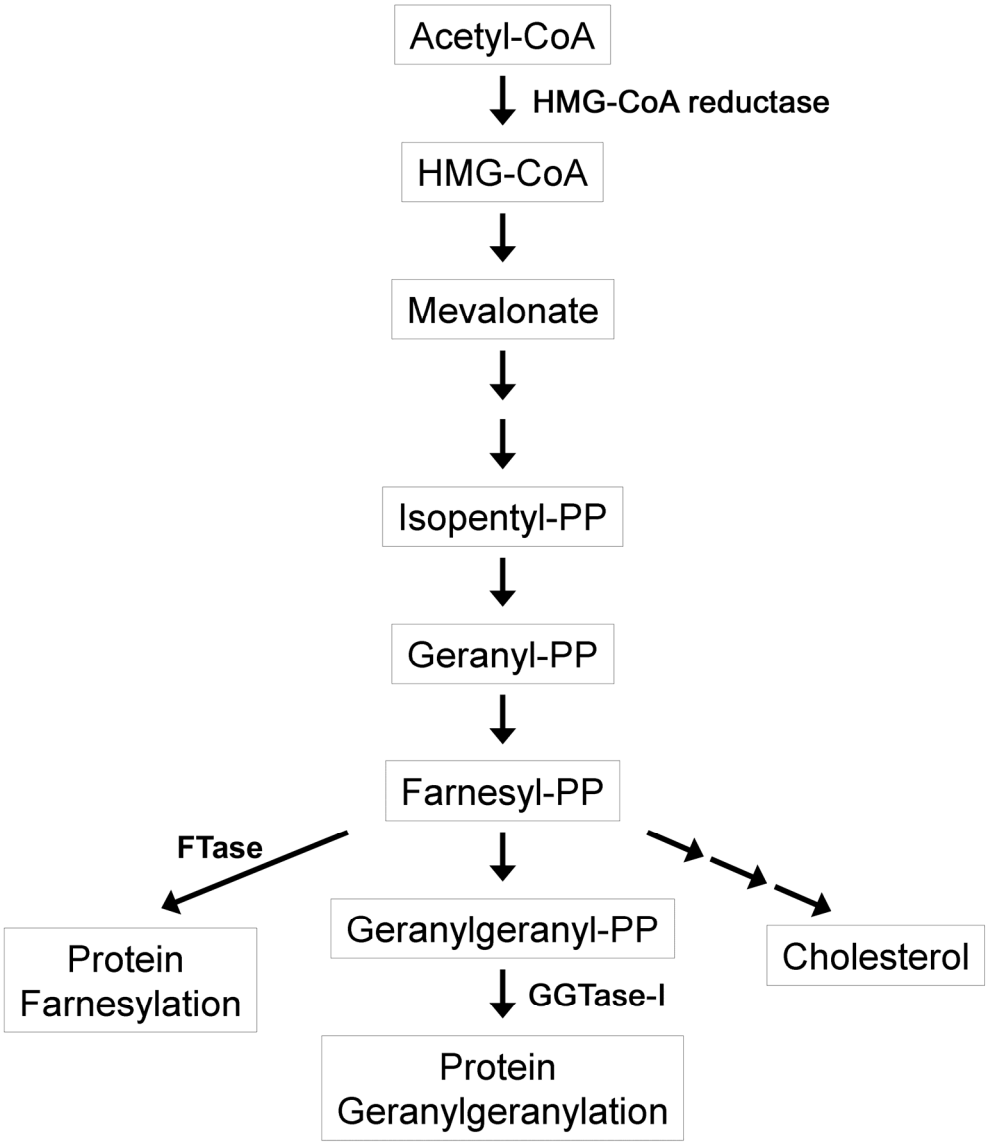


**Figure 1.3: The cholesterol biosynthesis pathway.** Cholesterol is synthesized starting from acetyl-CoA. Farnesyldiphosphate and geranylgeranyldiphosphate substrates originate from this pathway.

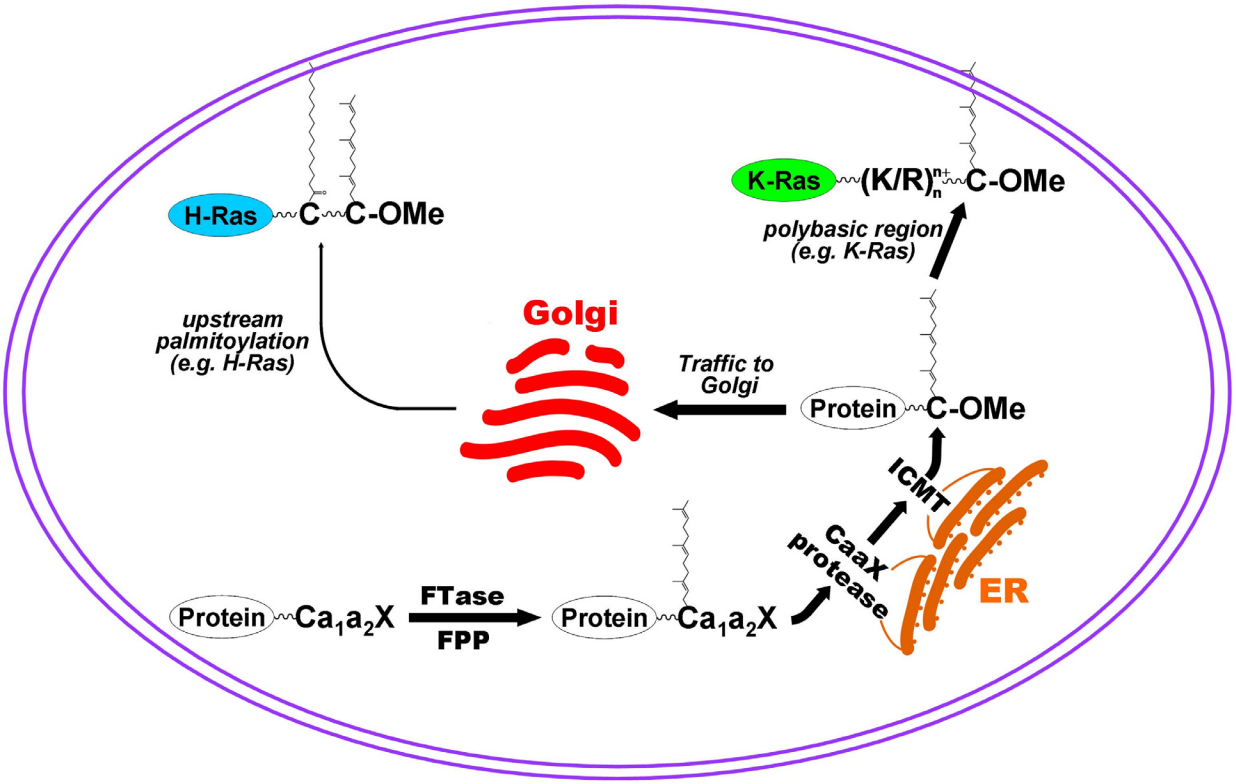


**Figure 1.4**. **The *in vivo* prenylation pathway.** Proteins can be prenylated in the cytosol, proteolyzed by the CaaX proteases zinc metalloprotease Ste24 (ZMPSTE24) or Ras-converting enzyme 1 (RCE1) at the endoplasmic reticulum, and carboxy methylated by isoprenylcysteine methyl transferase (ICMT). Besides these modifications, some proteins (like the Ras proteins) are palmitoylated or contain an upstream polybasic region of the CaaX sequence that aid in membrane association.

*Lamin A and Hutchinson-Gilford progeria syndrome*

The nuclear lamina provides scaffolding for the nucleus and is involved in a multitude of processes ranging from gene transcription to protein trafficking (*91-95*). Some major players in the nuclear lamina are the lamins, which include lamin A, lamins B1 and B1, and lamin C, and all but lamin C are modified by FTase at some stage in the protein processing (*91*). Lamin A is unique in that it exists in two forms: prelamin A and lamin A. The prelamin A is farnesylated, proteolyzed by RCE1, and methylated by ICMT like many other FTase substrates. It is then presumably trafficked to the nucleus where it is cleaved between residues Y646 and L647 by ZMPSTE24 to clip off the C-terminal tail, which includes the farnesyl group, to generate the mature lamin A (Figure 1.5 A (*85, 96*)). Hutchinson-Gilford progeria syndrome (HGPS) is caused by a mutation in the prelamin A that removes the cleavage site for the final proteolysis step (*96*). This devastating premature aging disease is marked by alopecia, slow growth, short stature, osteoporosis, osteolysis, loss of subcutaneous fat, thin skin, and limited joint mobility, resulting in premature death around 7-20 years typically by heart disease or stroke (*97*). HGPS is thought to be caused by the build-up of the farnesylated mutant pre-lamin A which is called "progerin" (Figure 1.5 B). On the cellular level, the progerin likely disrupts the structure of the nucleus, and visually, causes misshapen nuclei with blebbing of the membranes. Treatment with FTIs (farnesyltransferase inhibitors) seems to ameliorate this effect (*98*), improves symptoms in mouse models of the disease (*99*), and may be a method of treating the disease in the future for patients. In fact, a clinical study with the FTI lonafarnib is currently underway with twenty-seven children (*100, 101*).

*Ras GTPases, prenylation, and palmitoylation: interplay of lipid modifications*

As mentioned above, the Ras family of proteins is a substrate of the prenylation pathway (Figure 1.6), is involved in cellular proliferation, and mutations in this family that cause GTPase activation are found in cancer (*102, 103*). The Ras superfamily is a well-studied system that exemplifies the interplay between prenylation and palmitoylation. H-Ras, N-Ras, and K-Ras isoforms 4A and 4B are all farnesylated, but under treatment with an FTI, N-Ras and K-Ras4B can become geranylgeranylated (*104-106*). Prenylation is sufficient for association with the ER and Golgi membranes; however an additional "secondary signal" is necessary for targeting to the plasma membrane. Palmitoylation is required for N-Ras and H-Ras to be located to the membrane, but K-Ras has a polybasic region which directly aids in membrane association (*107-109*). Prenyl modifications are stable, but palmitoylation is a method of regulating the localization back and forth to the plasma membrane as this modification is readily reversible (*110-112*). The PATs (protein acyl transferases) catalyze the attachment of the palmitoyl group, while the APTs (protein acylthioesterases) catalyze removal of this group (Figure 1.6, left (*21-24*)). The dually prenylated and palmitoylated Ras is localized to the plasma membrane where the activity is regulated by bound nucleotides. When GDP is bound to Ras, the protein is in the "off" state; however, a guanine nucleotide exchange factor (GEF) can aid in exchanging the GDP for GTP, leading to active Ras•GTP. Conversely, a GTPase activating protein (GAP) can enhance the GTPase activity of Ras, causing the protein to form the inactive, Ras•GDP state. The active Ras•GTP state binds effectors and activates a kinase cascade, eventually leading to gene expression in the nucleus and cell proliferation (as shown in the simplified diagram

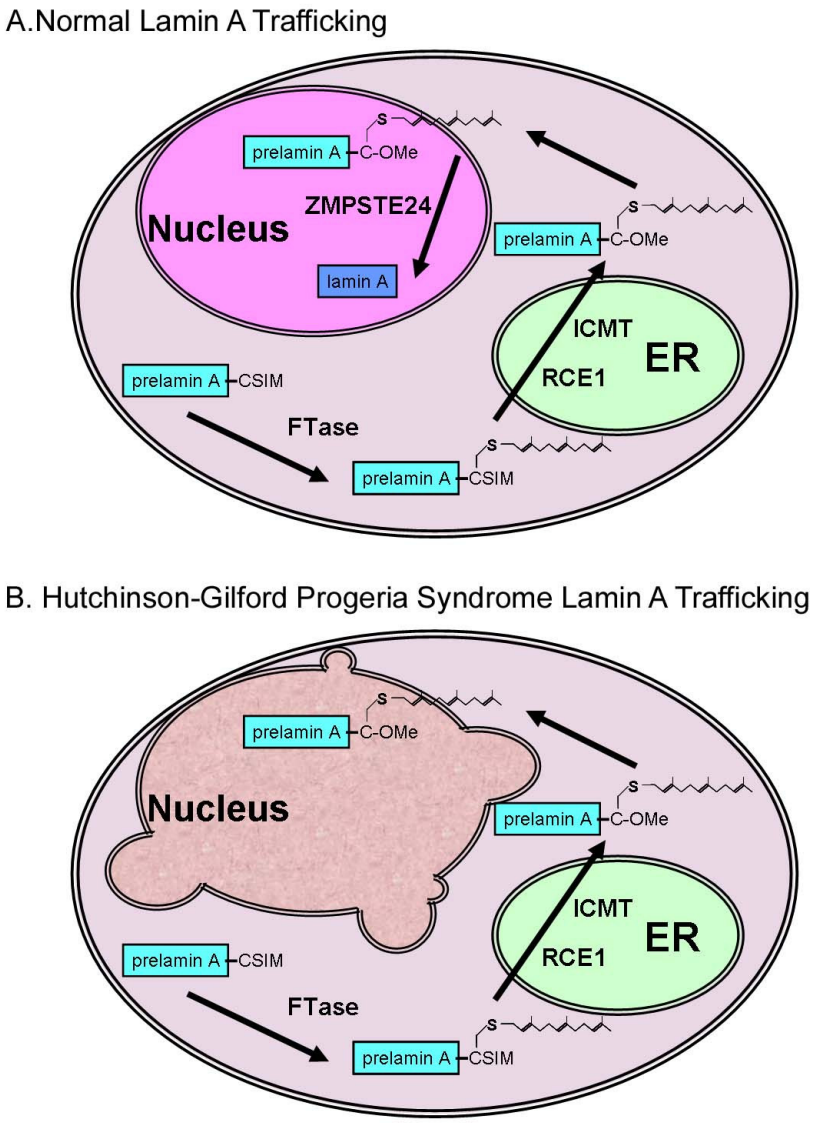


**Figure 1.5. Trafficking of Lamin A in normal and Hutchinson-Gilford Progeria Syndrome cells. A)** Normal Lamin A trafficking. Prelamin A is farnesylated in the cystosol by FTase, trafficked to the ER membrane where the three C-terminal amino acids are removed and methylated at the C-terminus. Then, the modified prelamin A is localized to the nucleus where it is proteolyzed between internal sites, removing the farnesyl group and becoming the mature lamin A protein. Lamin A is then thought to locate to the nucleoplasm. **B)** Trafficking of Lamin A in Hutchinson-Gilford Progeria Syndrome cells. All of the trafficking is the same, until the final proteolysis step. There is a mutation in the prelamin-A which does not allow the ZMPSTE24 to cleave the protein; therefore, the prelamin A retains the lipid modification. The membrane localized prelamin A is proposed to cause nuclear blebs and contribute to development of the disease.

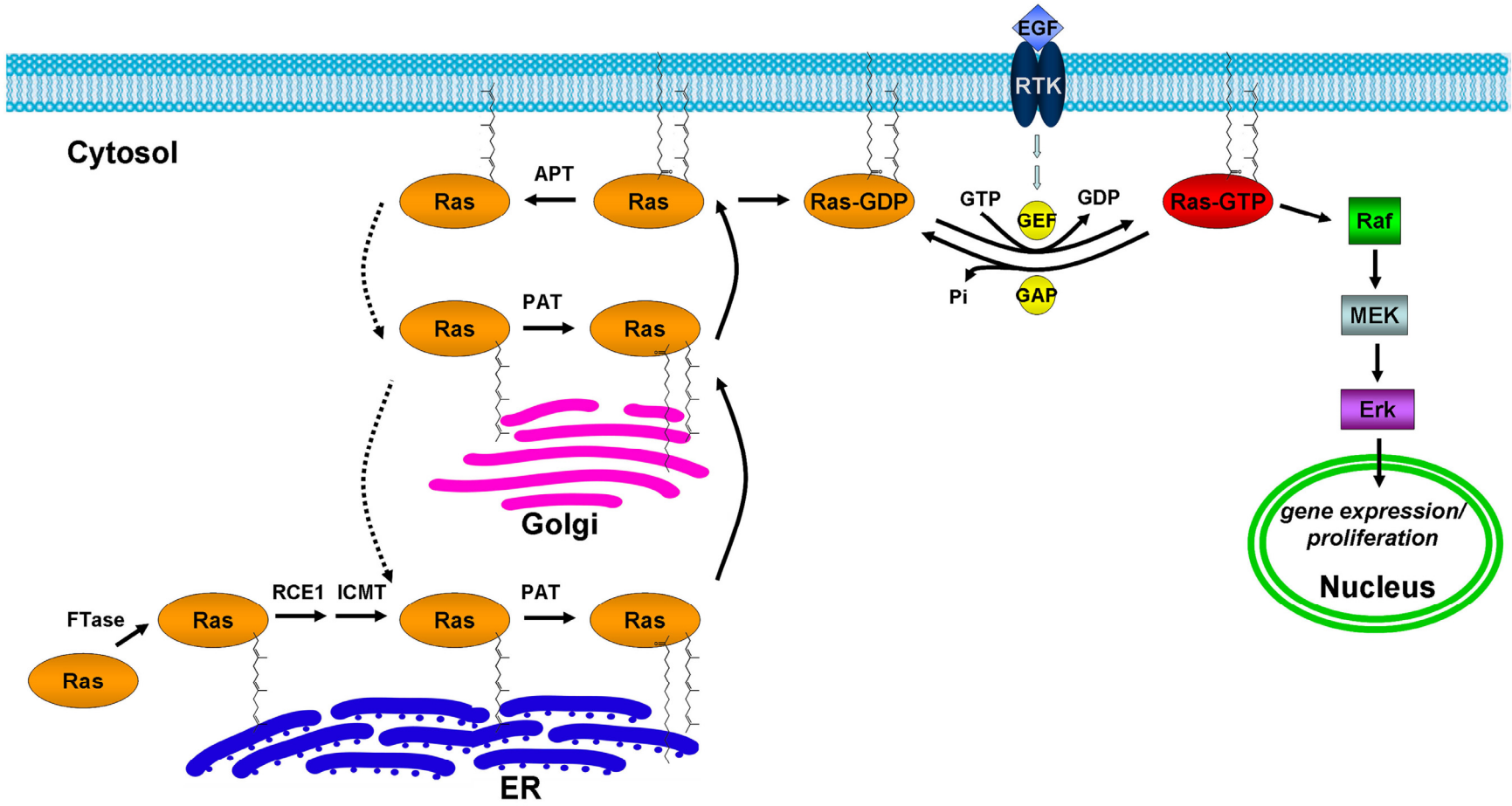


**Figure 1.6**. **Ras family prenylation and palmitoylation.** Members of the Ras family must be prenylated and palmitoylated to localize to the cellular membrane. The presence or absence of the palmitoyl group allows the protein to cycle back and forth between the ER, Golgi, and outer membrane. At the cellular membrane, the Ras proteins are regulated, with active Ras binding to effectors to stimulate a kinase cascade for gene expression.

in Figure 1.6 (*25, 48, 103*)). The Ras family shows the importance and effect of localization on the type and combination of lipidation.

*Parasitic prenyltransferases*

An emerging field is the study of prenyltransferases from pathogenic and parasitic organisms as potential therapeutic targets (*113-116*). For instance, the pathogenic yeast strain *Candida albicans*, known to affect immunocompromised patients, utilizes both protein farnesyltransferase and geranylgeranyltransferase-I enzymes to catalyze prenylation of substrate proteins (*117, 118*). The specificity of these two enzymes is currently being studied to define the complement of prenylation of proteins in this organism, since it is possible that inhibitors of the prenyltransferases could be used as antifungal therapeutics (*117, 119*). Additionally, inhibitors are being developed as therapeutics for the treatment of parasitic diseases such as malaria, caused by *Plasmodium falciparum* (*120*). These FTase inhibitors, based upon ethylenediamine (*115*) and tetrahydroquinoline (*114*) scaffolds, are thought to block *P. falciparum* FTase activity, which is essential for the organism. Based upon modeling and resistance mutation studies (*121*), these inhibitors may chelate the catalytic $Zn^{2+}$ and bind to the enzyme in the lipid substrate grove. Recently, it has also been discovered that the gram negative bacterium, *Legionella pneumophila*, hijacks the host prenyltransferase machinery to catalyze prenylation of bacterial proteins, raising the possibility that current mammalian prenyltransferase inhibitors could also be used as antibiotics (*122, 123*).

## Prenyltransferase Substrate Specificity and the Prenylated Proteome

Defining the extent of prenylation within the proteome of an organism can be approached using two complementary tactics: 1) direct *in vivo* determination of proteins that are prenylated; or 2) definition of the recognition elements used by FTase and GGTase-I to select peptide and protein substrates. This next section summarizes both approaches. The summary of the direct *in vivo* identification modes includes a brief discussion of methods for detecting prenylation on a small or large scale. Additionally, methods used to definite molecular recognition in prenyltransferases are described, including structure/function studies, peptide library studies, and computational predictive methods.

*Methods for discovering and predicting prenyltransferase substrates: identification of substrates one-by-one*

In the past, the methodology to determine prenyltransferase substrates has been difficult, and typically, was carried out by studying one protein at a time. The prenyl modification status of a particular protein in many cases can be assessed through incubation of cells or lysate with radioactive ($^{3}H$ or $^{14}C$) molecules, including: FPP or GGPP (*124-126*); a metabolic precursor of FPP or GGPP such as mevalonate (*124, 125, 127*); or an alcohol precursor of GGPP and FPP, such as geranylgeraniol (GGOH) or farnesol (FOH) (*126, 128*), that is phosphorylated *in vivo* (*129*). Treatment of cells with these compounds allows the protein of interest to be radiolabeled upon prenylation. A significant limitation of this method is the low signal from the prenylated proteins arising at least partly from the low specific activity of the radioactive molecules typically used

(*27, 126, 130*). Control experiments to eliminate false positive results from radiolabeling studies include mutation of the cysteine of the $Ca_1a_2X$ in the target protein to serine to block prenylation and/or incubation with prenyltransferase inhibitors (*130, 131*). These studies are difficult to perform on a large pool of protein targets since there is not a facile method to pull-down proteins containing prenyl groups. Antibodies have been raised to the prenyl modifications for detection using immunoblotting (*132, 133*); however, these antibodies can exhibit problematic cross-reactivity with other lipid modifications (*134*) and are not able to distinguish between farnesyl and geranylgeranyl modifications (*89, 133, 134*).

*FPP and GGPP analogs: aiding in prenyl group detection*

Using synthetic organic chemistry, multiple research groups have developed FPP and GGPP donor analogs with properties that enhance prenylated protein isolation and identification (*43, 135-137*). FTase and GGTase recognize various prenyl donor analogs as substrates for incorporating recognition tags onto prenylated proteins, both *in vitro* and *in vivo,* allowing for parallel identification of multiple substrates within the available pool of proteins as well as the monitoring of prenylation status changes upon inhibitor treatment. We will briefly focus on three recently reported classes of analogs: immunogenic analogs, analogs with functional groups allowing for chemoselective bioorthogonal labeling following prenylation, and analogs that contain an affinity tag such as biotin (Figure 1.7).

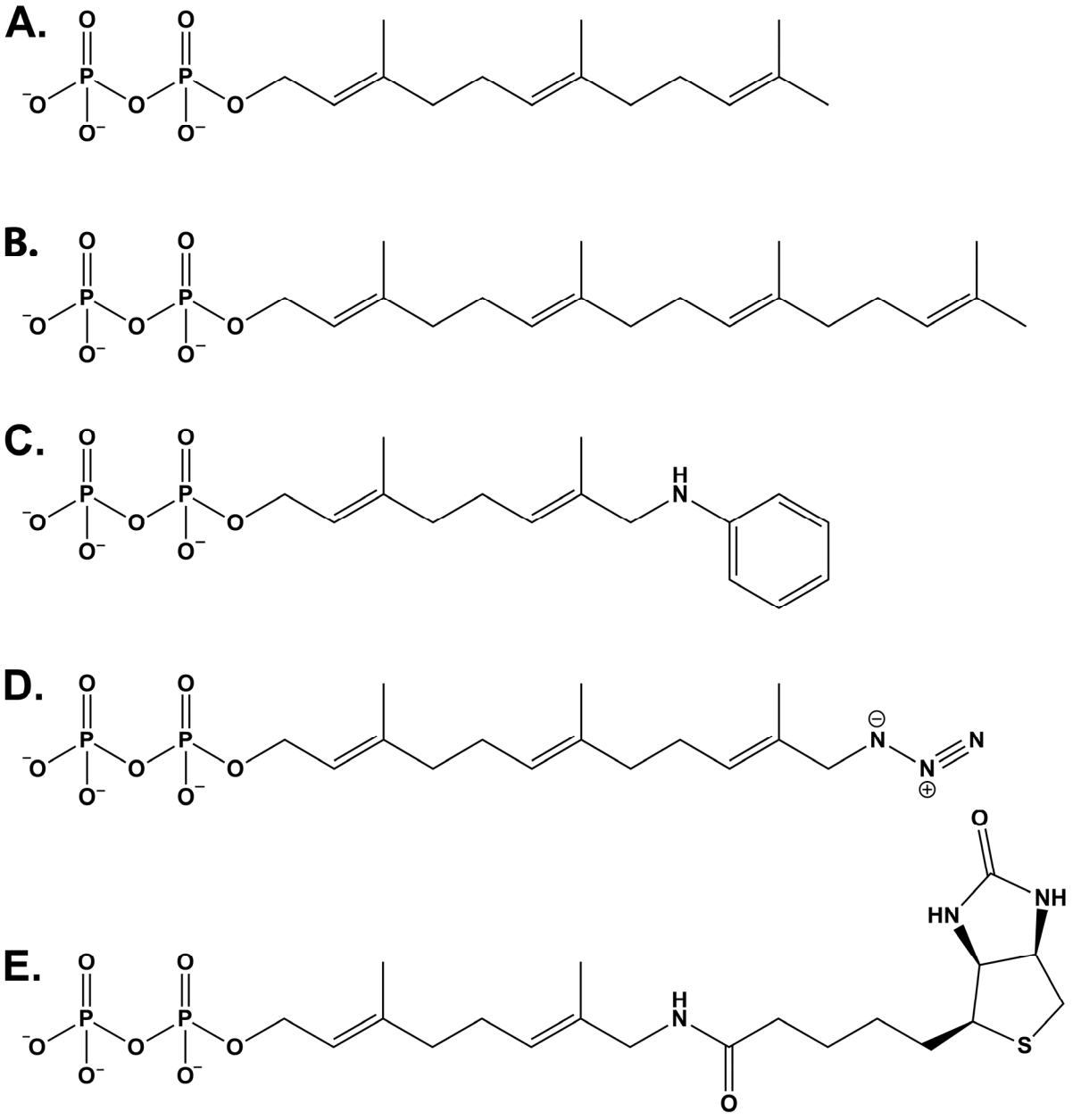


**Figure 1.7. Lipid structures.** Structures of: A) farnesyldiphosphate (FPP); B) geranylgeranyldiphosphate (GGPP); C) 8-anilinogeranyl diphosphate (AGPP); D) azido-FPP; E) biotin-geranyl pyrophosphate (BGPP) (*43, 136, 138*).

Immunogenic analogs incorporate a chemical moiety that can serve as the epitope for an analog-specific antibody. An example of such an analog is anilinogeraniol, developed by the Spielmann lab (*135*). Anilinogeraniol (AGOH), which is converted to the 8-anilinogeranyl diphosphate (AGPP) substrate *in vivo* (Figure 1.7 C) (*138*), replaces the terminal isoprene unit of FPP with an aniline moiety that serves as an epitope. In studies employing AGOH as an immunogenic tag, cells are treated with the analog, lysed, and proteins fractionated by 1- or 2-D gel electrophoresis. The farnesylated proteins are detected using Western blotting with an antibody raised to recognize the AG-KLH moiety (*135, 139*). This method detected several known FTase substrates using 1-D gel electrophoresis, but failed to identify many small GTPases known to be FTase substrates; however, coupling the Western blot analysis with 2-D gel electrophoresis increases the efficiency of prenylated protein identification (*139*).

An emerging class of FPP and GGPP analogs employs bioorthogonal ligation methods, wherein analogs containing an azido or alkyne group on the terminal isoprenoid of FPP or GGPP are attached to substrate proteins. These functional groups are amenable to Staudinger ligation or Huisgen 1,3-dipolar cycloaddition ("click" chemistry) (*140, 141*), which allows for chemoselective attachment of reporter molecules such as a fluorophore or biotin following protein prenylation. The first use of an azido analog for identification of prenylated proteins was reported by Kho and coworkers (*43*). They treated COS-1 cells with azido-FPP (Figure 1.7 D) or azido-farnesol, and demonstrated *in vivo* incorporation of this analog onto substrates catalyzed by FTase. Following cell lysis, the azido-farnesylated proteins were labeled with a biotin-containing phosphine capture reagent using the Staudinger ligation, purified using streptavidin beads and identified

using mass spectrometry (*43*). This method was used to detect eighteen FTase substrates, including known farnesylated proteins such as H-Ras and Rheb. More recently, azido-GG alcohol has been used to identify geranylgeranylated substrates using click chemistry with TAMRA-alkyne, 2-D gels, and mass spectrometry (*142*) and to label substrates in prenylation deficient mouse tissue (*143*). Additionally, alkyne derivatives of FPP and GGPP are gaining popularity for identification of prenylated proteins *in vitro* and *in vivo* (*144-146*).

An alternative method, the "affinity tag" approach, uses a biotin-geranyl diphosphate analog (BGPP, Figure 1.7 E) (*136*). This analog serves as an efficient substrate for GGTase-II, but not FTase or GGTase-I. To expand the utility of this analog, Alexandrov and coworkers "engineered" FTase and GGTase-I enzymes with altered substrate selectivity by mutating residues in the active site (W102T/Y154T or W102T/Y154T/Y205T for FTase, and F53Y/Y126 or F52Y/F53Y/Y126 for GGTase-I), allowing the engineered enzymes to catalyze modification of substrate proteins using the BGPP analog. Compactin was used to block native FPP and GGPP synthesis, increasing the pool of unmodified FTase and GGTase-I substrates. The compactin-treated cell lysates were incubated *in vitro* with BGPP and mutant or wild-type prenyltransferases, pulled down using streptavidin beads, and modified proteins were identified using mass spectrometry. Many Rab proteins were detected and quantified as GGTase-II substrates, while various molecular weight substrates were identified from the engineered FTase- and GGTase-I-treated lysates.

Although there has been much success using analogs to identify prenylated substrates, there are caveats to these approaches. First of all, the FPP and GGPP analogs

may alter the protein specificity of the enzyme (*147-149*) since the prenyl group forms a portion of the protein substrate binding site (*46*). In particular, for use of the biotin analogs the FTase and GGTase-I enzymes were engineered at positions in the active site that can alter the enzyme specificity (*150*). Another concern is the "hit rate" for these analogs, as treatment with several of these analogs did not identify known substrates as prenylated proteins; these issues with false negatives may indicate that these analogs may not be incorporated into substrates at a high enough level to allow identification. One method to obtain higher incorporation of the analogs is to treat the cells with an inhibitor of FPP or GGPP synthesis, such as compactin (*136, 151*); however this treatment may disrupt the homeostasis of the cell, not allowing for true identification of substrates under normal conditions. Additionally, the BGPP analog was incubated with cell lysates in an *in vitro* context, identifying potential substrates, but perhaps not identifying substrates that are prenylated under native *in vivo* conditions.

*Identifying prenyltransferase substrates through structural and structure-function biochemical studies*

This section reviews the findings from crystallographic data and structure-activity studies that combine to provide a functional and structural picture of the interactions important for prenyltransferase selectivity. Such studies have yielded valuable insights suggesting that the prenylated proteome may be much richer in size and diversity than has been previously proposed. Future studies, focusing on identifying additional interactions and quantifying the energetic contribution of each enzyme-substrate

interaction, have the potential to provide the basis of a quantitative model for predicting the full complement of prenylated proteins within the cell.

*Structural studies*

Over the past 15 years, a series of crystallographic structures of mammalian FTase and GGTase-I have provided insight into the active site interactions and structural context of the peptide substrate binding site within prenyltransferases (*46, 51, 152*). Using these structures as a reference, the selectivity for each amino acid within the $Ca_1a_2X$ motif can be interpreted in light of potential contacts with active site residues. The binding site for the $a_1$ residue of the peptide substrate is exposed to solvent at the interface of the FTase α and β subunits (*46*), consistent with the relaxed specificity at the $a_1$ position observed in biochemical studies (*39-41, 45*). In contrast, the $a_2$ and X residue binding sites lie within the solvent-excluded active site, suggesting that these two positions may be primarily responsible for prenyltransferase selectivity. The $a_2$ site in FTase is mainly composed of the side chains of W102β, W106β, and Y361β (Figure 1.8 D), with analogous residues T49β, F53β, and L321β in GGTase-I. In addition, the prenyl donor co-substrate also contacts the $a_2$ residue in both enzymes (*46*). Structures of FTase complexed with peptides containing different X groups suggest the possibility of two different X-residue binding pockets: S, Q and M interact with Y131α, A98β, S99β, W102β and H149β (Figure 1.8 E), while F (and likely L, N and H) interacts with L96β, S99β, W102β, W106β and A151β (*46*). These X-group binding sites in FTase (the former list) generally confer a preference for prenylation of peptides with moderately polar amino acids such as Ser, Met, and Glu at the X position. In GGTase-I, amino acids

T45β, T49β, and M124β replace the A98β, W102β, and P152β side chains in the X-group binding pocket in FTase. Unexpectedly, this less hydrophobic X residue binding site in GGTase-I leads to a preference for catalyzing prenylation of peptides with more hydrophobic X residues, such as Leu and Phe. These contacts, acting in concert, suggest a set of preferences for FTase and GGTase-I substrates that can serve to guide studies for identifying novel prenyltransferase substrates.

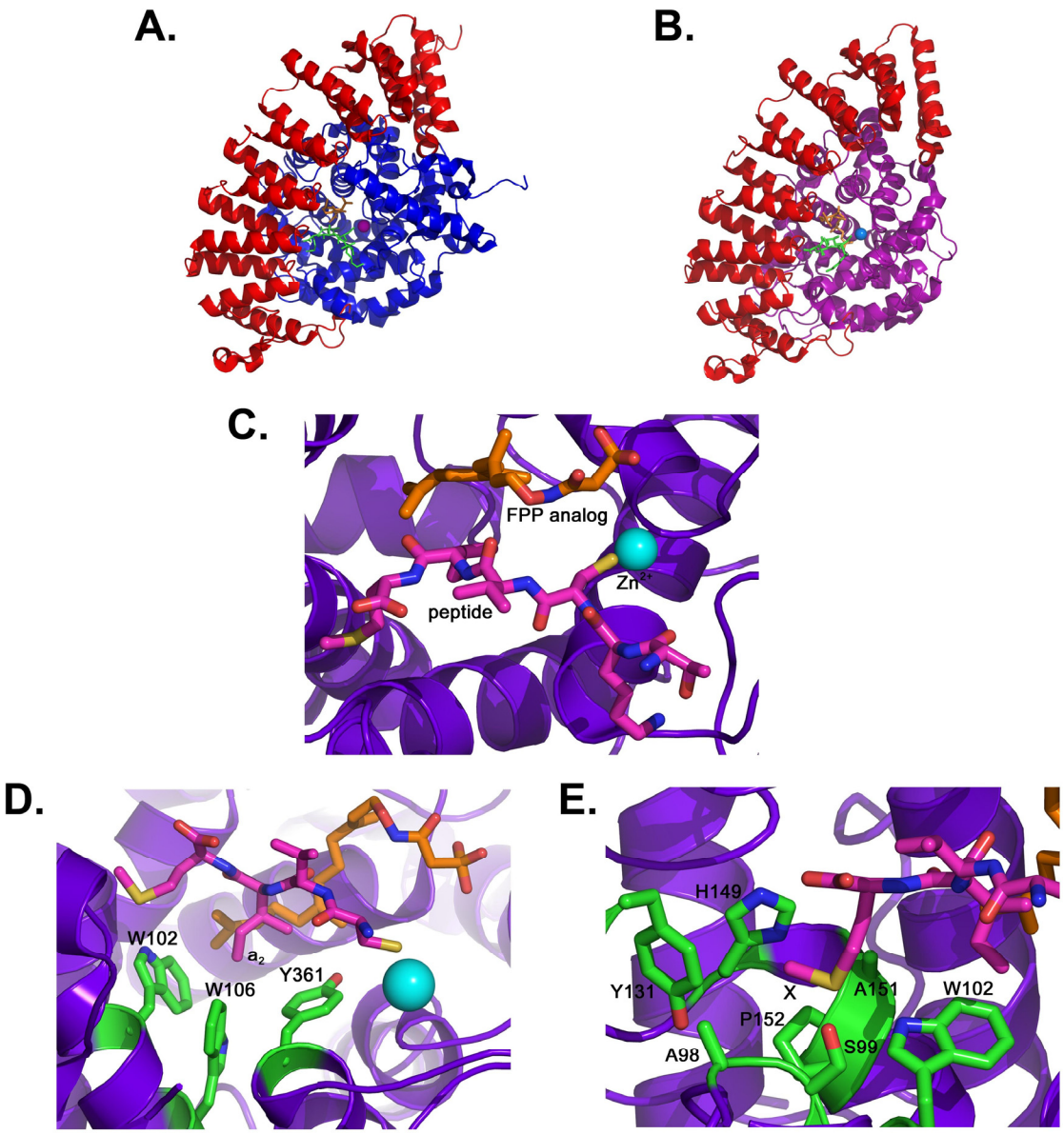


**Figure 1.8. Prenyltransferase structures and FTase active site.** A) The overall structure of FTase (PDB ID 1D8D). B) The overall structure of GGTase-I (PDB ID 1N4Q). The α subunits are shown in red, the β subunit for FTase is shown in blue and the β subunit of GGTase-I is shown in purple. Additionally, the peptide substrates are shown in green, the lipid substrates are shown in orange, and the catalytic zinc ions are shown as spheres. In C), D), and E), the FTase active site is shown (PDB ID 1D8D) with FTase residues in purple or green, peptide substrate KKKSKTKCVIM in pink, FPP analog (FPT Inhibitor II) in orange, and the catalytic $Zn^{2+}$ in light blue. For clarity, only a portion of the peptide substrate is shown: TKCVIM in C) and CVIM in D) and E). C) The structure of the rat FTase active site showing the position of the bound FPP analog and peptide; D) Structure of the contacts between the substrate $a_2$ residue (isoleucine) and FTase residues W102β, W106β, and Y361β (green) and the 3rd isoprene of FPP (orange); E) Structure of the contacts between the substrate X residue (methionine) and the X residue binding pocket in FTase, including residues Y131α, A98β, S99β, W102β, H149β, A151β, and P152β (green) (*46, 153*).

*Structure-function studies of peptide reactivity*

While structural studies provide insight into the interactions that engender FTase and GGTase-I peptide substrate specificity, the energetic benefit (or cost) of each interaction must be identified to ascertain the functional substrate preference at each position within the $Ca_1a_2X$ sequence. Structure-function studies of peptide reactivity, wherein changes in peptide reactivity are correlated with changes in amino acid properties such as hydrophobicity or steric size, provide one method for identifying the properties that are used by FTase or GGTase-I to recognize substrates. At the X position, Hartman and coworkers assayed the reactivity of a series of peptide substrates derived from the C-terminal sequence of K-Ras-4B (-TKCVIM) wherein the terminal amino acid was substituted with 14 amino acids, including the natural Met (*154*). This study illustrated that peptide selectivity for FTase versus GGTase-I arises from relative peptide reactivity, rather than relative binding affinity. Furthermore, peptide reactivity correlates strongly with hydrophobicity, with FTase and GGTase-I displaying inverse reactivity; FTase reactivity decreases and GGTase-I reactivity increases with increasing hydrophobicity at the X position. This "reciprocal" reactivity pattern yielded three groups of peptides: 1) peptides terminating in moderately polar X residues (*i.e.* S and Q) that exhibit exclusive reactivity with FTase; 2) peptides terminating in nonpolar X residues (*i.e.* L and I) that exhibit exclusive GGTase-I reactivity; and 3) peptides terminating in a subset of X residues, such as Met or Phe, that react efficiently with both FTase and GGTase-I. Furthermore, these studies identified a number of peptides that were rapidly prenylated under single turnover conditions but not under multiple turnover conditions, presumably due to slow product dissociation. Further studies of peptide

reactivity demonstrate that FTase is capable of catalyzing farnesylation of peptides with a variety of X-group residues, including leucine (*155*). Additionally, the functional importance of positively-charged residues upstream of $Ca_1a_2X$ motif, often referred to as the polybasic region, has been explored by comparing the reactivity of FTase and GGTase-I with the C-terminal sequence of K-Ras4B (KKKSKTKCVIM versus TKCVIM) (*42*). This work demonstrated that the upstream sequence enhances dual prenylation of substrates by decreasing the efficiency of FTase-catalyzed farnesylation to a level comparable to that of geranylgeranylation catalyzed by GGTase-I. These structure-function studies of X residue recognition indicate that both FTase and GGTase-I can recognize a much wider range of side chains at this position than had been previously proposed. In addition, the possibility for a class of substrates that can be prenylated by either enzyme underscores the potential role of "leaky prenylation" in affecting the makeup and biological role of the prenylated proteome. Taken together, these findings indicate that more than half of the 20 amino acids can serve as the X residue of a prenyltransferase-competent $Ca_1a_2X$ substrate sequence.

A similar structure-activity profile at the $a_2$ position both underscored predictions from structural work regarding $a_2$ sequence preferences and uncovered a previously unknown example of context-dependent substrate recognition at the $a_2$ position of the $Ca_1a_2X$ sequence (*156*). Analysis of the relative reactivity of a series of peptide substrates varying at the $a_2$ position (-$GCVa_2S$ and –$GCVa_2A$) indicated that FTase recognizes both the polarity and steric volume of the $a_2$ side chain simultaneously, discriminating against polar amino acids and both large and small amino acids at this position; maximal activity is observed for amino acids containing a steric volume near

that of valine. These preferences match those predicted by structural studies (*46*), providing a functional picture of the energetic contribution of the structurally predicted interactions to substrate selectivity. Surprisingly, when the reactivity of FTase with analogous peptides with different X residues (-GCV$a_2$M and –GCV$a_2$Q) was analyzed, substrate recognition at the $a_2$ position was predominantly due to polarity. In these peptides, the steric volume of the amino acid at the $a_2$ position did not significantly affect reactivity as long as that amino acid was either weakly polar or nonpolar. This context-dependent $a_2$ selectivity, wherein a larger range of $a_2$ residues can be present in FTase substrates when the X residue is Met or Gln compared to when X is Ser or Ala, suggests that FTase can catalyze farnesylation of proteins with a wide range of $a_2$ residues as efficient substrates provided that an appropriate X residue is present.

Structural studies of FTase and GGTase-I have provided an intricate picture of the interactions and active site microenvironment responsible for recognizing prenyltransferase substrates from among the milieu of all cellular proteins. Structure-function analysis of peptide reactivity indicates that both FTase and GGTase-I recognize a much wider range of protein sequences as substrates than was originally proposed. Furthermore, the specific chemical properties recognized by prenyltransferases at positions within the C$a_1a_2$X sequence have been characterized by correlation of peptide reactivity with amino acid properties such as size and polarity. These structural and functional insights will serve as essential foundations for development of models for comprehensive prediction of prenyltransferase substrates based on protein C-terminal sequence data.

*Peptide library studies: a high-throughput method of identifying potential prenyltransferase substrates and defining substrate specificity*

An additional method of determining prenyltransferase substrates is to define the molecular recognition elements of FTase and GGTase-I using peptide library studies. Short peptides, as small as tetra-peptides, are efficient substrates for FTase and GGTase-I with comparable affinity and reactivity to full-length proteins (*39, 40, 53, 157, 158*). Adding a dansyl group to the N-terminus of the peptide allows for continuous monitoring of the reaction using a fluorescence assay, as the dansyl group increases in fluorescence upon prenylation (*158, 159*). Conveniently, this assay can be carried out in a high-throughput manner using 96-well plates in a plate reader, and with large libraries, statistical analysis of peptide reactivity can be used to determine patterns of substrate recognition (*41*). The use of these peptide libraries to study prenyltransferase specificity is advantageous, since potential substrates can be screened quickly and efficiently, using wild-type enzymes and the natural lipid substrates FPP and GGPP. One limitation with this method is that in some cases the structure of protein substrates may alter recognition. Many research groups have tested small or non-homogenous libraries of peptides as a means to identify prenyltransferase substrates (*148, 160, 161*). For instance, groups have tested the $Ca_1a_2X$ paradigm using GCxxS and GCxxL libraries (*155, 161*), finding that their results generally correlate well with structural studies (*46*) and computational algorithms (*44, 45*).

To more completely define the scope of the prenylated proteome, a large-scale peptide library study of the substrate selectivity of FTase was carried out, using statistical analysis to analyze FTase preference patterns (*41*). The library peptides were screened for

reactivity with FTase under both multiple turnover (MTO), subsaturating peptide ($k_{cat}/K_M^{peptide}$) conditions and single turnover (STO) conditions, using the dansyl fluorescence assay or an *in vitro* radioactive assay with $^3$H-FPP (*41, 158, 159*). In experimental practice, the MTO reaction is performed under conditions with excess substrate such that [E] << [S] while the STO reaction includes excess enzyme and peptide substrate with limiting concentrations of FPP ([FPP] < [E]).

Two kinetic parameters, $k_{cat}/K_M^{peptide}$ and $k_{farnesylation}$, can be measured to describe peptide reactivity with FTase. The value of $k_{cat}/K_M^{peptide}$, measured under MTO conditions, is also termed the "specificity constant" (*162*) and is most representative of the reactivity of a particular substrate in a biological context where all of the protein substrates compete for prenylation catalyzed by FTase and GGTase-I (*162*). *In vivo* the relative rate of prenylation of a given substrate (and hence the selectivity) depends on both the concentration of the protein substrate and the value of $k_{cat}/K_M^{peptide}$. For the prenyltransferase reactions, $k_{cat}/K_M^{peptide}$ represents all of the reaction steps up to the first irreversible step in a reaction (*162*). Previous kinetic studies of FTase suggest the basic kinetic pathway shown in Figure 1.9 (*158, 163-167*). Substrate binding is functionally ordered, with FPP binding before peptide, followed by a conformational rearrangement of the first two isoprene units of FPP required to position the $C_1$ of FPP near the sulfur of the peptide substrate for facile reaction (*49-52*). After the chemical step, diphosphate dissociation is rapid (*165*). For FTase, the $k_{cat}/K_M$ parameter includes the rate constants for peptide binding to E•FPP (forming the ternary complex) through the formation of the prenylated peptide and pyrophosphate products, including the conformation change prior to chemistry and the chemical step ($k_{farnesylation}$, Figure 1.9) (*51, 52, 61, 165, 166*). Under

these conditions, dissociation of the diphosphate product is the first irreversible step (*165, 166*); therefore dissociation of the prenylated peptide product does not contribute to the observed value of $k_{cat}/K_M^{peptide}$. The MTO kinetic parameter measured at saturating concentrations of peptide and FPP, $k_{cat}$, includes all of the rate constants describing the formation of dissociated products (prenylated peptide and diphosphate) from the ternary complex (E•FPP•peptide). Under these conditions, dissociation of the farnesylated peptide is frequently the rate-limiting step (*165, 167*). Furthermore, dissociation of the prenylated product is enhanced by binding FPP, and possibly peptides, to the FTase•farnesylated-peptide complex (Figure 1.9 B) (*154, 167, 168*).

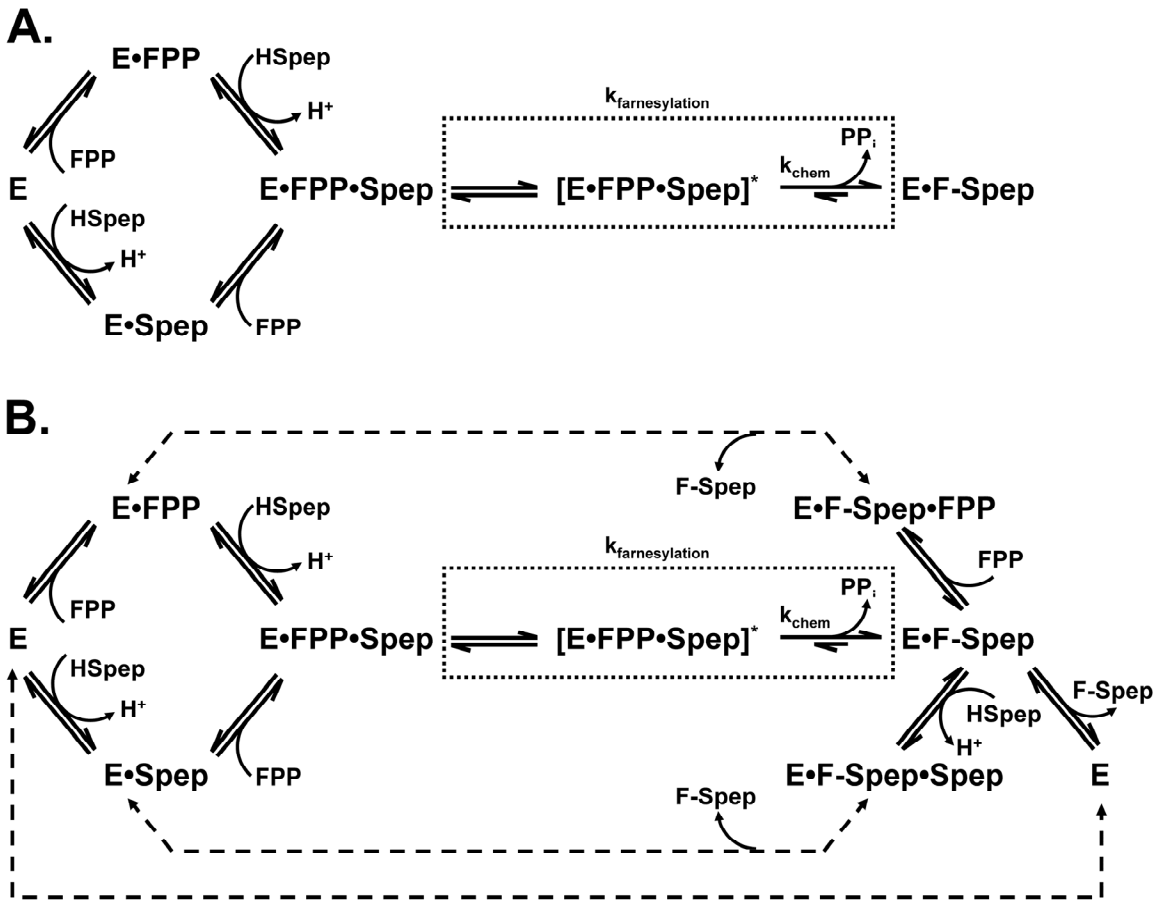


**Figure 1.9. FTase kinetic pathway.** The proposed catalytic cycle for FTase under A) single turnover conditions (STO, [E] > S); and B) multiple turnover conditions (MTO, [E] << [S]). The STO-only substrates are proposed to undergo farnesylation, but not product release. FTase catalyzes turnover of MTO substrates where binding of either an additional peptide or FPP molecule to the E•farnesylated-peptide complex may facilitate product release. The rate constant $k_{farnesylation}$ includes both the conformational change and chemistry steps (*41, 50, 163-167, 169*).

In addition, it is possible to directly measure the rate constant ($k_{farnesylation}$) for the formation of the farnesylated product by detecting the formation of diphosphate using a coupled assay under single-turnover (STO) conditions ([E] > [S]) (Figure 1.9 A) (*165*). The $k_{farnesylation}$ parameter for FTase includes rate constants for binding the peptide substrate to FTase•FPP, the conformation change before the chemistry step, the farnesylation step and the rapid dissociation of diphosphate, thereby including all of the reaction steps up to but not including release of the prenylated protein (*51, 52, 61, 165*). At saturating concentrations of FTase under STO conditions, $k_{farnesylation}$ reflects solely the conformational change and farnesylation steps. Previous studies have demonstrated that, in some cases, peptide substrates are rapidly farnesylated by FTase, as measured under STO conditions, but that product dissociation is so slow that MTO activity is not observable by the current assays (*154, 170*). Measurement of both MTO and STO activity provides additional information to probe the molecular recognition modes of FTase.

The reactivity of FTase with two peptide libraries was measured: an "initial library" of 213 peptides was designed based on the Cxxx sequences taken from the human proteome with a mild bias toward sequences predicted by the $Ca_1a_2X$ model, and a second library, the "targeted library" of 88 peptides, was similarly designed using human sequences but with a stronger bias towards sequences predicted to be FTase substrates based on the results from the initial library. The peptides in these libraries were of the form dansyl-TKCxxx, where x is any amino acid, and the upstream lysine was included to increase peptide solubility (*50*). Overall, FTase catalyzed the farnesylation of a surprisingly large number of peptides (*41*). FTase catalyzed prenylation of 77 peptides,

or 36%, of the initial library under MTO conditions, and an additional 85 peptides (40%) under STO conditions. Together, this means that 76% of the initial 213 peptides were FTase substrates. Additionally, in the targeted library 29 (33%) and 45 (51%) of the peptides were farnesylated by FTase under MTO and STO conditions, respectively. In both libraries, FTase catalyzed farnesylation of 78% of the peptides, including both MTO and STO conditions.

The values of $k_{cat}/K_M^{peptide}$ and $k_{farnesylation}$ were measured for a subset of FTase substrates (*41*). The $k_{cat}/K_M^{peptide}$ parameters for the MTO substrate subset demonstrate a variation of approximately 100-fold. Furthermore, the values for $k_{farnesylation}$ roughly correlate with $k_{cat}/K_M^{peptide}$, that is, more reactive substrates generally had higher $k_{farnesylation}$ values. For the STO substrates, the values for $k_{farnesylation}$ are comparable to those measured for the MTO substrates. Therefore, dissociation of the prenylated peptide must be sufficiently slow for the STO substrates such that MTO activity is not detectable. In this special case, product dissociation rates can alter substrate selectivity even in the presence of multiple competing substrates.

To further probe substrate recognition by FTase, the amino acid preferences at the $a_1$, $a_2$, and X positions of the MTO, STO, and non-substrate peptides from the initial library were compared using statistical analysis. For this analysis, the percentages of "canonical" and "non-canonical" $Ca_1a_2X$ sequences at the $a_2$ and X positions, for the overall library, MTO, STO, and NON pools of peptides (Figure 1.10, top panel) were evaluated (*41*). With "canonical" defined as V, I, L, M, and T for $a_2$, and A, S, M, Q, and F for X, the MTO peptide substrates were generally well-described by the $Ca_1a_2X$ paradigm; however, the STO substrates contain more varied sequences. A

hypergeometric distribution model (*41*) was used to determine statistically significant enrichment (overrepresentation) or depletion (underrepresentation) of a specific amino acid as compared to the overall library. Using $p \leq 0.02$, sequence preferences for FTase could be divided into three classes of peptides (MTO, STO, and NON). MTO substrates are relatively well described by the original $Ca_1a_2X$ paradigm: little sequence selectivity is observed at $a_1$; a nonpolar amino acid, like isoleucine and leucine, is preferred at $a_2$; and the X residue is preferentially phenylalanine, methionine, or glutamine. Additionally, cysteine and lysine were depleted at the $a_2$ position in the reactive MTO substrates. Conversely, STO substrates are not well described by the $Ca_1a_2X$ paradigm or other more recently published substrate prediction models (*44-46, 50, 154, 156*). For the STO peptide substrates: cysteine is enriched while leucine is depleted at the $a_1$ position; at the $a_2$ position, serine is enriched while isoleucine and lysine are depleted; and at the X position, no amino acid is enriched, while methionine is depleted. The anticorrelation of methionine for the MTO and STO substrates at the X position suggests that this amino acid enhances product release (*41*). For unreactive peptides: at the $a_1$ position no sequence preferences were observed; at the $a_2$ position D, K, and R were enriched while V, I, L, and T were depleted; and at the X site, P and R were enriched. In this case, enrichment reflects amino acids that decrease reactivity. These statistically-significant sequence preferences for FTase substrates derived from the human genome, along with other data (*41, 44, 45, 156*), further broadens and defines the $Ca_1a_2X$ paradigm describing FTase selectivity. Figure 1.10 (bottom panel) shows a summary of FTase sequence preferences for substrates at the $a_1$, $a_2$, and X positions derived from these peptide library studies.

The observation of a large number of peptide substrates farnesylated under single turnover but not multiple turnover conditions leads to the important question as to whether the STO substrates will likely be prenylated *in vivo*. Since slow dissociation of the farnesylated-peptide product is predicted to lead to STO behavior, it is possible that STO substrates could form long-lived product complexes with FTase in the cell, raising the possibility that these sequences function as inhibitors rather than biologically relevant substrates. However, at least two STO substrates, with C-terminal sequences corresponding to the peptides -CAVL and -CKAA (*43, 131*), have been demonstrated to be farnesylated *in vivo*, suggesting that these proteins can turnover. One model consistent with these data is that a MTO substrate (or another "release factor") could facilitate the dissociation of the farnesylated-STO protein, similar to the FPP-catalyzed dissociation of product (see kinetic scheme in Figure 1.9 B) (*51, 167*) and this may have a regulatory function within the cell. This model enjoys some support from studies of small peptide reactivity (*154, 167, 168*). Overall, these studies indicate that FTase may farnesylate a larger pool of proteins *in vivo* than would be predicted from the previously-described $Ca_1a_2X$ model.

*Computational work*

In addition to the biochemical and structural 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 help address this challenge of substrate identification. Various computational approaches have been introduced over the last several years to aid in large-scale prediction of prenylated proteins (*44, 45*), with features

that enable the user to rank the likelihood of a particular sequence being a substrate for FTase, GGTase-I, both, or neither. 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.

Several computer algorithms have been developed to help predict protein substrates for mammalian prenyltransferases. Among prediction software, one of the first tools built was Prosite protocol PS00294 which used the consensus pattern C-{DENQ}-[LIVM]-x> (http://www.expasy.org/prosite/, (*171*)). However, this tool is unable to distinguish between FTase and GGTase-I substrates, nor does it predict prenylation by GGTase-II enzyme. 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 (*46*). 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 the most recent algorithm developed to predict prenyltransferase substrates (*44*). To define this algorithm, the authors built a training set of known and homologous substrates (defined by specific rules) which resulted in a set of 692 FTase and 486 GGTase-I substrates. One of the difficulties in predicting prenylation substrates is the inherent complexity of substrate recognition motifs, which may extend beyond the $Ca_1a_2X$ box to include the upstream region of the protein. To address this additional complexity, the authors of PrePS included an eleven amino acid upstream region of the

$Ca_1a_2X$ motif to refine their algorithm, as well as expanding the list of acceptable amino acids for the $Ca_1a_2X$ motif. The upstream region typically consists of a flexible linker region that often has a compositional bias towards 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 fifteen amino acid sequence being a FTase, GGTase-I, and/or GGTase-II substrate. In cross-validation experiments, they were able to establish 92.6% and 98.6% true positive rate for FTase and GGTase-I substrates, respectively, with false positive rates of 0.11% and 0.02% for FTase and GGTase-I, 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 (*41*). 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.

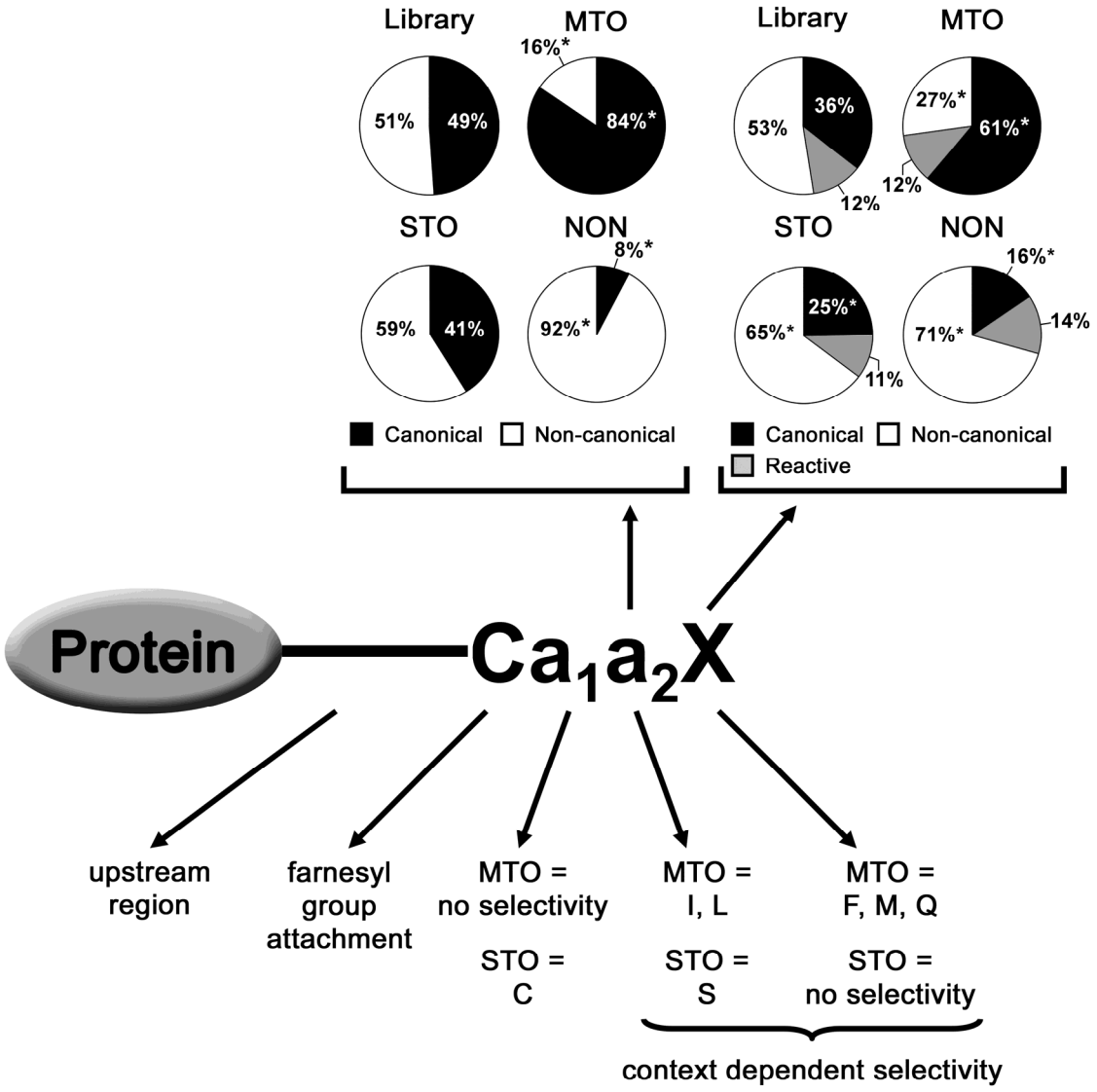


**Figure 1.10. Reactivity of FTase with a library of peptide substrates.** *Top panel:* The percentages of "canonical," "non-canonical," and "reactive" pools of peptides in either the total library, or the MTO, STO, or NON peptide pools catalyzed by FTase and sorted by the amino acid at either the $a_2$ and X position. For $a_2$, canonical = V, I, L, M, T, non-canonical = all remaining amino acids; for X, canonical = A, S, M, Q, F, reactive = C, N, T and non-canonical = I, L, R, Y, D, V, E, G, H, K, P, W. An asterisk (*) represents percentages which are statistically different ($p$<0.02) as compared to the library. *Bottom panel:* FTase sequence preferences for MTO and STO substrates at the $a_1$, $a_2$, and X positions. Upstream region selectivity can include residues that form a flexible linker region. The $a_2$ and X position amino acids are recognized by FTase in a context dependent manner (*41, 44, 45, 156*).

*Conclusions*

Multiple approaches have been employed to identify proteins that are prenylated *in vivo*, catalyzed by FTase and GGTase-I. Although prenyltransferase substrates can be identified and studied one at time, FPP and GGPP analogs are being developed to allow detection of farnesylated and geranylgeranylated proteins in a "high-throughput" fashion. Furthermore, structural and structure/function studies have provided insight into the interactions between substrates and the FTase and GGTase enzymes that are important for molecular recognition. Peptide library studies have supplied data regarding the amino acid composition of substrates and non-substrate pools, allowing the use of statistical analysis to determine patterns of FTase recognition and indicating that the $Ca_1a_2X$ paradigm does not sufficiently describe the recognition of substrates by FTase. Furthermore, these studies have uncovered the puzzle of the reactivity of single-turnover substrates. Computational studies have allowed prediction of substrates *in vivo*, based upon both the C-terminus of the protein as well as the upstream region. All of these studies in concert have helped to define the current pool of proteins known to be prenyltransferase substrates. Definition of the prenylated proteome will be essential to better understand the roles of prenylation in cellular signaling, disease processes, and the complex array of post-translational modifications within the cell.

*Objectives of this work*

Detailed substrate specificity studies for FTase have been carried out using peptide library studies, but less information about GGTase-I specificity is available. To begin, the expression conditions for recombinantly expressed GGTase-I were improved

using auto-induction media, in turn increasing the purification yield of GGTase-I 2.7-fold. The reactivity of a 402 dansyl-TKCxxx and dansyl-GCxxx peptide library with GGTase-I was tested. It was found that 111 of these peptides were substrates with GGTase-I (28% of the library) under MTO conditions and 178 (44%) peptides were active with GGTase-I under STO conditions. Statistical analysis revealed amino acid preferences for GGTase-I substrates in each pool (MTO, STO, and NON) of the peptides. For instance, for peptides with MTO activity, GGTase-I prefers I and L at the $a_2$ position, F, L, M, and V at the X position. Little sequence selectivity occurs at the $a_1$ position under MTO conditions. Further, peptides with STO activity with GGTase-I tend to have P and S at the X position and little sequence overrepresentation at the $a_1$ and $a_2$ positions. Although this analysis suggests that GGTase-I tends to recognize substrates with typical canonical CaaX residues, this work also indicates that a large percentage of peptides with non-canonical sequences can still react with GGTase-I.

In collaboration with Dr. Ora Schueler-Furman and Nir London at the Hebrew University of Jerusalem, a computational method called FlexPepBind was developed and tested to enhance the prediction of FTase substrates. This algorithm uses modeling based on structural data and the Rosetta suite, and calculates binding energy scores for CaaX sequences of potential FTase substrates. Twenty-nine good scoring peptides predicted from FlexPepBind were chosen to be tested *in vitro* and FTase catalyzed farnesylation of 26 peptides under MTO or STO conditions. FlexPepBind also enhanced prediction of non-canonical X residues of FTase substrates with a negatively charged X group, such as E or D, and also correctly identified a larger pool of the STO FTase substrates than previous algorithms. It is unclear at this time what physically measureable parameter

correlates with and explains the predictive power of the FlexPepBind algorithm, but nevertheless, the program is very successful at predicting FTase substrates.

Finally, it is unknown what substrates are modified in the prenylation pathway so more techniques are needed to identify farnesylated, geranylgeranylated, palmitoylated, proteolyzed, and methylated proteins *in vivo*. The lipid donor analogs azido-farnesol and azido-geranylgeraniol were used to tag and pull-down potential FTase and GGTase-I substrates. Overall, prenylated proteins were detected by this method, but the analogs were incorporated at a low level and there was high background. A different approach using a library of $His_6$-EGFP-TEV-$X_{11}$-CaaX genes in mammalian expression vectors, transfections, and fluorescence microscopy was used to evaluate FTase and GGTase-I specificity *in vivo*. The EGFP-fusion proteins that matched C-terminal sequences reactive with the prenyltransferases *in vitro* typically displayed membrane localization when overexpressed in cells, suggesting that the proteins undergo modification. Likely, mass spectrometry will be the most useful tool in the future to analyze the modifications on these fusion proteins. Together, this work furthers the understanding the substrate recognition of FTase and GGTase-I, helps to predict substrates *in vivo*, and will help to understand which signaling and metabolic pathways may be affected by treatment with FTase and GGTase-I inhibitors.

## References


(1) Nosjean, O., Briolay, A., and Roux, B. (1997) Mammalian GPI proteins: sorting, membrane residence and functions. *Biochim Biophys Acta 1331*, 153-86.

(2) Paulick, M. G., and Bertozzi, C. R. (2008) The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteinsâ€ *Biochemistry 47*, 6991-7000.

(3) Ferguson, M. A. (1999) The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. *J Cell Sci 112 ( Pt 17)*, 2799-809.

(4) Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G., and Eisenhaber, F. (2003) Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins. *Bioessays 25*, 367-85.

(5) Tiede, A., Bastisch, I., Schubert, J., Orlean, P., and Schmidt, R. E. (1999) Biosynthesis of glycosylphosphatidylinositols in mammals and unicellular microbes. *Biol Chem 380*, 503-23.

(6) Chesebro, B., Trifilo, M., Race, R., Meade-White, K., Teng, C., LaCasse, R., Raymond, L., Favara, C., Baron, G., Priola, S., Caughey, B., Masliah, E., and Oldstone, M. (2005) Anchorless prion protein results in infectious amyloid disease without clinical scrapie. *Science 308*, 1435-9.

(7) Benson, R. A., Lowrey, J. A., Lamb, J. R., and Howie, S. E. (2004) The Notch and Sonic hedgehog signalling pathways in immunity. *Mol Immunol 41*, 715-25.

(8) Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell 91*, 85-97.

(9) Buglino, J. A., and Resh, M. D. (2008) Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J Biol Chem 283*, 22076-88.

(10) Cooper, M. K., Porter, J. A., Young, K. E., and Beachy, P. A. (1998) Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science 280*, 1603-7.

(11) Towler, D. A., Eubanks, S. R., Towery, D. S., Adams, S. P., and Glaser, L. (1987) Amino-terminal processing of proteins by N-myristoylation. Substrate specificity of N-myristoyl transferase. *J Biol Chem 262*, 1030-6.

(12) Boutin, J. A. (1997) Myristoylation. *Cell Signal 9*, 15-35.

(13) Farazi, T. A., Waksman, G., and Gordon, J. I. (2001) The biology and enzymology of protein N-myristoylation. *J Biol Chem 276*, 39501-4.

(14) Martin, D. D., Beauchamp, E., and Berthiaume, L. G. (2011) Post-translational myristoylation: Fat matters in cellular life and death. *Biochimie 93*, 18-31.

(15) Zha, J., Weiler, S., Oh, K. J., Wei, M. C., and Korsmeyer, S. J. (2000) Posttranslational N-myristoylation of BID as a molecular switch for targeting mitochondria and apoptosis. *Science 290*, 1761-5.

(16) Resh, M. D. (2006) Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat Chem Biol 2*, 584-90.

(17) Fukata, Y., and Fukata, M. (2010) Protein palmitoylation in neuronal development and synaptic plasticity. *Nat Rev Neurosci 11*, 161-75.

(18) Nadolski, M. J., and Linder, M. E. (2007) Protein lipidation. *FEBS Journal 274*, 5202-5210.

(19) Smotrys, J. E., and Linder, M. E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Annu Rev Biochem 73*, 559-87.

(20) Resh, M. D. (2006) Palmitoylation of ligands, receptors, and intracellular signaling molecules. *Sci STKE 2006*, re14.

(21) Lobo, S., Greentree, W. K., Linder, M. E., and Deschenes, R. J. (2002) Identification of a Ras palmitoyltransferase in Saccharomyces cerevisiae. *J Biol Chem 277*, 41268-73.

(22) Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R., 3rd, and Davis, N. G. (2006) Global analysis of protein palmitoylation in yeast. *Cell 125*, 1003-13.

(23) Tsutsumi, R., Fukata, Y., and Fukata, M. (2008) Discovery of protein-palmitoylating enzymes. *Pflugers Arch 456*, 1199-206.

(24) Duncan, J. A., and Gilman, A. G. (1998) A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21(RAS). *J Biol Chem 273*, 15830-7.

(25) Wright, L. P., and Philips, M. R. (2006) Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras. *J Lipid Res 47*, 883-91.

(26) Yanai, A., Huang, K., Kang, R., Singaraja, R. R., Arstikaitis, P., Gan, L., Orban, P. C., Mullard, A., Cowan, C. M., Raymond, L. A., Drisdel, R. C., Green, W. N., Ravikumar, B., Rubinsztein, D. C., El-Husseini, A., and Hayden, M. R. (2006)

Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function. *Nat Neurosci 9*, 824-31.

(27) Benetka, W., Koranda, M., and Eisenhaber, F. (2006) Protein Prenylation: An (Almost) Comprehensive Overview on Discovery History, Enzymology, and Significance in Physiology and Disease. *Monatshefte für Chemie / Chemical Monthly 137*, 1241-1281.

(28) Casey, P. J., and Seabra, M. C. (1996) Protein Prenyltransferases. *Journal of Biological Chemistry 271*, 5289-5292.

(29) Marshall, C. J. (1993) Protein prenylation: a mediator of protein-protein interactions. *Science 259*, 1865-6.

(30) Casey, P. J. (1994) Lipid modifications of G proteins. *Curr Opin Cell Biol 6*, 219-25.

(31) Zhang, F. L., and Casey, P. J. (1996) Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem 65*, 241-69.

(32) Armstrong, S. A., Seabra, M. C., Sudhof, T. C., Goldstein, J. L., and Brown, M. S. (1993) cDNA cloning and expression of the alpha and beta subunits of rat Rab geranylgeranyl transferase. *J Biol Chem 268*, 12221-9.

(33) Horiuchi, H., Kawata, M., Katayama, M., Yoshida, Y., Musha, T., Ando, S., and Takai, Y. (1991) A novel prenyltransferase for a small GTP-binding protein having a C-terminal Cys-Ala-Cys structure. *J Biol Chem 266*, 16981-4.

(34) Seabra, M. C., Goldstein, J. L., Sudhof, T. C., and Brown, M. S. (1992) Rab geranylgeranyl transferase. A multisubunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys. *J Biol Chem 267*, 14497-503.

(35) Anant, J. S., Desnoyers, L., Machius, M., Demeler, B., Hansen, J. C., Westover, K. D., Deisenhofer, J., and Seabra, M. C. (1998) Mechanism of Rab geranylgeranylation: formation of the catalytic ternary complex. *Biochemistry 37*, 12559-68.

(36) Casey, P. J. (1995) Protein Lipidation in Cell Signaling. *Science 268*, 221-225.

(37) Caplin, B. E., Hettich, L. A., and Marshall, M. S. (1994) Substrate characterization of the Saccharomyces cerevisiae protein farnesyltransferase and type-I protein geranylgeranyltransferase. *Biochim Biophys Acta 1205*, 39-48.

(38) Omer, C. A., Kral, A. M., Diehl, R. E., Prendergast, G. C., Powers, S., Allen, C. M., Gibbs, J. B., and Kohl, N. E. (1993) Characterization of recombinant human farnesyl-protein transferase: cloning, expression, farnesyl diphosphate binding, and functional homology with yeast prenyl-protein transferases. *Biochemistry 32*, 5167-76.

(39) Reiss, Y., Stradley, S. J., Gierasch, L. M., Brown, M. S., and Goldstein, J. L. (1991) Sequence requirement for peptide recognition by rat brain p21ras protein farnesyltransferase. *Proc Natl Acad Sci U S A 88*, 732-6.

(40) Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L., and Gibbs, J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem 266*, 14603-10.

(41) Hougland, J. L., Hicks, K. A., Hartman, H. L., Kelly, R. A., Watt, T. J., and Fierke, C. A. (2010) Identification of Novel Peptide Substrates for Protein Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J Mol Biol 395*, 176-90.

(42) Hicks, K. A., Hartman, H. L., and Fierke, C. A. (2005) Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I. *Biochemistry 44*, 15325-33.

(43) Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004) A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci U S A 101*, 12479-84.

(44) Maurer-Stroh, S., and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biol 6*, R55.

(45) Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F. L., and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Comput Biol 3*, e66.

(46) Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol 343*, 417-33.

(47) Seabra, M. C., Reiss, Y., Casey, P. J., Brown, M. S., and Goldstein, J. L. (1991) Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit. *Cell 65*, 429-34.

(48) Lane, K. T., and Beese, L. S. (2006) Thematic review series: lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type I. *J Lipid Res 47*, 681-99.

(49) Bowers, K. E., and Fierke, C. A. (2004) Positively charged side chains in protein farnesyltransferase enhance catalysis by stabilizing the formation of the diphosphate leaving group. *Biochemistry 43*, 5256-65.

(50) Long, S. B., Hancock, P. J., Kral, A. M., Hellinga, H. W., and Beese, L. S. (2001) 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 98*, 12948-53.

(51) Long, S. B., Casey, P. J., and Beese, L. S. (2002) Reaction path of protein farnesyltransferase at atomic resolution. *Nature 419*, 645-50.

(52) Pickett, J. S., Bowers, K. E., Hartman, H. L., Fu, H. W., Embry, A. C., Casey, P. J., and Fierke, C. A. (2003) Kinetic studies of protein farnesyltransferase mutants establish active substrate conformation. *Biochemistry 42*, 9741-8.

(53) Hightower, K. E., Huang, C. C., Casey, P. J., and Fierke, C. A. (1998) H-Ras peptide and protein substrates bind protein farnesyltransferase as an ionized thiolate. *Biochemistry 37*, 15555-62.

(54) Saderholm, M. J., Hightower, K. E., and Fierke, C. A. (2000) Role of metals in the reaction catalyzed by protein farnesyltransferase. *Biochemistry 39*, 12398-405.

(55) Pickett, J. S., Bowers, K. E., and Fierke, C. A. (2003) Mutagenesis studies of protein farnesyltransferase implicate aspartate beta 352 as a magnesium ligand. *J Biol Chem 278*, 51243-50.

(56) Yang, Y., Chakravorty, D. K., and Merz, K. M., Jr. (2010) Finding a needle in the haystack: computational modeling of Mg2+ binding in the active site of protein farnesyltransferase. *Biochemistry 49*, 9658-66.

(57) Hartman, H. L., Bowers, K. E., and Fierke, C. A. (2004) Lysine beta311 of protein geranylgeranyltransferase type I partially replaces magnesium. *J Biol Chem 279*, 30546-53.

(58) Dolence, J. M., and Poulter, C. D. (1995) A mechanism for posttranslational modifications of proteins by yeast protein farnesyltransferase. *Proc Natl Acad Sci U S A 92*, 5008-11.

(59) Huang, C., Hightower, K. E., and Fierke, C. A. (2000) Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state. *Biochemistry 39*, 2593-602.

(60) Edelstein, R. L., Weller, V. A., Distefano, M. D., and Tung, J. S. (1998) Stereochemical Analysis of the Reaction Catalyzed by Yeast Protein Farnesyltransferase. *The Journal of Organic Chemistry 63*, 5298-5299.

(61) Pais, J. E., Bowers, K. E., and Fierke, C. A. (2006) Measurement of the alpha-secondary kinetic isotope effect for the reaction catalyzed by mammalian protein farnesyltransferase. *J Am Chem Soc 128*, 15086-7.

(62) Kamiya, Y., Sakurai, A., Tamura, S., Takahashi, N., Abe, K., Tsuchiya, E., Fukui, S., Kitada, C., and Fujino, M. (1978) Structure of rhodotorucine A, a novel

lipopeptide, inducing mating tube formation in Rhodosporidium toruloides. *Biochemical and Biophysical Research Communications 83*, 1077-1083.

(63) Schmidt, R. A., Schneider, C. J., and Glomset, J. A. (1984) Evidence for post-translational incorporation of a product of mevalonic acid into Swiss 3T3 cell proteins. *Journal of Biological Chemistry 259*, 10175-10180.

(64) Sinensky, M., and Logel, J. (1985) Defective macromolecule biosynthesis and cell-cycle progression in a mammalian cell starved for mevalonate. *Proceedings of the National Academy of Sciences 82*, 3257-3261.

(65) Powers, S., Michaelis, S., Broek, D., Santa Anna, S., Field, J., Herskowitz, I., and Wigler, M. (1986) RAM, a gene of yeast required for a functional modification of RAS proteins and for production of mating pheromone a-factor. *Cell 47*, 413-22.

(66) Winter-Vann, A. M., and Casey, P. J. (2005) Post-prenylation-processing enzymes as new targets in oncogenesis. *Nat Rev Cancer 5*, 405-12.

(67) Barrowman, J., and Michaelis, S. (2009) ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders. *Biological Chemistry 390*, 761-773.

(68) Trueblood, C. E., Boyartchuk, V. L., Picologlou, E. A., Rozema, D., Poulter, C. D., and Rine, J. (2000) The CaaX proteases, Afc1p and Rce1p, have overlapping but distinct substrate specificities. *Mol Cell Biol 20*, 4381-92.

(69) Dolence, J. M., Steward, L. E., Dolence, E. K., Wong, D. H., and Poulter, C. D. (2000) Studies with recombinant Saccharomyces cerevisiae CaaX prenyl protease Rce1p. *Biochemistry 39*, 4096-104.

(70) Otto, J. C., Kim, E., Young, S. G., and Casey, P. J. (1999) Cloning and characterization of a mammalian prenyl protein-specific protease. *J Biol Chem 274*, 8379-82.

(71) Jang, G. F., and Gelb, M. H. (1998) Substrate specificity of mammalian prenyl protein-specific endoprotease activity. *Biochemistry 37*, 4473-81.

(72) Hollander, I., Frommer, E., and Mallon, R. (2000) Human ras-converting enzyme (hRCE1) endoproteolytic activity on K-ras-derived peptides. *Anal Biochem 286*, 129-37.

(73) Hollander, I. J., Frommer, E., Aulabaugh, A., and Mallon, R. (2003) Human Ras converting enzyme endoproteolytic specificity at the P2' and P3' positions of K-Ras-derived peptides. *Biochim Biophys Acta 1649*, 24-9.

(74) Schmidt, W. K., Tam, A., Fujimura-Kamada, K., and Michaelis, S. (1998) Endoplasmic reticulum membrane localization of Rce1p and Ste24p, yeast

proteases involved in carboxyl-terminal CAAX protein processing and amino-terminal a-factor cleavage. *Proc Natl Acad Sci U S A 95*, 11175-80.

(75) Bergo, M. O., Wahlstrom, A. M., Fong, L. G., and Young, S. G. (2008) Genetic analyses of the role of RCE1 in RAS membrane association and transformation. *Methods Enzymol 438*, 367-89.

(76) Coats, S. G., Booden, M. A., and Buss, J. E. (1999) Transient palmitoylation supports H-Ras membrane binding but only partial biological activity. *Biochemistry 38*, 12926-34.

(77) Laude, A. J., and Prior, I. A. (2008) Palmitoylation and localisation of RAS isoforms are modulated by the hypervariable linker domain. *J Cell Sci 121*, 421-7.

(78) Dudler, T., and Gelb, M. H. (1996) Palmitoylation of Ha-Ras facilitates membrane binding, activation of downstream effectors, and meiotic maturation in Xenopus oocytes. *J Biol Chem 271*, 11541-7.

(79) Jackson, J. H., Li, J. W., Buss, J. E., Der, C. J., and Cochrane, C. G. (1994) Polylysine domain of K-ras 4B protein is crucial for malignant transformation. *Proc Natl Acad Sci U S A 91*, 12730-4.

(80) Heo, W. D., Inoue, T., Park, W. S., Kim, M. L., Park, B. O., Wandless, T. J., and Meyer, T. (2006) PI(3,4,5)P3 and PI(4,5)P2 lipids target proteins with polybasic clusters to the plasma membrane. *Science 314*, 1458-61.

(81) Samuel, F., and Hynds, D. L. (2010) RHO GTPase signaling for axon extension: is prenylation important? *Mol Neurobiol 42*, 133-42.

(82) Rusinol, A. E., and Sinensky, M. S. (2006) Farnesylated lamins, progeroid syndromes and farnesyl transferase inhibitors. *J Cell Sci 119*, 3265-72.

(83) Basso, A. D., Kirschmeier, P., and Bishop, W. R. (2006) Lipid posttranslational modifications. Farnesyl transferase inhibitors. *J Lipid Res 47*, 15-31.

(84) Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2008) Farnesyltransferase inhibitors: a detailed chemical view on an elusive biological problem. *Curr Med Chem 15*, 1478-92.

(85) Young, S. G., Meta, M., Yang, S. H., and Fong, L. G. (2006) Prelamin A farnesylation and progeroid syndromes. *J Biol Chem 281*, 39741-5.

(86) Nallan, L., Bauer, K. D., Bendale, P., Rivas, K., Yokoyama, K., Horney, C. P., Pendyala, P. R., Floyd, D., Lombardo, L. J., Williams, D. K., Hamilton, A., Sebti, S., Windsor, W. T., Weber, P. C., Buckner, F. S., Chakrabarti, D., Gelb, M. H., and Van Voorhis, W. C. (2005) Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. *J Med Chem 48*, 3704-13.

(87) Sepp-Lorenzino, L., Ma, Z., Rands, E., Kohl, N. E., Gibbs, J. B., Oliff, A., and Rosen, N. (1995) A peptidomimetic inhibitor of farnesyl:protein transferase blocks the anchorage-dependent and -independent growth of human tumor cell lines. *Cancer Res 55*, 5302-9.

(88) Nonaka, M., Uota, S., Saitoh, Y., Takahashi, M., Sugimoto, H., Amet, T., Arai, A., Miura, O., Yamamoto, N., and Yamaoka, S. (2009) Role for protein geranylgeranylation in adult T-cell leukemia cell survival. *Experimental Cell Research 315*, 141-150.

(89) Chiba, Y., Sato, S., Hanazaki, M., Sakai, H., and Misawa, M. (2009) Inhibition of geranylgeranyltransferase inhibits bronchial smooth muscle hyperresponsiveness in mice. *American Journal of Physiology-Lung Cellular and Molecular Physiology 297*, L984-L991.

(90) Ridker, P. M., Danielson, E., Fonseca, F. A., Genest, J., Gotto, A. M., Jr., Kastelein, J. J., Koenig, W., Libby, P., Lorenzatti, A. J., Macfadyen, J. G., Nordestgaard, B. G., Shepherd, J., Willerson, J. T., and Glynn, R. J. (2008) Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein. *N Engl J Med*.

(91) Burke, B., and Stewart, C. L. (2002) Life at the edge: the nuclear envelope and human disease. *Nat Rev Mol Cell Biol 3*, 575-85.

(92) Muchir, A., and Worman, H. J. (2004) The nuclear envelope and human disease. *Physiology (Bethesda) 19*, 309-14.

(93) Worman, H. J., and Bonne, G. (2007) "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res 313*, 2121-33.

(94) Mounkes, L. C., Burke, B., and Stewart, C. L. (2001) The A-type lamins: nuclear structural proteins as a focus for muscular dystrophy and cardiovascular diseases. *Trends Cardiovasc Med 11*, 280-5.

(95) Wilson, K. L. (2000) The nuclear envelope, muscular dystrophy and gene expression. *Trends Cell Biol 10*, 125-9.

(96) Mallampalli, M. P., Huyer, G., Bendale, P., Gelb, M. H., and Michaelis, S. (2005) Inhibiting farnesylation reverses the nuclear morphology defect in a HeLa cell model for Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci U S A 102*, 14416-21.

(97) Kieran, M. W., Gordon, L., and Kleinman, M. (2007) New approaches to progeria. *Pediatrics 120*, 834-41.

(98) Toth, J. I., Yang, S. H., Qiao, X., Beigneux, A. P., Gelb, M. H., Moulson, C. L., Miner, J. H., Young, S. G., and Fong, L. G. (2005) Blocking protein

farnesyltransferase improves nuclear shape in fibroblasts from humans with progeroid syndromes. *Proc Natl Acad Sci U S A 102*, 12873-8.

(99) Fong, L. G., Frost, D., Meta, M., Qiao, X., Yang, S. H., Coffinier, C., and Young, S. G. (2006) A protein farnesyltransferase inhibitor ameliorates disease in a mouse model of progeria. *Science 311*, 1621-3.

(100) Merideth, M. A., Gordon, L. B., Clauss, S., Sachdev, V., Smith, A. C., Perry, M. B., Brewer, C. C., Zalewski, C., Kim, H. J., Solomon, B., Brooks, B. P., Gerber, L. H., Turner, M. L., Domingo, D. L., Hart, T. C., Graf, J., Reynolds, J. C., Gropman, A., Yanovski, J. A., Gerhard-Herman, M., Collins, F. S., Nabel, E. G., Cannon, R. O., 3rd, Gahl, W. A., and Introne, W. J. (2008) Phenotype and course of Hutchinson-Gilford progeria syndrome. *N Engl J Med 358*, 592-604.

(101) Bishop, W. R., Doll, R., and Kirschmeier, P. (2011) Farnesyl Transferase Inhibitors: From Targeted Cancer Therapeutic to a Potential Treatment for Progeria, in *The Enzymes* (Tamanoi, F., Hrycyna, C. A., and Bergo, M. O., Eds.) pp 275-303, Academic Press.

(102) Karnoub, A. E., and Weinberg, R. A. (2008) Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol 9*, 517-31.

(103) Vojtek, A. B., and Der, C. J. (1998) Increasing complexity of the Ras signaling pathway. *J Biol Chem 273*, 19925-8.

(104) Whyte, D. B., Kirschmeier, P., Hockenberry, T. N., Nunez-Oliva, I., James, L., Catino, J. J., Bishop, W. R., and Pai, J. K. (1997) K- and N-Ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors. *J Biol Chem 272*, 14459-64.

(105) Zhang, F. L., Kirschmeier, P., Carr, D., James, L., Bond, R. W., Wang, L., Patton, R., Windsor, W. T., Syto, R., Zhang, R., and Bishop, W. R. (1997) Characterization of Ha-ras, N-ras, Ki-Ras4A, and Ki-Ras4B as in vitro substrates for farnesyl protein transferase and geranylgeranyl protein transferase type I. *J Biol Chem 272*, 10232-9.

(106) Lerner, E. C., Zhang, T. T., Knowles, D. B., Qian, Y., Hamilton, A. D., and Sebti, S. M. (1997) Inhibition of the prenylation of K-Ras, but not H- or N-Ras, is highly resistant to CAAX peptidomimetics and requires both a farnesyltransferase and a geranylgeranyltransferase I inhibitor in human tumor cell lines. *Oncogene 15*, 1283-8.

(107) Choy, E., Chiu, V. K., Silletti, J., Feoktistov, M., Morimoto, T., Michaelson, D., Ivanov, I. E., and Philips, M. R. (1999) Endomembrane trafficking of ras: the CAAX motif targets proteins to the ER and Golgi. *Cell 98*, 69-80.

(108) Chiu, V. K., Bivona, T., Hach, A., Sajous, J. B., Silletti, J., Wiener, H., Johnson, R. L., 2nd, Cox, A. D., and Philips, M. R. (2002) Ras signalling on the endoplasmic reticulum and the Golgi. *Nat Cell Biol 4*, 343-50.

(109) Hancock, J. F., Paterson, H., and Marshall, C. J. (1990) A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell 63*, 133-9.

(110) Rocks, O., Peyker, A., Kahms, M., Verveer, P. J., Koerner, C., Lumbierres, M., Kuhlmann, J., Waldmann, H., Wittinghofer, A., and Bastiaens, P. I. (2005) An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. *Science 307*, 1746-52.

(111) Magee, A. I., Gutierrez, L., McKay, I. A., Marshall, C. J., and Hall, A. (1987) Dynamic fatty acylation of p21N-ras. *Embo J 6*, 3353-7.

(112) Goodwin, J. S., Drake, K. R., Rogers, C., Wright, L., Lippincott-Schwartz, J., Philips, M. R., and Kenworthy, A. K. (2005) Depalmitoylated Ras traffics to and from the Golgi complex via a nonvesicular pathway. *J Cell Biol 170*, 261-72.

(113) Gelb, M. H., Van Voorhis, W. C., Buckner, F. S., Yokoyama, K., Eastman, R., Carpenter, E. P., Panethymitaki, C., Brown, K. A., and Smith, D. F. (2003) Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. *Mol Biochem Parasitol 126*, 155-63.

(114) Bulbule, V. J., Rivas, K., Verlinde, C. L., Van Voorhis, W. C., and Gelb, M. H. (2008) 2-Oxotetrahydroquinoline-based antimalarials with high potency and metabolic stability. *J Med Chem 51*, 384-7.

(115) Fletcher, S., Cummings, C. G., Rivas, K., Katt, W. P., Horney, C., Buckner, F. S., Chakrabarti, D., Sebti, S. M., Gelb, M. H., Van Voorhis, W. C., and Hamilton, A. D. (2008) 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 51*, 5176-97.

(116) Smalera, I., Williamson, J. M., Baginsky, W., Leiting, B., and Mazur, P. (2000) Expression and characterization of protein geranylgeranyltransferase type I from the pathogenic yeast Candida albicans and identification of yeast selective enzyme inhibitors. *Biochim Biophys Acta 1480*, 132-44.

(117) Song, J. L., and White, T. C. (2003) RAM2: an essential gene in the prenylation pathway of Candida albicans. *Microbiology 149*, 249-59.

(118) Mazur, P., Register, E., Bonfiglio, C. A., Yuan, X., Kurtz, M. B., Williamson, J. M., and Kelly, R. (1999) Purification of geranylgeranyltransferase I from Candida

albicans and cloning of the CaRAM2 and CaCDC43 genes encoding its subunits. *Microbiology 145 ( Pt 5)*, 1123-35.

(119) Kelly, R., Card, D., Register, E., Mazur, P., Kelly, T., Tanaka, K. I., Onishi, J., Williamson, J. M., Fan, H., Satoh, T., and Kurtz, M. (2000) Geranylgeranyltransferase I of Candida albicans: null mutants or enzyme inhibitors produce unexpected phenotypes. *J Bacteriol 182*, 704-13.

(120) Eastman, R. T., Buckner, F. S., Yokoyama, K., Gelb, M. H., and Van Voorhis, W. C. (2006) Thematic review series: lipid posttranslational modifications. Fighting parasitic disease by blocking protein farnesylation. *J Lipid Res 47*, 233-40.

(121) Eastman, R. T., White, J., Hucke, O., Yokoyama, K., Verlinde, C. L., Hast, M. A., Beese, L. S., Gelb, M. H., Rathod, P. K., and Van Voorhis, W. C. (2007) Resistance mutations at the lipid substrate binding site of Plasmodium falciparum protein farnesyltransferase. *Mol Biochem Parasitol 152*, 66-71.

(122) Ivanov, S. S., Charron, G., Hang, H. C., and Roy, C. R. (2010) Lipidation by the host prenyltransferase machinery facilitates membrane localization of Legionella pneumophila effector proteins. *J Biol Chem 285*, 34686-98.

(123) Price, C. T., Al-Quadan, T., Santic, M., Jones, S. C., and Abu Kwaik, Y. (2010) Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of Legionella pneumophila. *J Exp Med 207*, 1713-26.

(124) Hancock, J. F. (1995) Reticulocyte lysate assay for in vitro translation and posttranslational modification of Ras proteins. *Methods Enzymol 255*, 60-5.

(125) Wilson, A. L., and Maltese, W. A. (1995) Coupled Translation/Prenylation of Rab Proteins in-Vitro, in *Lipid Modifications of Proteins* pp 79-91, Academic Press Inc, San Diego.

(126) Gibbs, B. S., Zahn, T. J., Mu, Y., Sebolt-Leopold, J. S., and Gibbs, R. A. (1999) Novel farnesol and geranylgeraniol analogues: A potential new class of anticancer agents directed against protein prenylation. *J Med Chem 42*, 3800-8.

(127) Peter, M., Chavrier, P., Nigg, E. A., and Zerial, M. (1992) Isoprenylation of Rab Proteins on Structurally Distinct Cysteine Motifs. *Journal of Cell Science 102*, 857-865.

(128) Corsini, A., Farnsworth, C. C., McGeady, P., Gelb, M. H., and Glomset, J. A. (1999) Incorporation of radiolabeled prenyl alcohols and their analogs into mammalian cell proteins. A useful tool for studying protein prenylation. *Methods Mol Biol 116*, 125-44.

(129) Andres, D. A., Crick, D. C., Finlin, B. S., and Waechter, C. J. (1999) Rapid identification of cysteine-linked isoprenyl groups by metabolic labeling with [3H]farnesol and [3H]geranylgeraniol. *Methods Mol Biol 116*, 107-23.

(130) Benetka, W., Koranda, M., Maurer-Stroh, S., Pittner, F., and Eisenhaber, F. (2006) Farnesylation or geranylgeranylation? Efficient assays for testing protein prenylation in vitro and in vivo. *BMC Biochem 7*, 6.

(131) Liu, Z. H., Meray, R. K., Grammatopoulos, T. N., Fredenburg, R. A., Cookson, M. R., Liu, Y. C., Logan, T., and Lansbury, P. T. (2009) Membrane-associated farnesylated UCH-L1 promotes alpha-synuclein neurotoxicity and is a therapeutic target for Parkinson's disease. *Proc. Natl. Acad. Sci. U. S. A. 106*, 4635-4640.

(132) Baron, R., Fourcade, E., Lajoie-Mazenc, I., Allal, C., Couderc, B., Barbaras, R., Favre, G., Faye, J. C., and Pradines, A. (2000) 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 97*, 11626-31.

(133) Lin, H. P., Hsu, S. C., Wu, J. C., Sheen, I. J., Yan, B. S., and Syu, W. J. (1999) Localization of isoprenylated antigen of hepatitis delta virus by anti-farnesyl antibodies. *J Gen Virol 80 ( Pt 1)*, 91-6.

(134) Liu, X. H., Suh, D. Y., Call, J., and Prestwich, G. D. (2004) Antigenic prenylated peptide conjugates and polyclonal antibodies to detect protein prenylation. *Bioconjug Chem 15*, 270-7.

(135) Troutman, J. M., Roberts, M. J., Andres, D. A., and Spielmann, H. P. (2005) Tools to analyze protein farnesylation in cells. *Bioconjug Chem 16*, 1209-17.

(136) Nguyen, U. T., Guo, Z., Delon, C., Wu, Y., Deraeve, C., Franzel, B., Bon, R. S., Blankenfeldt, W., Goody, R. S., Waldmann, H., Wolters, D., and Alexandrov, K. (2009) Analysis of the eukaryotic prenylome by isoprenoid affinity tagging. *Nat Chem Biol 5*, 227-35.

(137) Nguyen, U. T. T., Goody, R. S., and Alexandrov, K. (2010) Understanding and Exploiting Protein Prenyltransferases. *Chembiochem 11*, 1194-1201.

(138) Chehade, K. A., Andres, D. A., Morimoto, H., and Spielmann, H. P. (2000) Design and synthesis of a transferable farnesyl pyrophosphate analogue to Ras by protein farnesyltransferase. *J Org Chem 65*, 3027-33.

(139) Onono, F. O., Morgan, M. A., Spielmann, H. P., Andres, D. A., Subramanian, T., Ganser, A., and Reuter, C. W. (2010) A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with Western blotting. *Mol Cell Proteomics 9*, 742-51.

(140) Saxon, E., and Bertozzi, C. R. (2000) Cell surface engineering by a modified Staudinger reaction. *Science 287*, 2007-2010.

(141) Moses, J. E., and Moorhouse, A. D. (2007) The growing applications of click chemistry. *Chem Soc Rev 36*, 1249-62.

(142) Chan, L. N., Hart, C., Guo, L., Nyberg, T., Davies, B. S., Fong, L. G., Young, S. G., Agnew, B. J., and Tamanoi, F. (2009) A novel approach to tag and identify geranylgeranylated proteins. *Electrophoresis 30*, 3598-606.

(143) Berry, A. F., Heal, W. P., Tarafder, A. K., Tolmachova, T., Baron, R. A., Seabra, M. C., and Tate, E. W. (2010) Rapid multilabel detection of geranylgeranylated proteins by using bioorthogonal ligation chemistry. *Chembiochem 11*, 771-3.

(144) Charron, G., Tsou, L. K., Maguire, W., Yount, J. S., and Hang, H. C. (2011) Alkynyl-farnesol reporters for detection of protein S-prenylation in cells. *Mol Biosyst 7*, 67-73.

(145) DeGraw, A. J., Palsuledesai, C., Ochocki, J. D., Dozier, J. K., Lenevich, S., Rashidian, M., and Distefano, M. D. (2010) Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. *Chem Biol Drug Des 76*, 460-71.

(146) Labadie, G. R., Viswanathan, R., and Poulter, C. D. (2007) Farnesyl diphosphate analogues with omega-bioorthogonal azide and alkyne functional groups for protein farnesyl transferase-catalyzed ligation reactions. *J Org Chem 72*, 9291-7.

(147) Reigard, S. A., Zahn, T. J., Haworth, K. B., Hicks, K. A., Fierke, C. A., and Gibbs, R. A. (2005) Interplay of isoprenoid and peptide substrate specificity in protein farnesyltransferase. *Biochemistry 44*, 11214-23.

(148) Krzysiak, A. J., Rawat, D. S., Scott, S. A., Pais, J. E., Handley, M., Harrison, M. L., Fierke, C. A., and Gibbs, R. A. (2007) Combinatorial modulation of protein prenylation. *ACS Chem Biol 2*, 385-9.

(149) Troutman, J. M., Subramanian, T., Andres, D. A., and Spielmann, H. P. (2007) Selective modification of CaaX peptides with ortho-substituted anilinogeranyl lipids by protein farnesyl transferase: Competitive substrates and potent inhibitors from a library of farnesyl diphosphate analogues. *Biochemistry 46*, 11310-11321.

(150) Terry, K. L., Casey, P. J., and Beese, L. S. (2006) Conversion of protein farnesyltransferase to a geranylgeranyltransferase. *Biochemistry 45*, 9746-55.

(151) Luckman, S. P., Hughes, D. E., Coxon, F. P., Russell, R. G. G., and Rogers, M. J. (1998) Nitrogen-containing bisphosphonates inhibit the mevalonate pathway and prevent post-translational prenylation of GTP-binding proteins, including Ras. *Journal of Bone and Mineral Research 13*, 581-589.

(152) Taylor, J. S., Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2003) Structure of mammalian protein geranylgeranyltransferase type-I. *Embo J 22*, 5963-74.

(153) Long, S. B., Casey, P. J., and Beese, L. S. (2000) The basis for K-Ras4B binding specificity to protein farnesyltransferase revealed by 2 A resolution ternary complex structures. *Structure 8*, 209-22.

(154) Hartman, H. L., Hicks, K. A., and Fierke, C. A. (2005) Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity. *Biochemistry 44*, 15314-24.

(155) Krzysiak, A. J., Aditya, A. V., Hougland, J. L., Fierke, C. A., and Gibbs, R. A. (2009) Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates. *Bioorganic & Medicinal Chemistry Letters 20*, 767-770.

(156) Hougland, J. L., Lamphear, C. L., Scott, S. A., Gibbs, R. A., and Fierke, C. A. (2009) Context-dependent substrate recognition by protein farnesyltransferase. *Biochemistry 48*, 1691-701.

(157) Goldstein, J. L., Brown, M. S., Stradley, S. J., Reiss, Y., and Gierasch, L. M. (1991) Nonfarnesylated tetrapeptide inhibitors of protein farnesyltransferase. *J Biol Chem 266*, 15575-8.

(158) Pompliano, D. L., Gomez, R. P., and Anthony, N. J. (1992) Intramolecular Fluorescence Enhancement - a Continuous Assay of Ras Farnesyl - Protein Transferase. *Journal of the American Chemical Society 114*, 7945-7946.

(159) Cassidy, P. B., Dolence, J. M., and Poulter, C. D. (1995) Continuous fluorescence assay for protein prenyltransferases. *Methods Enzymol 250*, 30-43.

(160) Boutin, J. A., Marande, W., Petit, L., Loynel, A., Desmet, C., Canet, E., and Fauchere, J. L. (1999) Investigation of S-farnesyl transferase substrate specificity with combinatorial tetrapeptide libraries. *Cell Signal 11*, 59-69.

(161) Krzysiak, A. J., Scott, S. A., Hicks, K. A., Fierke, C. A., and Gibbs, R. A. (2007) Evaluation of protein farnesyltransferase substrate specificity using synthetic peptide libraries. *Bioorg Med Chem Lett 17*, 5548-51.

(162) Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W.H. Freeman and Company, New York, NY.

(163) Pompliano, D. L., Schaber, M. D., Mosser, S. D., Omer, C. A., Shafer, J. A., and Gibbs, J. B. (1993) Isoprenoid diphosphate utilization by recombinant human farnesyl:protein transferase: interactive binding between substrates and a preferred kinetic pathway. *Biochemistry 32*, 8341-7.

(164) Zimmerman, K. K., Scholten, J. D., Huang, C. C., Fierke, C. A., and Hupe, D. J. (1998) High-level expression of rat farnesyl:protein transferase in Escherichia coli as a translationally coupled heterodimer. *Protein Expr Purif 14*, 395-402.

(165) Pais, J. E., Bowers, K. E., Stoddard, A. K., and Fierke, C. A. (2005) A continuous fluorescent assay for protein prenyltransferases measuring diphosphate release. *Anal Biochem 345*, 302-11.

(166) Furfine, E. S., Leban, J. J., Landavazo, A., Moomaw, J. F., and Casey, P. J. (1995) Protein farnesyltransferase: kinetics of farnesyl pyrophosphate binding and product release. *Biochemistry 34*, 6857-62.

(167) Tschantz, W. R., Furfine, E. S., and Casey, P. J. (1997) Substrate binding is required for release of product from mammalian protein farnesyltransferase. *J Biol Chem 272*, 9989-93.

(168) Troutman, J. M., Andres, D. A., and Spielmann, H. P. (2007) Protein farnesyl transferase target selectivity is dependent upon peptide stimulated product release. *Biochemistry 46*, 11299-309.

(169) Pompliano, D. L., Rands, E., Schaber, M. D., Mosser, S. D., Anthony, N. J., and Gibbs, J. B. (1992) Steady-state kinetic mechanism of Ras farnesyl:protein transferase. *Biochemistry 31*, 3800-7.

(170) Spence, R. A., Hightower, K. E., Terry, K. L., Beese, L. S., Fierke, C. A., and Casey, P. J. (2000) Conversion of Tyr361 beta to Leu in mammalian protein farnesyltransferase impairs product release but not substrate recognition. *Biochemistry 39*, 13651-9.

(171) Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K., and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res 30*, 235-8.

# CHAPTER II

# INVESTIGATION OF PROTEIN GERANYLGERANYLTRANSFERASE-I SUBSTRATE SPECIFICITY USING PEPTIDE LIBRARY STUDIES[1]

## Introduction

Protein geranylgeranyltransferase-I (GGTase-I) and protein farnesyltransferase (FTase) catalyze the attachment of a 20-carbon or 15-carbon geranylgeranyl or farnesyl group, respectively, to the C-terminus of a substrate protein (*1, 2*). These hydrophobic modifications help to target substrates to the cellular membranes and promote protein/protein interactions (*3, 4*) and are involved in a myriad of cellular processes and pathways. Therefore, inhibitors towards GGTase-I and FTase are of interest for the treatment of diseases such as cancer (*5*), Hutchinson-Gilford Progeria syndrome (*6*), and parasitic diseases like malaria (*7*). Some examples of prenylated proteins include the Ras and Rho GTPase superfamilies (*8, 9*) and the nuclear lamins (*10*) although the full complement of *in vivo* substrates of GGTase-I and FTase are unknown. Because of the impact of prenylation on a host of diseases, it is necessary to define the set of *in vivo* substrates of the prenyltransferase enzymes to allow for development of targeted

[1] A portion of the screening and kinetic analysis of peptide reactivity with GGTase-I that is discussed in this work was carried out by Hartman, H.A., Janik, L.S., Kelley, R.A., and Hendershot, J. The statistical analysis was carried out in collaboration with Watt, T.J. Dansyl-GCxxx peptides were provided by Aditya, A..V. and Gibbs, R.A. at Purdue University.

therapeutics against these prenylation-dependent pathways. Both FTase and GGTase-I were proposed to recognize a canonical "$Ca_1a_2X$" sequence at the C-terminus of protein substrates (*2, 8, 11*), where "C" is cysteine, "$a_1$" and "$a_2$" are aliphatic amino acids, and "X" is typically methionine, alanine, glutamate or serine for FTase and leucine or phenylalanine for GGTase-I (*12-15*). Although the $Ca_1a_2X$ sequence can act as model for some substrates, it is becoming increasingly evident that substrate recognition by FTase and GGTase-I is more complex than previously proposed and that the CaaX paradigm should be expanded (*16*). *In vitro,* short peptides can be utilized as prenyltransferase substrates with similar activity and affinity to full-length proteins (*14, 15, 17-19*), allowing for the investigation and expansion of the CaaX paradigm. Previous studies have shown that FTase and GGTase-I selectivity arises from relative peptide reactivity rather than binding affinity (*20*). Furthermore, FTase and GGTase show inverse correlation with substrate X group hydrophobicity (*20*); FTase tends to be less reactive as the hydrophobicity of X increases (preferring residues such as S and Q), whereas GGTase-I is more reactive with increasingly nonpolar X residues like I and L (*20*). A group of X residues with optimum hydrophobicity can act as dual substrates for both FTase and GGTase-I (i.e., M and F) (*20*). Additionally, *in vivo* many proteins contain a polybasic region upstream of the CaaX sequence, and peptide studies have shown that in FTase, a polybasic region increases peptide binding affinity but decreases reactivity, while in GGTase-I there appears to be no effect (*21*). Therefore, dual substrates of FTase or GGTase-I could also arise because of similar reactivity (*21*). Further studies have also characterized the relationship of the X residue to the $a_2$ residue

in substrates for FTase, finding that the FTase recognizes substrates in a context-dependent manner (*22*).

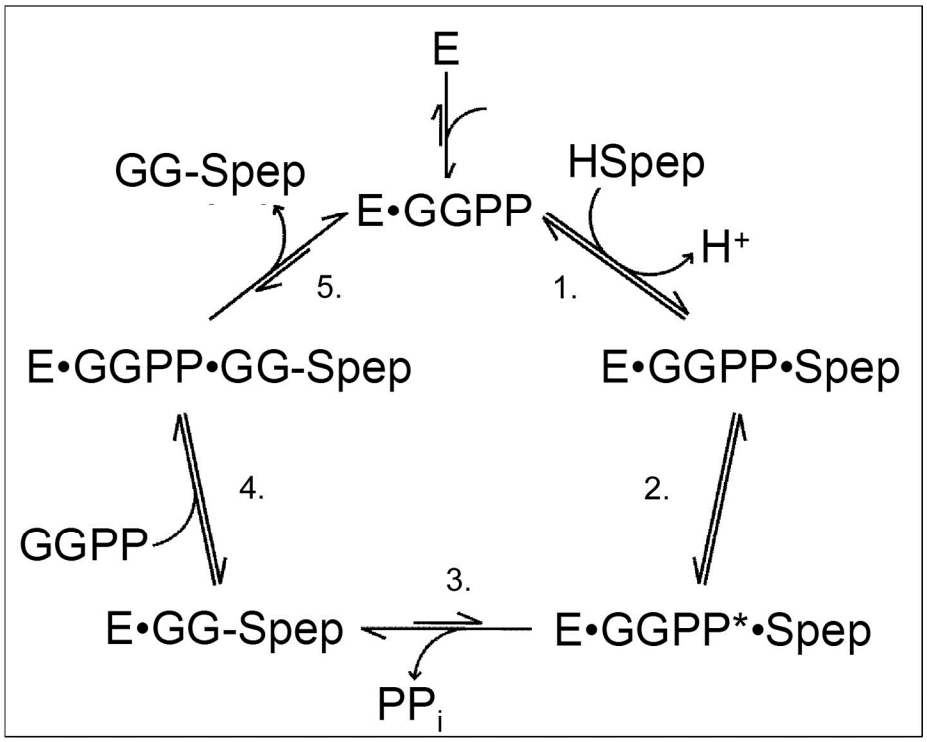


**Figure 2.1. Preferred kinetic mechanism of GGTase-I.** This is a simplified diagram of the kinetic pathway. The GGTase-I enzyme binds the lipid donor GGPP and the peptide or protein substrate. In order for the reaction to occur, the prenyl chain of the GGPP must rotate to move the C1 of GGPP closer to the thiolate of the peptide. Another substrate binds to help facilitate product release. MTO peptides are able to proceed through steps 1-5 whereas STO substrates only go through chemistry, steps 1-3. The E•GGPP and E•GG-Spep complexes can also slowly dissociate; however, for clarity these pathways are not shown. See Chapter 1 Figure 1.9 for a more detailed scheme.

Both of the CaaX prenyltransferases have been proposed to proceed through the kinetic scheme shown in Figure 2.1. The mechanism is thought to be kinetically ordered with FPP and GGPP binding to FTase or GGTase-I before the peptide or protein (*23-25*), consistent with the binding affinities of FPP and GGPP being in the low nanomolar range (*26, 27*). Additionally, for FTase, formation of the E•FPP complex enhances the affinity for the peptide (*19*). Once the ternary complex is formed, the C1 of FPP and GGPP are about 7Å away from the catalytic $Zn^{2+}$, so the first and second isoprene units rotate such that the C1 is near the sulfur of the peptide or protein substrate (*28-31*). Then, the reaction occurs, forming the prenylated product and diphosphate, and the diphosphate substrate is rapidly released (*32*). For the prenyltransferase reactions under subsaturating substrate and initial rate conditions, the dissociation of the diphosphate product is the first irreversible step (*24, 32*), so $k_{cat}/K_M^{peptide}$ includes the rate constants from peptide binding to E•GGPP (or E•FPP) through the conformational change and the chemistry step, but does not include the rate constants for the release of the farnesylated peptide product (*24, 28, 31-33*). Conversely, $k_{cat}$ conditions (saturating substrates) include the rate constants after the formation of E•GGPP•peptide through product release and product release is frequently the rate limiting step (*32, 34*). Lastly, the rate of dissociation of the prenylated peptide is enhanced by the binding of GGPP (or FPP) to the product•enzyme complex (*20, 34, 35*).

Recently, large peptides libraries have been used to determine elements of molecular recognition for FTase, using statistical analysis to determine FTase sequence preferences. Two libraries were tested: a library based on sequences derived from the human proteome and a targeted library that was designed based on results from the

original library. FTase was found to catalyze prenylation of two pools of substrates: 1) multiple turnover substrates (MTO) and 2) single turnover-only substrates (STO), peptides that are farnesylated but for which multiple turnovers do not occur (*16, 20*). In short, MTO substrates are able to proceed through the whole kinetic cycle (steps 1-5 in Figure 2.1), but modified STO substrates are not released from the enzyme, only progressing through the chemical step (steps 1-3 in Figure 2.1). The MTO pool of substrates followed the $Ca_1a_2X$ paradigm more closely, while STO substrates had a wider range in sequences. Additionally, GCxxS and GCxxL libraries (where x is any amino acid) were tested with FTase by the Gibbs lab at Purdue University, finding that FTase can modify many proteins in both libraries (*36, 37*), contrary to the prediction that a leucine at the terminal amino acid confers specificity for GGTase-I.

Less information is available to describe how GGTase-I recognizes its substrates. Peptide libraries studies are a useful approach, since a large number of sequences can be screened quickly and efficiently using a 96-well plate assay, using the wild-type enzyme and the natural lipid donor, GGPP. Statistical analysis can then be applied to indentify elements important for substrate recognition for GGTase-I. In this chapter, I design a diverse library of peptides and measure the reactivity of GGTase-I with these peptides to evaluate the determinants of substrate recognition. These studies provide useful information for predicting potential *in vivo* GGTase-I substrates, and analyzing the potential for dual prenylation of substrates. This work will also aid in identifying the effects of GGTase-I and FTase inhibitors on prenylation of *in vivo* proteins and metabolic pathways.

## Experimental Procedures

*GGTase-I expression and purification*

The *E. coli* expression vector for rat GGTase-I in the pET23a-GGPT vector was constructed by previous Fierke lab members as described in (*38*). The pET23a-GGPT plasmid was transformed into BL21(DE3) *E. coli* cells and grown overnight at 37ºC on LB/agar plates with 100 μg/mL ampicillin. To improve and test GGTase-I expression conditions, sterile 24-well plates were prepared with 0.9 mL per well of LB containing 10% glycerol, 100 μg/mL ampicillin, isopropyl β-D-1-thiogalactopyranoside (IPTG, 0, 50, 100, or 400 μM) or zinc (0, 100 μM $ZnSO_4$ or 100 μM $ZnSO_4$/67 μM $NH_4Cl$). Additionally, protein production was measured using autoinduction media (Overnight Express Autoinduction System, Novagen) with varying zinc. An inoculating loop was used to pick one colony from the LB/agar plate with ampicillin, and then was dipped into three media conditions to inoculate the cultures. The cultures were grown for 24 hours at 28°C shaking at 350 rpm. A sample of 0.75 mL was removed from each culture and transferred to a microcentrifuge tube. The cells were harvested by centrifugation at 10,000xg for 5 minutes and lysed by adding 113 μL of B-per Reagent (Thermo) spiked with 0.2 μL benzonase (Sigma). The soluble and insoluble fractions were separated by centrifugation at 15,000xg for 5 minutes, and the insoluble fraction was dissolved in 113 μL 6M urea. Undissolved material was removed from the insoluble fraction by centrifugation at 15,000xg for 5 minutes. To visualize protein expression, both soluble and insoluble fractions were fractionated on an SDS-PAGE gel and protein bands were visualized using Coomassie dye. As the autoinduction media plus $ZnCl_2$ showed the best expression, for large scale protein purification, the pET23a-GGPT plasmid was

transformed into BL21(DE3) *E. coli* cells like before, and one colony was inoculated into 5 mL of 2xYT media with 100 μg/mL ampicillin. This culture was grown for 4 hours at 37ºC shaking at 225 rpm until it was cloudy. Then the 5 mL starter culture was added to 1L of autoinduction media and 100 μg/mL ampicillin and incubated at 28ºC for 24 hours, shaking at 185 rpm. Cells were harvested by centrifugation (15 minutes, 4°C, 6,000 rpm) and resuspended in 20 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) pH 7.8, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 10 ng/mL phenylmethanesulfonyl fluoride (PMSF), and 1 ng/mL $N_{\alpha}$-*p*-Tosyl-L-arginine methyl ester hydrochloride (TAME) and frozen at -80ºC. For purification, the cells were thawed on ice, lysed using a microfluidizer (Microfluidics, Newton, MA), and the nucleic acids were precipitated by addition of 1/10 volume of 10% w/v streptomycin sulfate followed by centrifugation at 12,000 rpm for 45 min at 4°C. The GGTase-I protein was purified from the supernatant of this cleared lysate (generated from a 1L original culture) on approximately 100-150 mL DEAE-cellulose column (Whatman) using a 500 mL gradient of 0.1 to 0.5 M NaCl in 50 mM HEPES pH 7.8, 10 μM $ZnCl_2$, and 2 mM TCEP (HTZ buffer). After analyzing the fractions from the column for GGTase-I using SDS-PAGE, the fractions containing GGTase-I were pooled and concentrated to 10 mL using Amicon Ultra centrifugal filter devices with a 10,000-MW cutoff filter (Millipore) and dialyzed against 2x4L of HTZ buffer. Next, the protein was further fractionated by loading the entire 10 mL of concentrated partially purified GGTase-I on a 20 mL HQ POROS column (Applied Biosystems) using a gradient 0 to 0.5 M NaCl in 750 mL HTZ buffer. Fractions containing GGTase-I were pooled and concentrated to using Amicon Ultra centrifugal

filter devices with a 10,000-MW cutoff filter and dialyzed against 50 mM HEPES pH 7.8 and 2 mM TCEP. Aliquots of 250-500 μM GGTase-I (5-20 μL) were flash frozen and stored at -80ºC. An active site titration was performed to determine the protein concentration by monitoring fluorescence resonance energy transfer (FRET, 280 nm excitation/496 nm emission) of the dansyl-GCVLL peptide and GGTase-I when the peptide is titrated into a stock of GGTase-I alone or GGTase-I/3-aza-GGPP (an inactive GGPP analog) complex.

*Peptide libraries*

To increase the diversity of amino acids at the $a_1$, $a_2$, and X positions of the lab's peptide libraries, a "diversity set" of peptides was added to the overall peptide library. In collaboration with Dr. Terry Watt, human protein sequences from the Swiss-Prot database were downloaded that contained a cysteine four amino acids from the C-terminus and all redundant and viral sequences were removed. Eighty-three additional CaaX sequences were chosen from this unique human sequence list such that each of the 20 amino acids would be represented in a minimum of 10 peptides at each of the $a_1$, $a_2$, and X positions in the final combined library. Peptides of the form dns-TKCxxx (where x is any amino acid, and dns is the dansyl fluorophore) were purchased from Sigma Genosys (The Woodland, TX) in the PEPscreen™ format , to form the "diversity" library and are >70% pure. The peptides were dissolved in 10% DMSO in EtOH and stored at -80ºC and peptide concentration was calculated by measuring free thiols from the absorbance change after reaction with 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) (*39*) using an extinction coefficient of 14,150 $M^{-1}cm^{-1}$ at 412 nm. Additionally, peptides of the

| 1. Original Library | | | | | | 2. Targeted Library | | 3. Diversity Library | | | 4. "G" Library |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CKAA | CTVF | CIVK | CGIM | CGCR | CVTT | CPLA | CRLR | CWLA | CMMK | CAWV | CIHF |
| CGCA | CSKG | CHDL | CNIM | CTCR | CNVT | CSLA | CTLR | CTQA | CWNK | CLYV | CSHF |
| CEDA | CFLG | CSDL | CSIM | CVIR | CVVT | CVPA | CDTR | CEVA | CVYK | CARY | CCIF |
| CQDA | CPLG | CTDL | CVIM | CTKR | CACV | CAAF | CFTR | CYYA | CFFL | | CFIF |
| CQEA | CSLG | CSFL | CKKM | CAVR | CCFV | CFAF | CLTR | CRGC | CDGL | | CKIF |
| CSEA | CFMG | CGGL | CSKM | CPAS | CTHV | CFIF | CGVR | CRKC | CWGL | | CVIF |
| CGIA | CSSG | CIHL | CLLM | CGCS | CVIV | CDLF | CTFS | CDRC | CWKL | | CSKF |
| CVIA | CATG | CLHL | CNLM | CKCS | CTKV | CFTF | CFPS | CGSC | CRLL | | CDLF |
| CKKA | CCTG | CAIL | CSLM | CTDS | CWQV | CFVF | CRPS | CPSC | CVPL | | CGLF |
| CTKA | CCAH | CCIL | CTLM | CAES | CPFW | CLAG | CEVS | CMTC | CLSL | | CQPF |
| CCLA | CVAH | CGIL | CVQM | CGHS | CIGW | CNFG | CVLT | CTVC | CAYL | | CGQF |
| CTLA | CGCH | CIIL | CLVM | CVHS | CRLW | CHLG | CKIV | CQED | CGEM | | CTVF |
| CVLA | CLEH | CLIL | CYVM | CKIS | CGTW | CPVG | CTIV | CLFD | CWEM | | CGCI |
| CVLA | CCHH | CNIL | CNFN | CRIS | CEVW | CHTH | CGLV | CEGD | CKWM | | CHCI |
| CCNA | CGKH | CPIL | CGGN | CSIS | CQVW | CFII | CKLW | CALD | CIGN | | CPSI |
| CCTA | CCLH | CSIL | CVIN | CVIS | CRVW | CVII | CAAY | CPRD | CYQN | | CNTI |
| CQTA | CVLH | CVIL | CRKN | CAKS | CSVW | CHLK | CRLY | CSSD | CQWN | | CHVI |
| CIVA | CQRH | CEKL | CKLN | CQKS | CWDY | CSLK | CYTY | CRGE | CWRP | | CCWI |
| CNWA | CACI | CFKL | CLLN | CFLS | CDIY | CDTK | CAVY | CCIE | CAGQ | | CGCL |
| CSAC | CGCI | CALL | CCSN | CQLS | CVIY | CETK | CYVY | CFLE | CWLQ | | CTIL |
| CTGC | CKFI | CCLL | CYSN | CVLS | CTKY | CDVK | | CLME | CAMQ | | CQLL |
| CVIC | CIHI | CILL | CVVN | CTMS | CGLY | CKGL | | CENE | CNNQ | | CRLL |
| CSLC | CGII | CSLL | CSIP | CFPS | CVLY | CMIL | | CKSE | CFDR | | CVLL |
| CGRC | CIII | CTLL | CKKP | CFSS | CNSY | CPLL | | CYSE | CMKR | | CMPL |
| CPED | CSII | CVLL | CLKP | CSSS | CSVY | CPVL | | CGTE | CMVR | | CVPL |
| CNHD | CTII | CLML | CLMP | CVSS | CTVY | CCLM | | CAVE | CRLS | | CHSL |
| CVID | CVII | CLNL | CAQP | CLTS | CIYY | CHVM | | CKYE | CDMS | | CVSL |
| CYPD | CNKI | CQNL | CCRP | CNTS | | CYPN | | CHDF | CKNS | | CCVL |
| CSVD | CQKI | CGQL | CSGQ | CSVS | | CFVN | | CMGF | CLPS | | CLVL |
| CIHF | CTKI | CSQL | CVIQ | CEYS | | CRVN | | CADG | CMPS | | CVIM |
| CSHF | CQNI | CISL | CVKQ | CLIT | | CLIP | | CPMG | CQPS | | CKWM |
| CCIF | CSPI | CVSL | CNLQ | CVIT | | CHLP | | CCNG | CNRS | | CCFV |
| CKIF | CGQI | CTTL | CTLQ | CCKT | | CGVP | | CWVG | CMSS | | CLLV |
| CVIF | CNTI | CAVL | CVLQ | CLKT | | CHVP | | CGYG | CLWS | | CPPV |
| CSKF | CHVI | CCVL | CKQQ | CALT | | CEIQ | | CRYG | CQWS | | CTSV |
| CTKF | CSVI | CNVL | CITQ | CHLT | | CHIQ | | CMKI | CLET | | |
| CGLF | CTAK | CSVL | CQTQ | CLLT | | CDLR | | CCWI | CGNT | | |
| CLLF | CEEK | CSFM | CVTQ | CSPT | | CFLR | | CHWI | CRRT | | |
| CNLF | CWHK | CVFM | CAVQ | CSRT | | CMLR | | CKWI | CTMV | | |
| CGQF | CVKK | CLHM | CACR | CKTT | | CPLR | | CRWI | CGNV | | |

**Table 2.1. List of total peptide libraries.** Libraries 1, 2, and 3 are dns-TKCxxx peptides and library 4 is dns-GCxxx peptides, where x is any amino acid. 402 peptides are in the total library. Some of the Cxxx sequences in library 4 are identical those in libraries 1-3.

form dns-GCxxx (G library) were synthesized by Animesh Aditya at Purdue University as described in (*37*). A full list of the peptides tested is shown in Table 2.1. The "Original" library (Library 1) and the "Targeted" library (Library 2) were previously tested with FTase (*16*), and the "Diversity" library (Library 3) and the "G" library (Library 4) are new.

*Multiple turnover assay screen*

Peptides were screened for multiple turnover (MTO) activity with GGTase. The "Original" and "Targeted" libraries of dns-TKCxxx peptides were already screened for MTO activity catalyzed by GGTase-I by previous lab members Dr. Heather Hartman and Dr. Rebekah Kelley, but the remainder of the library was screened as follows: 3 μM dns-peptide, 20 nM GGTase-I, and 10 μM geranylgeranyl diphosphate (GGPP) were incubated together in 50 mM N-(2-Hydroxyethyl)piperazine-N'-(2-hydroxypropanesulfonic acid) sodium salt (HEPPSO) pH 7.8 and 5 mM TCEP at 25ºC. Peptides were incubated in buffer for at least 20 minutes to reduce any disulfide bonds to free thiols before the reaction was initiated. Prenylation of a dansyl peptide leads to an enhancement in fluorescence that was monitored continuously ($\lambda_{ex}$=340 nm, $\lambda_{em}$=520 nm, (*40*)) in a 96-well plate (Costar, non-binding surface, Corning, Corning, NY) using a POLARstar Galaxy plate reader (BMG Labtechnologies, Durham, NC) (Figure 2.2). Any peptide that showed an increase in fluorescence compared to a no enzyme control background after 12 hours at 25ºC was considered a MTO substrate. Using the assumptions that [peptide] < $K_M$, [GGPP] is saturating, and 10% product is formed the lower limit of this assay is approximately 400 $M^{-1}s^{-1}$, similar to other work (*16*).

(Reaction conditions in Drs. Hartman and Kelley's work included 50 mM Tris-HCl pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 10 μM $ZnCl_2$, 10 μM GGPP, and 10 μM dns-TKCxxx peptide. Reactions were initiated by addition of 200 nM GGTase-I and monitored using the fluorescence assay described above. Peptides that showed a ≥15% change in fluorescence after 1 hour were designated a MTO substrate. Using the same assumptions as above, the lower limit of this assay is 200 $M^{-1}s^{-1}$). Therefore, under both conditions the lower limit of the assay is 200-400 $M^{-1}s^{-1}$.

*Single turnover assay screen*

Peptides that were not prenylated under MTO conditions were screened for STO reactivity using either a radioactive or fluorescent assay. For the radioactive assay, 1 μM GGTase-I, 0.8 μM $^3$[H]-GGPP, and 3 μM dns-TKCxxx or dns-GCxxx peptide were incubated for 30 minutes in 50 mM HEPPSO pH 7.8 and 5 mM TCEP at 25ºC before the reaction was quenched by the addition of an equal volume of 80:20 isopropanol:acetic acid. The reactions were spotted on a silica TLC plate, fractionated using 8:1:1 isopropanol:ammonium hydroxide: water solvent, and the radioactive products on the plates were visualized by autoradiography. Peptides were considered a STO substrate if at least 10% of the $^3$[H]-GGPP reacts with the peptide after 30 minutes (by comparing to a sample of 1/10 $^3$[H]-GGPP reacted with GGTase-I and dns-TKCVLL). The lower limit of the peptide reactivity in this assay is estimated to be ≥ 0.0009 $s^{-1}$ and is a similar cutoff as in other work with FTase (*16*).

For the STO fluorescent assay, 384-well plates were used to screen for STO activity with GGTase-I using the POLARStar plate reader. Reactions contained 0, 1 or 2

μM GGTase-I, 0.8 μM GGPP, and 3 μM dns-TKCxxx or dns-GCxxx peptide in 50 mM HEPPSO and 5 mM TCEP at 25ºC. Peptides were incubating in buffer for at least 20 minutes before the reaction was initiated in order to reduce any disulfide bonds. If a significant increase in fluorescence over the no enzyme control was observed for both reactions containing GGTase-I (1 or 2 μM) after 30 minutes - 1 hour, the peptide was designated a STO substrate. The lower limit of this assay is also $\geq 0.0009\ s^{-1}$.

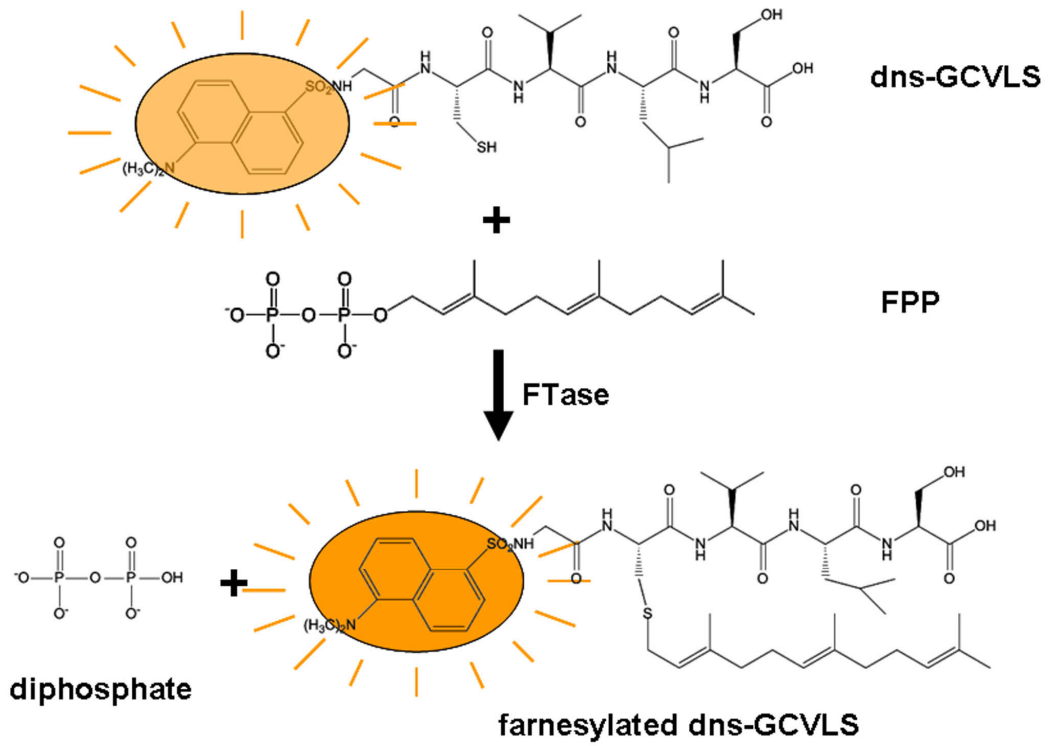


**Figure 2.2. Continuous fluorescence assay to monitor the prenyltransferase reaction.** Short dansylated peptides can be used as prenyltransferase substrates, and the dansyl group increases in fluorescence upon modification due to the enhanced hydrophobic environment of the dansyl group in the prenylated product (*18, 40*). The farnesyltransferase reaction is used here as an example.

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

In collaboration with Dr. Terry Watt (Xavier University of Louisiana), GGTase-I substrate preferences were analyzed using statistical analysis. A Chi-square test was used to determine the significance of canonical vs. non-canonical sequences in the MTO and STO pools of GGTase-I substrates at the $a_1$, $a_2$ and X positions. Next, hypergeometric square model analysis (Equation 1) was used to determine if any amino acid sequences were over-represented or under-represented compared to the overall library at the $a_1$, $a_2$, and X positions of the peptide substrates. *p* is the probability that a particular amino acid occurring at a certain frequency at a position in the substrate pool due to random chance, *N* is the size of the total library, *P* is the number of peptides in the total library with a given amino acid at the position ($a_1$, $a_2$, X), *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 (*16*).

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

*Steady state kinetics*

The steady state parameter $k_{cat}/K_M^{peptide}$ was measured for a subset of active GGTase-I MTO substrates using the fluorescent assay described above, and rotation student Jenna Hendershot aided in measuring the reactivity of several of these peptides. Reactions contained 0.2 – 40 μM dns-peptide, 20 -100 nM GGTase-I, and 10 μM GGPP in 50 mM HEPPSO pH 7.8 and 5 mM TCEP at 25ºC. The concentration of dns-peptide was maintained at a concentration ≥ 5 times the concentration of GGTase-I. All curve

fitting was performed using Graphpad Prism (Graphpad Software, San Diego, CA). Reactions were monitored such that both an initial slope correlating to the initial velocity and the reaction endpoint could be observed. To convert fluorescence units ("fl") to product concentration, the change in fluorescence to reach the reaction endpoint was divided by peptide concentration. Several peptide concentrations were used to calculate an average amplitude conversion factor ($Amp_{conv}$). The initial slope, in fl/s, was converted to velocity (μM product/s) as in equation 2, where $V$ is velocity (μM/s), $R$ is the initial slope (fl/s), and $Amp_{conv}$ is the conversion factor (fl/μM).

$$V = \frac{R}{Amp_{conv}} \tag{2}$$

To obtain values for $k_{cat}/K_M^{peptide}$, the Michaelis-Menton equation (Equation 3).was fit to the dependence of the initial velocity (V/[E]) on the substrate concentration. In some cases the values for $k_{cat}$ and $K_M$ were also obtained from these fits.

$$\frac{V}{[E]} = \frac{k_{cat}[peptide]}{K_M + [peptide]} \tag{3}$$

## Results

*Improvement of GGTase-I expression conditions*

The rat GGTase-I (and FTase) genes are encoded in the pET23a vector, where the gene is behind a strong T7 promoter (*41*). Traditionally, the two subunits of the prenyltransferase proteins were expressed recombinantly by growth of BL21(DE3) *E. coli* cells containing the expression plasmid to mid-log phase, and then inducing by addition of IPTG, a lactose analog. The IPTG works to relieve the lac repressor on the lac

operon, allowing the chromosomal T7 polymerase under the lac operon to be synthesized and the prenyltransferase genes to be transcribed and translated (*41, 42*). Using this method yielded 15 mg of GGTase-I per 1 L of culture, as reported in (*38*). Overall yield from this new purification protocol was 40 mg of GGTase-I/L of culture, which is a 2.7-fold improvement. Recently, autoinduction media has become a convenient method for protein expression, and in some cases even enhances expression. This method works by providing the cells with media containing optimal levels of glucose, glycerol, and lactose, such that the cells metabolize glucose and glycerol initially for optimal growth with lactose metabolized last. When the cells utilize the lactose, the expression is induced automatically ("autoinduction") because the lac repressor is relieved (*43, 44*). Similarly, rather than inducing protein expression by addition of a high concentration of IPTG to a culture of cells in late log phase, growing the cells in media that contain a low level of IPTG throughout the growth helps to slowly express proteins. Both of these types of expression media are convenient, as they do not require monitoring of cellular density for induction.

Different media conditions were tested to improve GGTase-I expression. Five types of base media were used: either 0, 50, 100 or 400 μM ITPG in LB with 10% glycerol and 100 μg/mL ampicillin or autoinduction media with 100 μg/mL ampicillin. These media were tested alone, supplemented with 100 μM $ZnSO_4$, or supplemented with 100 μM $ZnSO_4$/67 μM $NH_4Cl$. The soluble fraction of the lysates of cell grown in each of these expression conditions was separated using SDS-PAGE as shown in Figure 2.3. On this gel, the 48 kDa α and 43 kDa β subunits of GGTase-I (*45*) run between 40 and 50 kDa bands of the ladder (Lanes *A, B,* and *C*). ITPG concentrations of 50 or 100 μM in

LB/glycerol (lanes 2, 3, 7, 8, 12, and 13) and autoinduction media (lanes 5, 10, and 15) appeared to show improved expression of GGTase-I with $ZnSO_4$ supplemented cultures having the best expression overall (lanes 7, 8, and 10). Overall, the cleanest expression appears to be lane 10, where autoinduction media was supplemented with $ZnSO_4$. After this, a large scale growth in autoinduction media was used for GGTase-I purification, using the procedure described in the Experimental Procedures. Overall, this new method for expressing GGTase-I helped improved the total yield of GGTase-I protein 2.7-fold.

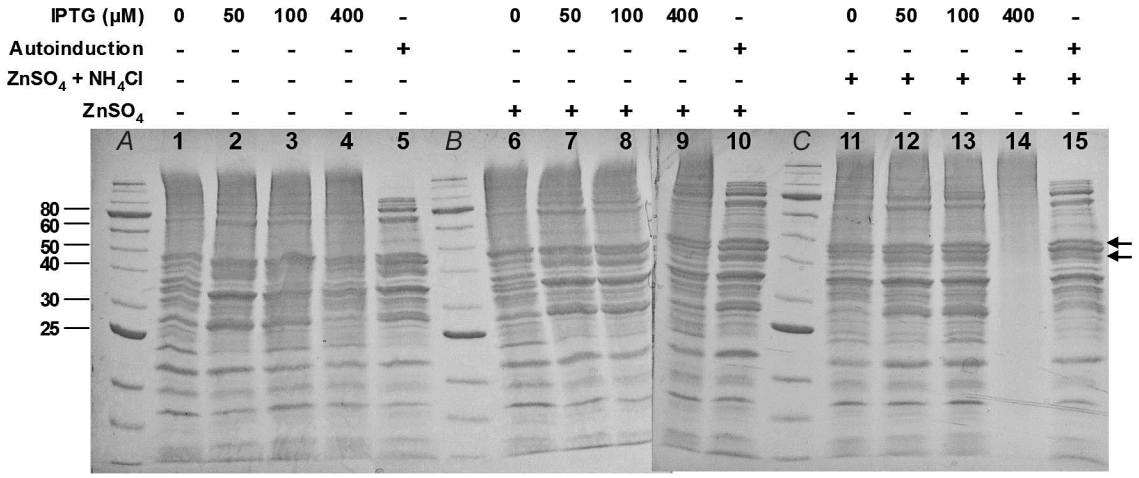


**Figure 2.3. Optimization of growth conditions for GGTase-I expression.**
Cultures were grown in 24-well plates in 0.9 mL of either LB media with 10% glycerol and 100 μg/mL ampicillin supplemented with IPTG or autoinduction media with 100 μg/mL ampicillin. Additionally, cultures were supplemented with 100 μM $ZnSO_4$ or 100 μM $ZnSO_4$/67 μM $NH_4Cl$. Cells were grown for 24 hours at 28ºC shaking at 350 rpm and harvested, lysed with B-per Reagent, clarified by centrifugation, and the soluble fractions were fractionated on an SDS-PAGE gel. Lanes *A, B* and *C* are molecular mass markers, and the two subunits of GGTase-I run between 40 and 50 kDa and are indicated by the arrows. Cells grown in media containing 50 or 100 μM IPTG in LB/glycerol (lanes 2, 3, 7, 8, 12, and 13) and autoinduction media (lanes 5, 10, and 15) appeared to have improved expression, with $ZnSO_4$ supplemented cultures having the best expression overall (lanes 7, 8, and 10). The culture in Lane 14 did not grow properly. The darkness/contrast has been adjusted in this figure for clearer visualization of the bands over the gel.

*Peptide substrates of GGTase-I*

Overall, the reactivity of purified GGTase-I for catalyzing geranylgeranylation of 402 peptides was tested under both multiple turnover (MTO) and single turnover (STO) conditions (Table 2.2). In this analysis, GGTase-I catalyzes prenylation of 111 or 28% of the peptides under MTO substrates conditions ([S] > [E]). Additionally, under STO conditions ([E] > [GGPP]), GGTase-I catalyzes geranylgeranylation of an additional 178 peptides, or 44% of the library. STO activity indicates that the peptides bind to GGTase-I and are prenylated, but GGTase-I does not react with a second peptide, presumably due to slow product dissociation. Therefore, GGTase-I can recognize and catalyze geranylgeranylation to varying extents of 72% of the peptides in the library. Table 2.3 lists the peptides in the MTO and STO substrate pools.

| Pool | Number of Peptides |
|---|---|
| Library | 402 |
| Multiple Turnover (MTO) | 111 (28%) |
| Single Turnover (STO) | 178 (44%) |

**Table 2.2. Substrates of GGTase-I.** 402 peptides were screened for activity with GGTase-I (a mix of dns-TKCxxx and dns-GCxxx peptides, where x is any amino acid). GGTase-I catalyzes prenylation of 72% of the peptides in this library.

*Sequence analysis of MTO and STO GGTase-I substrates*

With the help of Dr. Terry Watt, we analyzed overall patterns in the peptide sequences for the MTO, STO, and non-substrate (NON) pools of GGTase-I substrates using a Chi-squared test. The peptides were separated into canonical or non-canonical $Ca_1a_2X$ sequences and the sequence preferences were analyzed at the $a_2$ and X positions. A canonical $a_2$ residue for GGTase-I was defined as I or L (with the other 18 amino acids being non-canonical) and canonical GGTase-I residues at the X position were defined as

| Multiple Turnover | | | Single Turnover | | | | Non-substrates | | |
|---|---|---|---|---|---|---|---|---|---|
| CWLA | CMIL | CWLQ | CKAA | CPMG | CSFM | CSIS | CGCA | CIVK | CRIS |
| CVIC | CNIL | CLIT | CGIA | CSSG | CLHM | CVIS | CEDA | CHDL | CAKS |
| CTVC | CPIL | CVLT | CVIA | CATG | CKKM | CQKS | CQDA | CTDL | CFLS |
| CFAF | CSIL | CCFV | CPLA | CCTG | CSKM | CQLS | CQEA | CDGL | CDMS |
| CCIF | **CTIL** | **CCFV** | CTLA | CWVG | CLVM | CRLS | CSEA | CGGL | CFPS |
| **CCIF** | CVIL | CKIV | CVLA | CGYG | CYVM | CVLS | CKKA | CIHL | CFSS |
| CFIF | CWKL | CTIV | CVPA | CRYG | CKWM | CTMS | CTKA | CEKL | CSSS |
| **CFIF** | CALL | CVIV | CTQA | CCAH | **CKWM** | CKNS | CCLA | CFKL | CEYS |
| CKIF | CCLL | CGLV | CCTA | CVAH | CIGN | CFPS | CSLA | CQNL | CCKT |
| **CKIF** | CILL | **CLLV** | CEVA | CGCH | CVIN | CLPS | CVLA | CGQL | CLKT |
| CVIF | CPLL | CTMV | CIVA | CCHH | CKLN | CMPS | CCNA | CISL | CSPT |
| **CVIF** | **CQLL** | CAWV | CYYA | CGKH | CLLN | CQPS | CQTA | CVSL | CSRT |
| CTKF | CRLL | CLYV | CSAC | CCLH | CYPN | CRPS | CNWA | CGEM | CKTT |
| CDLF | **CRLL** | CKLW | CTGC | CVLH | CYQN | CNRS | CRGC | CVQM | CACV |
| **CDLF** | CSLL | CRLW | CRKC | CHTH | CRVN | CMSS | CDRC | CNFN | CTHV |
| CGLF | CTLL | CVIY | CSLC | CACI | CLIP | CVSS | CGRC | CGGN | CTKV |
| **CGLF** | CVLL | CGLY | CPSC | **CGCI** | CSIP | CLTS | CGSC | CRKN | CGNV |
| CNLF | **CVLL** | CRLY | CPED | **CHCI** | CLKP | CNTS | CMTC | CCSN | CWQV |
| **CQPF** | CLML | CVLY | CQED | CKFI | CHLP | CEVS | CVID | CYSN | CIGW |
| CFTF | **CMPL** | CAVY | CLFD | CMKI | CAQP | CSVS | CALD | CVVN | CGTW |
| CFVF | CVPL | CYVY | CEGD | CNKI | CCRP | CLWS | CSVD | CQWN | CEVW |
| CTVF | **CVPL** | | CNHD | CQNI | CWRP | CQWS | CCIE | CKKP | CQVW |
| **CTVF** | **CVSL** | | CYPD | CGQI | CGVP | CLET | CENE | CLMP | CRVW |
| CFII | CAVL | | CPRD | **CPSI** | CHVP | CVIT | CAVE | CAGQ | CSVW |
| CGII | CCVL | | CSSD | CNTI | CVIQ | CALT | CAAF | CSGQ | CTKY |
| CIII | **CCVL** | | CRGE | CCWI | CVKQ | CHLT | CSKF | CEIQ | CARY |
| CSII | **CLVL** | | CFLE | CRWI | CNLQ | CLLT | CLAG | CAMQ | |
| CTII | CNVL | | CLME | CTAK | CTLQ | CGNT | CADG | CNNQ | |
| CVII | CPVL | | CKSE | CSLK | CVLQ | CRRT | CNFG | CKQQ | |
| CVII | CSVL | | CYSE | CMMK | CITQ | CVTT | CSKG | CVTQ | |
| **CNTI** | CAYL | | CGTE | CWNK | CQTQ | CNVT | CPLG | CAVQ | |
| CHVI | CVFM | | CKYE | CDTK | CTCR | CVVT | CFMG | CACR | |
| **CHVI** | CGIM | | CHDF | CETK | CVIR | **CPPV** | CCNG | CGCR | |
| CSVI | CNIM | | CMGF | CDVK | CMKR | **CTSV** | CPVG | CFDR | |
| **CCWI** | CSIM | | CIHF | CVYK | CTKR | CPFW | CLEH | CDLR | |
| CHWI | CVIM | | **CIHF** | **CGCL** | CPLR | CAAY | CQRH | CFLR | |
| CKWI | **CVIM** | | CSHF | CSDL | CRLR | CWDY | CGCI | CMLR | |
| CFFL | CCLM | | **CSHF** | CKGL | CAVR | CDIY | CIHI | CTLR | |
| CSFL | CLLM | | **CSKF** | CLHL | CGVR | CNSY | CQKI | CDTR | |
| CWGL | CNLM | | CLLF | CLNL | CMVR | CYTY | CTKI | CFTR | |
| CAIL | CSLM | | CGQF | CSQL | CTDS | CSVY | CSPI | CLTR | |
| CCIL | CTLM | | **CGQF** | **CHSL** | CAES | CTVY | CEEK | CPAS | |
| CGIL | CHVM | | CFLG | CLSL | CTFS | CIYY | CWHK | CGCS | |
| CIIL | CFVN | | CHLG | CTTL | CGHS | | CVKK | CKCS | |
| CLIL | CHIQ | | CSLG | CWEM | CVHS | | CHLK | CKIS | |

**Table 2.3. MTO and STO substrates of GGTase-I.** Substrates are a mix of dns-TKCxxx (regular) and dns-GCxxx peptides (**bold**). They are ordered by X group.

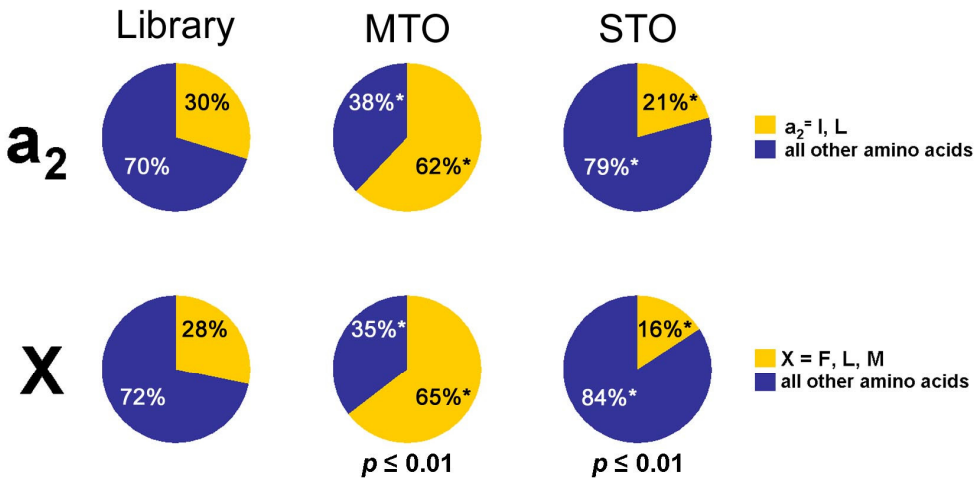


**Figure 2.4. The percentages of canonical and non-canonical residues for peptides in the overall library, and in the pools of peptides that are substrates for GGTase-I under multiple turnover (MTO) and single turnover (STO) conditions.** At the $a_2$ position, canonical is defined as I and L whereas at the X position, canonical residues are defined as F, L, and M (*12-15, 20*). The sequences of the peptides in the MTO and STO pools are significantly different from the library ($p \leq 0.01$, Chi-squared test). The sequences of the MTO substrates generally follow typical CaaX predictions; however, the sequences of the STO peptides are not well described by the canonical CaaX paradigm.

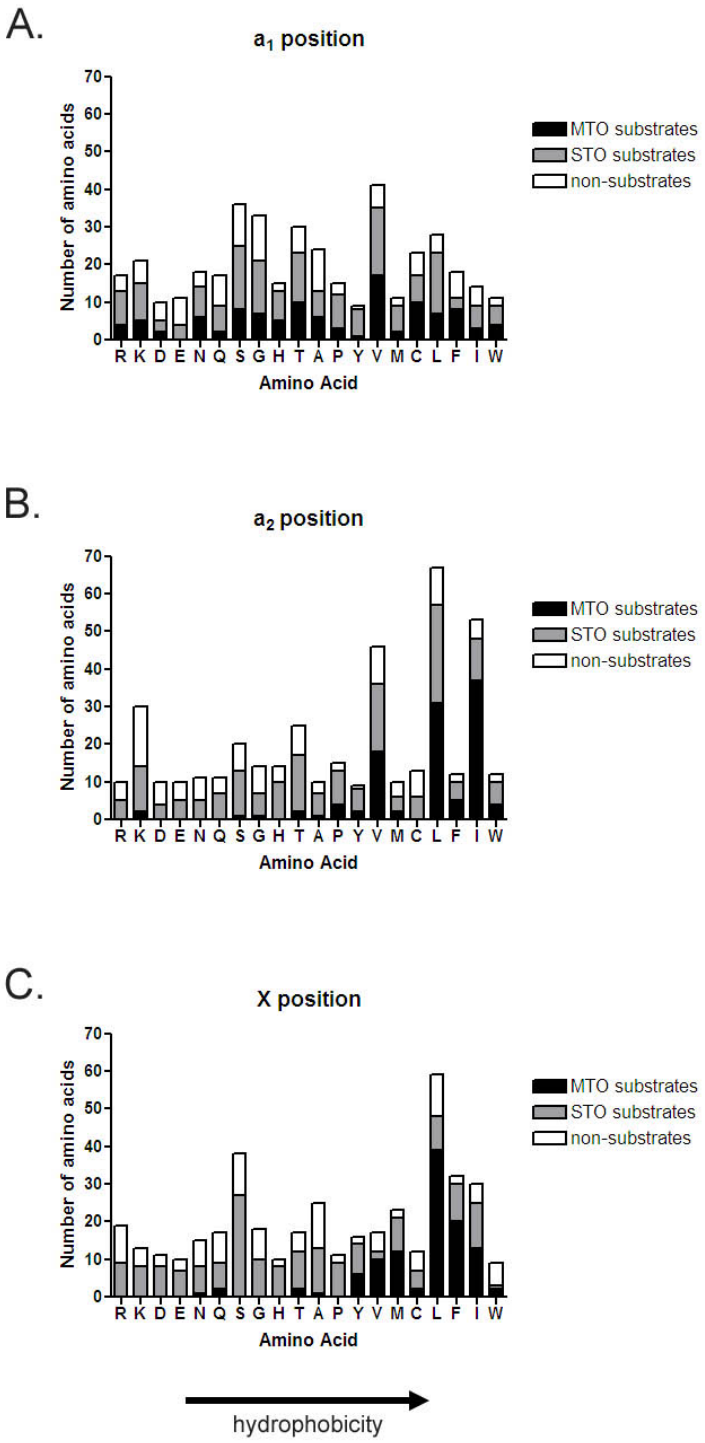


**Figure 2.5. Distribution of amino acids in peptides at the $a_1$, $a_2$, and X positions that are substrates for GGTase-I under MTO and STO conditions or are not reactive.** Each amino acid was tallied at the $a_1$ (A), $a_2$ (B), and X (C) position for each pool of substrates and graphed by amino acid count vs. amino acid identity. Peptides with MTO activity with GGTase-I are in black, peptides with STO activity are in gray, and non-substrates are in white. Additionally, the amino acids are organized on the X-axis by increasing hydrophobicity based on $\Delta G_{transfer}$ values (kcal/mol) from reference (*46*).

F, L, or M (defining the other 17 amino acids as non-canonical) (*12-15, 20*). Figure 2.4 shows the percentages of canonical and non-canonical amino acids contained in the library, the MTO pool, and the STO pool at the $a_2$ and X positions.

The MTO and STO pools of GGTase-I 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 the X positions as compared to the library. For instance, at the X position in the MTO pool vs. the library, the percentage of peptides with a canonical residue (F, L, or M) increases from 28% to 65%. This shows that in general, peptides that are active with GGTase-I under multiple turnover conditions are likely to have a canonical residue at the $a_2$ or X position. Although many substrates contain canonical sequences at $a_2$ and X, it is worth noting that many substrates *do not*. Figure 2.5 shows the frequency of each amino acid in each type of GGTase-I substrate at the $a_1$, $a_2$, and X positions of the peptides (Figure 2.5 A, B, and C, respectively). In general, at the $a_2$ position, substrates that have MTO activity with GGTase-I (Figure 2.5 B, black bars) can be hydrophobic, contain large residues such as W and Y, or even hydrophilic residues like K and S. The same trend is observed for the X position for GGTase-I substrates with MTO activity. Generally, GGTase-I recognizes more substrates containing hydrophobic amino acids (Figure 2.5 C, black bars), but also catalyzes geranylgeranylation of peptides with a variety of amino acids at the X position including hydrophilic ones like N and Q. Additionally, GGTase-I appears to recognize all amino acids at the $a_1$ position for substrates with MTO activity.

Conversely, in the STO pool, the substrates for GGTase-I are more varied. Comparing the library peptides to the single turnover pool, there is actually a significant

increase in the non-canonical sequences at the $a_2$ and X positions in the substrates of GGTase-I with single turnover activity (Figure 2.4). This trend is also illustrated in Figure 2.5. Each of the 20 amino acids is present at the $a_1$, $a_2$, and X positions in the substrates with STO activity with GGTase-I (gray bars Figure 2.5 B and C). Overall, this study suggests that GGTase-I is able to prenylate a wider range of substrates than originally proposed.

To more carefully examine these trends, a hypergeometric distribution model was used to analyze GGTase-I substrates for single amino acid preferences at the $a_1$, $a_2$ and X positions (with the help of Dr. Terry Watt). This analysis determines whether a single amino acid appears in a substrate more or less often compared to its appearance in the overall library; therefore, it shows whether a certain amino acid is overrepresented or underrepresented in the MTO, STO, and NON pools than the overall library. Table 2.4 shows the final results of this analysis. For amino acids in bold, $p \leq 0.02$, and for amino acids in italics, $0.02 \leq p \leq 0.05$.

Overall, the results of this analysis reveal that, in general, GGTase-I shows a preference for reactivity with peptides containing canonical CaaX sequences at the $a_2$ and X positions and shows little selectivity at the $a_1$ position. However, the CaaX model should be expanded. At the $a_1$ position, GGTase-I appears to slightly prefer V in substrates with MTO activity ($p \leq 0.05$) and Y for substrates with STO activity ($p \leq 0.05$), and selects against E ($p \leq 0.05$) and F ($p \leq 0.02$) for MTO and STO substrates, respectively. Furthermore, peptides that are not active with GGTase-I often contain an A ($p \leq 0.05$) or an E ($p \leq 0.02$) at the $a_1$ position. The historical CaaX paradigm states that $a_1$ should be small and aliphatic amino acids (and the overrepresentation of V in the MTO substrates

fits that definition); however, the structural data indicate the $a_1$ residue is exposed to solvent in the FTase•FPP analog•peptide complex consistent with biochemical data suggesting broader selectivity. (*14-16, 47, 48*). In focusing on the $a_2$ position, GGTase-I recognizes peptides that contain the residues I and L ($p \leq 0.02$) and to a lesser extent, V ($p \leq 0.05$) under MTO conditions. Furthermore, 10 amino acids that are hydrophilic or charged are underrepresented at the $a_2$ position in peptides with MTO activity with GGTase-I. For substrates with STO activity with GGTase-I, the residue H ($p \leq 0.05$) appears more often and I ($p \leq 0.02$) appears less often at the $a_2$ position. For peptides with no activity with GGTase-I, the amino acids C and D ($p \leq 0.05$) as well as K ($p \leq 0.02$) are overrepresented at the $a_2$ position. Therefore, GGTase-I prefers substrates containing an I, L, and V (small and aliphatic) at the $a_2$ position as the CaaX paradigm predicts and selects against amino acids with charged and hydrophilic side chains (like H, C, D, and K) at the $a_2$ position as demonstrated in the STO and NON substrate pools. Finally, GGTase-I prefers substrates with residues like F, I, L, M and V at the X position under MTO conditions, while selecting against peptides with X residues like A, D, E, G, H, K, P, R and S, indicating that GGTase-I selects for hydrophobic residues and generally discriminates against smaller and charged residues at the X position. The analysis of substrates with STO activity with GGTase-I shows more non-canonical sequence preferences at the X position since H, P, and S are overrepresented and L and V are underrepresented. Additionally, small, large, and charged amino acids like A, R, and W are overrepresented in the peptides that are inactive with GGTase-I. Together, these data suggest that the canonical X-group residues for GGTase-I should be expanded to include V and I, and H, P, and S for substrates with STO activity. Although these data indicate

substrates preferences at each of the $a_1$, $a_2$, and X side chains; however, there is little strict discrimination against any one amino acid since almost all amino acids were observed at each position in at least one substrate with at least single turnover activity.

An interesting trend in this analysis is that many amino acids that are overrepresented in a particular substrate pool for GGTase-I are underrepresented in another substrate pool. For instance, F, L, M, and V are overrepresented in the MTO substrates at the X position but L and V are underrepresented in the STO substrates while F, L, and M are underrepresented in the NON substrates. Another example is I and L at the $a_2$ position. These two residues are preferred in MTO substrates at $a_2$ while I and I/L are depleted in the STO and NON substrates, respectively. These anticorrelation trends validate the preferences as they are depleted for the STO and/or NON pools.

*Steady state parameters for GGTase-I: importance of upstream regions*

Steady state parameters were measured for a subset of MTO substrates. The kinetics for GGTase-I catalyzed geranylgeranylation of the dns-TKCxxx peptides were measured by previous Fierke lab members Dr. Hartman and undergraduate student L.S. Janik; however, dns-GCxxx peptides were measured in this work. Table 2.5 shows values for $k_{cat}/K_M^{peptide}$ for peptides with identical CaaX sequences but varying upstream regions. In general, the values for $k_{cat}/K_M^{peptide}$ for the reaction of GGTase-I with dns-GCxxx peptides are one to three times higher than the dns-TKCxxx peptides, correlating with similar observations with FTase (*16, 22*).

| position | $a_1$ | | $a_2$ | | X | |
|---|---|---|---|---|---|---|
| | **over-represented** | **under-represented** | **over-represented** | **under-represented** | **over-represented** | **under-represented** |
| **GGTase-I MTO** | *V* | *E* | **I, L,** *V* | **C,** *D, E,* **H, K,** *N, Q, R,* **S, T** | **F,** *I,* **L, M, V** | **A,** *D, E,* **G,** *H,* **K,** *P,* **R, S** |
| **GGTase-I STO** | *Y* | **F** | *H* | **I** | *H,* **P, S** | **L, V,** *W* |
| **GGTase-I NON** | *A,* **E** | *V* | *C, D,* **K** | **I, L** | *A,* **R, W** | **F,** *L, M* |
| **FTase MTO** | *L* | *C* | **I, L**, *V* | **C,** *D,* **K** | **F, M, Q** | None |
| **FTase STO** | *A,* **C** | **L** | *A*, **S** | **I, K** | None | **M,** *P, Q* |
| **FTase NON** | *K* | *I* | **D**, *E,* **K, R** | **V, I, L, T** | **P, R** | *F, Q* |

**Table 2.4. Amino acids that are overrepresented or underrepresented in 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, $0.02 \leq p \leq 0.05$. FTase data is from reference (*16*).

| Sequence | $k_{cat}/K_M$ ($mM^{-1}s^{-1}$) | Sequence | $k_{cat}/K_M$ ($mM^{-1}s^{-1}$) | Fold Difference |
|---|---|---|---|---|
| GCRLL | 54±4 | TKCRLL | 21±3 | 2.6 |
| GCTIL | 53±8 | | | |
| GCVLL | 50±7 | TKCVLL | 38.6±0.4 | 1.3 |
| GCVIM | 48±5 | | | |
| GCVIF | 42±6 | | | |
| GCCIF | 40±5 | | | |
| GCCIF | 24±3 | TKCCIF | 15±1 | 1.6 |
| GCCVL | 20±1 | TKCCVL | 11.0±0.5 | 1.8 |
| GCFIF | 19±1 | | | |
| GCKIF | 17±2 | TKCKIF | 5.0±0.7 | 3.3 |
| GCLVL | 13.5±0.9 | | | |
| GCLLV | 10±2 | | | |
| GCMPL | 4.6±0.3 | | | |
| GCTVF | 3.3±0.7 | TKCTVF | 2.9±0.1 | 1.1 |

**Table 2.5. Steady state parameters of peptides with GGTase-I.** $k_{cat}/K_M$ ($mM^{-1}s^{-1}$) was measured for active MTO dns-GCxxx peptides and compared to dns-TKCxxx peptides (measured by previous Fierke lab members). Dns-GCxxx peptides are generally 1-3 fold more active than dns-TKCxxx peptides.

*Limits of this analysis*

The lower limits of the reactivity of peptides with GGTase-I under MTO and STO conditions in this work are estimated to be ≥200-400 $M^{-1}s^{-1}$ and ≥ 0.0009 $s^{-1}$, respectively. It is possible that some peptides classified as STO may in fact have MTO activity with GGTase-I, but that the MTO activity is below the detection limit under the conditions chosen in this work. Further, non-substrate peptides could perhaps have low activity with GGTase-I under STO conditions after longer incubation time. However, the cutoffs chosen in this work are similar to the peptide library studies done with FTase (*16*) to allow for comparison of the two studies. Overall, cutoffs must be chosen to classify the 400 peptides into different pools of reactivity; however, any one particular peptide might display a different type of reactivity with GGTase-I under a different set of assay conditions or time course.

## Discussion

*MTO and STO substrate specificity*

Different steps of the prenyltransferase reactions likely contribute to specificity for the MTO and STO substrates. For MTO peptides, selectivity arises from everything up to the first irreversible step under $k_{cat}/K_M^{peptide}$ conditions; this means selectivity is before or at the chemistry step of the reaction since dissociation of the diphosphate product is irreversible (*22, 49*). For peptides that are reactive with GGTase-I under MTO conditions, it is likely that the rate limiting step of the reaction is the chemical step, which can include the conformational change or the actual chemistry. Recently, mean force QM/MM studies with FTase suggest that the activity of FTase arises more from the chemistry than the conformational change (Yue Yang, Bing Wang, Melek N. Ucisik, Guanglei Cui, Carol A. Fierke, and Kenneth M. Merz, Jr, unpublished data). Similar trends may occur with GGTase-I, but have not been elucidated. Additionally, under $k_{cat}$ conditions, the release of the prenylated product can be rate limiting for some peptides, at least with FTase (*32, 34*). For substrates with activity with GGTase-I under STO conditions, a step after chemistry such as the dissociation of the product, is slow and rate-limiting. Therefore, under both MTO and STO conditions, GGTase-I binds to the peptides and catalyzes geranylgeranylation. For MTO substrates, product dissociation occurs sufficiently rapidly to observe turnover. However, for STO substrates, product dissociation is slow. This could reflect additional interactions leading to a high affinity product or could be due to the inability of the geranylgeranylated peptide to move from the binding pocket to the exit groove (*28, 50*).

*Comparison of FTase and GGTase-I substrates*

This work and peptide library studies with FTase have revealed similarities and differences in selectivity of FTase and GGTase-I for substrates. Table 2.5 shows the results from the statistical analysis of previous work done with FTase in reference (*16*) and the work described above with GGTase-I.

At the $a_2$ position for substrates, it is striking that FTase and GGTase-I both select for residues I, L, and (to a lesser extent) V in the MTO pool. However, there are not overlapping residues for STO substrates; H is overrepresented for GGTase-I substrates while A and S are overrepresented for FTase. For both prenyltransferases, C, D, and K are underrepresented at $a_2$ in MTO substrates, I in STO substrates, and I and L for non-substrates. Crystallographic studies have shown that the $a_2$ position of the substrate is contacted by residues W102β, W106β, Y361β, and the 3$^{rd}$ isoprene of FPP for FTase and T49β, F53β, and L320β, and the 3$^{rd}$ and 4$^{th}$ isoprene of GGPP and the X residue for GGTase-I (*48*). These mainly hydrophobic pockets were predicted to recognize a range of small and hydrophobic residues, preferring I and L, but also accepting V, F, Y, T, and M (*48*). The current analysis is consistent with the crystallographic work and provides further definition of this selectivity. Although these predictions can define many FTase and GGTase-I substrates, these rules are not absolute. A large group of substrates poorly defined by this paradigm are reactive with FTase and GGTase-I (*16*). For instance, when defining the canonical $a_2$ residues as I, L, and V for GGTase-I substrates, this paradigm describes 78 % of the pool of MTO substrates but only describes 31 % of the pool of STO substrates.

For the X position, the library studies in reference (*16*) with FTase and the studies here with GGTase-I indicate that F and M are preferred substrates for both FTase and GGTase-I under MTO conditions, demonstrating overlapping specificity. However, the selectivity is not identical as FTase prefers Q while GGTase-I also prefers L, V, and to a lesser extent, I. These substrate preferences are consistent with previous biochemical studies (*20*) as well as structural work (at least for the phenylalanine) (*48*). Crystallographic studies have shown that the X binding pocket for FTase is more hydrophilic in nature, whereas the X pocket for GGTase-I is more hydrophobic leading to FTase preferring more polar residues in general and GGTase-I preferring more hydrophobic residues (*48*). The differences in amino acid preferences in the peptide library studies reinforce the differences in the specificity of FTase and GGTase-I. Though this work and other studies define distinct sequence preferences at the X position for GGTase-I and FTase substrates, there are still many substrates for these prenyltransferases that do not contain a canonical X side chain and even canonical FTase X residues can be recognized by GGTase-I and vice versa. Biochemical work with FTase has indicated that selectivity can be context-dependent, meaning that the $a_2$ and X residue can be recognized together, suggesting that the identity of one amino acid in the CaaX sequence can impact the recognition of another (*22*). It is possible that similar patterns may occur with GGTase-I. Figure 2.5 is a summary of how the CaaX paradigm can be updated for general GGTase-I substrate preferences. Overall, if canonical X residues for GGTase-I substrates are defined as F, I, L, M, and V, this updated paradigm describes 85% of the MTO pool of substrates and 24% of the STO pool of substrates.

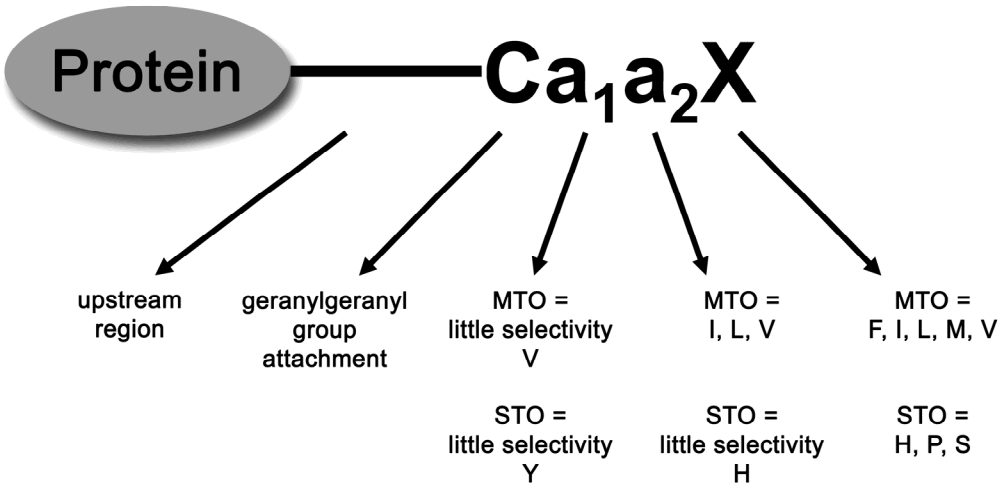


**Figure 2.6. Updated $Ca_1a_2X$ paradigm for GGTase-I.** Based upon peptide library studies and statistical analysis where $p \leq 0.05$, there is selectivity at the $a_2$ and X position for MTO substrates and selectivity in X for STO substrates. The upstream region may or may not be important for reactivity.

*Implications of the residues upstream of the CaaX sequence*

In the cell, the regions of the C-terminus of prenyltransferase substrates are variable and often contain a polybasic region (*47, 51*). In this work we compare the reactivity of dns-TKCxxx vs. dns-GCxxx peptides, a charged upstream region vs. an uncharged upstream region. Past biochemical work has shown that for GGTase-I, multiple lysines have little effect on the reactivity or binding affinity of peptides, but for FTase, a polybasic region decreases reactivity but increases affinity (*21*). As seen in the Results section and Table 2.5, G peptides generally have slightly higher values for $k_{cat}/K_M^{peptide}$ than the TK peptides. As the proposed rate limiting step is chemistry for the peptides reacting with GGTase-I under these conditions, the upstream lysine could form long range electrostatic interactions with the GGTase-I enzyme which may alter the position of the peptide affecting the reactivity. This upstream polybasic sequence could have further implications *in vivo*, as the region could contain amino acids that are post-

translationally modified and polybasic regions have been shown to enhance the interactions with the plasma membrane for some prenylated proteins (*52-54*).

*In vivo implications*

A wider range of peptide sequences are accepted as GGTase-I substrates, including many non-canonical sequences, especially in the STO pool. Are STO substrates truly prenylated *in vivo*? Since the STO products are not rapidly released from the enzyme under *in vitro* conditions, the product may form a complex with the modified protein and inhibit the enzyme; however, there are at least two known examples of *in vivo* STO substrates, corresponding to C-terminal sequences –CAVL (active with FTase under STO conditions) and –CKAA (active with GGTase-I and FTase under STO conditions) (*55, 56*). *In vitro* work has shown that an additional substrate can bind to the product complex to help facilitate product release (*20, 34, 35*), and in most cases this additional substrate is the FPP or GGPP molecule (*34*), although peptides have also been proposed (*20, 35*). Therefore, in the cell it is possible that product release for the STO substrates is facilitated by another substrate, protein (i.e., a release factor), or small molecule. Potentially, facilitation of product dissociation could be a mechanism of regulating prenylation. Furthermore, this could be a mechanism to deliver modified proteins to particular cellular membranes. Overall, this work suggests that many more proteins are geranylgeranylated than originally proposed and will aid us in understanding what pathways include geranylgeranylated proteins and may therefore be affected by inhibitors of GGTase-I.

## References


(1) Benetka, W., Koranda, M., and Eisenhaber, F. (2006) Protein Prenylation: An (Almost) Comprehensive Overview on Discovery History, Enzymology, and Significance in Physiology and Disease. *Monatshefte für Chemie / Chemical Monthly 137*, 1241-1281.

(2) Casey, P. J., and Seabra, M. C. (1996) Protein Prenyltransferases. *Journal of Biological Chemistry 271*, 5289-5292.

(3) Marshall, C. J. (1993) Protein prenylation: a mediator of protein-protein interactions. *Science 259*, 1865-6.

(4) Casey, P. J. (1994) Lipid modifications of G proteins. *Curr Opin Cell Biol 6*, 219-25.

(5) Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2008) Farnesyltransferase inhibitors: a detailed chemical view on an elusive biological problem. *Curr Med Chem 15*, 1478-92.

(6) Young, S. G., Meta, M., Yang, S. H., and Fong, L. G. (2006) Prelamin A farnesylation and progeroid syndromes. *J Biol Chem 281*, 39741-5.

(7) Nallan, L., Bauer, K. D., Bendale, P., Rivas, K., Yokoyama, K., Horney, C. P., Pendyala, P. R., Floyd, D., Lombardo, L. J., Williams, D. K., Hamilton, A., Sebti, S., Windsor, W. T., Weber, P. C., Buckner, F. S., Chakrabarti, D., Gelb, M. H., and Van Voorhis, W. C. (2005) Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. *J Med Chem 48*, 3704-13.

(8) Zhang, F. L., and Casey, P. J. (1996) Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem 65*, 241-69.

(9) Samuel, F., and Hynds, D. L. (2010) RHO GTPase signaling for axon extension: is prenylation important? *Mol Neurobiol 42*, 133-42.

(10) Rusinol, A. E., and Sinensky, M. S. (2006) Farnesylated lamins, progeroid syndromes and farnesyl transferase inhibitors. *J Cell Sci 119*, 3265-72.

(11) Casey, P. J. (1995) Protein Lipidation in Cell Signaling. *Science 268*, 221-225.

(12) Caplin, B. E., Hettich, L. A., and Marshall, M. S. (1994) Substrate characterization of the Saccharomyces cerevisiae protein farnesyltransferase and type-I protein geranylgeranyltransferase. *Biochim Biophys Acta 1205*, 39-48.

(13) Omer, C. A., Kral, A. M., Diehl, R. E., Prendergast, G. C., Powers, S., Allen, C. M., Gibbs, J. B., and Kohl, N. E. (1993) Characterization of recombinant human farnesyl-protein transferase: cloning, expression, farnesyl diphosphate binding,

and functional homology with yeast prenyl-protein transferases. *Biochemistry 32*, 5167-76.

(14) Reiss, Y., Stradley, S. J., Gierasch, L. M., Brown, M. S., and Goldstein, J. L. (1991) Sequence requirement for peptide recognition by rat brain p21ras protein farnesyltransferase. *Proc Natl Acad Sci U S A 88*, 732-6.

(15) Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L., and Gibbs, J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem 266*, 14603-10.

(16) Hougland, J. L., Hicks, K. A., Hartman, H. L., Kelly, R. A., Watt, T. J., and Fierke, C. A. (2010) Identification of Novel Peptide Substrates for Protein Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J Mol Biol 395*, 176-90.

(17) Goldstein, J. L., Brown, M. S., Stradley, S. J., Reiss, Y., and Gierasch, L. M. (1991) Nonfarnesylated tetrapeptide inhibitors of protein farnesyltransferase. *J Biol Chem 266*, 15575-8.

(18) Pompliano, D. L., Gomez, R. P., and Anthony, N. J. (1992) Intramolecular Fluorescence Enhancement - a Continuous Assay of Ras Farnesyl - Protein Transferase. *Journal of the American Chemical Society 114*, 7945-7946.

(19) Hightower, K. E., Huang, C. C., Casey, P. J., and Fierke, C. A. (1998) H-Ras peptide and protein substrates bind protein farnesyltransferase as an ionized thiolate. *Biochemistry 37*, 15555-62.

(20) Hartman, H. L., Hicks, K. A., and Fierke, C. A. (2005) Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity. *Biochemistry 44*, 15314-24.

(21) Hicks, K. A., Hartman, H. L., and Fierke, C. A. (2005) Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I. *Biochemistry 44*, 15325-33.

(22) Hougland, J. L., Lamphear, C. L., Scott, S. A., Gibbs, R. A., and Fierke, C. A. (2009) Context-dependent substrate recognition by protein farnesyltransferase. *Biochemistry 48*, 1691-701.

(23) Pompliano, D. L., Schaber, M. D., Mosser, S. D., Omer, C. A., Shafer, J. A., and Gibbs, J. B. (1993) Isoprenoid diphosphate utilization by recombinant human farnesyl:protein transferase: interactive binding between substrates and a preferred kinetic pathway. *Biochemistry 32*, 8341-7.

(24) Furfine, E. S., Leban, J. J., Landavazo, A., Moomaw, J. F., and Casey, P. J. (1995) Protein farnesyltransferase: kinetics of farnesyl pyrophosphate binding and product release. *Biochemistry 34*, 6857-62.

(25) Yokoyama, K., McGeady, P., and Gelb, M. H. (1995) Mammalian protein geranylgeranyltransferase-I: substrate specificity, kinetic mechanism, metal requirements, and affinity labeling. *Biochemistry 34*, 1344-54.

(26) Huang, C., Hightower, K. E., and Fierke, C. A. (2000) Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state. *Biochemistry 39*, 2593-602.

(27) Yokoyama, K., Zimmerman, K., Scholten, J., and Gelb, M. H. (1997) Differential prenyl pyrophosphate binding to mammalian protein geranylgeranyltransferase-I and protein farnesyltransferase and its consequence on the specificity of protein prenylation. *J Biol Chem 272*, 3944-52.

(28) Long, S. B., Casey, P. J., and Beese, L. S. (2002) Reaction path of protein farnesyltransferase at atomic resolution. *Nature 419*, 645-50.

(29) Bowers, K. E., and Fierke, C. A. (2004) Positively charged side chains in protein farnesyltransferase enhance catalysis by stabilizing the formation of the diphosphate leaving group. *Biochemistry 43*, 5256-65.

(30) Long, S. B., Casey, P. J., and Beese, L. S. (2000) The basis for K-Ras4B binding specificity to protein farnesyltransferase revealed by 2 A resolution ternary complex structures. *Structure 8*, 209-22.

(31) Pickett, J. S., Bowers, K. E., Hartman, H. L., Fu, H. W., Embry, A. C., Casey, P. J., and Fierke, C. A. (2003) Kinetic studies of protein farnesyltransferase mutants establish active substrate conformation. *Biochemistry 42*, 9741-8.

(32) Pais, J. E., Bowers, K. E., Stoddard, A. K., and Fierke, C. A. (2005) A continuous fluorescent assay for protein prenyltransferases measuring diphosphate release. *Anal Biochem 345*, 302-11.

(33) Pais, J. E., Bowers, K. E., and Fierke, C. A. (2006) Measurement of the alpha-secondary kinetic isotope effect for the reaction catalyzed by mammalian protein farnesyltransferase. *J Am Chem Soc 128*, 15086-7.

(34) Tschantz, W. R., Furfine, E. S., and Casey, P. J. (1997) Substrate binding is required for release of product from mammalian protein farnesyltransferase. *J Biol Chem 272*, 9989-93.

(35) Troutman, J. M., Andres, D. A., and Spielmann, H. P. (2007) Protein farnesyl transferase target selectivity is dependent upon peptide stimulated product release. *Biochemistry 46*, 11299-309.

(36) Krzysiak, A. J., Scott, S. A., Hicks, K. A., Fierke, C. A., and Gibbs, R. A. (2007) Evaluation of protein farnesyltransferase substrate specificity using synthetic peptide libraries. *Bioorg Med Chem Lett 17*, 5548-51.

(37) Krzysiak, A. J., Aditya, A. V., Hougland, J. L., Fierke, C. A., and Gibbs, R. A. (2009) Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates. *Bioorganic & Medicinal Chemistry Letters 20*, 767-770.

(38) Hartman, H. L., Bowers, K. E., and Fierke, C. A. (2004) Lysine beta311 of protein geranylgeranyltransferase type I partially replaces magnesium. *J Biol Chem 279*, 30546-53.

(39) Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch Biochem Biophys 82*, 70-7.

(40) Cassidy, P. B., Dolence, J. M., and Poulter, C. D. (1995) Continuous fluorescence assay for protein prenyltransferases. *Methods Enzymol 250*, 30-43.

(41) Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology 189*, 113-130.

(42) Studier, F. W., Rosenberg, A. H., Dunn, J. J., Dubendorff, J. W., and David, V. G. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, in *Methods in Enzymology* pp 60-89, Academic Press.

(43) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif 41*, 207-34.

(44) Grabski, A., Mehler, M., and Drott, D. (2005) The Overnight Express Autoinduction System: High-density cell growth and protein expression while you sleep. *Nature Methods 2*, 234-235.

(45) Lane, K. T., and Beese, L. S. (2006) Thematic review series: lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type I. *J Lipid Res 47*, 681-99.

(46) Karplus, P. A. (1997) Hydrophobicity regained. *Protein Sci 6*, 1302-7.

(47) Maurer-Stroh, S., and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biol 6*, R55.

(48) Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol 343*, 417-33.

(49) Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W.H. Freeman and Company, New York, NY.

(50) Taylor, J. S., Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2003) Structure of mammalian protein geranylgeranyltransferase type-I. *Embo J 22*, 5963-74.

(51) Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F. L., and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Comput Biol 3*, e66.

(52) Laude, A. J., and Prior, I. A. (2008) Palmitoylation and localisation of RAS isoforms are modulated by the hypervariable linker domain. *J Cell Sci 121*, 421-7.

(53) Heo, W. D., Inoue, T., Park, W. S., Kim, M. L., Park, B. O., Wandless, T. J., and Meyer, T. (2006) PI(3,4,5)P3 and PI(4,5)P2 lipids target proteins with polybasic clusters to the plasma membrane. *Science 314*, 1458-61.

(54) Jackson, J. H., Li, J. W., Buss, J. E., Der, C. J., and Cochrane, C. G. (1994) Polylysine domain of K-ras 4B protein is crucial for malignant transformation. *Proc Natl Acad Sci U S A 91*, 12730-4.

(55) Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004) A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci U S A 101*, 12479-84.

(56) Liu, Z. H., Meray, R. K., Grammatopoulos, T. N., Fredenburg, R. A., Cookson, M. R., Liu, Y. C., Logan, T., and Lansbury, P. T. (2009) Membrane-associated farnesylated UCH-L1 promotes alpha-synuclein neurotoxicity and is a therapeutic target for Parkinson's disease. *Proc. Natl. Acad. Sci. U. S. A. 106*, 4635-4640.

# CHAPTER III

# PREDICTING FTASE SUBSTRATES: DEVELOPMENT OF A NOVEL COMPUTATIONAL METHOD[1]

## Introduction

Protein farnesyltransferase (FTase) and geranylgeranyltransferase-I (GGTase-I) catalyze the addition of a farnesyl or geranylgeranyl group to a cysteine near the C-terminus of a substrate protein (*1, 2*). These hydrophobic modifications help to locate substrates to cellular membranes and promote protein-protein interactions (*3, 4*). FTase and GGTase-I were proposed to recognize a "$Ca_1a_2X$" motif on substrates (*2, 5, 6*), where "C" is the cysteine where the prenyl group is attached via a thioether bond, "$a_1$" and "$a_2$" are small and aliphatic amino acids, and "X" confers specificity between the two enzymes, with FTase typically preferring methionine, alanine, glutamate or serine and GGTase-I preferring leucine or phenylalanine (*7-10*). Although the $Ca_1a_2X$ motif can serve as a model for specificity, many studies have indicated that not all substrates are described by this paradigm and that the $Ca_1a_2X$ model should be expanded (Chapter 2, (*11, 12*)).

[1] A portion of this work is published in London, N., Lamphear, C. L., Hougland, J. L., Fierke, C. A., and Schueler-Furman, O. (2011) Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. *PLoS Comput Biol* 7, e1002170.

Inhibitors towards these two prenyltransferases are of great interest for the treatment of a multitude of diseases. For example, Ras was an early discovery as a prenyltransferase substrate (*13*) and is mutated in 30% of all human cancers (*14*). Prenyltransferase inhibitors were initially developed to target Ras protein signaling pathways implicated in cancer, but it was later determined that inhibitor efficacy is the result of modulating prenylation of non-Ras proteins (*15*). Therefore, inhibitors are being studied for the treatment of other diseases besides cancer such as Hutchinson-Gilford Progeria syndrome (*16*) and parasitic infections (*17*). Because of the potential implications of prenylation in cell biology and potential disease treatment, it is necessary to define the complete set of *in vivo* prenylated proteins. Traditional methods to detect prenylation *in vivo*, such as the use of radioactivity and antibodies, have met limited success (*1, 18-20*); therefore, computational methods to predict prenylated proteins are of great value.

Past computational approaches to predict prenylation have been developed based upon already known substrates. The most recent example of an algorithm for prenylation prediction is PrePS, developed by the Eisenhaber group (*21, 22*). They generated a learning set of known and homologous substrates, and also included the region of 11 amino acids upstream of the $Ca_1a_2X$ sequence in their analysis. Therefore, PrePS predicts the likelihood of prenylation of the C-terminal 15 amino acids of a protein. In recently published work of peptide library studies with FTase, PrePS had very low false positive rate for farnesylated peptides, but had a high false negative rate, about 40% (*11*). Therefore, this method misses a large group of potential substrates. Other predictions

have been done with structural studies (*23*), but overall a more robust prediction program is needed.

In this work, we collaborated with Nir London and Dr. Ora Schueler-Furman at the Hebrew University of Jerusalem to test a program that predicts farnesylation of $Ca_1a_2X$ sequences (*24*). The Scheuler-Furman lab developed the FlexPepBind method using the FTase peptide library work as a training set (*11*). To test the predictive power of this method, we chose a set of 29 peptides predicted to be FTase substrates to test for farnesylation activity. I measured the reactivity of FTase with this set of 29 peptides, measuring the value of $k_{cat}/K_M^{peptide}$ for peptides with reactivity with FTase under MTO conditions. Overall, FlexPepBind is a robust program for predicting a wide range of FTase substrates, which is an exciting breakthrough; however, at this time it is unclear what catalytic parameter may correlates with FlexPepBind score.

## Experimental Methods

*Development of FlexPepBind*

This program was developed by our collaborators Nir London and Ora Scheuler-Furman at the Hebrew University of Jerusalem; more specific details of the program are in reference (*24*). A brief summary of the development of this program is as follows. A crystal structure of FTase in complex with the peptide CNIQ and a farnesyl diphosphate analog (([(3,7,11-trimethyl-dodeca-2,6,10-trienyloxycarbamoyl)-methyl]-phosphonic acid) was used as a template (PDB ID: 1tn6, (*23*)). Varied peptide sequences were threaded onto the peptide backbone and the side chains were packed in the best rotamer configuration. Three bonds were held constant to preserve the structure of the binding

site: 1) the position of sulfur atom of the CaaX cysteine relative to nearby FTase sidechains; 2) the structurally conserved hydrogen bond from the C-terminal carboxylate of the peptide to FTase Q167α; and 3) the hydrogen bond between the backbone carbonyl of the $a_2$ residue to the sidechain of FTase R202β. The arrows in Figure 3.1 show these three constraints for the calculations. The Rosetta FlexPepDock protocol was previously developed by Scheuler-Furman lab to model protein-protein interactions (*25*) using the Rosetta modeling suite (*26*) and based on this, the Shueler-Furman lab developed a simpler and less computationally expensive protocol called FlexPepBind to model FTase/peptide interactions. The training set for development of FlexPepBind to distinguish between FTase substrates and non-substrates was from reactivity of FTase with a peptide library (*11*) using 77 MTO peptide sequences and 51 non-reactive peptide sequences. For each peptide, the energy score from the best configuration was chosen. Energy cutoff scores of -0.4 and -1.1 were used validate the training set of peptide reactivity with FTase. This protocol was then applied to other test sets as detailed in (*24*) and the results section, and applied to the total 8,000 possible CaaX sequences to predict FTase substrates and non-substrates.

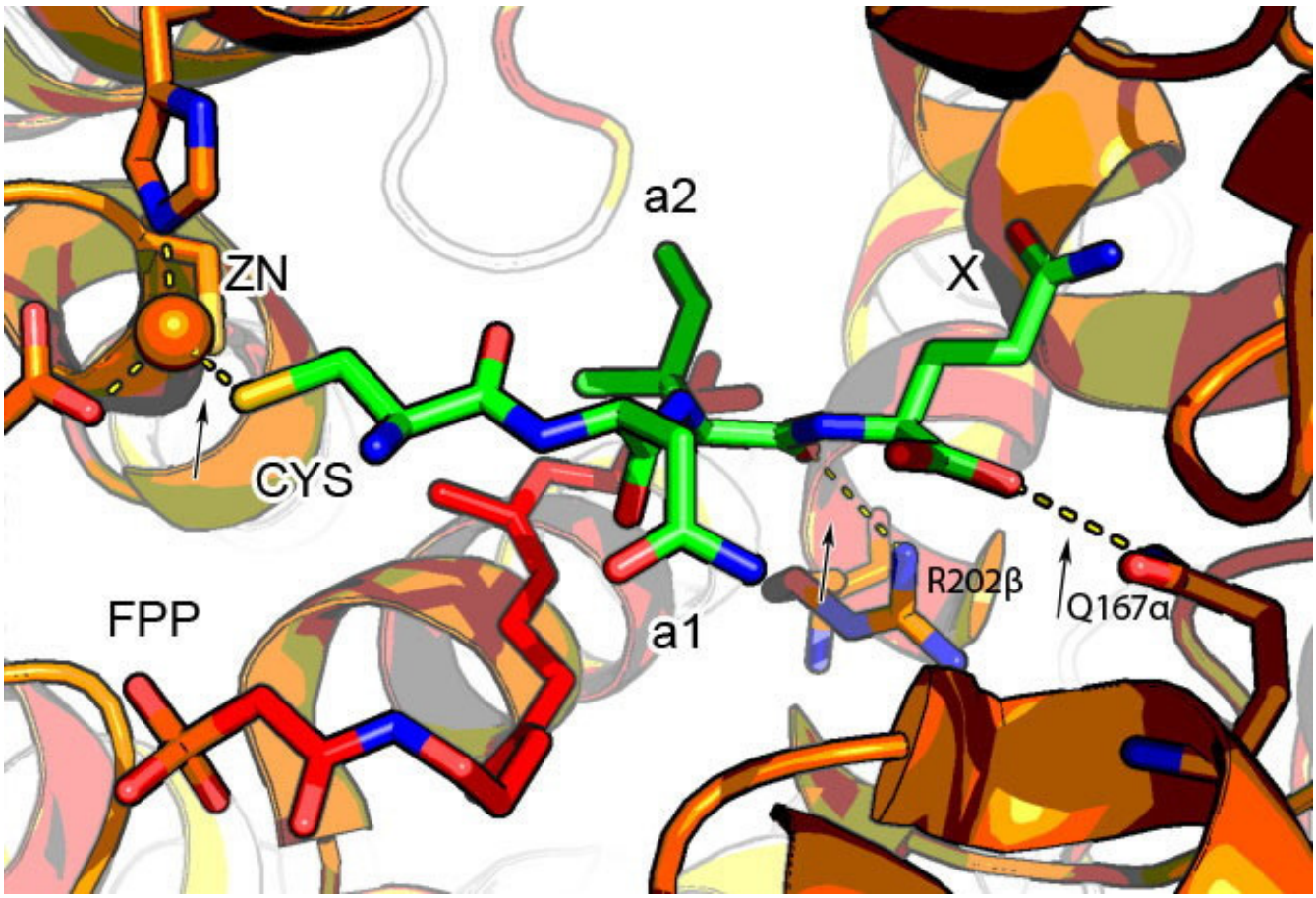


**Figure 3.1**. **Crystal Structure of FTase**. Crystal structure of FTase bound to CNIQ peptide (green) and an FPP analog (red) (PDB ID 1tn6 (*23*)). Three bonds were constrained: 1) the position of sulfur atom of the CaaX cysteine relative to nearby FTase sidechains; 2) the hydrogen bond between the X residue carboxylate and sidechain of FTase residue Q167α; and 3) the hydrogen bond between the $a_2$ backbone carbonyl oxygen to the side chain of FTase R202β.

*Choosing peptides for experimental validation of FlexPepBind*

The best scoring 29 sequences of peptides whose reactivity with FTase had not previously been measured were chosen to be tested for activity *in vitro;* 16 of these peptides correspond to the C-terminal sequences of human proteins. These peptides were purchased from Sigma Genosys (The Woodland, TX) in the PEPscreen™ format in the form dns-TKCxxx, where x is any amino acid. The peptides were dissolved in 10%

DMSO in EtOH and stored at -80ºC. The peptide concentration was calculated by measuring free thiols using a DTNB assay with an extinction coefficient of 14,150 $M^{-1}cm^{-1}$ at 412 nm (*27*).

*Multiple turnover screen for FTase activity*

3 μM dansylated-peptide (dns-TKCxxx) was incubated with 1 μM $^{3}$H-farnesyldiphosphate and 25 nM rat FTase in 50 mM HEPPSO, pH 7.8, 5 mM TCEP, 5 mM $MgCl_2$ at 25ºC for two hours. The reaction was quenched with 80:20 isopropanol:acetic acid and fractionated on a silica TLC plate (8:1:1 isopropanol:ammonium hydroxide: water). The TLC plates were visualized by autoradiography. Peptides that were observed to be at least 10–20% farnesylated, as evaluated compared to $^{3}$H-farnesyl-dns-GCVLS, were labeled MTO substrates. Using the assumptions that [peptide] < $K_M$ and [FPP] is saturating, the lower limit of this assay 200–400 $M^{-1}s^{-1}$, similar to the limits observed in previous work (*11*).

*Single turnover screen for FTase activity*

Single turnover assays were carried out as described above for the MTO assays except that the enzyme concentration was increase to 1 μM FTase and incubated with 0.8 μM $^{3}$H-FPP and 3 μM dns-TKCxxx peptide for one hour at 25ºC before the reaction was quenched. Peptides were considered a STO substrate if at least 10 - 20% of the 3H-FPP reacted with the peptide after one hour. The lower limit of the peptide reactivity in this assay is estimated to be ≥ 0.0004-0.0006 $s^{-1}$ and is a similar cutoff as in other work with FTase (*11*).

*Steady state kinetics*

Steady state kinetic parameters for the FTase catalyzed farnesylation of the MTO peptides predicted by FlexPepBind were using the fluorescent assay described in Chapter 2. In this case, reactions contained 0.2 – 40 μM dns-peptide, 20 -100 nM FTase, and 10 μM FPP in 50 mM HEPPSO pH 7.8, 5 mM TCEP and 5 mM $MgCl_2$ at 25ºC. All curve fitting was performed using Graphpad Prism (Graphpad Software, San Diego, CA). Both the initial rate and the endpoint of the reactions were measured and the fluorescence was converted to concentration using a conversion factor, which was generated by dividing the change in fluorescence to reach the reaction endpoint by peptide concentration and taking the average at several peptide concentrations. The values for the kinetic parameters were determined from the fit of the Michaelis-Menten equation (Equation 1) to the peptide dependence of $V_{initial}/[E]$.

$$\frac{V}{[E]} = \frac{k_{cat}[peptide]}{K_M + [peptide]} \quad (1)$$

*Correlation of parameters with FlexPepBind scores*

To determine if various physical parameters correlate with FlexPepBind scores, the parameter was graphed with FlexPepBind score. A nonparametric (Spearman) correlation analysis was used to determine if the values correlate using Graphpad Prism (Equation 2, Graphpad Software, San Diego, CA). This analysis assumes that there is not a Gaussian distribution of values and that both the variables are independent. A correlation coefficient is obtained from this analysis where $r_s = 0$ if the variables are not correlated and $r_s = 1$ or -1 if the variables are perfectly correlated or anti-correlated.

Further, if $-1 < r_s < 0$, for example, one variable increases as the other decreases. A two tailed *p* value was also obtained with the nonparametric (Spearman) correlation, to assess if the correlation is real and not due to chance. In the nonparametric (Spearman) correlation analysis, $r_s$ is the correlation coefficient, N is the number of pairs (XY), and D is the difference between each pair (X-Y) (*28*).

$$r_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \quad (2)$$

## Results

*Development of FlexPepBind: a good predictor of FTase substrates*

FlexPepBind was created by taking structural data for FTase•peptide complexes (*23*), threading different sequences onto the backbone, and finding the optimal rotamers using a modified protocol of FlexPepDock (which utilizes the Rosetta modeling suite) that is simpler and faster (*24*). Three bonds were held constant to maintain the integrity of the active site: 1) the cysteine-$Zn^{2+}$ bond (by maintaining the position of the sulfur position in the active site); 2) the hydrogen bond between the X group carboxylate and the side chain of FTase Q167α; and 3) the hydrogen bond between the $a_2$ backbone carbonyl and the side chain of FTase R202β. Energy scores were calculated for the binding affinity of FTase for the training set peptides including 77 MTO and 51 NON (non-substrate) peptide sequences previously tested for reactivity with FTase (*11*). A threshold of -0.4 energy score was set as a cutoff; any peptide below this threshold was designated a peptide that would presumably bind and react with FTase while any peptide above this threshold was designated a non-binder and hence not a substrate for FTase. Overall, this computational analysis of the training set correctly classified 69% of the

substrates in the proper reactive or non-reactive pools with FTase (true positive rate); and incorrectly categorized 8% of the training set (false positive rate). A more stringent cutoff of -1.1 energy score yields a 44% true positive rate and a 2% false positive rate with the training set. As the false positive rate of 8% with the -0.4 cutoff is quite low, this threshold was used for further work.

After FlexPepBind was developed, the program was tested on other published data sets (see reference (*24*) for more details). First, the program was tested on the reactivity of a second "targeted" library data set, published in reference (*11*), that contained peptides whose sequences were biased towards canonical $a_2$ and X sequences. In this library, 29 peptides were MTO and 15 peptides were non-substrates (NON); applying the FlexPepBind protocol yielded an 86% true positive rate and a 12.5% false positive rate (a slightly higher true positive rate than the test set, but also a slightly higher false positive rate). The FlexPepBind protocol was also applied to analyze an additional set of 41 peptides of the form CxxL (a canonical GGTase-I library) where the reactivity with FTase was published (*29*). FlexPepBind yielded an 88% true positive rate and a 12% false positive rate when using a cutoff of -0.4. Figure 3.2 shows the distribution of peptide scores for this CxxL library grouped into the two categories identified by FlexPepBind: binders/substrates (left side) and non-binders/non-substrates (right side). Any substrate peptide with a FlexPepBind score >-0.4 and any non-substrate peptide with a score <-0.4 (indicated by the arrows) contributes to the 12% false positive rate. The opposite results, substrate <-0.4 and non-substrate >-0.4, contribute to the 88% true positive rate. Interestingly, FlexPepBind is able to accurately predict FTase peptide substrates even when the X group is Leu, the canonical GGTase-I X group. Overall,

FlexPepBind gives similar results for predicting FTase substrates tested against a variety of peptide sequences.

*Application of FlexPepBind to all possible CaaX sequences and analysis of STO sequences*

To identify possible additional FTase substrates and to further explore FTase molecular recognition, FlexPepBind was applied to obtain the energy score for all of the 8,000 possible CaaX sequences. The distribution of peptide scores for this analysis is shown in Figure 3.3. Also graphed is the distribution of the MTO, STO, and NON substrates from reference (*11*). Overall, there are 2309 peptides with scores less than the threshold of -0.4 that are predicted to be FTase substrates (29% of the 8,000 possible sequences). In particular, a new proposed sequence motif was identified, -CaaX sequences that contain an acidic (D or E) X residue. For the data sets of known reactivity with FTase from reference (*11*), the scores of MTO and STO pools of peptides are skewed to more negative energy scores, as expected for substrates (Figure 3.3, green and red lines). The FlexPepBind algorithm identifies 47% of the STO peptides as FTase substrates, which is significantly better than the PrePS algorithm, which predicted only 14% of those sequences as substrates for FTase.

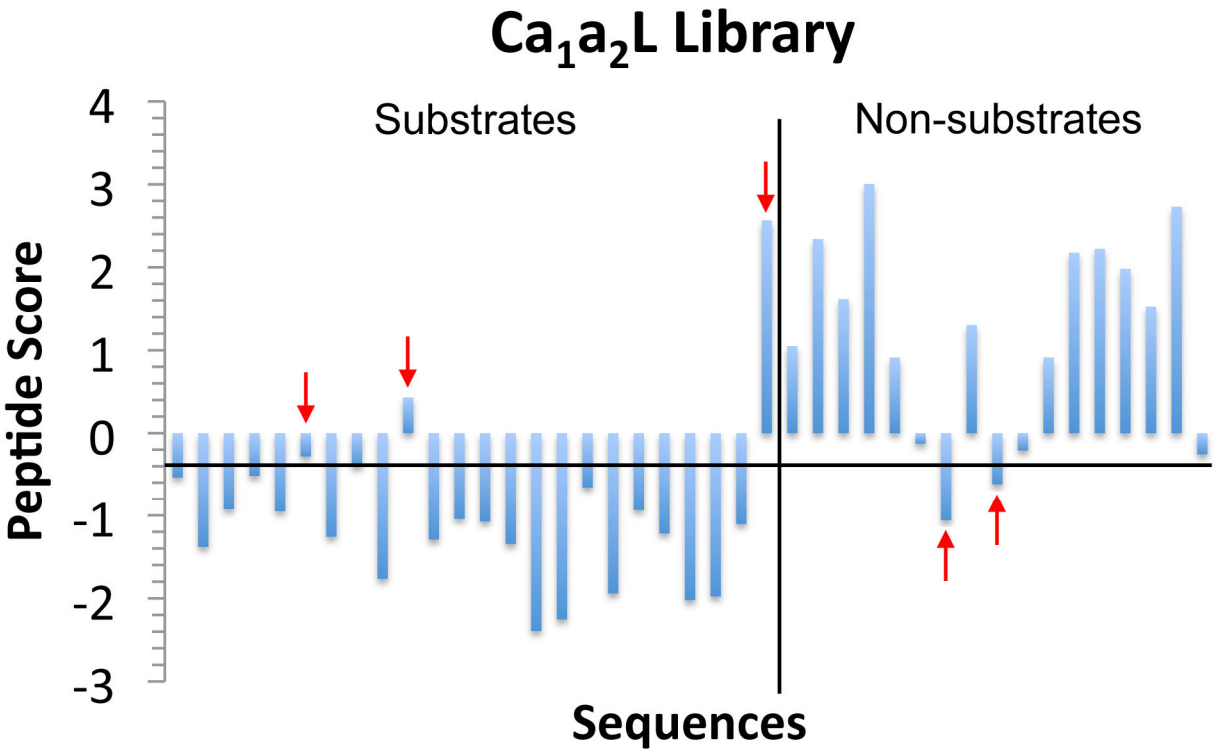


**Figure 3.2. The distribution of FlexPepBind scores of a set of $Ca_1a_2L$ peptides tested with FTase.** The left section of the graph is substrates with FTase while the right section of the graph is non-substrates. The horizontal black line is the -0.4 threshold. In general, FTase substrates have a score less than -0.4. Overall, there is an 88% true positive rate and a 12% false positive rate, with the peptides that contribute to the 12% false positive rate noted by arrows.

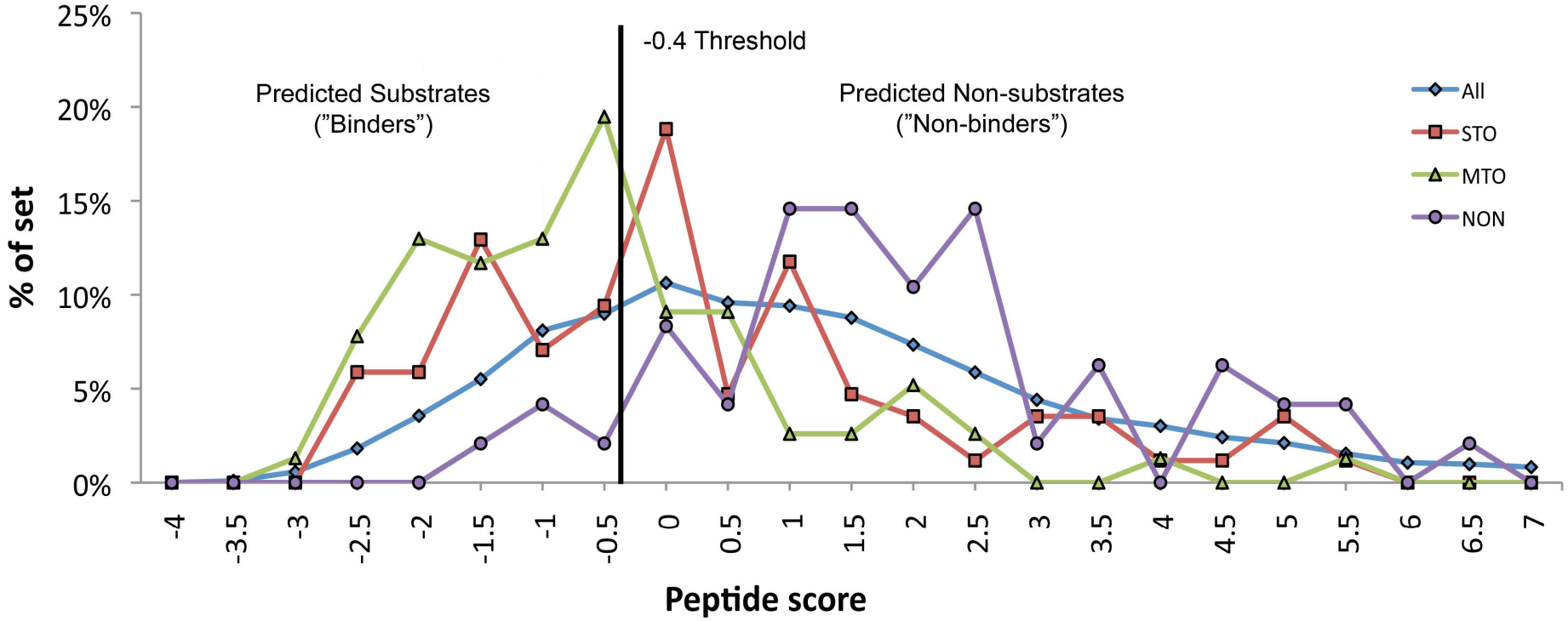


**Figure 3.3. The energy distribution scores as calculated by FlexPepBind of all CaaX sequences, as well as the MTO, STO, and NON data sets from reference (*11*).** The black line indicates the cutoff FlexPepBind score of -0.4. Anything less than this threshold is predicted to be a substrate, while those above are predicted to be non-substrates. The scores for the peptides in the MTO and STO pools are skewed towards more negative energy values (FTase substrates) while the scores for the peptides in the NON pool are skewed towards higher values.

*Testing substrates of FlexPepBind in vitro*

To further test the predictive power of the FlexPepBind algorithm, the reactivity of FTase was tested with 29 peptides having high FlexPepBind scores. The best scoring 13 sequences of previously uncharacterized peptides were chosen as well as 16 high scoring peptides corresponding to real proteins in the human proteome that have not been previously identified as FTase substrates. Seven of these peptides contain an acidic X residue, a class of non-canonical X groups. The peptides were screened for MTO and STO reactivity with FTase and the results are shown in Table 3.1.

FTase catalyzed farnesylation of 26 out of the 29 peptides predicted to be FTase substrates using FlexPepBind, despite the inclusion of non-canonical amino acids, such as X = D or E. Nineteen of these peptides were MTO substrates of FTase, seven were STO substrates, and only three were non-reactive with FTase under the conditions of the screen. Generally, a more negative FlexPepBind score correlates with MTO or STO FTase reactivity (Table 3.1). Furthermore, six of the seven peptides with acidic X residues (bold in Table 3.1) are FTase substrates with three farnesylated under MTO conditions (CYVE, CYIE, CYLD) and three are farnesylated under STO conditions (CYLE, CFIE, CTTE). Only CLFE was a non-substrate. The MTO, STO, and NON pools of substrates had average energy scores of -2.8 ± 0.8, -2.7 ± 0.8, and -1.6 ± 0.1, respectively. The modeled structures of peptide substrates with a negatively charged X group bound to FTase suggest that this X group is stabilized by an electrostatic interaction with FTase residue H149β and hydrogen bonds with W102β and S99β (Figure 3.5). Overall, this analysis further validates shows that FlexPepBind as a robust method

| Sequence | | | | Score | Protein |
|---|---|---|---|---|---|
| **MTO Substrates:** | | | | | |
| C | Y | L | I | -3.96 | |
| C | Y | L | V | -3.60 | |
| C | F | L | V | -3.60 | |
| C | L | I | I | -3.51 | |
| **C** | **Y** | **V** | **E** | -3.43 | |
| **C** | **Y** | **I** | **E** | -3.40 | |
| C | L | I | V | -3.33 | NACHT and WD repeat domain-containing protein 1 (and isoform 2) |
| C | Y | L | L | -3.24 | |
| **C** | **Y** | **L** | **D** | -3.13 | |
| C | W | L | V | -3.01 | |
| C | Y | V | A | -2.88 | Isoform 5 of Zinc finger protein 64 homolog, isoforms 3 and 4. |
| C | F | L | T | -2.74 | |
| C | W | L | S | -2.46 | Isoform 3 of Ankyrin repeat and BTB/POZ domain-containing protein BTBD11 |
| C | C | L | S | -2.37 | Isoform 3 of Intersectin-2 |
| C | K | L | A | -2.06 | Isoform 6 of Target of rapamycin complex 2 subunit MAPKAP1 |
| C | W | T | C | -1.94 | Isoform 3 of Folliculin |
| C | S | L | I | -1.90 | Isoform 5 of Rho GTPase-activating protein 19 |
| C | G | V | G | -1.65 | Putative uncharacterized protein ENSP00000347057 |
| C | V | C | V | -1.12 | |
| **STO Substrates:** | | | | | |
| **C** | **Y** | **L** | **E** | -3.82 | |
| **C** | **F** | **I** | **E** | -3.34 | |
| C | W | V | I | -3.03 | |
| C | A | F | I | -2.62 | Proton-coupled amino acid transporter 1 |
| **C** | **T** | **T** | **E** | -2.14 | Isoform 2 of Decaprenyl-diphosphate synthase subunit 1 |
| C | H | F | H | -2.14 | Isoform 2 of Solute carrier family 7 member 13 |
| C | P | F | F | -1.69 | Homeobox protein ESX1 |
| **Non-substrates:** | | | | | |
| **C** | **L** | **F** | **E** | -1.77 | Isoform 2 of C-type lectin domain family 2 member D (& isoforms 4&6) |
| C | F | D | I | -1.59 | Phosphatidate phosphatase PPAPDC1B |
| C | H | C | I | -1.56 | Growth/differentiation factor 15. |

**Table 3.1. Substrates of FTase predicted by FlexPepBind**. Thirteen top scoring, previously uncharacterized peptides and 16 high scoring peptides corresponding to untested human protein sequences were chosen to be tested for reactivity with FTase *in vitro*. The synthesized peptides of the form dns-TKCxxx and peptides reactive with FTase were identified using a radioactive assay under a stopped MTO or STO reaction that detects $^{3}$H-farnesyl-peptide products fractioned on TLC plates (see Experimental Methods). Overall, FTase catalyzed farnesylation of 19 peptides under MTO conditions, 7 under STO conditions, and 3 were unreactive with FTase under the conditions of the assay. Additionally, 6 out of 7 peptides with an acidic X group (bold) were FTase substrates.

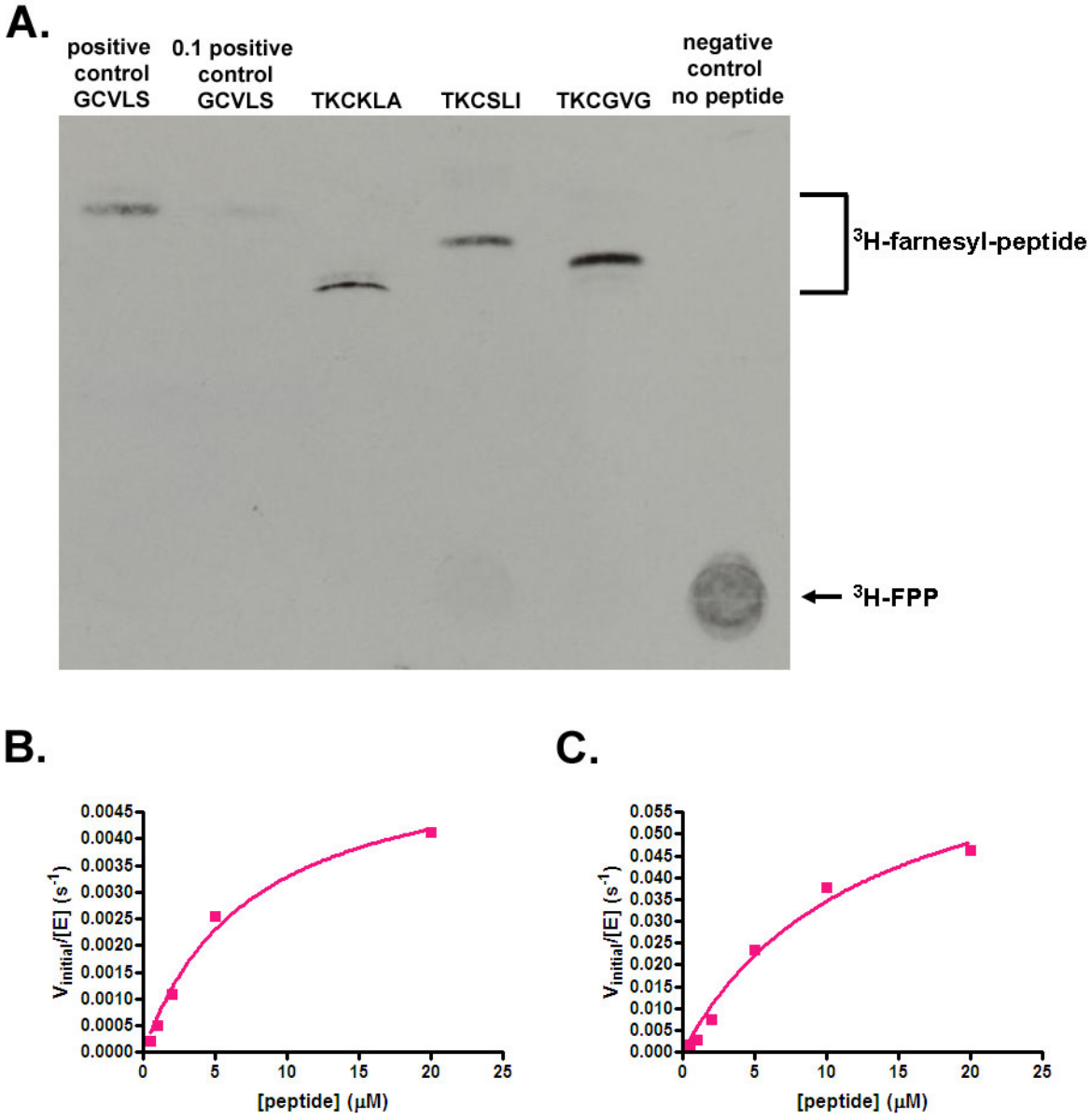


**Figure 3.4. Reactivity of peptides with FTase. A)** Dns-TKCxxx peptides predicted to be substrates by FlexPepBind were screened for reactivity with FTase under MTO and STO conditions using a radioactive assay. For the MTO screen, 3 μM dansylated-peptide (dns-TKCxxx) was incubated with 1 μM $^3$H- FPP and 25 nM rat FTase in 50 mM HEPPSO, pH 7.8, 5 mM TCEP, 5 mM $MgCl_2$ at 25ºC for 2 hours. For the STO screen, 1 μM FTase was incubated with 0.8 μM $^3$H-FPP and 3 μM dns-TKCxxx peptide for one hour at 25ºC in the same buffer. Reactions were fractionated on a TLC plate, and peptides were designated MTO or STO substrates if they were at least 10-20% reacted as compared to a 10 or 20% positive control. A) shows a TLC plate illustrating the separation of $^3$H-FPP and $^3$H-farnesylated peptide under STO conditions; peptides TKCKLA, TKCSLI, AND TKCGVG are reactive with FTase under STO conditions**. B)** and **C)** The steady state parameters were measured for peptides reactive with FTase under MTO conditions. In the same conditions as in A), 0.2 – 40 μM dns-peptide, 20 -100 nM FTase, and 10 μM FPP were incubated together, and the initial velocities of the reaction were measured from the time dependant change in fluorescence ($\lambda_{ex}$ = 340 nm, $\lambda_{em}$ = 520 nm). The Michaelis-Menten equation was fit to the data to determine the values of $k_{cat}/K_M^{peptide}$ (reported in Table 3.2). B) and C) show the plots for dns-TKCYLL and dns-

TKCWTC, respectively.

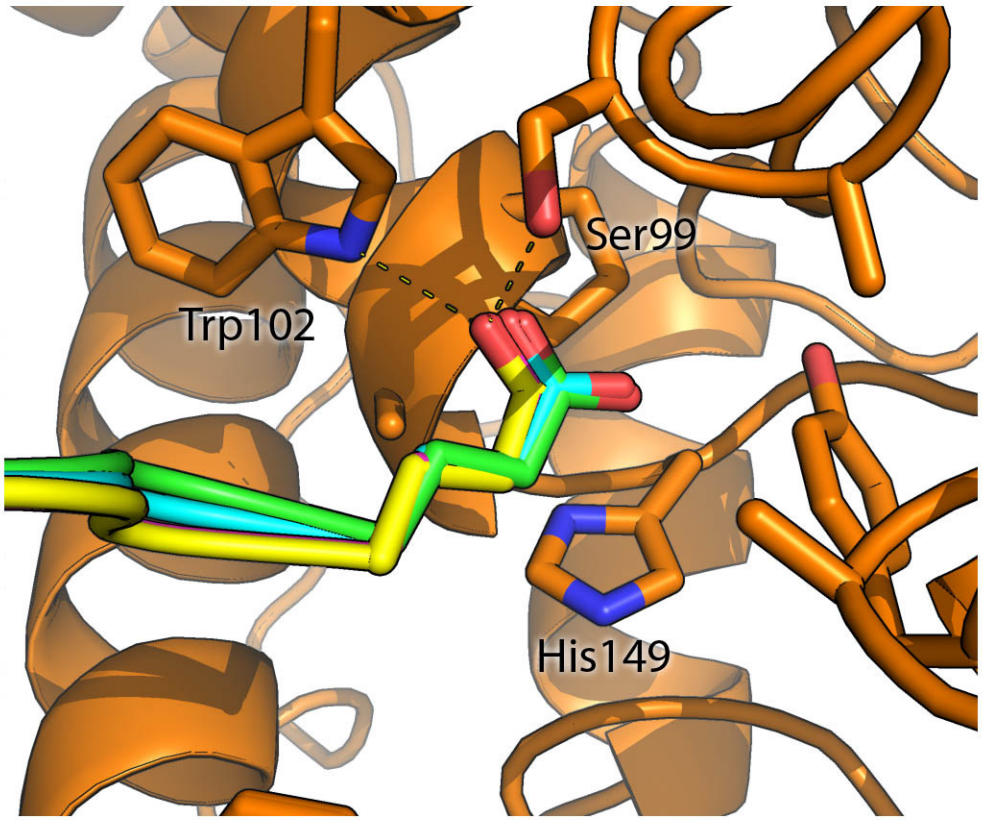


**Figure 3.5. Proposed structure of peptides terminating in acidic residues bound to FTase.** Peptides CYLE (green), CYVE (cyan), CYIE (magenta), and CFIE (yellow) are modeled into the binding pocket of FTase (orange). The negatively charged X group of E may be stabilized by interaction with positively charged H149β and the formation of hydrogen bonds with W102β and S99β.

of predicting FTase substrates, and suggests a new class of FTase substrates containing an acidic X residue.

*Steady state kinetic analysis*

Next, we measured the steady state kinetic parameters of the novel MTO substrates predicted by FlexPepBind with FTase (Tables 3.1 and 3.2) using the enhancement in fluorescence of the dansyl group in the dns-TKCxxx peptides upon farnesylation. Using the assay, we obtained kinetic parameters for 12 of the peptides with MTO activity with FTase. The remaining 5 peptides were below the threshold of activity measured by the fluorescence assay. Table 3.2 shows a summary of the $k_{cat}/K_M^{peptide}$ values measured for farnesylation of these peptides catalyzed by FTase obtained by this analysis.

| Peptide | $k_{cat}/K_M^{peptide}$ ($mM^{-1}s^{-1}$) | Score: |
|---|---|---|
| **CYLV** | 1.43 ± 0.09 | -3.6 |
| **CLII** | 1.6 ± 0.3 | -3.51 |
| **CYIE** | 0.25 ± 0.01 | -3.4 |
| **CLIV** | 3.3 ± 0.2 | -3.33 |
| **CYLL** | 0.8 ± 0.1 | -3.24 |
| **CYVA** | 32 ± 6 | -2.88 |
| **CWLS** | 21 ± 2 | -2.46 |
| **CCLS** | 39 ± 19 | -2.37 |
| **CKLA** | 2.7 ± 0.2 | -2.06 |
| **CWTC** | 6.2 ± 1.0 | -1.94 |
| **CGVG** | 1.0 ± 0.1 | -1.65 |
| **CVCV** | 0.8 ± 0.2 | -1.12 |

**Table 3.2. Steady state parameters for a subset of MTO peptides predicted by FlexPepBind.** $k_{cat}/K_M^{peptide}$ was measured for FTase catalyzed farnesylation of 12 dns-TKCxxxx peptides by incubating 0.2 – 40 μM dns-peptide with 20 -100 nM FTase, 10 μM FPP in 50 mM HEPPSO pH 7.8, 5 mM TCEP and 5 mM $MgCl_2$ at 25ºC. The FlexPepBind score for each sequence is also shown.

## Discussion

*FlexPepBind vs. PrePS*

The FTase substrate reactivity predicted by PrePS, a computational method previously developed by the Eisenhaber lab (*21*), was compared to FlexPepBind. In a study of the reactivity of a peptide library with FTase, PrePS had very low false positive rate for farnesylated peptides, but had a high false negative rate, about 40% (*11*). It also poorly identifies peptides with STO reactivity with FTase (14%). One difference between the methodologies of these algorithms is that PrePS analyzes the C-terminal 15 amino acids to take into account the region upstream of the last four terminal amino acids and the flexibility of the C-terminal tail. To examine this, we compared the PrePS and FlexPepBind scores for the 29 novel peptide sequences identified by FlexPepBind (Table 3.3). In column C, the 30 amino acid C-terminal sequence of the 16 peptides corresponding to human proteins was scored by PrePS. In all cases, PrePS does not predict that these proteins will be FTase substrates (as indicated by "-", "--", or "---" symbols); however, *in vitro* analysis shows that 13 out of 19 sequences are farnesylated by FTase *in vitro* (Table 3.3). If a canonical CaaX (the H-Ras CVLS sequence) replaces the last four amino acids of each sequence (Figure 3.3 column D), then 9 of the 15 sequences are predicted to be substrates by PrePS (as indicated by "+" or "++"). Furthermore, if the 26 upstream amino acids are replaced by the H-Ras sequence, then 8 of the 16 sequences are predicted to be farnesylated (Figure 3.3 column E). This analysis suggests that PrePS is discriminating against these sequences based on both the –CaaX sequence and the upstream sequence. The role of the upstream sequence in substrate selectivity and determinants of MTO and STO is unclear, and the relationship of the

Cxxx sequence and the upstream region is further explored as in Chapter 4 with *in vivo* work. This issue could be further clarified by additional *in vitro* studies to characterize the relationship between the Cxxx sequence, the upstream region, and FTase reactivity by measuring the reactivity of FTase with longer peptides containing varied upstream sequences, such as those shown in Table 3.3.

*Steady state parameters, $K_D$, and FlexPepBind score*

The FlexPepBind algorithms are based on the prediction of the affinity of FTase for each peptide with the assumption that the high affinity peptides would also react readily with FTase. However, the training set was based on a library of peptides that react readily with FTase rather than those that bind tightly. To evaluate the physical parameters that determine the FlexPepBind score we analyzed the correlation of this score with two experimental parameters, $k_{cat}/K_M^{peptide}$ and $K_D$. The $k_{cat}/K_M^{peptide}$ parameter reflects the reactivity of a peptide with FTase•FPP complex and indicates the ability of a peptide to compete with other substrates, such as the situation in the cell. Therefore, this parameter is termed the "specificity constant" (*31*). For FTase, $k_{cat}/K_M^{peptide}$ includes all of the reaction steps up to the first irreversible step in the reaction, so it includes the rate constants of the peptide binding to E•FPP, the FPP conformation change, and the formation of the prenylated peptide and pyrophosphate products (*32-36*). Under these conditions, the release of the diphosphate is the first irreversible step (*33, 34*), so the release of the prenylated peptide does not contribute to the value of $k_{cat}/K_M^{peptide}$. The training set identifies substrates based on reactivity under $k_{cat}/K_M^{peptide}$ (MTO) conditions. Therefore, we examined whether the FlexPepBind score correlates with the value of

| Sequence (A) | Derived from protein (B) | PrePS prediction | | | Score (F) | Activity (G) |
|---|---|---|---|---|---|---|
| | | Full (C) | $x_{26}$-CVLS (D) | H-Ras-Cxxx (E) | | |
| CYLI | | | | - | -3.96 | STO |
| CYLE | | | | - | -3.82 | MTO |
| CYLV | | | | - | -3.60 | MTO |
| CFLV | | | | - | -3.60 | STO |
| CLII | | | | ++ | -3.51 | MTO |
| CYVE | | | | - | -3.43 | MTO |
| CYIE | | | | - | -3.40 | MTO |
| CFIE | | | | - | -3.34 | STO |
| CLIV | | | | ++ | -3.33 | MTO |
| CYLL | | | | - | -3.24 | MTO |
| CYLD | | | | - | -3.13 | MTO |
| CWVI | | | | - | -3.03 | MTO |
| CWLV | | | | - | -3.01 | MTO |
| CYVA | Q9NTW7-3 | - | - | + | -2.88 | STO |
| CFLT | O75037 | -- | - | + | -2.74 | STO |
| CAFI | Q7Z2H8 | -- | + | - | -2.62 | MTO |
| CWLS | A6QL63-3 | - | + | + | -2.46 | MTO |
| CCLS | Q9NZM3-3 | -- | -- | ++ | -2.37 | MTO |
| CTTE | Q5T2R2-2 | -- | - | - | -2.14 | MTO |
| CHFH | Q8TCU3-2 | --- | + | -- | -2.14 | MTO |
| CKLA | Q9BPZ7-6 | - | - | + | -2.06 | MTO |
| CWTC | Q8NFG4-3 | - | ++ | - | -1.94 | None |
| CSLI | Q14CB8-5 | - | + | ++ | -1.90 | MTO |
| CLFE | Q9UHP7-3 | -- | + | -- | -1.77 | STO |
| CPFF | Q8N693 | --- | - | -- | -1.69 | MTO |
| CGVG | A6NHS1 | - | - | + | -1.65 | None |
| CFDI | Q8NEB5 | -- | ++ | -- | -1.59 | None |
| CHCI | Q99988 | -- | + | - | -1.56 | STO |
| CVCV | O75391 | - | + | + | -1.12 | MTO |

**Table 3.3. Peptide sequences, score, PrePS prediction, and *in vitro* activity of peptides. A)** Peptide sequence; **B)** Uniprot identifier of human proteins containing the C-terminal CaaX sequence (*30*); **C)** PrePS (*21*) predictions for the C-terminal 30 amino acids; **D)** PrePS predictions of the 26 amino acids from Column C appended to C-terminal CVLS (H-Ras CaaX sequence); **E)** H-Ras C-terminal region with the last four amino acids replaced with the Cxxx sequence tested here. **F)** FlexPepBind peptides score; **G)** *in vitro* activity with FTase. In summary, Column D shows the amenability of farnesylation of the upstream sequence with a good CaaX sequence and Column E shows the amenability of farnesylation with a good upstream region with Cxxx sequences predicted by FlexPepBind.

$k_{cat}/K_M^{peptide}$. Additionally, $K_D$ is the dissociation constant of the dns-TKCxxx peptide from the E•FPP analog•dns-TKCxxx complex and is indicative of how tightly the peptide binds to this complex. As FlexPepBind is based on an analysis of protein-peptide interactions in the E E•FPP analog•dns-TKCxxx complex, we proposed to see a correlation between the FlexPepBind scores and the affinity measurements. However, the FlexPepBind algorithm only analyzes interactions of the Cxxx sequence.

Figure 3.6 shows a graph of the correlation of $k_{cat}/K_M^{peptide}$ values and FlexPepBind score, with $k_{cat}/K_M^{peptide}$ graphed on a log scale. $k_{cat}/K_M^{peptide}$ values were obtained from this work (Table 3.2, Figure 3.6 triangles) as well as the previously published peptide library studies with FTase (squares, reference (*11*)). Overall, there is little correlation between $k_{cat}/K_M^{peptide}$ and FlexPepBind score, regardless of the origin of the peptide sequence. The reactivity of the peptides generally range between 100-50,000 $M^{-1}s^{-1}$ across the range of FlexPepBind scores (-4-6). The majority of the substrates do have FlexPepBind scores ≤ -0.04 (44 out of 58 peptides). Furthermore, there appears to be "sweet spot" of high reactivity for a few peptides with FlexPepBind score between -2.5 and -3.0; however, there are plenty of peptides with lower reactivity with the range of scores as well. Also, comparing the rate constant for farnesylation under single turnover conditions ($k_{farnesylation}$) from reference (*11*) to the FlexPepBind scores does not correlate with the FlexPepBind scores for either MTO or STO substrates ($r_s$ = -0.26, $p$ = 0.053, not significant, data not shown).

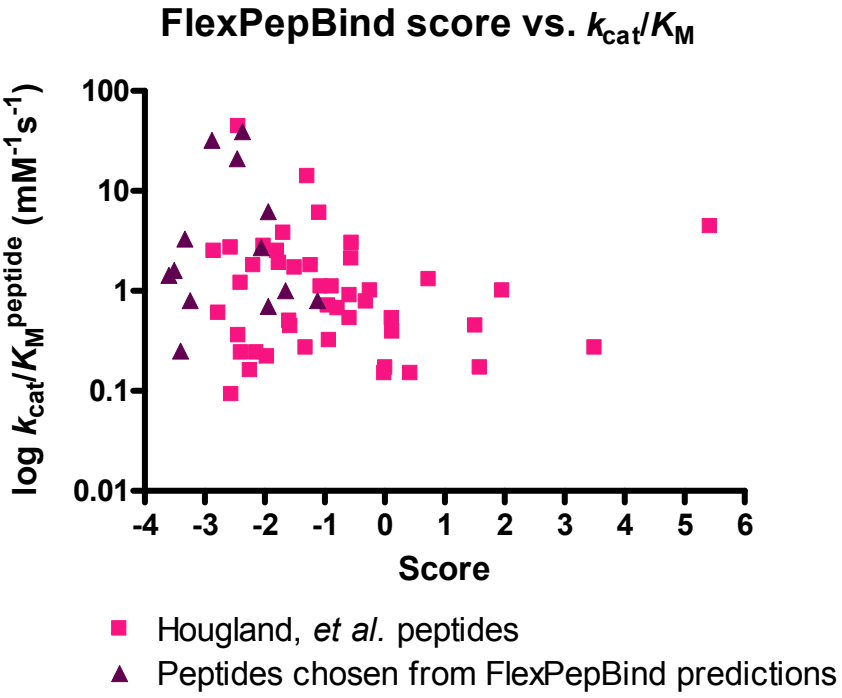


**Figure 3.6. Comparison of $k_{cat}/K_M^{peptide}$ to FlexPepBind scores.** The $k_{cat}/K_M^{peptide}$ values were obtained from reference (*11*) (squares) or were obtained in this work from peptides predicted by FlexPepBind predictions (triangles, see Table 3.2). Overall, there is little correlation between FlexPepBind score and the value of $k_{cat}/K_M^{peptide}$ ($r_s = -0.26$, $p = 0.053$, not significant).

Finally, we analyzed the correlation between log $K_D^{peptide}$ and FlexPepBind score, as shown in Figure 3.7. The $K_D$ values measuring the dissociation of the dns-TKCxxx peptide from the E•FPP analog•dns-TKCxxx complex are taken from reference (*11*). Unexpectedly, there is also no readily identifiable correlation between peptide affinity and FlexPepBind score ($r_s$ = 0.14, $p$ = 0.407, not significant). As FlexPepBind was developed by calculating an "energy score" that correlates with peptide binding to the FTase active site, we were surprised by this result; however there may be several reasons for the lack of a trend. A caveat of the measurement of the $K_D$ values experimentally is that the binding of the peptide was to an enzyme complexed with an unreactive FPP analog. It is possible that the affinity under these conditions is not reflective of the affinity using native substrates (but this is impossible to measure using FPP as the peptide would undergo modification); however, the FlexPepBind algorithm was developed from a structure with this analog (*23*). Further, the binding affinity studies are carried out with peptides containing an N-terminal dansyl moiety, which may dominate binding affinity. These $K_D$ values are also for peptides reactive with FTase. In the future, we would like to measure the affinity of *non-reactive* peptides with FTase complexed to the FPP analog. It is possible that there may be a shift in FlexPepBind scores, where poor scoring peptides may have higher $K_D$ values, such that in general there may be a slightly upward slope in points going from negative scores to positive scores.

These results show that there is no significant correlation between $k_{cat}/K_M^{peptide}$ or $K_D$ of peptides with FlexPepBind scores, and it is unclear what experimental parameter the FlexPepBind score represents. Nonetheless, the data indicate that FlexPepBind is a good method of scoring sequences for farnesylation by FTase, yielding a cutoff for

prediction of substrates (MTO or STO) or non-substrates. However, the significance of the actual value of the FlexPepBind score is currently unclear, as the score does not differentiate well between levels of reactivity. Figure 3.3 illustrates this trend, since there is overlap in the scores with classes of substrates. For instance, when comparing non-substrates (NON, purple line) to MTO substrates (green line), there is overlap in sequence scores from -1.5 to 0. FlexPepBind is clearly not a perfect prediction program, but overall it improves the current computational techniques by enhancing the prediction of the reactivity of FTase with non-canonical C-terminal Cxxx sequences as well as with the STO peptide substrates. This is highly useful for identifying potential *in vivo* substrates as well as understanding what metabolic pathways might contain farnesylated proteins and therefore be affected by inhibitor treatment.

*GGTase-I and computational techniques*

In the future, now that a larger pool of known GGTase-I MTO substrates, STO substrates, and non-substrates is available (Chapter 2), it would be interesting to modify and apply the FlexPepBind program to predicting GGTase-I substrates. Specifically, it would be interesting to compare the score of a particular peptide for FTase vs. GGTase-I to see if the program can enhance the prediction of overlapping substrate selectivity between the two enzymes. We are currently working with Dr. Ora Scheuler-Furman and her lab to score and test GGTase-I to improve GGTase-I substrate predictions.

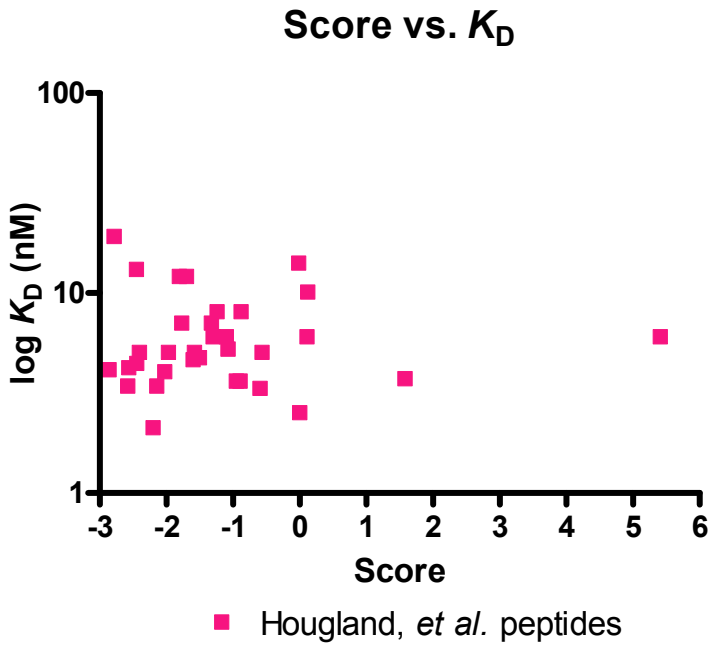


**Figure 3.7. Comparison of log $K_D$ to FlexPepBind scores.** $K_D$ values were obtained from reference (*11*). Overall, there does not appear a correlation between $K_D$ and FlexPepBind scores with FTase ($r_s$= 0.14, $p$ = 0.407, not significant).

## References


(1) Benetka, W., Koranda, M., and Eisenhaber, F. (2006) Protein Prenylation: An (Almost) Comprehensive Overview on Discovery History, Enzymology, and Significance in Physiology and Disease. *Monatshefte für Chemie / Chemical Monthly 137*, 1241-1281.

(2) Casey, P. J., and Seabra, M. C. (1996) Protein Prenyltransferases. *Journal of Biological Chemistry 271*, 5289-5292.

(3) Marshall, C. J. (1993) Protein prenylation: a mediator of protein-protein interactions. *Science 259*, 1865-6.

(4) Casey, P. J. (1994) Lipid modifications of G proteins. *Curr Opin Cell Biol 6*, 219-25.

(5) Zhang, F. L., and Casey, P. J. (1996) Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem 65*, 241-69.

(6) Casey, P. J. (1995) Protein Lipidation in Cell Signaling. *Science 268*, 221-225.

(7) Caplin, B. E., Hettich, L. A., and Marshall, M. S. (1994) Substrate characterization of the Saccharomyces cerevisiae protein farnesyltransferase and type-I protein geranylgeranyltransferase. *Biochim Biophys Acta 1205*, 39-48.

(8) Omer, C. A., Kral, A. M., Diehl, R. E., Prendergast, G. C., Powers, S., Allen, C. M., Gibbs, J. B., and Kohl, N. E. (1993) Characterization of recombinant human farnesyl-protein transferase: cloning, expression, farnesyl diphosphate binding, and functional homology with yeast prenyl-protein transferases. *Biochemistry 32*, 5167-76.

(9) Reiss, Y., Stradley, S. J., Gierasch, L. M., Brown, M. S., and Goldstein, J. L. (1991) Sequence requirement for peptide recognition by rat brain p21ras protein farnesyltransferase. *Proc Natl Acad Sci U S A 88*, 732-6.

(10) Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L., and Gibbs, J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem 266*, 14603-10.

(11) Hougland, J. L., Hicks, K. A., Hartman, H. L., Kelly, R. A., Watt, T. J., and Fierke, C. A. (2010) Identification of Novel Peptide Substrates for Protein Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J Mol Biol 395*, 176-90.

(12) Hougland, J. L., Lamphear, C. L., Scott, S. A., Gibbs, R. A., and Fierke, C. A. (2009) Context-dependent substrate recognition by protein farnesyltransferase. *Biochemistry 48*, 1691-701.

(13) Hancock, J. F., Magee, A. I., Childs, J. E., and Marshall, C. J. (1989) All ras proteins are polyisoprenylated but only some are palmitoylated. *Cell 57*, 1167-77.

(14) Cox, A. D. (2001) Farnesyltransferase inhibitors: potential role in the treatment of cancer. *Drugs 61*, 723-32.

(15) Sepp-Lorenzino, L., Ma, Z., Rands, E., Kohl, N. E., Gibbs, J. B., Oliff, A., and Rosen, N. (1995) A peptidomimetic inhibitor of farnesyl:protein transferase blocks the anchorage-dependent and -independent growth of human tumor cell lines. *Cancer Res 55*, 5302-9.

(16) Young, S. G., Meta, M., Yang, S. H., and Fong, L. G. (2006) Prelamin A farnesylation and progeroid syndromes. *J Biol Chem 281*, 39741-5.

(17) Nallan, L., Bauer, K. D., Bendale, P., Rivas, K., Yokoyama, K., Horney, C. P., Pendyala, P. R., Floyd, D., Lombardo, L. J., Williams, D. K., Hamilton, A., Sebti, S., Windsor, W. T., Weber, P. C., Buckner, F. S., Chakrabarti, D., Gelb, M. H., and Van Voorhis, W. C. (2005) Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. *J Med Chem 48*, 3704-13.

(18) Gibbs, B. S., Zahn, T. J., Mu, Y., Sebolt-Leopold, J. S., and Gibbs, R. A. (1999) Novel farnesol and geranylgeraniol analogues: A potential new class of anticancer agents directed against protein prenylation. *J Med Chem 42*, 3800-8.

(19) Benetka, W., Koranda, M., Maurer-Stroh, S., Pittner, F., and Eisenhaber, F. (2006) Farnesylation or geranylgeranylation? Efficient assays for testing protein prenylation in vitro and in vivo. *BMC Biochem 7*, 6.

(20) Lin, H. P., Hsu, S. C., Wu, J. C., Sheen, I. J., Yan, B. S., and Syu, W. J. (1999) Localization of isoprenylated antigen of hepatitis delta virus by anti-farnesyl antibodies. *J Gen Virol 80 ( Pt 1)*, 91-6.

(21) Maurer-Stroh, S., and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biol 6*, R55.

(22) Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F. L., and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Comput Biol 3*, e66.

(23) Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol 343*, 417-33.

(24) London, N., Lamphear, C. L., Hougland, J. L., Fierke, C. A., and Schueler-Furman, O. (2011) Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. *PLoS Comput Biol 7*, e1002170.

(25) Raveh, B., London, N., and Schueler-Furman, O. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins 78*, 2029-40.

(26) Das, R., and Baker, D. (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem 77*, 363-82.

(27) Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch Biochem Biophys 82*, 70-7.

(28) Cann, A. J. (2003) *Maths from Scratch for Biologists*, John Wiley and Sons, West Sussex, England.

(29) Krzysiak, A. J., Aditya, A. V., Hougland, J. L., Fierke, C. A., and Gibbs, R. A. (2009) Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates. *Bioorganic & Medicinal Chemistry Letters 20*, 767-770.

(30) Bairoch, A., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Argoud-Puy, G., Axelsen, K., Baratin, D., Blatter, M.-C., Boeckmann, B., Bolleman, J., Bollondi, L., Boutet, E., Braconi Quintaje, S., Breuza, L., Bridge, A., deCastro, E., Ciapina, L., Coral, D., Coudert, E., Cusin, I., Delbard, G., Dornevil, D., Duek Roggli, P., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gehant, S., Farriol-Mathis, N., Ferro, S., Gasteiger, E., Gateau, A., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hulo, N., James, J., Jimenez, S., Jungo, F., Junker, V., Kappler, T., Keller, G., Lachaize, C., Lane-Guermonprez, L., Langendijk-Genevaux, P., Lara, V., Lemercier, P., Le Saux, V., Lieberherr, D., de Oliveira Lima, T., Mangold, V., Martin, X., Masson, P., Michoud, K., Moinat, M., Morgat, A., Mottaz, A., Paesano, S., Pedruzzi, I., Phan, I., Pilbout, S., Pillet, V., Poux, S., Pozzato, M., Redaschi, N., Reynaud, S., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.-L., Yip, L., Zuletta, L., Apweiler, R., Alam-Faruque, Y., Antunes, R., Barrell, D., Binns, D., Bower, L., Browne, P., Chan, W. M., Dimmer, E., Eberhardt, R., Fedotov, A., Foulger, R., Garavelli, J., Golin, R., Horne, A., Huntley, R., Jacobsen, J., Kleen, M., et al. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res 37*, D169-74.

(31) Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W.H. Freeman and Company, New York, NY.

(32) Long, S. B., Casey, P. J., and Beese, L. S. (2002) Reaction path of protein farnesyltransferase at atomic resolution. *Nature 419*, 645-50.

(33) Pais, J. E., Bowers, K. E., Stoddard, A. K., and Fierke, C. A. (2005) A continuous fluorescent assay for protein prenyltransferases measuring diphosphate release. *Anal Biochem 345*, 302-11.

(34) Furfine, E. S., Leban, J. J., Landavazo, A., Moomaw, J. F., and Casey, P. J. (1995) Protein farnesyltransferase: kinetics of farnesyl pyrophosphate binding and product release. *Biochemistry 34*, 6857-62.

(35) Pickett, J. S., Bowers, K. E., Hartman, H. L., Fu, H. W., Embry, A. C., Casey, P. J., and Fierke, C. A. (2003) Kinetic studies of protein farnesyltransferase mutants establish active substrate conformation. *Biochemistry 42*, 9741-8.

(36) Pais, J. E., Bowers, K. E., and Fierke, C. A. (2006) Measurement of the alpha-secondary kinetic isotope effect for the reaction catalyzed by mammalian protein farnesyltransferase. *J Am Chem Soc 128*, 15086-7.

# CHAPTER IV

# SPECIFICITY STUDIES OF THE PRENYLATION PATHWAY *IN VIVO*: METHODS TO DETECT *IN VIVO* POST-TRANSLATIONAL MODIFICATIONS[1]

## Introduction

Prenylation, an important post-translational modification, aids in targeting modified proteins to the membrane and helps to promote protein-protein interactions (*1, 2*). Prenylation comes in two forms: the 15-carbon farnesylation and the 20-carbon geranylgeranylation and attachment of the lipid to the C-terminus of proteins is catalyzed by protein farnesyltransferase (FTase) from the lipid donor farnesyldiphosphate (FPP) and protein geranylgeranyltransferase-I (GGTase-I) from geranylgeranyldiphosphate, respectively (*3, 4*). After prenylation, substrates can be further modified *in vivo*. The enzymes zinc metalloprotease Ste24 (ZMPSTE24) or Ras converting enzyme 1 (RCE1) catalyze proteolysis of the three C-terminal amino acids of a prenylated protein, and isoprenylcysteine methyltransferase (ICMT) can catalyze methylation of the carboxy terminus of the substrate (Figure 4.1, (*5-7*)). These additional modifications can aid in targeting the protein to the membrane; however, the full extent of prenylation,

[1] Construction of several of the vectors and the corresponding microscopy was carried out by Jenna Hendershot. The *in vitro* proof of principle experiments were carried out by Megan Novak.

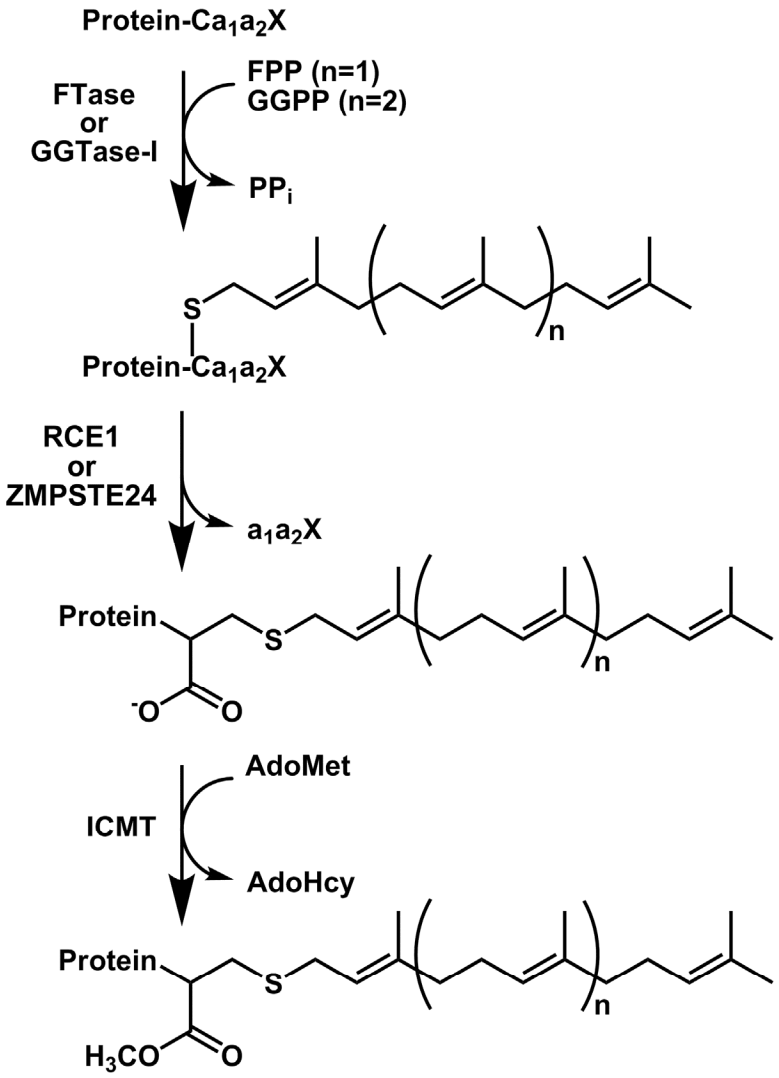


**Figure 4.1. The prenylation pathway.** A protein with a cysteine four amino acids from the C-terminus is modified, with a farnesyl or geranylgeranyl group, catalyzed by either FTase or GGTase-I. Next the last three amino acids of the protein can be proteolyzed catalyzed by RCE1 or ZMPSTE24, and the carboxy terminus can be methylated, catalyzed by ICMT.

proteolysis, and methylation of the prenylated proteome *in vivo* is unknown at this time (*8-12*).

Besides prenylation, the post-translational modification of S-palmitoylation can aid in membrane localization of proteins as well as target proteins to various cellular organelles or lipid domains like lipid rafts (*13-16*). Unlike prenylation, palmitoylation forms a thioester and is therefore readily reversible; the addition of palmitate from palmitoyl-CoA to a cysteine via a thioester linkage can be catalyzed by protein acyl-transferases (PATs) and removal can be catalyzed by protein acylthioesterases (APTs, (*17, 18*)). There is no specific consensus sequence for palmitoylation; however, mass spectrometry and genetic studies have revealed some trends, including that palmitoylation can occur near a prenylation site (*19*) and that PATs have overlapping specificity (*19*). Therefore, understanding which proteins are modified by both the prenyltransferases and the PATs will allow us to understand the role of lipidation in cellular processes.

The specificity of the enzymes in the prenylation pathway has been studied, but more work is needed to further refine the modes of recognition by these enzymes *in vivo*. FTase and GGTase-I were proposed to recognize a canonical "$Ca_1a_2X$" motif at the C-terminus of a substrate protein (*4, 20, 21*), whereby "C" is the cysteine where the prenyl group is attached, "$a_1$" and "$a_2$" are aliphatic amino acids, and "X" is typically methionine, alanine, glutamate, or serine for FTase and leucine or phenylalanine for GGTase-I. Although the CaaX motif describes many substrates, there is evidence that this model excludes a number of important substrates (*12, 22, 23*), and that a pool of substrates undergo modification catalyzed by both enzymes, i.e., "dual substrates" (*24*).

Specificity studies of RCE1 and ZMPSTE24 have been more limited, but it is known these enzymes require a prenylcysteine for recognition and that the substrate selectivity may overlap (*7, 25-31*). Understanding the molecular recognition of these enzymes will aid in determining the modifications that occur *in vivo*.

Methods to determine global prenylation within the cell have been met with limited success thus far. Historically, substrates of FTase and GGTase-I were primarily determined by studying one protein at a time. One method for identifying prenylated proteins is to treat cells with a radiolabeled metabolite, such as mevalonate which is a precursor of FPP and GGPP (*32-34*) and then identify the radiolabeled proteins; however this method yields low signal (*3, 35, 36*). Antibodies have been raised to the farnesyl and geranylgeranyl modifications for detection by Western blotting but these antibodies can cross react with other lipid modifications and are unable to distinguish between the two types of prenyl groups (*37-40*). FPP and GGPP donor analogs, which are specifically recognized by FTase or GGTase-I as substrates, have been utilized to aid in modified protein isolation and identification. These analogs have either immunogenic properties (*41, 42*), an affinity tag (*43*), or a group amenable to bioorthogonal ligation (*8, 44*). Although the techniques using these analogs show promise, the whole complement of prenylated proteins *in vivo* still remains to be identified.

In this work, we are developing methods to identify prenylated proteins *in vivo.* We began using the FPP and GGPP analogs azido-farnesol and azido-geranylgeraniol to label and then identify the labeled proteins. This method led to only a subset of proteins tagged *in vivo*. We then developed a procedure that involves transfecting a vector library into mammalian cells corresponding to potentially modified proteins, using fluorescence

microscopy to observe protein localization, and mass spectrometry to detect modifications. We first generated a library of tagged green fluorescent protein (GFP) constructs attached to C-terminal sequences hypothesized to signal prenylation. We transfected these constructs into several mammalian tissue culture cell lines and observed localization, demonstrating that the GFP localized to different parts of the cell depending on the C-terminal sequence and, presumably, modification status. In the future, the specific modifications on the GFP-fusion proteins can be analyzed by mass spectrometry techniques.

## Experimental Methods

*Vector construction*

To begin preparing a library of mammalian expression vectors, the $His_6$-EGFP-TEV-$X_{11}$-Cxxx gene was constructed using two rounds of primer extension polymerase chain reaction (PCR) to add a BamHI site and a $His_6$ tag to the 5' end of the gene and a TEV cleavage site followed by a C-terminal tail flanked by a KasI site and an EcoRI site to the 3' end of the EGFP (Figure 4.2 A). The KasI site and EcoRI site allow for easy interchange of the C-terminal amino acids of the gene. Unless noted, all enzymes used in this work were purchased from New England BioLabs (Ipswich, MA). For the parent vector construction, the overall gene was $His_6$-EGFP-TEV-myc. The myc sequence is EQKLISEEDL; this sequence can be detected by Western blotting with an anti-myc antibody and is not prenylated, serving as a negative control. The template for the EGFP sequence was the CS2+EGFP vector and was a gift from Dr. Anne Vojtek's lab. A FailSafe PCR kit (Epicentre, Madison, WI), which contains multiple solutions that vary

PCR conditions, was used with the following thermocycler program for the first round of primer extension PCR: 1) 94ºC for 2 minutes; 2) 30 cycles of 94ºC for 45 seconds, 55ºC for 45 seconds, 72ºC for 1 minute; 3) 72ºC for 4 minutes; and 4) hold at 4ºC. Each tube contained 200 ng of the CS2+eGFP template, 50 μM primer A (Figure 4.3), and 50 μM primer B (Figure 4.3) in 1X Fail Safe Enzyme Mix and 1X Fail Safe Buffer (each A through L). This method yielded the correct DNA fragment around 850 base pairs with FailSafe buffers A, B and C, and these PCR products were gel purified. This PCR product encodes the BamHI site, the $His_6$ tag, the EGFP gene, the TEV site, and a portion of the myc tag. A second round of primer extension PCR was carried out using PFU Turbo enzyme (0.4 U/μL, Stratagene, Santa Clara, CA), 1 X PFU buffer, 0.8 mM dNTP mix, 0.4 ng/μL PCR product, and 2.5 ng/μL primers C and D (Figure 4.3). The same thermocycler program as above was used and a band around 900 bases was seen on a gel which encodes the entire gene as seen in Figure 4.2 and 4.3. The band was cut out, gel purified and the inserts were double digested with EcoR1 and BamHI, and then ligated using T4 DNA ligase into the pcDNA4/TO mammalian expression vector (Invitrogen) digested with the same restriction enzymes (Figure 4.2 B, with the $Ras_{15}$ tail as an example). The vector sequence was verified by sequencing (DNA Sequencing Core, University of Michigan).

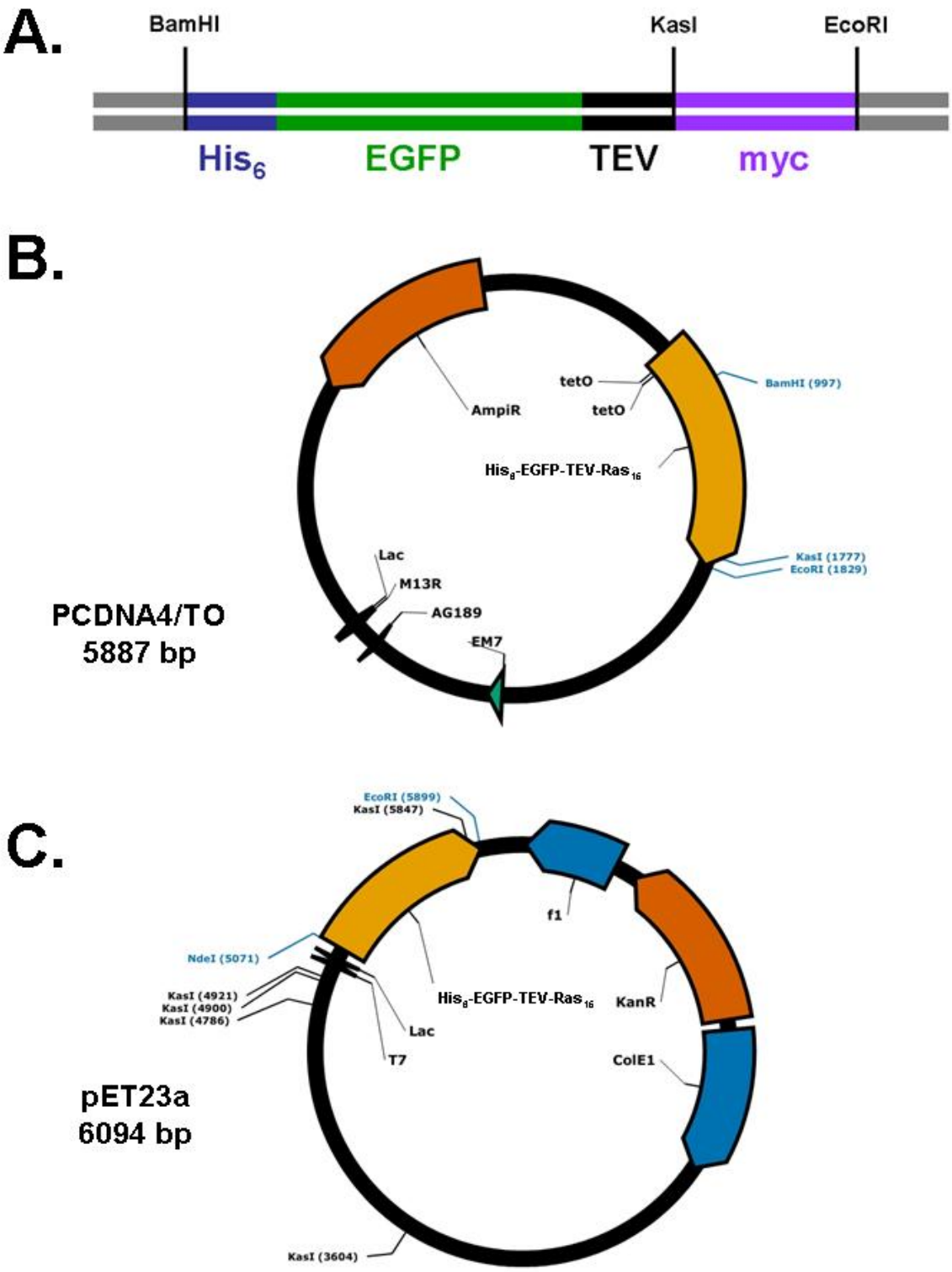


**Figure 4.2. Gene insert scheme and plasmid maps. A)** To construct a parent vector for the $His_6$-EGFP-TEV-$X_{11}$-CaaX expression library, primer extension PCR was used to generate the above insert, flanked by a BamHI site and an EcoR1 site. A KasI site was added between the TEV protease site and the C-terminal tail to allow for the C-terminal sequences to easily be swapped. The plasmid maps for the $His_6$-EGFP-TEV-$Ras_{15}$ gene are shown in the PCDNA4/TO mammalian vector **(B)** and the pET23a *E. coli* expression vector **(C)**. The maps were generated using the PlasmaDNA program.

## $His_6$-EGFP-TEV-myc gene sequence

ATGAAACATCACCATCACCATCACGGCTCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCG
GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT
CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA
AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCC
GAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTC
AAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACAT
CGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGGGCGGTGGAGAGAATCTTTATTTTCAGGGCGCCGAGCAGA
AGCTCATCAGCGAGGAGGACCTGTGAATTC

**Key**:
1. **ATG**AAA is the sequence for start codon (coding methionine) and a lysine linker
2. CATCACCATCACCATCACGGCTCC is the $His_6$-tag plus and an extra linker sequence of glycine and serine
3. ATGGTGAGC …. is the EGFP gene
4. GAGAATCTTTATTTTCAGGGCGCC is the TEV protease site sequence and an alanine linker
5. GAG……CTG TGA is the C-terminal amino acids of gene encoding the myc tag and the stop codon
6. GGCGCC is the KasI site
7. GAATTC is the EcoRI site
8. The lavender bases correspond to vector sequence.

| | |
|---|---|
| Primer A | TGCAGGATCCATGAAACATCACCATCACCATCACGGCTCCATGGTGAGCAAGGGCGAG |
| Primer B | TGATGAGCTTCTGCTCGGCGCCCTGAAAATAAAGATTCTCTCCACCGCCCTTGTACAG |
| Primer C | TGCAGGATCCATGAAACATCA |
| Primer D | CCTTGAATTCTGATCACAGGTCCTCCTCGCTGATGAGCTTCTGCTCGGC |

**Figure 4.3. The DNA sequence of the $His_6$-EGFP-TEV-myc gene and primers for primer extension PCR.** Two rounds of primer extension PCR using primers A-D (table) following the procedure detailed in the experimental methods section yielded the correct product for the $His_6$-EGFP-TEV-myc gene. The sequence for the gene and the key for the color coding of each portion of the gene are above.

Next, a library of expression pcDNA4/TO-EGFP expression vectors was created by altering the C-terminal tails, and part of this work was done in collaboration with a rotation student, Jenna Hendershot. The parent vector was double digested with KasI and EcoRI to create two sticky ends. The C-terminal tail insertions were constructed from two synthetic complementary DNA oligonucleotides (Integrated DNA Technologies), such that when two strands are annealed together, the oligonucleotides have overhangs that correspond to KasI and EcoRI sites sot that this sequence can be readily ligated with the EcoRI/KasI digested parent vector (see Figure 4.8). For this procedure, oligonucleotides were dissolved in water and diluted to 5 pmols/μL in 1X Quick Ligase buffer. The oligonucleotides were annealed by heating to 94ºC and slowly cooling them using a thermocycler and the program shown in Table 4.1. The annealing reactions were then diluted 100-fold in 1X Quick T4 DNA Ligase buffer and water. The ligation reaction included 50 pmols total of the annealed insert was incubated in 1X Ligase buffer, 50 ng of digested pCDNA4/TO $His_6$-EGFP-TEV-myc vector, and 1 μL Quick T4 Ligase (per 20 μL). The ligation reaction was incubated at room temperature for 5 minutes and then stored on ice until it was transformed into XL1-Blue chemically competent or Z-competent *E. coli* cells. If the cells used were chemically competent, 5 μL of the ligation mix was added to 100 μL of cells, incubated on ice 30 minutes, heated for 45 seconds at 42ºC, and incubated on ice for 2 minutes. Then 250μL of 2xYT media was added to the tubes and incubated 37ºC for 1 hour. Finally, 100 μL of cells were added to a pre-warmed LB agar plate with 100 μg/mL ampicillin and incubated at 37ºC overnight. If the cells were Z-competent, 5 μL of the ligation mix was added to 100 μL of cells and incubated on ice for 20 minutes. Then, 100 μL of the cells were added to LB agar plate with 100

ug/mL ampicillin and incubated at 37ºC overnight. The cells from single colonies on the plates were inoculated into 5 mL of 2xYT media with 100 μg/mL ampicillin, grown overnight at 37ºC shaking at 225 rpm, and then the DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen) or a Wizard® Plus Minipreps DNA Purification System (Promega). The sequences of the EGFP fusion genes were verified by sequencing (DNA Sequencing Core, University of Michigan).

| Segment | Temperature | Time | Number of Cycles |
|---|---|---|---|
| 1 | 94ºC | 5 minutes | 1 |
| 2 | 94ºC<br>80ºC | 30 seconds<br>5 minutes | 2 |
| 3 | 72ºC | 7 minutes | 1 |
| 4 | 66ºC | 7 minutes | 1 |
| 5 | 60ºC | 3 minutes | 1 |
| 6 | 52ºC | 3 minutes | 1 |
| 7 | 46ºC | 3 minutes | 1 |
| 8 | 37ºC | 20 minutes | 1 |
| 9 | 4ºC | hold | |

**Table 4.1. Annealing program.** The thermocycler (Eppendorf) program to anneal two complementary DNA oligonucleotides together.

Additionally, the $His_6$-EGFP-TEV-$Ras_{15}$ gene was cloned into the *E. coli* expression vector pET24a. The $Ras_{15}$ sequence is the C-terminal 15 amino acids of DESGPGCMSCKCVLS in the H-Ras gene (DNA sequence: GATGAGAGTGGCCCCGGCTGCATGAGCTGCAAGTGTGTGCTCTCCTGA) This work was done in collaboration with an undergraduate student, Megan Novak. Using

similar primer extension PCR techniques as outlined above catalyzed by PFU Turbo polymerase, BamHI and NdeI sites were added to the 5' end of the gene while an EcoRI site was added to the 3' end of the gene. After PCR amplification, the DNA band was cut out, gel purified, and digested with NdeI and EcoRI, it was ligated with the NdeI/EcoRI digested pET23a vector catalyzed by Quick T4 DNA Ligase (Figure 4.2C). The sequence of the $His_6$-EGFP-TEV-$Ras_{15}$ gene was verified by DNA sequencing (DNA Sequencing Core, University of Michigan).

Additionally, dual mammalian expression vectors were constructed that contain genes for both FTase and $His_6$-EGFP-TEV-$X_{11}$-CaaX sequences (see Figure 4.14). Previously, Dr. James Hougland constructed a mammalian expression vector starting form the pACT vector that contained the two subunits of FTase either under the CMV or SV40 promoters and a red fluorescent protein under the other promoter. The plasmid pCAF1 contains the FTase gene under the SV40 promoter (with the Tag-RFP gene under the CMV promoter) and pCAF2 plasmid contains the FTase gene under the CMV promoter (with the TagRFP under the SV40 promoter). Primer extension PCR was used to add restriction sites to the EGFP genes. To insert the $His_6$-EGFP-TEV-$X_{11}$-CaaX gene into pCAF1, NheI and BamHI sites were added to the 5' end and EcoR1 and NotI sites were added to the 3' end. Similarly, to insert the EGFP fusion gene into pCAF2, HindIII and BamHI sites were added to the 5' end of the gene and EcoRI and SacII sites were added to the 3' end. Insert 1 and the pCAF1 vector were digested with NheI and NotI while the insert 2 and pCAF2 were digested with HindIII and SacII. These digested DNA fragments were gel purified and then insert 1/pCAF1 and insert 2/pCAF2 were incubated with Quick T4 DNA Ligase to ligate the genes with the vector. These plasmids were

transformed into XL-Blue *E. coli* cells, grown, purified, and sequenced as described above.

*Tissue culture*

HeLa, HEK-293 and NIH-3T3 cells were purchased from ATCC (Manassas, VA) and unless otherwise noted, all medium and supplements were purchased from Gibco (Life Technologies, Grand Island, NY). HEK-293 and HeLa cells were grown in 12 mL DMEM medium supplemented with 10% FBS in flasks at 37ºC in 5% $CO_2$, and NIH/3T3 cells were grown in 12 mL DMEM medium with 10% calf serum in flasks at 37ºC in 5% $CO_2$. If needed, the medium was also supplemented with 1% penicillin/streptomycin. To passage cells, the medium was removed and 2 mL of 0.25% (w/v) Trypsin-EDTA was incubated with the cells for 10 minutes at 37ºC to remove the cells from the plate. Then the cells were diluted to 12 mL in fresh medium in a flask. For the analog work, HeLa cells were split into 6-well plates, with 50,000 cells per well. After growing for 48 hours, the cells were incubated for 24 hours with either 40 μM azido-farnesol or azido-geranylgeraniol, DMSO, or nothing. The medium was removed and cells were lysed by addition of 300 μL lysis buffer (1.0% NP-40, 150 mM NaCl, and 50 mM Tris pH 8.0 with a mini protease inhibitor tablet (1 tablet/10 mL buffer, Roche)) to each of the wells and rocking the plates for 30 minutes at 4ºC. The lysates were transferred to chilled microcentrifuge tubes, centrifuged at 12,000 rpm at 4ºC for 20 minutes to remove the cellular debris, and the lysates were stored at -20ºC.

*Click chemistry, fluorescence scanning, and Western blotting*

Click chemistry of azido-farnesol or azido-geranylgeraniol treated lysates was carried out using the Click-iT® Kit (Invitrogen, Carlsbad, CA) to conjugate the azido modified proteins to tetramethylrhodamine-alkyne (TAMRA-alkyne, Invitrogen, Carlsbad, CA). The analogs were obtained as a gift from Invitrogen. A Bradford assay was used to normalize the protein concentration in each sample. The click reaction, which uses a copper catalyst to form a triazole conjugate from the reaction of an azide with an alkyne (Figure 4.5), was carried out as outlined in the kit protocol. To prepare samples for SDS-PAGE analysis, a methanol/chloroform extraction was used to remove the reaction components and precipitate the proteins. The samples were dried and resolubilized in SDS-PAGE loading buffer and fractionated on 12% Tris-HCl gels.

To visualize the fluorescently labeled proteins, a Typhoon phosphoimager was used. The gel was scanned using the green laser (532 nm excitation, 580 nm emission). The Typhoon PMT was set between 600 and 800 V and 200 pixels were used with normal to medium sensitivity.

Western blotting was also used in some cases to visualize the azide modified proteins. The anti-TAMRA antibody (Invitrogen) and anti-rabbit-HRP (Thermo) were used as primary and secondary antibodies. SDS-PAGE, transferring proteins to nitrocellulose membranes from the gel, Western blotting, and film developing were carried out using standard conditions (*45-47*).

*Transfections and fluorescence microscopy*

HeLa, HEK-293, and NIH/3T3 cells were used for transfections. For all transfections, cells were split, allowed to grow for 24 hours, transfected, and then visualized 24 hours later by microscopy (or if necessary, Western blotting). There was no improvement in expression if the cells were grown 48 hours after transfection. For the analog work, the Effectene reagent (Qiagen) was used for transfections of the myc-Rheb protein (in the pRK5 vector, gift from Dr. Kun-Liang Guan). For a 6-well plate with 80% confluent cells at the time of transfection, 0.4 μg DNA: 3.2 μL enhancer and 10 μL Effectene were added to each well and cells were grown at 37ºC for 24 hours. For a negative control, a "mock" transfection was carried out using Effectene and the enhancer, but without addition of the DNA. To enhance the EGFP-fusion expression in each of the cell lines, the FuGENE 6 reagent (Promega) was used. For most pcDNA4/TO based plasmids, the optimal FuGENE 6 to DNA ratio was 6:1, although with some constructs the transfection efficiency was higher using a 3:2 FuGENE 6:DNA ratio. Generally, cells were split to 40,000 – 80,000 cells per well for a 12-well plate and 100,000 cells per well for a 6-well plate 24 hours before transfection and the cells were grown for 24 hours after transfection.

Fluorescence microscopy was used to visualize the localization of the EGFP fusion proteins. A Nikon Eclipse TE2000 U inverted epifluorescence microscope with the MetaMorph software (Molecular Devices) was used to obtain fluorescent images. The 470 nm emission and a 525 emission cubes were used with a 5-100 ms exposure to obtain EGFP fluorescent protein localization images.

*In vitro farnesylation reaction: labeling a peptide*

Rat FTase was recombinantly expressed in BL21(DE3)•pET23aPFT cells and purified as described previously (*22, 48, 49*). The dns-SKTKCVIM peptide (10 μM) was incubated with 40 nM FTase, 10 μM FPP, 50 mM HEPPSO pH 7.8, 5 mM TCEP, and 5 mM $MgCl_2$ at 25ºC for 3.5 hours. The peptide was estimated to be greater than >90% farnesylated according to the change in fluorescence signal. The farnesylated peptide was stored at -20ºC.

*Mass spectrometry of a farnesylated peptide*

In collaboration with Dr. Eric Simon and Dr. Phil Andrews at the University of Michigan, mass spectrometry was used to analyze the farnesyl modification on the prenylated peptide dns-SKTKC(F)VIM. Contaminants were removed form the peptide solution using a ZipTip (Millipore). A Maldi-TOF-TOF instrument was used to detect a farnesyl-cysteine residue with a 307 m/z.

*Recombinant $His_6$-EGFP-TEV-$Ras_{15}$ expression and purification*

The $His_6$-EGFP-TEV-$Ras_{15}$ gene in the pet23a vector was transformed into BL21(DE3) cells and grown in a 1 L culture of autoinduction media at 30ºC for 24 hours. Cells were pelleted and resuspended in 50 mM Tris pH 8.0 and 100 mM KCl. Cells were lysed and clarified as described in Chapter 2.

A batch purification method with TALON cobalt affinity resin (Clontech, Mountain View, CA) was used to purify the protein. In short, the lysate was incubated

with resin (1 mL of packed resin per 15 mg total protein as estimated by absorbance at 280 nm) for 20 minutes agitating at room temperature and washed three times with 10 column volumes of Buffer A (50 mM Tris pH 8.0, 100 mM KCl, and 2 mM imidazole). To elute the protein, the resin was transferred to a gravity column, washed with 5 column volumes of Buffer A, and eluted in fractions with 5 column volumes of Buffer B (50 mM Tris pH 8.0, 100 mM KCl, and 200 mM imidazole). Fractions were analyzed by SDS-PAGE and Coomassie blue staining, pooled, concentrated using 10,000 MWCO Amicon filter devices (Millipore), and dialyzed against 2 L of 20 mM Tris pH 8.0 for about 2.5 hours at 4ºC and then 4 L of 20 mM Tris pH 8.0 with 1 mM TCEP overnight. The protein was then separated into aliquots, frozen using a Nalgene "Mr. Frosty" Freezing device containing isopropanol (which cools the samples 1ºC/min when placed in -80ºC freezer), and stored at -80ºC.

*Digestion with TEV protease*

Conditions for proteolysis of the recombinantly purified $His_6$-EGFP-TEV-$Ras_{15}$ protein were optimized by Megan Novak. The best results were obtained by incubating 588 μg of purified $His_6$-EGFP-TEV-$Ras_{15}$ with 360 μg TEV protease in 50 mM Tris pH 8.0 and 1 mM TCEP for 3 hours at room temperature. Purified $His_6$-EGFP-TEV-$Ras_{15}$ alone and TEV protease alone were used as negative controls. To verify digestion by a gel shift, samples were run on a 12% SDS-PAGE gel and stained with Coomassie blue dye.

## Results

*Azido FPP and GGPP analogs to detect in vivo prenylated proteins*

Prenylated proteins are difficult to detect using traditional radio-labeling methods. A method that has been gaining popularity to detect prenylation and other post-translational modifications (PTM) is to incorporate functional groups into the PTM that are amendable to bio-orthogonal conjugation chemistry (also called "tagging-via-substrate" (*8*)). Staudinger ligation or "click" chemistry reactions allow for the selective labeling of an azide group to an alkyne (*50, 51*). We chose to use the "click" chemistry approach, as it is conveniently carried out at room temperature and in aqueous solution (*51*).

FTase and GGTase-I have been previously shown to catalyze the modification of substrates onto proteins using azido-FPP or azido-GGPP (*8, 44*) (Figure 4.4). For labeling in cells, the azido-farnesol and azido-geranylgeraniol readily cross the cellular membrane and are converted to FPP or GGPP *in vivo (8, 52)*. After cell lysis, the azido-modified proteins are labeled using tetramethylrhodamine-alkyne (TAMRA-alkyne, Figure 4.5) which can be detected by fluorescence or Western blotting.

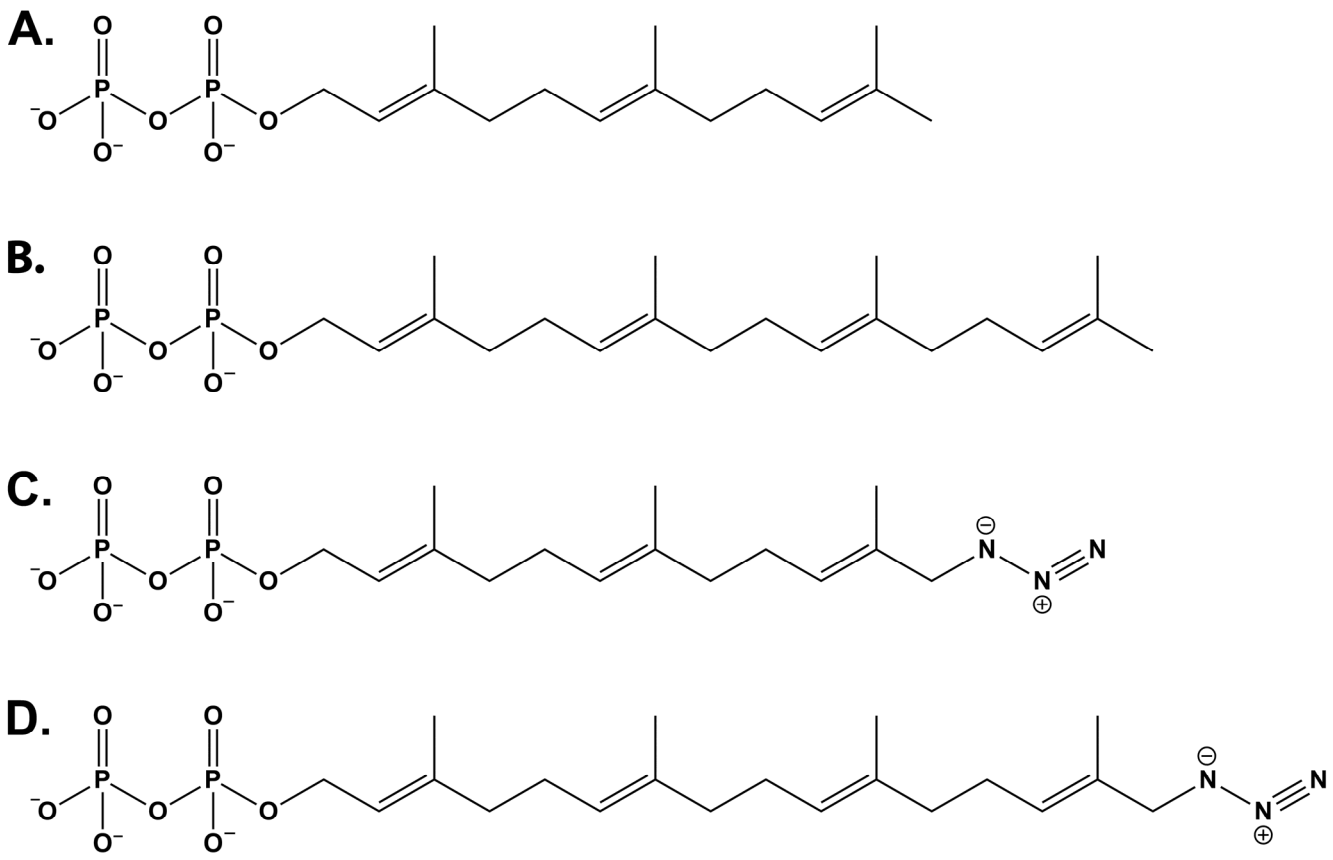


**Figure 4.4. Structures of lipid substrates and analogs.** A) farnesyl diphosphate (FPP); B) geranylgeranyl diphosphate (GGPP); C) azido-FPP; D) azido-GGPP.

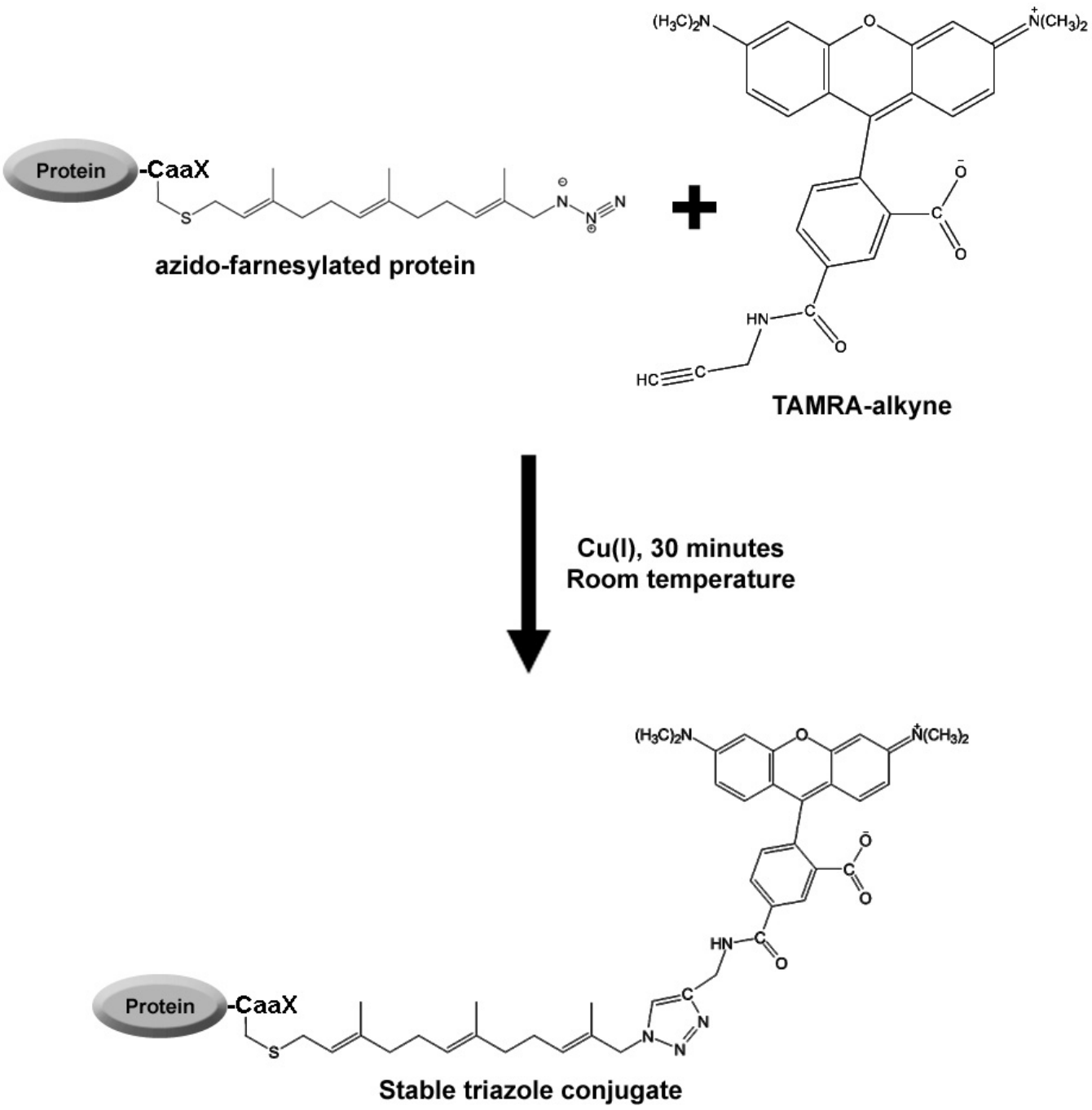


**Figure 4.5. The "click" reaction.** FTase catalyzed the labeling of proteins *in vivo* with the azido-farnesyl analog. In the presence of copper(I), the azido-farnesyl group on the proteins reacts with an alkyne attached to tetramethylrhodamine (TAMRA) to form a stable triazole conjugate.

A number of cellular proteins were detected using this method (Figure 4.6A). However, the background was high and a quantity of known prenylated proteins of low molecular mass was not observed. Figure 4.6 A) shows an anti-TAMRA Western blot of azido-farnesol treated HeLa cells with or without a myc-Rheb transfections. Rheb is a known substrate of FTase (*53*), so it should become azido-modified. Overall, slightly more bands corresponding to farnesylated proteins are observed in azido-farnesol treated cells (lane 3) compared to the control (lane 1). Furthermore, the analog was transferred to a known FTase substrate (myc-Rheb) that was overexpressed from a plasmid transfected into the cells (lane 4). A high background is observed in the control lane. This experiment was then repeated using either azido-farnesol ($N_3$-FOH) or azido-geranylgeraniol ($N_3$-GGOH) (Figure 4.6 B). In general, the Western blots did not indicate that additional proteins were labeled in the azido-analog treated cells (lanes 3 and 9). Fluorescence scanning was used to analyze labeling (Figure 4.6 B) and the method was more sensitive and faster. Lanes 6 and 12 show that proteins are modified with azido-farnesyl and azido-geranylgeranyl modifications (arrows) compared to the untreated cells; however, there is still high background. It was promising that low molecular weight proteins that likely correspond to known prenylated proteins such as the GTPases were modified with azido-geranylgeranyl groups. This behavior was not observed for the azido-farnesol modified group in our experiments. As this was not a robust method to detect prenylated proteins, we chose to employ a different method to understand the specificity of FTase and GGTase-I *in vivo*.

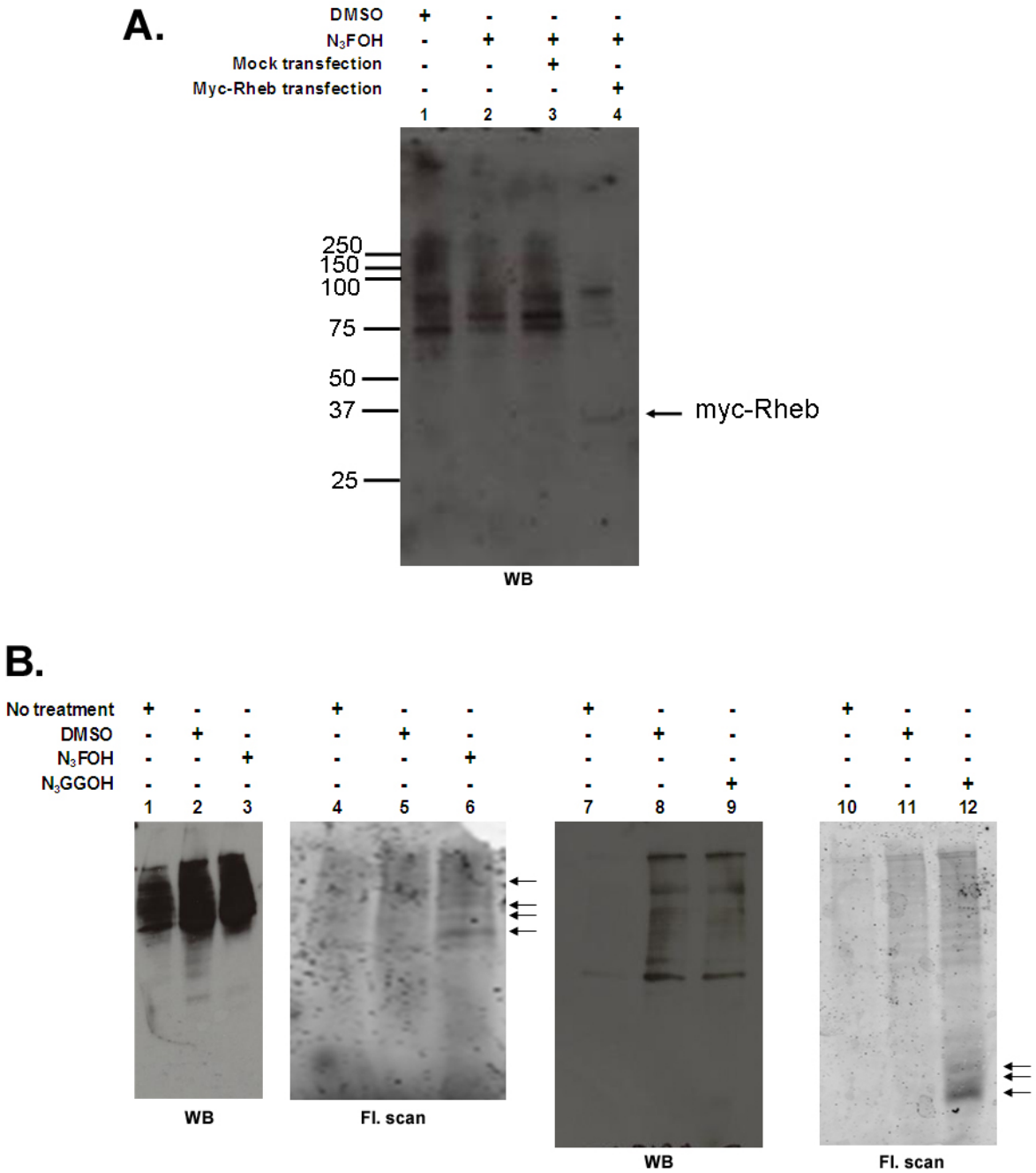


**Figure 4.6. Fluorescence scans and Western blotting. A)** HeLa cells were treated with 40 μM azidofarnesol ($N_3$-FOH) or DMSO for 24 hours before cell lysis. Using "click" chemistry, azido-farnesylated proteins were conjugated to tetramethylrhodamine-alkyne (TAMRA-alkyne). The TAMRA-farnesyl conjugated proteins were fractionated using SDS-PAGE and visualized using Western blotting with an anti-TAMRA antibody. There is a slight increase in bands observed in the $N_3$FOH treated cells (lanes 2-4) compared to the DMSO treated control (lane 1). The overexpressed myc-Rheb protein was also labeled (arrow, lane 4). **B)** HeLa cells were treated with nothing, DMSO, 40 μM azidofarnesol, or 40 μM azido-geranylgeraniol ($N_3$-GGOH) and "click" chemistry was used to conjugate the azide to TAMRA-alkyne. The proteins are fractionated using SDS-PAGE and visualized by either anti-TAMRA Western blot (lanes 1-3 and 7-9) or fluorescence scans (lanes 4-6 and 10-12, excitation 532 nm, emission 580 nm). In this case, little increase in labeling of cells was observed by Western blot in the lysates treated with $N_3$-FOH or $N_3$-GGOH compared to the control (lanes 2/3 and lanes 8/9). The fluorescence scans look more promising as several protein bands are apparent in $N_3$-FOH or $N_3$-GGOH lysates compared to the control (lanes 5/6 and 11/12, respectively; arrows indicate bands of labeled proteins).

*A new method: transfections, microscopy, and mass spectrometry*

We developed a new and hopefully more informative method to test the specificity of the prenylation pathway *in vivo*, by combining transfection of a library of expression vectors, fluorescence microscopy, and mass spectrometry techniques. Figure 4.7 shows the overall scheme. Libraries of mammalian expression vectors encoding proteins with variable CaaX sequences are transfected into various mammalian cells. These expression vectors contain a $His_6$ tag appended to the N-terminus of the enhanced green fluorescence protein (EGFP). Attached to the C-terminus of this EGFP are 15 amino acids that correspond to a proposed prenylated protein that contains both the putative CaaX sequence and upstream regions, since it is proposed that the upstream flexible linker region may contribute to selectivity. Between the C-terminal sequence and the EGFP is a TEV protease cleavage site, so overall, the protein is $His_6$-EGFP-TEV-$X_{11}$-CaaX. After this plasmid is transfected into mammalian tissue culture cells and the protein is expressed, the localization of fusion protein is observed by fluorescence microscopy, which may indicate its prenylation status. Additionally, the protein can be purified from the lysates by utilizing the $His_6$ tag and, potentially, the 15 amino acid modified tail can be cleaved from the protein by TEV protease. The modification status of either the entire protein or the cleaved peptide can then be determined using mass spectrometry. This analysis provides information about whether a given CaaX sequence is prenylated, proteolyzed, and/or methylated (Figure 4.1) as well as determining whether a farnesyl, geranylgeranyl, or either moiety is incorporated. Finally, some prenylated proteins are also palmitoylated at a cysteine near the prenylation site (*19*) and upstream lysines may also be acetylated or methylated, so mass spectrometry will indicate whether

these modifications occur as well. The expected mass changes for some of the various modifications are shown in Table 4.2.

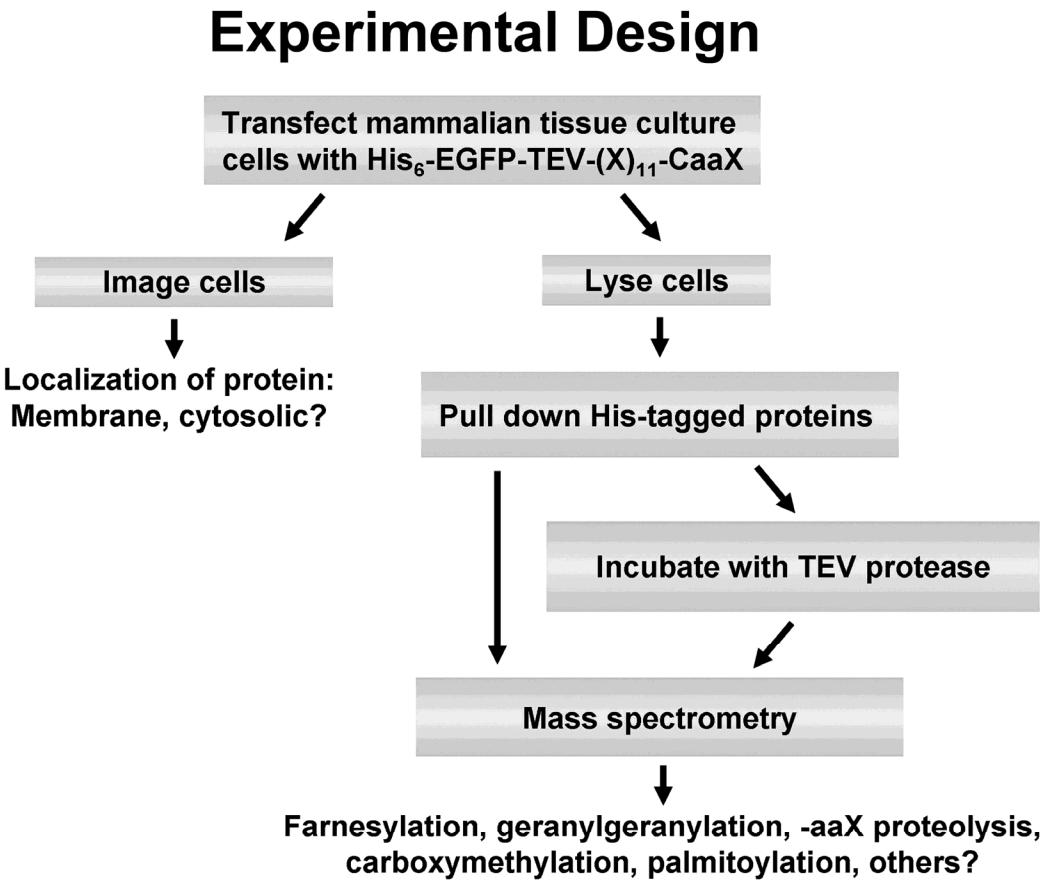


**Figure 4.7. Overall method scheme.** Cells are transfected with vectors containing $His_6$-EGFP-TEV-$X_{11}$-CaaX genes under the CMV promoter and expression and localization of EGFP fusion proteins are analyzed by fluorescence microscopy. If a protein is localized to the membranes, this suggests that it is modified. Additionally, cells can be lysed and the proteins can be pulled down using the $His_6$ tag and digested by incubation with TEV protease. The 15 amino acid tail or full length protein can then be analyzed by mass spectrometry to determine the various post-translational modifications.

| Mass changes with associated modifications | |
|---|---|
| **Modification** | **Mass change (Da)** |
| **Farnesylation** | **204** |
| **Geranylgeranylation** | **272** |
| **Palmitoylation** | **238** |
| **Methylation** | **14** |
| **Proteolysis** | **213-558** |

**Table 4.2. Post-translational modification mass changes.** The expected mass changes for various post-translational modifications in the prenylation pathway when using mass spectrometry.

*Development of the method*

The first step in this technique was to quickly and efficiently generate a library of mammalian expression vector constructs. To this end, we developed a molecular biology method using restriction enzymes and DNA oligonucleotides to easily alter the sequence of the 15 amino acid C-terminal tails of the fusion proteins (Figure 4.8). In a parent vector, the C-terminal region was flanked by KasI and EcoRI restriction enzyme site; digestion of the plasmid with these two restriction enzymes creates a linear vector with two overhangs. To add the tails, DNA oligonucleotides were purchased corresponding to sequences of interest and were designed such that when two oligonucleotides are annealed together, overhangs are formed that match the KasI and EcoRI sites of the parent vector. The vector and annealed oligonucleotides are then ligated, transformed into *E. coli*, and purified to obtain a library of vector constructs. We evaluated this method on one test vector, and as it was successful, we could apply the method to readily obtain all of the new vectors.

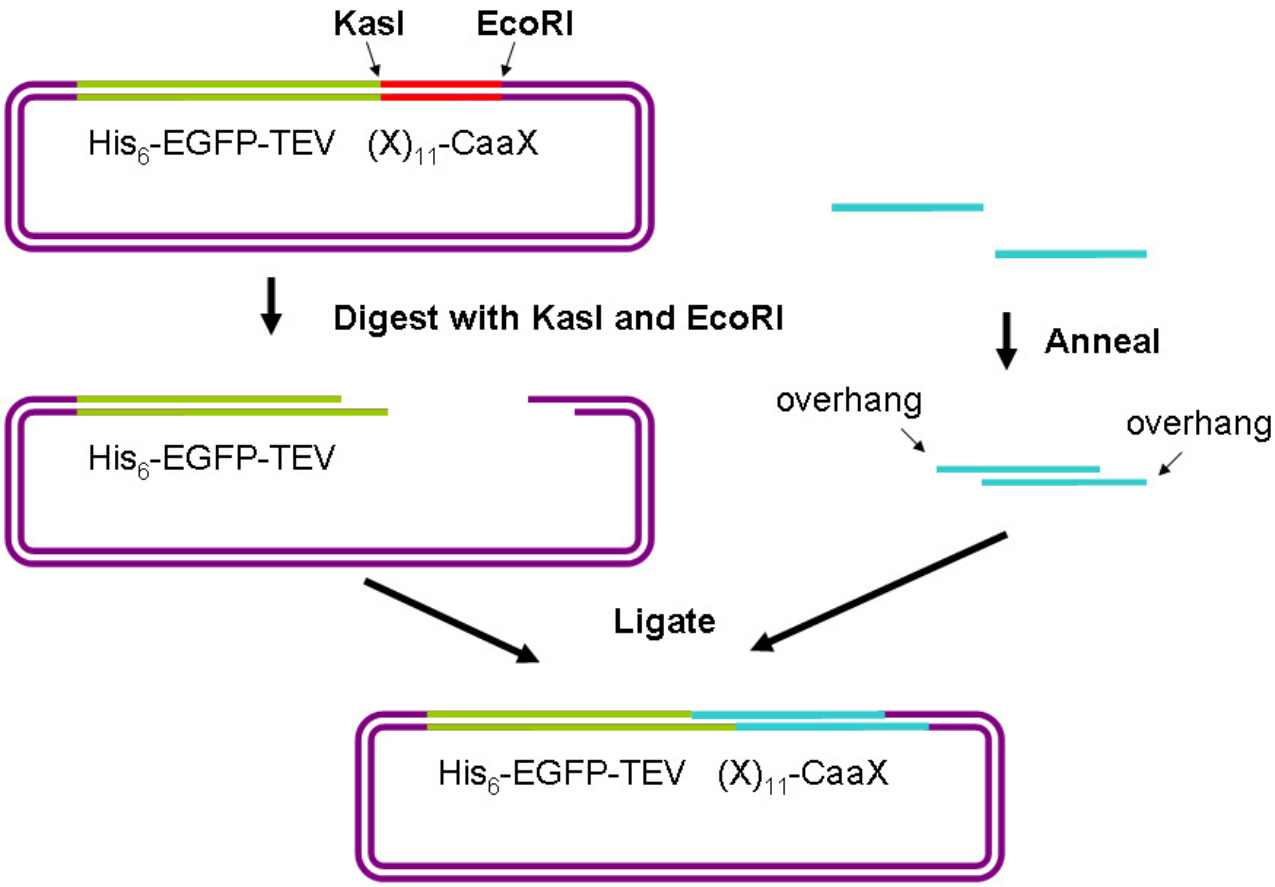


**Figure 4.8. Ligation scheme.** A parent vector was constructed ($His_6$-EGFP-TEV-myc) such that the C-terminal region is flanked by KasI and EcoRI restriction sites, and when the vector is digested with these enzymes, sticky ends are created. Two complementary DNA oligonucleotides that correspond to a 15 amino acid C-terminal tail are purchased (blue). The oligonucleotides are designed such that when the two strands are annealed together, overhangs are created. Then, the annealed oligonucleotides and digested parent vector can be ligated to form a unique complete vector. This method is an efficient way to create a library of different mammalian expression vectors.

Next, we verified that mass spectrometry techniques could be used to detect a farnesyl modification in collaboration with Dr. Eric Simon and Dr. Phil Andrews at the University of Michigan. The peptide dns-SKTKCVIM was farnesylated to ≥90% by incubation with FPP and FTase *in vitro*. This peptide was purified from other reaction components using a ZipTip, and a MALDI-TOF-TOF mass spectrometer was used to analyze the mass of the peptide. Figure 4.9 shows the spectrum obtained, and there is a mass difference of around 307 m/z between two peaks, corresponding to a farnesyl modification on a cysteine residue on the peptide. This suggests that mass spectrometry can be applied to modified proteins and peptides obtained from *in vivo* samples.

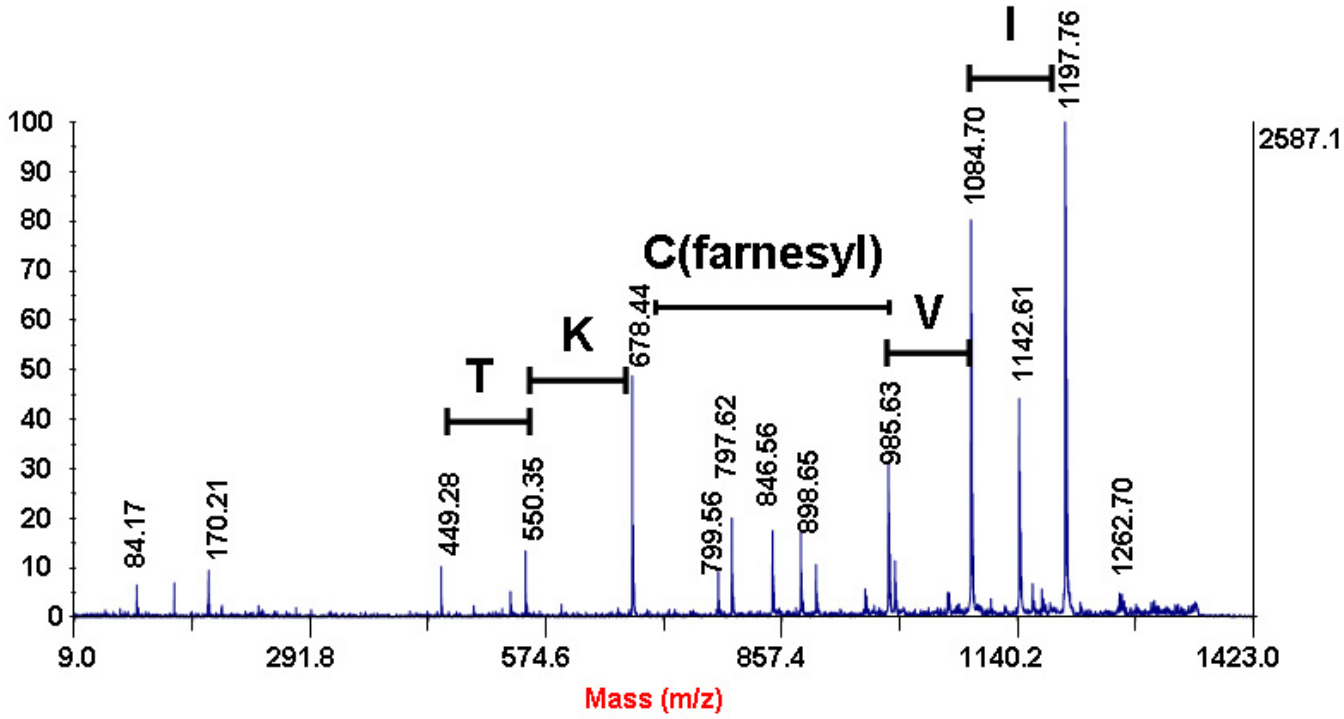


**Figure 4.9. Mass spectrometry of a farnesylated peptide.** The peptide dns-SKTKCVIM was labeled *in vitro* with a farnesyl group. A MALDI-TOF-TOF instrument was used to measure the mass of the protein. This spectrum shows a change in mass of about 307 m/z which corresponds to the farnesyl (204 m/z) modification on a cysteine residue.

Additionally, a $His_6$-EGFP-TEV-$Ras_{15}$ protein was recombinantly expressed and purified in BL21(DE3) *E. coli* cells and this work was done in collaboration with an undergraduate student in the lab, Megan Novak. By expressing and purifying this protein, large amounts of this protein could be obtained as a negative (unmodified) control and the TEV digestion could be optimized. Megan was able to successfully express and purify the $His_6$-EGFP-TEV-$Ras_{15}$ protein (Figure 4.10 A, fractions 2-4). She was also able to successfully digest the protein as indicated by a gel shift in Figure 4.10 B, lane 1. Since the library construction, mass spectrometry, and TEV digestion protocols were each verified in various proof of principle tests, I then began to optimize the whole protocol for use in mammalian tissue culture cells.

*Expression vector library construction*

A library of vectors that corresponds to previously tested *in vitro* peptides was constructed, and rotation student Jenna Hendershot helped with some of this work. As controls, sequences corresponding to the C-terminus of H-Ras (a known farnesylated and palmitoylated protein) and the myc tag (non-substrate) were generated as positive and negative controls. Then, CaaX sequences were chosen that had a variety of *in vitro* activity, including MTO FTase substrates, GGTase-I MTO substrates, dual substrates for both FTase and GGTase-I, FTase STO substrates, and non-substrates (*12*). Additionally, the Rab-18 is suggested to be palmitoylated at a cysteine four amino acids upstream of the CaaX sequence (by similarity, (*54*)). The last 15 amino acids of the protein (including the CaaX sequence) were appended to the C-terminus of the $His_6$-EGFP-TEV proteins. Additional substrates were chosen based on high scoring peptides predicted to be FTase

substrates from the FlexPepBind studies in Chapter 3. Table 4.3 shows a summary of the sequences in the C-terminal tail and the reactivity of the peptide with FTase and GGTase-I *in vitro*.

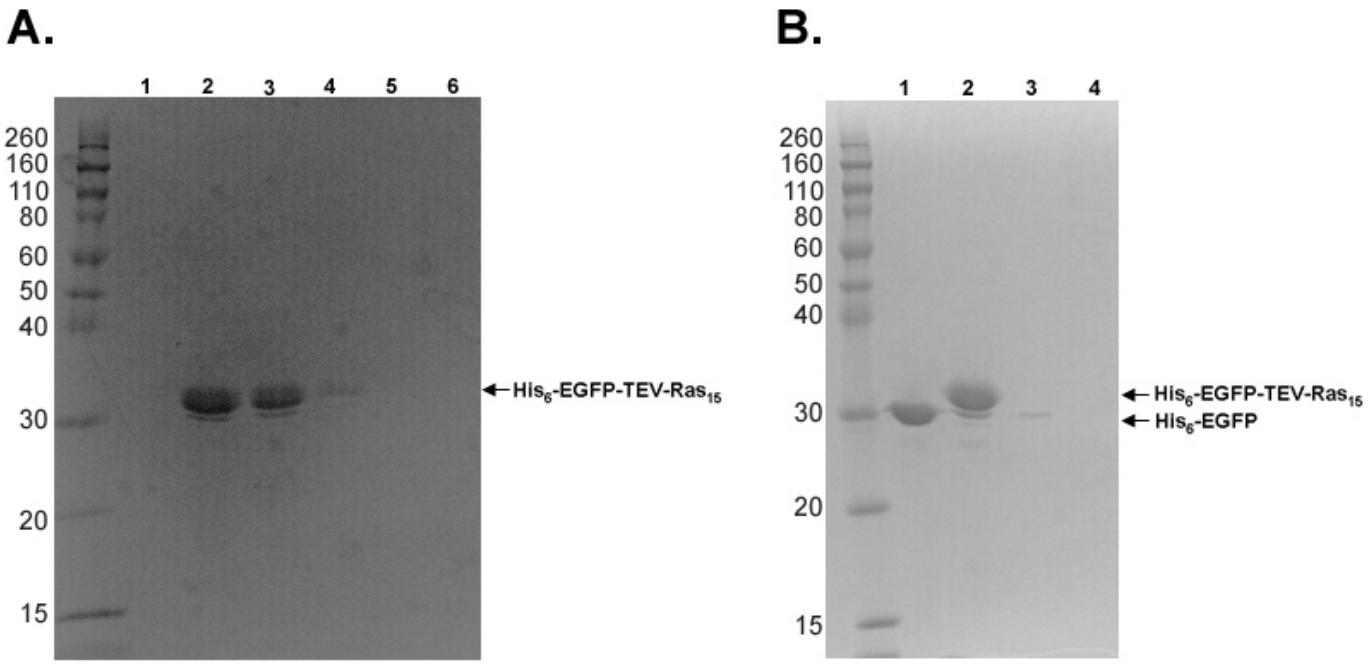


**Figure 4.10. Purification and digestion of the $His_6$-EGFP-TEV-$Ras_{15}$ protein**. **A)** SDS-PAGE gel showing fractions eluted from a Talon Cobalt metal affinity column using imidazole. Purification of the $His_6$-EGFP-TEV-$Ras_{15}$ recombinant protein (30,760 Da) eluted in fractions 2-4 and runs between 30 and 40 kDa ladder bands on an SDS-PAGE gel stained with Coomassie blue dye. The gel has been darkened to more easily visualize the bands. **B)** SDS-PAGE gel analyzing TEV–protease digestion of $His_6$-EGFP-TEV-$Ras_{15}$ recombinant protein. Lane 1 is the digestion of $His_6$-EGFP-TEV-$Ras_{15}$ plus TEV protease, lane 2 is undigested $His_6$-EGFP-TEV-$Ras_{15}$, and lane 3 is TEV protease alone. Lane 4 is blank. The TEV digestion alters the apparent mass of the $His_6$-EGFP protein (Lane 1).

| High activity *in vitro* peptides | | | | | |
|---|---|---|---|---|---|
| **Protein** | **Sequence (Last 15 aa)** | **CaaX** | **GGTase-I $k_{cat}/K_M^{peptide}$ ($mM^{-1}s^{-1}$)[a]** | **FTase $k_{cat}/K_M^{peptide}$ ($mM^{-1}s^{-1}$)[b]** | ***In vitro* activity[a,b]** |
| Cell Division Control Protein 42 Homolog | AALEPPETQPKRKCCIF | CCIF | 15 ± 1 | 3.0 ± 0.3 | Dual substrate |
| Type I inositol- 1,4,5-trisphosphate 5-phosphatase | AGKPHAHVHKCCVVQ | CVVQ | | | MTO farnesylated |
| Aldehyde dehydrogenase family 3 member B1 | RMLLVAMEAQGCSCTLL | CTLL | 14 ± 4 | | MTO Geranylgeranylated |
| Neuronal membrane glycoprotein M6-b (transmembrane) | NYAVLKFKSREDCCTKF | CTKF | 1.12 ± 0.07 | 4.4 ± 0.4 | Dual substrate |
| Interferon-induced guanylate-binding protein 2 | WDIQMRSKSLEPICNIL | CNIL | 33 ± 2 | | MTO Geranylgeranylated |
| Many possibilities. Ras-related protein Rab-8A | PPSAPRKKGGGCPCVLL | CVLL | 38.6 ± 0.4 | | Dual substrate |
| Rho-related GTP-binding protein Rho6 | LGNSSPRTQSPQNCSIM | CSIM | 1.22 ± 0.08 | 14 ± 1 | Dual substrate |
| Ras-related protein Rab-18 | REEGQGGGACGGYCSVL | CSVL | 4 ± 4 | | MTO Geranylgeranylated + Palmitoylation |
| Extracellular superoxide dismutase [Cu-Zn] | EHSERKKRRRESECKAA | CKAA | | | STO farnesylated |
| Rho-related GTP-binding protein RhoC | AGLQVRKNKRRRGCPIL | CPIL | | | STO farnesylated |
| Zinc Finger Protein 124 | KHKKTHTGEKPYKCKKM | CKKM | | | CAAX non-substrate |
| Forkhead box protein C2 | PPLYRHAAPYSYDCTKY | CTKY | | | CAAX non-substrate |
| **Peptides predicted by FlexPepBind** | | | | | |
| **Protein** | **Sequence (Last 15 aa)** | **CaaX** | **GGTase-I $k_{cat}/K_M^{peptide}$ ($mM^{-1}s^{-1}$)** | **FTase $k_{cat}/K_M^{peptide}$ ($mM^{-1}s^{-1}$)** | ***In vitro* activity[c]** |
| Zinc finger protein 64 homolog, isoform 5 | LTVHLRSHTGCCYVA | CYVA | | | MTO farnesylated |
| Intersectin-2 (Isoform 3) | WRLLLASSRGICCLS | CCLS | | | MTO farnesylated |
| Folliculin isoform 3 | RLPCPELREESCWTC | CWTC | | | MTO farnesylated |
| Homeobox protein ESX1 | TWAPVINSYYACPFF | CPFF | | | STO farnesylated |
| Decaprenyl-diphosphate synthase subunit 1 (2nd isoform) | FAWRCRQSSTVCTTE | CTTE | | | STO farnesylated |
| Growth/differentiation factor 15 | LQTYDDLLAKDCHCI | CHCI | | | CAAX non-substrate |
| Phosphatidate phosphatase PPAPDC1B | LSTAQKPGDSYCFDI | CFDI | | | CAAX non-substrate |
| K-Ras (2B isoform) | GKKKKKKSKTKCVIM | CVIM | | | MTO farnesylated control |

**Table 4.3. Mammalian expression vector sequences and *in vitro* activity.** $His_6$-EGFP-TEV-$X_{11}$-CaaX genes were constructed in the pCDNA4/TO vector. The last 15 amino acid tail and corresponding *in vivo* protein is shown on the left. If available, the *in vitro* steady state parameters are indicated. Steady state data and activity is either from [a]unpublished data, from [b]reference (*12*), or from [c]Chapter 3..

Overall, 19 vectors and 3 vector controls were successfully made in the pcDNA4/TO expression system. The ligation method developed by using DNA oligonucleotides proved very successful. For example, for the "high activity" vectors, the construction of 13 plasmids was initiated, and 12/13 plasmids were successfully generated on the first attempt (a 92% success rate).

*Fluorescence microscopy*

Each vector in the library was transfected into mammalian tissue cultures cells to visualize the fusion protein localization. If a protein localizes to cellular membranes, this observation suggests that the protein is modified; however, it is not known for certain that all prenylated proteins localize to the membranes. To begin, the positive and negative controls of H-Ras and myc tag EGFP fusion proteins were transfected into NIH/3T3 and HEK-293 cells. As predicted, the H-Ras EGFP fusion protein localizes to the membrane where the myc fusion showed diffuse localization, showing that these are good controls (Figure 4.11 shows images from the HEK-293 cells transfected with the control plasmids). Next, both the NIH/3T3 and HEK-293 cell lines were transfected with plasmids expressing EGFP fusion proteins containing C-terminal tails that showed high *in vitro* MTO activity. Some of this work was carried out by Jenna Hendershot. Figure 4.12 shows a representation of some of the fluorescence images obtained for a subset of the vectors, and Table 4.4 (the top panel) also shows a summary of the observed localization, with "n/a" meaning that the vector did not transfect efficiently.

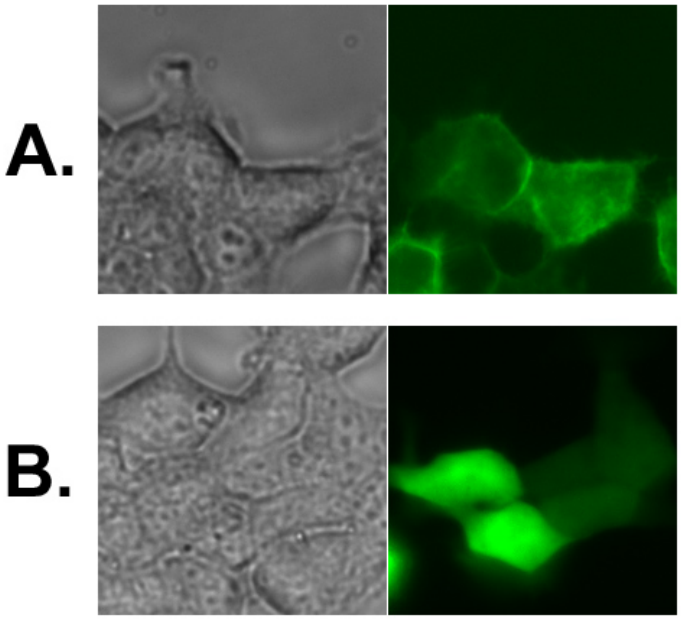


**Figure 4.11. Brightfield and GFP fluorescence microscopy of HEK-293 cells transfected with positive and negative control plasmids.** A Nikon Eclipse TE2000 U inverted epifluorescense microscope (470 excitation and 525 emission cubes) was used to obtain fluorescent images of the EGFP. **A)** Positive control: The plasmid encodes $His_6$-EGFP-TEV-$Ras_{15}$ (with DESGPGCMSCKCVLS corresponding to the C-terminus of H-Ras, a known farnesylated and palmitoylated protein). The fusion protein localizes to the plasma membrane. **B)** Negative control: The plasmid encodes $His_6$-EGFP-TEV-myc (sequence EQKLISEEDL is the myc tag). The fluorescence is diffuse and evenly localizes throughout the cell, consistent with cytosolic localization and no lipid modification on the protein.

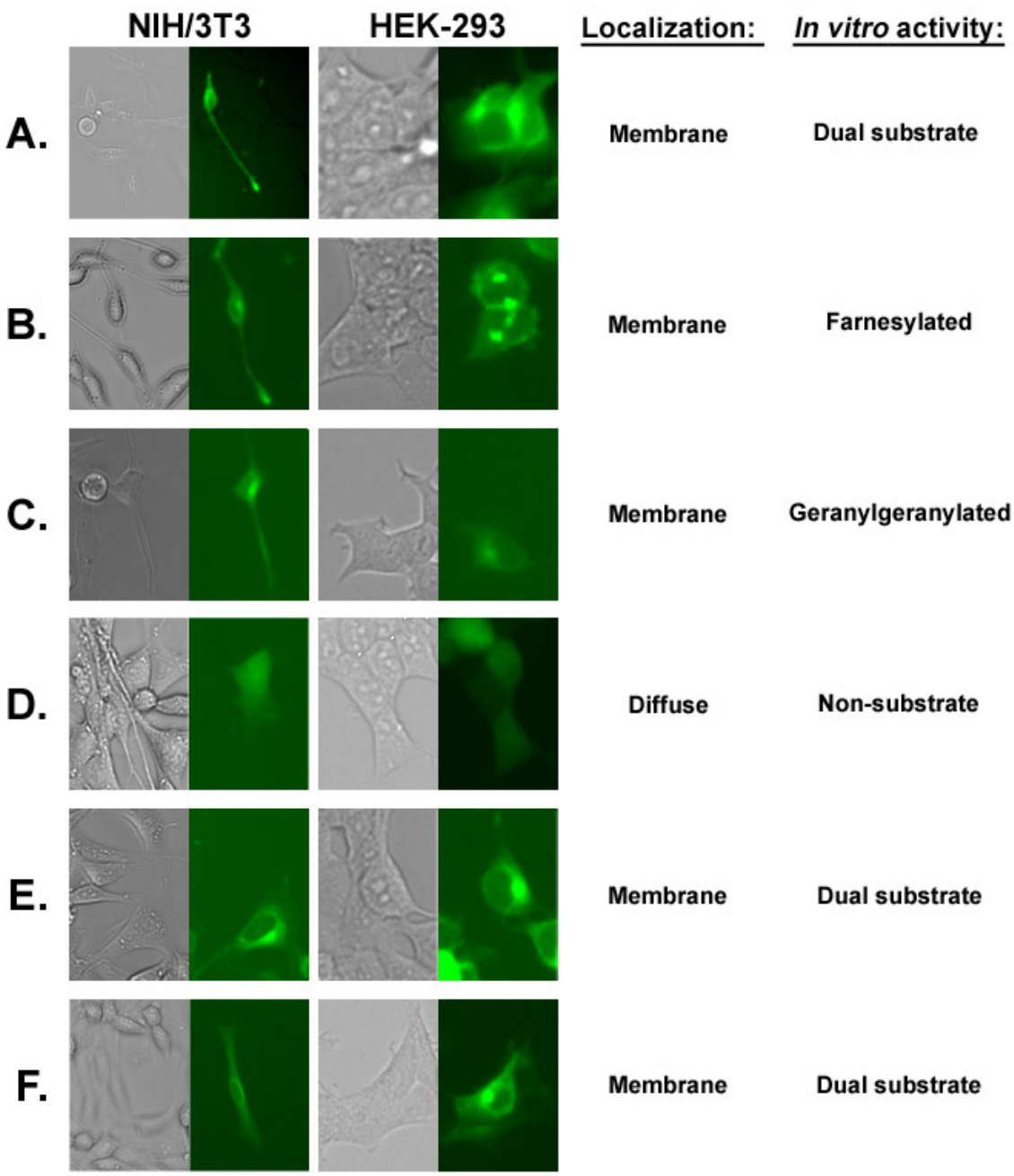


**Figure 4.12. Brightfield and GFP fluorescence microscopy of $His_6$-EGFP-TEV-$X_{11}$-CaaX fusion proteins.** The left two panels are images of NIH/3T3 cells and the right panels are images of HEK-293 cells including both brightfield and fluorescence images. In each row, the transfected plasmid contains a different sequence from: **A)** Cell Division Control Protein 42 Homolog, CaaX=CCIF, predicted dual substrate; **B)** Type I inositol-1,4,5-trisphosphate 5-phosphatase, CaaX=CVVQ, predicted farnesylated; **C)** Interferon-induced guanylate-binding protein 2, CaaX=CNIL, predicted geranylgeranylated; **D)** Forkhead box protein C2, CaaX=CTKY, predicted non-substrate; **E)** Ras-related protein Rab-8A, CaaX=CVLL, predicted dual substrate; **F)** Rho-related GTP-binding protein Rho6, CaaX=CSIM, predicted dual substrate.

| High activity *in vitro* peptides | | | | | | |
|---|---|---|---|---|---|---|
| **Protein** | **Sequence (Last 15 aa)** | **PrePS FTase prediction** | **PrePS GGTase-I prediction** | **FTase FlexPepBind score (≤ -0.4 predicted substrate)** | ***In vitro* activity** | ***In vivo* localization** |
| Cell Division Control Protein 42 Homolog | AALEPPETQPKRKCCIF | ++ | ++ | -0.55 | Dual substrate | membrane |
| Type I inositol- 1,4,5-trisphosphate 5-phosphatase | AGKPHAHVHKCCVVQ | ++ | --- | -1.32 | MTO farnesylated | membrane |
| Aldehyde dehydrogenase family 3 member B1 | RMLLVAMEAQGCSCTLL | - | + | -2.02 | MTO Geranylgeranylated | diffuse |
| Neuronal membrane glycoprotein M6-b (transmembrane) | NYAVLKFKSREDCCTKF | -- | --- | 5.42 | Dual substrate | diffuse |
| Interferon-induced guanylate-binding protein 2 | WDIQMRSKSLEPICNIL | - | ++ | -0.28 | MTO Geranylgeranylated | membrane |
| Many possibilities. Ras-related protein Rab-8A | PPSAPRKKGGGCPCVLL | ++ | +++ | -0.66 | Dual substrate | membrane |
| Rho-related GTP-binding protein Rho6 | LGNSSPRTQSPQNCSIM | ++ | +++ | -1.29 | Dual substrate | membrane |
| Ras-related protein Rab-18 | REEGQGGGACGGYCSVL | + | +++ | -1.22 | MTO Geranylgeranylated + Palmitoylation | membrane |
| Extracellular superoxide dismutase [Cu-Zn] | EHSERKKRRRESECKAA | - | --- | -0.45 | STO farnesylated | n/a |
| Rho-related GTP-binding protein RhoC | AGLQVRKNKRRRGCPIL | + | +++ | -0.94 | STO farnesylated | membrane |
| Zinc Finger Protein 124 | KHKKTHTGEKPYKCKKM | -- | --- | 2.92 | CAAX non-substrate | n/a |
| Forkhead box protein C2 | PPLYRHAAPYSYDCTKY | -- | --- | 4.83 | CAAX non-substrate | diffuse |
| **Peptides predicted by FlexPepBind** | | | | | | |
| **Protein** | **Sequence (Last 15 aa)** | **PrePS FTase prediction** | **PrePS GGTase-I prediction** | **FTase FlexPepBind score (≤ -0.4 predicted substrate)** | ***In vitro* activity** | ***In vivo* localization** |
| Zinc finger protein 64 homolog, isoform 5 | LTVHLRSHTGCCYVA | - | --- | -2.88 | MTO farnesylated | diffuse |
| Intersectin-2 (Isoform 3) | WRLLLASSRGICCLS | - | --- | -2.37 | MTO farnesylated | possible membrane |
| Folliculin isoform 3 | RLPCPELREESCWTC | - | --- | -1.94 | MTO farnesylated | diffuse |
| Homeobox protein ESX1 | TWAPVINSYYACPFF | --- | --- | -1.69 | STO farnesylated | diffuse |
| Decaprenyl-diphosphate synthase subunit 1 (2nd isoform) | FAWRCRQSSTVCTTE | -- | --- | -2.14 | STO farnesylated | diffuse |
| Growth/differentiation factor 15 | LQTYDDLLAKDCHCI | -- | --- | -1.56 | CAAX non-substrate | diffuse |
| Phosphatidate phosphatase PPAPDC1B | LSTAQKPGDSYCFDI | -- | --- | -1.59 | CAAX non-substrate | diffuse |
| K-Ras (2B isoform) | GKKKKKKSKTKCVIM | +++ | +++ | -0.37 | MTO farnesylated control | membrane |

**Table 4.4. *In vivo* localization of fusion proteins and computational algorithm predictions.** $His_6$-EGFP-TEV-$X_{11}$-CaaX genes in the pCDNA4/TO vector were transfected into cells, and localization was observed by fluorescence microscopy. The "high *in* vitro activity" library was transfected into both NIH/3T3 cells and HEK-293 cells, while the "FlexPepBind" library was solely transfected into HEK-293 cells. A summary of the observed localization for each gene is shown. "Membrane" indicates either plasma membrane or plasma membrane/organelle membrane localization. If a gene did not transfect well, it is indicated by "n/a." PrePS predictions are from reference (*9*) and FlexPepBind predictions are from reference (*23*) and Chapter 3. A "+", "++," or "+++" symbol from the PrePS algorithm indicates that a protein is predicted to be prenylated; a score ≤ -0.4 indicates FlexPepBind prediction of farnesylation.

Overall for the peptides characterized as having "high *in vitro* activity," the EGFP-CaaX protein localized to the membranes in both cell lines, despite being overexpressed. For instance, in Figure 4.12 row A, which corresponds to the C-terminus of CDC42 homolog with a CaaX sequence of CCIF, shows non-diffuse localization at the cellular membranes. Additionally, Figure 4.12 B which corresponds to Type I inositol-1,4,5-trisphosphate 5-phosphatase (CaaX=CVVQ, predicted to be farnesylated) appears to localize to both the plasma membrane and punctuate spots within the cells. Figure 4.12 rows C, E, and F which are predicted to be geranylgeranylated or dual substrates appear to localize to a bright spot, presumably outside of the nucleus as well as perhaps the nuclear membranes. Two proteins that did not follow the trend were the CTLL (images not shown) and CTKF proteins (not shown), that showed diffuse localization. These C-terminal tails were predicted to be geranylgeranylated, or dual substrates, respectively based on the *in vitro* reactivity with FTase and GGTase-I. However, the lack of localization does not prove that the proteins are not modified. A protein that is expected to be a non-substrate, Forkhead box protein C2 (CaaX=CTKY), shows completely diffuse and even localization throughout the cell, as predicted (Figure 4.12 D). To further investigate the localization of these EGFP-fusion proteins, confocal microscopy, inhibitor, and co-localization experiments with organelle markers would reveal more information about the exact cellular localization and modification status of each of these proteins (more details in the Discussion section).

The plasmids containing CaaX sequences identified from the "FlexPepBind" analysis (Chapter 3) were also transfected into HEK-293 cells (Figure 4.13, Table 4.4 bottom panel). Of these EGFP-CaaX proteins, only intersectin-2 isoform 3 (CaaX =

CCLS) was observed to have some membrane localization (Figure 4.13 A). The rest of expressed proteins showed diffuse localization, including sequences proposed to be substrates and non-substrates. Figure 4.13 A shows localization of the intersectin-2 GFP fusion which appears to localizes to a spot within the cell. A predicted non-substrate Growth/differentiation factor 15 (CaaX = CHCI, Figure 4.12 B) showed diffuse localization. Row C) is a positive control, K-Ras, which clearly shows membrane localization. Again, the lack of membrane localization does not prove the protein is unmodified.

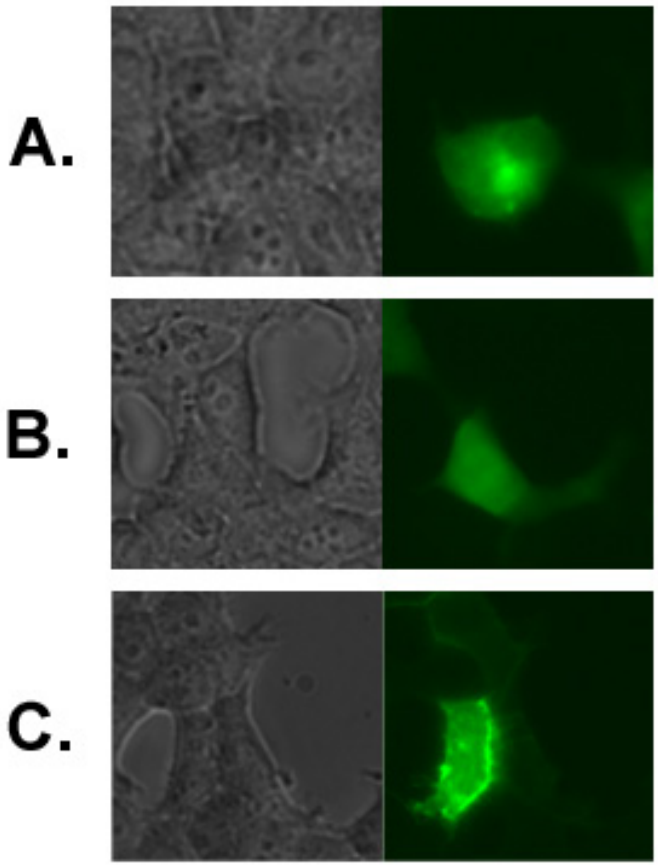


**Figure 4.13. Brightfield and GFP fluorescence microscopy of HEK-293 cells expressing $His_6$-EGFP-TEV-$X_{11}$-CaaX fusion proteins with C-terminal sequences predicted by FlexPepBind.** Each row is transfected with a plasmid containing different C-terminal sequences. **A)** Intersectin-2 (Isoform 3, CaaX = CCLS). The peptide has MTO activity with FTase. **B)** Growth/differentiation factor 15 (CaaX = CHCI). The peptide has no reactivity with FTase. **C)** K-Ras (2B isoform, CaaX = CVIM). The peptide has MTO reactivity with FTase. C) is a positive control.

These studies show that by changing the C-terminal sequence of the fusion proteins, the cellular localization can be altered. Even small changes in the last 15 amino acids can aid in localizing proteins to different regions of the cell. Also, the localization is maintained in multiple cells lines. In general, for the sequences obtained from peptides that react rapidly with FTase or GGTase-I, the expressed proteins localize to the membranes as predicted. For the "FlexPepBind" library, the results were less clear, as only one sequence localized to membranes. A potential explanation for this behavior is that as sequences from the first library have higher *in vitro* activity, these proteins are also more readily modified *in vivo*. The activity of the FlexPepBind peptides is likely significantly lower, as the activity was measured using a single timepoint radioactive assay which is highly sensitive. Possibly, a low percentage of the pool of the EGFP-CaaX protein is modified *in vivo*, but this level is below the level of detection in this whole cell fluorescence microscopy assay. Alternatively, it is possible that the non-canonical side-chains in the CaaX sequence (such as X= E) do not allow membrane localization even if the proteins are modified.

A route to potentially increase the fraction of membrane localization of the overexpressed substrate proteins (and perhaps improve transfection efficiency as well) would be to generate expression vectors that dually express both subunits of FTase or GGTase and the fluorescent protein substrate. If the level of the prenyltransferase is increased, the pool of modified protein may increase thereby enhancing detection of membrane localization. Preliminary work done by Dr. James Hougland, a previous postdoctoral fellow in the lab, has indicated that this type of vector may improve the membrane visualization and localization of proteins with weaker CaaX sequences A

scheme of vectors that are constructed and ready to be tested is shown in Figure 4.14. By switching the promoters the genes are under, the protein expression may be tuned.

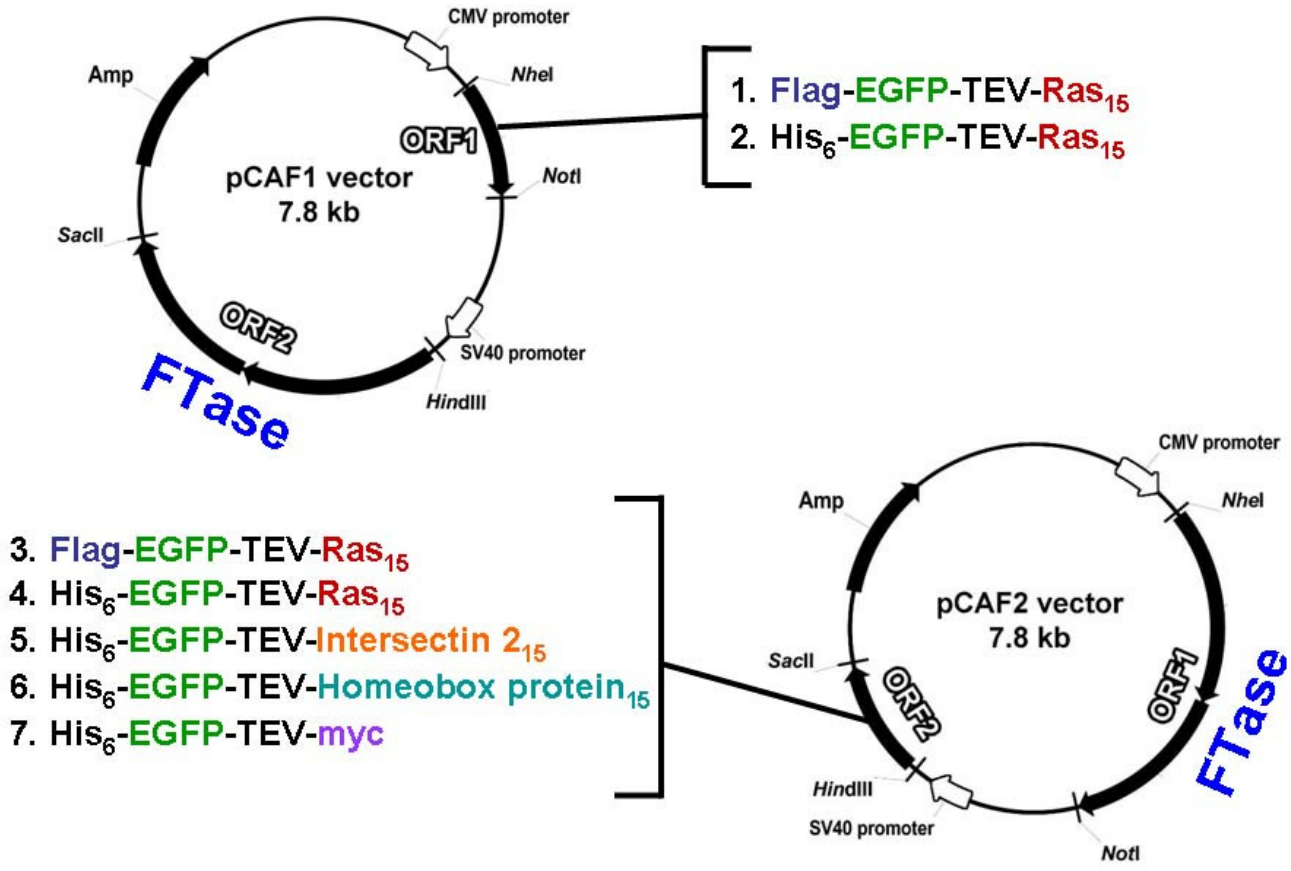


**Figure 4.14. Dual expression vectors**. Vectors were made that co-express both subunits of FTase and the $His_6$-EGFP-TEV-$X_{11}$CaaX or Flag-EGFP-TEV-$X_{11}$CaaX genes. The genes are under either the CMV or SV40 promoters.

## Discussion

*FPP and GGPP analogs*

In this study, we had only modest success detecting prenyl modifications using the azido-farnesol and azido-geranylgeraniol analogs. High background was observed, which could possibly be minimized by optimizing the "click" chemistry conditions. The larger issue, however, is the low incorporation of the analogs onto proteins. One way to increase analog incorporation would be to treat cells with an HMG-CoA reductase inhibitor such as lovastatin, that will block the cholesterol biosynthesis pathway and stop the production of endogenous FPP and GGPP (*55, 56*). If the levels of the native lipid substrates are decreased, it is more likely that the analogs will be incorporated onto the proteins. This method enjoys support from previous work in the literature (*8, 44*). Blocking FPP and GGPP synthesis would not fix this low incorporation issue if the modified proteins are in too low of a concentration to be visualized (or exist in an only partially modified pool).

A disadvantage of the use of FPP or GGPP analogs is that the analogs may alter the protein specificity of the prenyltransferases (*57-59*), as the prenyl group forms a portion of the protein substrate binding site (*11*). Crystal structures have revealed that the $a_2$ residue of the peptide/protein is contacted by the lipid substrate in both FTase and GGTase-I (*11*). Therefore, altering the FPP or GGPP moieties may change the interactions with the peptide or protein substrate and alter the overall enzyme selectivity. Studies using the FPP analog 3-MeBFPP (3-(3-Methyl-2-butenyl)-7,11-dimethyldodeca-2(Z),6(E),10-triene 1-Diphosphate) demonstrated this. When 3-MeBFPP was incubated with FTase and various peptides *in vitro*, different FTase reactivity was observed

depending on peptide sequence (*57*). Another study utilized biotin-FPP or GGPP analogs to identify potential prenyltransferase substrates in compactin-treated (unmodified) cell lysates. Interestingly, FTase and GGTase-I could not tolerate these lipid substrates unless the enzymes were engineered at positions W102T/Y154T or W102T/Y154T/Y205T for FTase and F53Y/Y126 or F52Y/F53Y/Y126 for GGTase-I (*43*). The residues mutated in this work include contacts to the peptide substrate, which are known to affect prenyltransferase selectivity (*22*). Taken together, recent work with azido-FPP, azido-GGPP, anilinogeraniol, and C10-alkyne to identify FTase and GGTase-I substrates *in vivo* has indicated that different analogs pull down different substrates (*43, 60-62*), which agrees with the hypothesis that alteration to the lipid donor can change selectivity. Overall, these studies explain why some expected and known substrates are not modified with the analogs and are not detected. This caveat should be taken into account when using analogs as FTase and GGTase-I substrates.

*Comparison to prenylation prediction programs*

Computational tools have aided in predicting prenylated proteins. PrePS was developed to predict the likelihood of prenylation by FTase and GGTase-I (as well as GGTase-I) based on a training set of known prenylated proteins and the linker amino acids upstream of the CaaX sequence (*9*). FlexPepBind, a program that predicts FTase substrates, was developed based on "energy" binding scores with structural modeling of CaaX peptide structures in the FTase active site (Chapter 3, (*23*)). Table 4.4 shows the PrePS and FlexPepBind scores for each C-terminal region of the $His_6$-EGFP-TEV-$X_{11}$-CaaX proteins. In general, the observed localization of the EGFP-CaaX proteins in this

work correlates with computational predictions. For the $His_6$-EGFP-TEV-$X_{11}$-CaaX sequences chosen for high peptide *in vitro* activity with FTase and GGTase-I, all substrates with MTO or STO *in vitro* prenyltransferase activity are predicted to be prenylated by PrePS and FlexPepBind; however exceptions are the neuronal membrane glycoprotein M6-b (CaaX = CTKF) and the extracellular superoxide dismutase [Cu-Zn] (CaaX = CKAA) fusion proteins which are both predicted to be non-substrates by PrePS and FlexPepBind. Additionally, the interferon-induced guanylate-binding protein 2 fusion protein (CaaX = CNIL) is predicted as a non-substrate by FlexPepBind. Furthermore, all of the $His_6$-EGFP-TEV-$X_{11}$-CaaX proteins constructed based on FlexPepBind scores were predicted as non-substrates by the PrePS algorithm, but had exceptionally good scores with FlexPepBind. Interestingly, the extracellular superoxide dismutase [Cu-Zn] (CaaX = CKAA) fusion protein poorly transfected and the neuronal membrane glycoprotein M6-b (CaaX = CTKF) fusion displayed diffuse localization, but the remainder of the EGFP-CaaX proteins predicted as substrates by PrePS and FlexPepBind localized to the membranes. Additionally, with the exception of the intersectin-2 fusion protein (CaaX = CCLS), all of the $His_6$-EGFP-TEV-$X_{11}$-CaaX proteins based on high FlexPepBind scores showed diffuse localization *in vivo*. Therefore, EGFP-CaaX proteins were more likely to display membrane localization if predicted to be prenylated by PrePS. Again, this suggests that high reactivity with the prenyltransferases and a certain level of modification must occur *in vivo* in order to observe membrane localization by this technique. It is likely that the PrePS algorithm predicts more reactive substrates with FTase and GGTase-I since it was developed based on known prenyltransferase substrates and the *in vitro* activity of peptides with good FlexPepBind scores is lower.

*Future in vivo work*

Now that a method of observing the localization of various CaaX sequences has been developed, further work can be done to gain more information about the behavior and modification status of proteins in the prenylation pathway. In the future, it would be interesting to carry out co-localization studies in order to determine to which cellular organelles the fusion proteins localize, since not all prenylated proteins localize solely to the plasma membrane to carry out their functions. For instance, Rheb 1 and Rheb 2 (known farnesylated proteins) which activate mTOR signaling have been shown to localize primarily to the ER and Golgi (*63*). The additional modifications of proteolysis and methylation after prenylation are required for proper Rheb localization, but are not essential for activation of the signaling pathway (potentially due to an unknown protein activating the pathway) (*63*). Comparison of Rheb studies with the work here suggests that several of the EGFP constructs (specifically Figure 4.12 C CaaX=CNIL, Figure 4.12 E CaaX=CVLL, and Figure 4.12 F CaaX=CSIM) may also locate to the Golgi and/or the ER. Co-localization studies using confocal microscopy, EGFP-CaaX transfections, and cellular compartment stains to the endoplasmic reticulum (ER-Tracker Red, Life Technologies), Golgi complex (BODIPY® TR C5 ceramide complexed to BSA, Life Technologies) and the plasma membrane (Alexa Fluor® 350 conjugate of WGA, Life Technologies) will aid in determining the localization of EGFP proteins. Furthermore, some proteins require a "secondary signal" in order for proper cellular localization. For example, the prenylated N-Ras and H-Ras proteins localize to the ER and Golgi membranes instead of the plasma membrane unless they are palmitoylated; conversely K-Ras contains an upstream polybasic region that directly aids in membrane localization

(*64-66*). It is possible that the potential organelle localization of EGFP fusion proteins in this work are not modified with the necessary secondary modifications after prenylation like palmitoylation, proteolysis, and methylation by the respective enzymes due to overexpression of the EGFP proteins or because they lack the necessary consensus sequence for the modification. Using inhibitors against FTase, GGTase-I, RCE1, ZMPSTE24, ICMT, and the PATs will help to isolate the steps in the prenylation pathway that impact protein localization, but it is possible that some modified proteins do not display membrane localization. Further, studies using immunostaining against the specific endogenous proteins that are potentially prenylated would help in understanding the protein localization under native conditions.

Overall, more optimization is still needed to help determine and understand the substrates of the prenylation pathway. Likely the most informative tool to use in the future will be mass spectrometry, as the various modifications of farnesylation, geranylgeranylation, proteolysis, methylation, and palmitoylation can be observed. It will be very interesting to see if all prenylated proteins get proteolyzed and methylated, as this is currently unknown. Additionally, it will be interesting to observe the level of modification as well as the combination of modifications. For instance, some substrates may exist in a pool that is only partially modified *in vivo*. Using this method to investigate prenylation will also enhance the understanding of the pathways affected by the FTIs and GGTIs and which proteins are alternatively modified under treatment with the prenyltransferase inhibitors. Pull-downs using the tag and TEV-digestion of *in vivo* samples remain to be optimized, but these methods will aid in enriching concentration of the modified protein in the samples. Overall, this work is a good first step towards

detecting and observing post-translational modifications and understanding the prenylation pathway *in vivo*.

## References


(1) Marshall, C. J. (1993) Protein prenylation: a mediator of protein-protein interactions. *Science 259*, 1865-6.

(2) Casey, P. J. (1994) Lipid modifications of G proteins. *Curr Opin Cell Biol 6*, 219-25.

(3) Benetka, W., Koranda, M., and Eisenhaber, F. (2006) Protein Prenylation: An (Almost) Comprehensive Overview on Discovery History, Enzymology, and Significance in Physiology and Disease. *Monatshefte für Chemie / Chemical Monthly 137*, 1241-1281.

(4) Casey, P. J., and Seabra, M. C. (1996) Protein Prenyltransferases. *Journal of Biological Chemistry 271*, 5289-5292.

(5) Winter-Vann, A. M., and Casey, P. J. (2005) Post-prenylation-processing enzymes as new targets in oncogenesis. *Nat Rev Cancer 5*, 405-12.

(6) Barrowman, J., and Michaelis, S. (2009) ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders. *Biological Chemistry 390*, 761-773.

(7) Trueblood, C. E., Boyartchuk, V. L., Picologlou, E. A., Rozema, D., Poulter, C. D., and Rine, J. (2000) The CaaX proteases, Afc1p and Rce1p, have overlapping but distinct substrate specificities. *Mol Cell Biol 20*, 4381-92.

(8) Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004) A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci U S A 101*, 12479-84.

(9) Maurer-Stroh, S., and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biol 6*, R55.

(10) Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F. L., and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Comput Biol 3*, e66.

(11) Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol 343*, 417-33.

(12) Hougland, J. L., Hicks, K. A., Hartman, H. L., Kelly, R. A., Watt, T. J., and Fierke, C. A. (2010) Identification of Novel Peptide Substrates for Protein

Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J Mol Biol 395*, 176-90.

(13) Huang, K., and El-Husseini, A. (2005) Modulation of neuronal protein trafficking and function by palmitoylation. *Curr Opin Neurobiol 15*, 527-35.

(14) Resh, M. D. (2006) Palmitoylation of ligands, receptors, and intracellular signaling molecules. *Sci STKE 2006*, re14.

(15) Smotrys, J. E., and Linder, M. E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Annu Rev Biochem 73*, 559-87.

(16) Linder, M. E., and Deschenes, R. J. (2007) Palmitoylation: policing protein stability and traffic. *Nat Rev Mol Cell Biol 8*, 74-84.

(17) Magee, T., and Seabra, M. C. (2005) Fatty acylation and prenylation of proteins: what's hot in fat. *Curr Opin Cell Biol 17*, 190-6.

(18) el-Husseini Ael, D., and Bredt, D. S. (2002) Protein palmitoylation: a regulator of neuronal development and function. *Nat Rev Neurosci 3*, 791-802.

(19) Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R., 3rd, and Davis, N. G. (2006) Global analysis of protein palmitoylation in yeast. *Cell 125*, 1003-13.

(20) Zhang, F. L., and Casey, P. J. (1996) Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem 65*, 241-69.

(21) Casey, P. J. (1995) Protein Lipidation in Cell Signaling. *Science 268*, 221-225.

(22) Hougland, J. L., Lamphear, C. L., Scott, S. A., Gibbs, R. A., and Fierke, C. A. (2009) Context-dependent substrate recognition by protein farnesyltransferase. *Biochemistry 48*, 1691-701.

(23) London, N., Lamphear, C. L., Hougland, J. L., Fierke, C. A., and Schueler-Furman, O. (2011) Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. *PLoS Comput Biol 7*, e1002170.

(24) Hicks, K. A., Hartman, H. L., and Fierke, C. A. (2005) Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I. *Biochemistry 44*, 15325-33.

(25) Dolence, J. M., Steward, L. E., Dolence, E. K., Wong, D. H., and Poulter, C. D. (2000) Studies with recombinant Saccharomyces cerevisiae CaaX prenyl protease Rce1p. *Biochemistry 39*, 4096-104.

(26) Otto, J. C., Kim, E., Young, S. G., and Casey, P. J. (1999) Cloning and characterization of a mammalian prenyl protein-specific protease. *J Biol Chem 274*, 8379-82.

(27) Jang, G. F., and Gelb, M. H. (1998) Substrate specificity of mammalian prenyl protein-specific endoprotease activity. *Biochemistry 37*, 4473-81.

(28) Hollander, I., Frommer, E., and Mallon, R. (2000) Human ras-converting enzyme (hRCE1) endoproteolytic activity on K-ras-derived peptides. *Anal Biochem 286*, 129-37.

(29) Hollander, I. J., Frommer, E., Aulabaugh, A., and Mallon, R. (2003) Human Ras converting enzyme endoproteolytic specificity at the P2' and P3' positions of K-Ras-derived peptides. *Biochim Biophys Acta 1649*, 24-9.

(30) Schmidt, W. K., Tam, A., Fujimura-Kamada, K., and Michaelis, S. (1998) Endoplasmic reticulum membrane localization of Rce1p and Ste24p, yeast proteases involved in carboxyl-terminal CAAX protein processing and amino-terminal a-factor cleavage. *Proc Natl Acad Sci U S A 95*, 11175-80.

(31) Bergo, M. O., Wahlstrom, A. M., Fong, L. G., and Young, S. G. (2008) Genetic analyses of the role of RCE1 in RAS membrane association and transformation. *Methods Enzymol 438*, 367-89.

(32) Hancock, J. F. (1995) Reticulocyte lysate assay for in vitro translation and posttranslational modification of Ras proteins. *Methods Enzymol 255*, 60-5.

(33) Wilson, A. L., and Maltese, W. A. (1995) Coupled Translation/Prenylation of Rab Proteins in-Vitro, in *Lipid Modifications of Proteins* pp 79-91, Academic Press Inc, San Diego.

(34) Peter, M., Chavrier, P., Nigg, E. A., and Zerial, M. (1992) Isoprenylation of Rab Proteins on Structurally Distinct Cysteine Motifs. *Journal of Cell Science 102*, 857-865.

(35) Gibbs, B. S., Zahn, T. J., Mu, Y., Sebolt-Leopold, J. S., and Gibbs, R. A. (1999) Novel farnesol and geranylgeraniol analogues: A potential new class of anticancer agents directed against protein prenylation. *J Med Chem 42*, 3800-8.

(36) Benetka, W., Koranda, M., Maurer-Stroh, S., Pittner, F., and Eisenhaber, F. (2006) Farnesylation or geranylgeranylation? Efficient assays for testing protein prenylation in vitro and in vivo. *BMC Biochem 7*, 6.

(37) Chiba, Y., Sato, S., Hanazaki, M., Sakai, H., and Misawa, M. (2009) Inhibition of geranylgeranyltransferase inhibits bronchial smooth muscle hyperresponsiveness in mice. *American Journal of Physiology-Lung Cellular and Molecular Physiology 297*, L984-L991.

(38) Baron, R., Fourcade, E., Lajoie-Mazenc, I., Allal, C., Couderc, B., Barbaras, R., Favre, G., Faye, J. C., and Pradines, A. (2000) 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 97*, 11626-31.

(39) Lin, H. P., Hsu, S. C., Wu, J. C., Sheen, I. J., Yan, B. S., and Syu, W. J. (1999) Localization of isoprenylated antigen of hepatitis delta virus by anti-farnesyl antibodies. *J Gen Virol 80 ( Pt 1)*, 91-6.

(40) Liu, X. H., Suh, D. Y., Call, J., and Prestwich, G. D. (2004) Antigenic prenylated peptide conjugates and polyclonal antibodies to detect protein prenylation. *Bioconjug Chem 15*, 270-7.

(41) Troutman, J. M., Roberts, M. J., Andres, D. A., and Spielmann, H. P. (2005) Tools to analyze protein farnesylation in cells. *Bioconjug Chem 16*, 1209-17.

(42) Chehade, K. A., Andres, D. A., Morimoto, H., and Spielmann, H. P. (2000) Design and synthesis of a transferable farnesyl pyrophosphate analogue to Ras by protein farnesyltransferase. *J Org Chem 65*, 3027-33.

(43) Nguyen, U. T., Guo, Z., Delon, C., Wu, Y., Deraeve, C., Franzel, B., Bon, R. S., Blankenfeldt, W., Goody, R. S., Waldmann, H., Wolters, D., and Alexandrov, K. (2009) Analysis of the eukaryotic prenylome by isoprenoid affinity tagging. *Nat Chem Biol 5*, 227-35.

(44) Chan, L. N., Hart, C., Guo, L., Nyberg, T., Davies, B. S., Fong, L. G., Young, S. G., Agnew, B. J., and Tamanoi, F. (2009) A novel approach to tag and identify geranylgeranylated proteins. *Electrophoresis 30*, 3598-606.

(45) Hames, B. D., and Rickwood, D. (1998) *Gel Electrophoresis of Proteins: A Practical Approach, Third Edition*, 3 ed., Oxford University Press.

(46) Harlow, E., and Lane, D. (1999) *Using antibodies: a laboratory manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

(47) Bolt, M. W., and Mahoney, P. A. (1997) High-efficiency blotting of proteins of diverse sizes following sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal Biochem 247*, 185-92.

(48) Bowers, K. E., and Fierke, C. A. (2004) Positively charged side chains in protein farnesyltransferase enhance catalysis by stabilizing the formation of the diphosphate leaving group. *Biochemistry 43*, 5256-65.

(49) Zimmerman, K. K., Scholten, J. D., Huang, C. C., Fierke, C. A., and Hupe, D. J. (1998) High-level expression of rat farnesyl:protein transferase in Escherichia coli as a translationally coupled heterodimer. *Protein Expr Purif 14*, 395-402.

(50) Saxon, E., and Bertozzi, C. R. (2000) Cell surface engineering by a modified Staudinger reaction. *Science 287*, 2007-2010.

(51) Moses, J. E., and Moorhouse, A. D. (2007) The growing applications of click chemistry. *Chem Soc Rev 36*, 1249-62.

(52) Andres, D. A., Crick, D. C., Finlin, B. S., and Waechter, C. J. (1999) Rapid identification of cysteine-linked isoprenyl groups by metabolic labeling with [3H]farnesol and [3H]geranylgeraniol. *Methods Mol Biol 116*, 107-23.

(53) Clark, G. J., Kinch, M. S., Rogers-Graham, K., Sebti, S. M., Hamilton, A. D., and Der, C. J. (1997) The Ras-related protein Rheb is farnesylated and antagonizes Ras signaling and transformation. *J Biol Chem 272*, 10608-15.

(54) Reid, H. M., Mulvaney, E. P., Turner, E. C., and Kinsella, B. T. (2010) Interaction of the human prostacyclin receptor with Rab11: characterization of a novel Rab11 binding domain within alpha-helix 8 that is regulated by palmitoylation. *J Biol Chem 285*, 18709-26.

(55) Sinensky, M., Beck, L. A., Leonard, S., and Evans, R. (1990) Differential inhibitory effects of lovastatin on protein isoprenylation and sterol synthesis. *J Biol Chem 265*, 19937-41.

(56) Kim, R., Rine, J., and Kim, S. H. (1990) Prenylation of mammalian Ras protein in Xenopus oocytes. *Mol Cell Biol 10*, 5945-9.

(57) Reigard, S. A., Zahn, T. J., Haworth, K. B., Hicks, K. A., Fierke, C. A., and Gibbs, R. A. (2005) Interplay of isoprenoid and peptide substrate specificity in protein farnesyltransferase. *Biochemistry 44*, 11214-23.

(58) Krzysiak, A. J., Rawat, D. S., Scott, S. A., Pais, J. E., Handley, M., Harrison, M. L., Fierke, C. A., and Gibbs, R. A. (2007) Combinatorial modulation of protein prenylation. *ACS Chem Biol 2*, 385-9.

(59) Troutman, J. M., Subramanian, T., Andres, D. A., and Spielmann, H. P. (2007) Selective modification of CaaX peptides with ortho-substituted anilinogeranyl lipids by protein farnesyl transferase: Competitive substrates and potent inhibitors from a library of farnesyl diphosphate analogues. *Biochemistry 46*, 11310-11321.

(60) DeGraw, A. J., Palsuledesai, C., Ochocki, J. D., Dozier, J. K., Lenevich, S., Rashidian, M., and Distefano, M. D. (2010) Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. *Chem Biol Drug Des 76*, 460-71.

(61) Onono, F. O., Morgan, M. A., Spielmann, H. P., Andres, D. A., Subramanian, T., Ganser, A., and Reuter, C. W. (2010) A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with Western blotting. *Mol Cell Proteomics 9*, 742-51.

(62) Galichet, A., and Gruissem, W. (2006) Developmentally controlled farnesylation modulates AtNAP1;1 function in cell proliferation and cell expansion during Arabidopsis leaf development. *Plant Physiol 142*, 1412-26.

(63) Hanker, A. B., Mitin, N., Wilder, R. S., Henske, E. P., Tamanoi, F., Cox, A. D., and Der, C. J. (2010) Differential requirement of CAAX-mediated posttranslational processing for Rheb localization and signaling. *Oncogene 29*, 380-91.

(64) Choy, E., Chiu, V. K., Silletti, J., Feoktistov, M., Morimoto, T., Michaelson, D., Ivanov, I. E., and Philips, M. R. (1999) Endomembrane trafficking of ras: the CAAX motif targets proteins to the ER and Golgi. *Cell 98*, 69-80.

(65) Chiu, V. K., Bivona, T., Hach, A., Sajous, J. B., Silletti, J., Wiener, H., Johnson, R. L., 2nd, Cox, A. D., and Philips, M. R. (2002) Ras signalling on the endoplasmic reticulum and the Golgi. *Nat Cell Biol 4*, 343-50.

(66) Hancock, J. F., Paterson, H., and Marshall, C. J. (1990) A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell 63*, 133-9.

# CHAPTER V

# SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

## Summary and Conclusions

In this work, we investigated the substrate specificity of GGTase-I and FTase using a variety of methods. The proteins modified by the prenyltransferases are involved in a variety of important signaling pathways within the cell, and inhibitors towards FTase and GGTase-I are being assessed for the treatment of diseases ranging from cancer to parasitic infections. Therefore, identifying the prenyltransferase substrates is an important goal. Two approaches can be used for identifying prenyltransferase substrates. First, we can strive to understand how an enzyme recognizes substrates by determining the elements of molecular recognition. We followed this approach using peptide library studies and by testing and developing a new computational method. Secondly, we can directly identify an enzyme's substrates *in vivo*; this method was carried out by using a vector expression library, transfections, and fluorescence microscopy.

*GGTase-I substrate recognition and peptide library studies*

The prenyltransferases can recognize peptides as substrates (*1*-*5*), allowing for the quick and efficient testing of a wide array of sequences. In this study, we improved the expression of mammalian GGTase-I using auto-induction media supplemented with

$ZnSO_4$. Then, 402 peptides of the form dns-TKCxxx or dns-GCxxx were tested for activity with mammalian GGTase-I. 111 of these peptides were substrates with GGTase-I (28% of the library) under MTO conditions. Additionally, 178 (44%) of the peptides displayed STO-only behavior with GGTase-I. 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, we found that GGTase-I prefers I and L at the $a_2$ position and F, L, M, and V at the X position. There is little sequence preference at the $a_1$ position. For STO-only substrates, GGTase-I prefers P and S at the X position and has little sequence preference at the $a_1$ and $a_2$ positions.

Interesting patterns appear when comparing the results of the sequence preferences for GGTase-I with similar studies of FTase selectivity (*6*). For MTO activity, FTase and GGTase-I both prefer substrates with residues I, L, and (to a lesser extent) V at the $a_2$ position; however, there are not overlapping preferences at the $a_2$ residue for STO substrates. Both enzymes contact the $a_2$ side chain with hydrophobic residues and were predicted to select a wide range of small and hydrophobic residues, as evaluated by crystallographic work (*7*), so these substrate analyses agree with and refine these studies. Further, at the X position for MTO substrates, FTase and GGTase-I are similar in their preference for both F and M, but differ in that FTase also prefers substrates terminating in I, L, and V and GGTase-I prefers substrates terminating in Q. This is additional evidence of dual substrates for the prenyltransferase enzymes and is in agreement with other work (*7, 8*).

Finally, a small comparison of a group of dns-GCxxx peptides and dns-TKCxxx peptides was carried out by measuring the steady state parameters. In general, dns-GCxxx

peptides were 1-3 times more reactive with GGTase-I than the dns-TKCxxx peptides. One explanation for this behavior is that the presence of a charged lysine may interact with the GGTase-I residues and affect the reactivity.

*Development and analysis of FlexPepBind: a computational method to predict FTase substrates*

In this work, we collaborated with Dr. Ora Schueler-Furman and Nir London at the Hebrew University of Jerusalem to develop and test a new computational method called FlexPepBind that predicts FTase substrates. The Schueler-Furman lab calculated binding scores based on the Rosetta modeling suite (*9, 10*) by threading various $Ca_1a_2X$ sequences onto a peptide backbone in the FTase active site, based on data from a crystal structure (*7*). Two hydrogen bonds between FTase and the peptide substrate as well as the $Zn^{2+}$-sulfur bond from the peptide cysteine were held constant. This method performed very well on predicting FTase substrates in test data sets. For example, FlexPepBind was applied to a CxxL library that was tested for activity with FTase (*11*) and gave an 88% true positive rate and a 12% false positive rate when using a cutoff of -0.4 for FlexPepBind scores. Another benefit of this method is that it better predicts STO substrates. For instance, 47% of STO peptides are predicted by FlexPepBind compared to the 14% prediction rate by the previous prenyltransferase prediction algorithm PrePS (*12, 13*).

After analysis of the affinity of all 8,000 possible Cxxx sequences, a group of 29 peptides predicted by FlexPepBind to bind to FTase was tested for activity with FTase *in vitro*. The peptides fell into two groups: 1) those that had exceptionally good

FlexPepBind scores and 2) those that correspond to *in vivo* proteins and were previously uncharacterized. Of these 29 peptides, FTase catalyzed farnesylation of 26 peptides under MTO or STO conditions. This group included peptides with a negatively charged X group, such as E or D, which had not previously been observed in any FTase substrates (*14*).

Further work was done to characterize these peptides and the relationship of physically measurable parameters with FlexPepBind score. The $k_{cat}/K_M^{peptide}$ values were measured for some of the MTO peptides. No correlation was observed between the value of $k_{cat}/K_M^{peptide}$ (*6*) and FlexPepBind scores. Additionally when comparing $K_D^{peptide}$ measurements for FTase substrates to FlexPepBind scores, no obvious pattern is observed Overall, it is unclear at this time what physically measureable parameter correlates with and explains the predictive power of the FlexPepBind algorithm. Nonetheless, this program is very successful at identifying a wider range of FTase substrates than previously developed algorithms.

*Investigation into the in vivo prenylation pathway*

The full complement of *in vivo* substrates of FTase and GGTase-I is unknown at this time. Furthermore, after prenylation catalyzed by FTase or GGTase-I, substrates can undergo additional modification catalyzed by enzymes ZMPSTE24 or RCE1 and ICMT to proteolyze and carboxy-methylate the C-terminus. Currently, it is also unknown whether all proteins undergo each of these modifications. It is important to understand the prenylation pathway *in vivo* to not only appreciate cellular signaling and localization

processes, but also to comprehend which proteins may modulate the effect of prenyltransferase inhibitors.

One approach to identifying FTase and GGTase-I substrates is to use FPP and GGPP analogs that allow for the tagging and pull-down of modified proteins. We evaluated the efficacy of treating cells with azido-farnesol and azido-geranylgeraniol substrates (*15, 16*). These analogs can be "tagged" using Click chemistry techniques which react the azide with an alkyne (*17*); in our case, we used TAMRA-alkyne. We then could detect azido-prenylated proteins using Western blotting and fluorescence scanning techniques. Overall, we were able to detect prenylated proteins; however the level of incorporation was very low, proteins of molecular weight that were known substrates were not labeled with high efficiency, and there was high background. Although the low analog incorporation and high background issues could likely be fixed by further optimization, a tremendous caveat to using these analogs is that they may change the specificity of the enzyme. We therefore chose to investigate the prenylation pathway using a different approach.

To explore the substrate selectivity in the prenylation pathway, we developed a method using a library of GFP mammalian expression vectors, transfections, and fluorescence microscopy. We prepared a vector system that contains $His_6$-EGFP-TEV-$X_{11}$-CaaX genes. The $His_6$ tag allows for protein pull-down, EGFP is the green fluorescent protein, TEV is a protease site, and the C-terminal tail contains 15 amino acids that correspond to a protein of interest. This tail contains the CaaX sequence as well as the upstream region which may be important for substrate selectivity (*12, 18*) and, in some cases, is a potential palmitoylation site. These vectors were also designed such that

the 15 C-terminal DNA sequences are flanked by KasI and EcoRI restriction enzyme sites. DNA oligonucleotides corresponding to C-terminal sequences of interest were annealed together and included overhangs that match the KasI and EcoRI sticky ends. The sequences in the C-terminal tail can be quickly interchanged using digestions and ligations to create each plasmid. Overall, the method is to transfect cells with the $His_6$-EGFP-TEV-$X_{11}$-CaaX genes, observe the localization using fluorescence microscopy, and ultimately pull-down the proteins, digest with TEV protease to obtain the modified C-terminal tail, and analyze the post-translational modifications using mass spectrometry.

We have made progress on this technique in both *in vitro* proof-of-principle experiments as well as some *in vivo* work. We successfully constructed a small library of $His_6$-EGFP-TEV-$X_{11}$-CaaX vectors as well as a bacterial expression vector with the $His_6$-EGFP-TEV-$Ras_{15}$ gene. As a test, we expressed the gene in *E. coli*, purified the protein, and tested TEV protease digestion conditions. We were able to successfully digest the recombinant protein. Additionally, we detected a farnesyl modification on an *in vitro* modified peptide using mass spectrometry. These experiments suggest that the overall method can be applied to a mammalian system.

We tested two small libraries of $His_6$-EGFP-TEV-$X_{11}$-CaaX genes *in vivo*. A library with C-terminal tail sequences that corresponded to high activity *in vitro* peptides generally showed membrane localization *in vivo*. Furthermore, the *in vivo* localization changed depending on the C-terminal sequence of the protein, with some localizing to the plasma membrane and others localizing to various organelles. A library of constructs that contain C-terminal tail sequences that correspond to peptides predicted to be substrates by FlexPepBind generally showed diffuse localization *in vivo*. A hypothesis explaining

this behavior is that the FlexPepBind peptides had lower activity *in vitro* with FTase, so a lower fraction of the protein is modified within the cell and therefore difficult to see localization. In the future, it is likely that the mass spectrometry will be the most informative technique in identifying the various modifications or combination of modifications on proteins. Overall, this work is a good first step towards understanding and testing the relationship of substrate specificity and the prenylation pathway.

## Future Directions

*Investigation of the region upstream of the CaaX sequence*

Preliminary experiments in this work with GGTase-I has indicated that the sequence upstream of the CaaX sequence may affect the peptide reactivity. In general, dns-GCxxx peptides were 1-3 times more reactive than dns-TKCxxx peptides. We have obtained more dns-GCxxx peptides that match the "CaaX" sequence of dns-TKCxxx peptides in our current library. A future experiment would be to test the reactivity of the dns-GCxxx peptides with GGTase-I. We could then have a larger sample size of peptides with matching CaaX sequences to compare reactivity and obtain a clearer picture of the overall trend. Furthermore, we could test longer peptides with longer upstream regions *in vitro* that more closely correspond to protein sequences to compare reactivity. As we have an *in vivo* system using the $His_6$-EGFP-TEV-$X_{11}$-CaaX vectors, we could easily test the amenability of prenylation for CaaX sequences with upstream regions in the cell. Different upstream sequences could be swapped with various CaaX sequences. Besides just testing the effect of the upstream region, we could also test the effect of palmitoylation on membrane localization since many palmitoylation sites are found near

prenylation sites (*19*). Palmitoylation has been suggested to be a method of regulating localization to and from the plasma membrane, so understanding the interplay of palmitoylation, prenylation, and localization would be interesting and informative. Furthermore, additional upstream modifications, such as acetylation and methylation, may regulate the reactivity of these substrates with FTase/GGTase-I.

*Computational methods: predicting GGTase-I substrates*

Now that the new and improved computational method, FlexPepBind, has been established (*14*) to predict FTase substrates, it would be useful to develop a similar method for the prediction of GGTase-I substrates. A large amount of *in vitro* reactivity data with GGTase-I is now available from this work, so training and test sets could be established. Crystal structures of the GGTase-I enzyme and substrates are also available for the threading experiments (*7, 20*). We are in fact currently working with the Scheuler-Furman lab to develop a program predicting GGTase-I substrate selectivity.

An interesting aspect of a program to predict GGTase-I substrates is that the outcome of FlexPepBind with FTase and GGTase-I could be compared to enhance the prediction of the overlap between substrates for the prenyltransferases. These predictions could then be tested *in vivo* and *in vitro*. We could also compare the overall number of modified proteins by each enzyme. Furthermore, it would be interesting to develop a similar computational method for the parasitic enzymes; crystal structures of GGTase-I from *C. albicans* and FTase from *C. neoformans* (both pathogenic fungi) have recently been solved (*21, 22*).

*Determining in vivo substrates of the prenylation pathway*

We have established a method using $His_6$-EGFP-TEV-$X_{11}$-CaaX mammalian vectors to probe the *in vivo* localization of proteins that may be modified by enzymes in the prenylation pathway. There is additional work that could be done to further investigate this system *in vivo*. As mentioned previously, mass spectrometry methods will be highly informative in identifying the post-translational modifications and combination of modifications of proteins; however, this method must still be optimized. Another method to potentially improve the expression and localization of low activity $His_6$-EGFP-TEV-$X_{11}$-CaaX genes is to use expression vectors that dually express both the FTase (or GGTase-I) protein as well as the $His_6$-EGFP-TEV-$X_{11}$-CaaX gene. A previous lab member had some success in improving the localization of a gene using this technique, so a similar method could be applied to this technique. These plasmids are already cloned; the next step is to optimize gene transfection and expression *in vivo*.

In the future, several deeper aspects of the prenylation pathway could be studied. Various inhibitors could be tested to evaluate which proteins are affected by blocking prenylation. Testing the effect of the inhibitors would aid in understanding what cellular pathways are affected with drug treatment. Further, it was recently discovered that the bacteria *Legionella pneumophila* hijacks the host prenyltransferase machinery to catalyze prenylation of bacterial proteins (*21, 23*). The effect of inhibitors on the prenylation of the bacterial machinery could be determined; inhibitors towards FTase and GGTase-I could be potential antibiotic drugs.

A long term goal is to determine the entire prenylated proteome, and this could be carried out by mass spectrometry as it would identify both the proteins and the

modifications. Advantages of this approach would be that native protein substrates and lipid donors would be used, and endogenous prenyltransferase enzymes would catalyze the prenylation. No inhibitors would be necessary for the incorporation of FPP or GGPP analogs. This method would entirely reflect the native conditions and states of the proteins and modification status in the cells. A technique to make this goal a reality would be to enrich for the prenylated proteins after cell lysis so that the concentration would be high enough for mass spectrometry detection. First, membrane cellular fractions could be isolated since presumably most of the prenylated proteins would reside in cellular membranes. Then, an affinity column that binds to the hydrophobic prenyl modifications could isolate the prenylated proteins. Specifically, a modified β-cyclodextrin resin has been shown to bind to prenyl groups (*24, 25*), so a column such as this could be used to further enrich for farnesylated and geranylgeranylated proteins from the membrane fractions. Then, proteins could be identified by mass spectrometry, only if, of course, they are expressed at a high enough level. Once this method is established, we could monitor the whole complement of prenylated proteins under inhibitor treatment, in different cells lines that model various diseases (such as the HGPS cells), and potentially move to identifying prenyl modifications in specific organs in animal systems. Prenylation plays a role in a variety of cellular pathways and in various diseases, so understanding these enzymes is incredibly important. The possibilities for further work in this system are endless.

## References


(1) Reiss, Y., Stradley, S. J., Gierasch, L. M., Brown, M. S., and Goldstein, J. L. (1991) Sequence requirement for peptide recognition by rat brain p21ras protein farnesyltransferase. *Proc Natl Acad Sci U S A 88*, 732-6.

(2) Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L., and Gibbs, J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem 266*, 14603-10.

(3) Hightower, K. E., Huang, C. C., Casey, P. J., and Fierke, C. A. (1998) H-Ras peptide and protein substrates bind protein farnesyltransferase as an ionized thiolate. *Biochemistry 37*, 15555-62.

(4) Goldstein, J. L., Brown, M. S., Stradley, S. J., Reiss, Y., and Gierasch, L. M. (1991) Nonfarnesylated tetrapeptide inhibitors of protein farnesyltransferase. *J Biol Chem 266*, 15575-8.

(5) Pompliano, D. L., Gomez, R. P., and Anthony, N. J. (1992) Intramolecular Fluorescence Enhancement - a Continuous Assay of Ras Farnesyl - Protein Transferase. *Journal of the American Chemical Society 114*, 7945-7946.

(6) Hougland, J. L., Hicks, K. A., Hartman, H. L., Kelly, R. A., Watt, T. J., and Fierke, C. A. (2010) Identification of Novel Peptide Substrates for Protein Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J Mol Biol 395*, 176-90.

(7) Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol 343*, 417-33.

(8) Hartman, H. L., Hicks, K. A., and Fierke, C. A. (2005) Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity. *Biochemistry 44*, 15314-24.

(9) Raveh, B., London, N., and Schueler-Furman, O. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins 78*, 2029-40.

(10) Das, R., and Baker, D. (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem 77*, 363-82.

(11) Krzysiak, A. J., Aditya, A. V., Hougland, J. L., Fierke, C. A., and Gibbs, R. A. (2009) Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates. *Bioorganic & Medicinal Chemistry Letters 20*, 767-770.

(12) Maurer-Stroh, S., and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biol 6*, R55.

(13) Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F. L., and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Comput Biol 3*, e66.

(14) London, N., Lamphear, C. L., Hougland, J. L., Fierke, C. A., and Schueler-Furman, O. (2011) Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. *PLoS Comput Biol 7*, e1002170.

(15) Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004) A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci U S A 101*, 12479-84.

(16) Chan, L. N., Hart, C., Guo, L., Nyberg, T., Davies, B. S., Fong, L. G., Young, S. G., Agnew, B. J., and Tamanoi, F. (2009) A novel approach to tag and identify geranylgeranylated proteins. *Electrophoresis 30*, 3598-606.

(17) Moses, J. E., and Moorhouse, A. D. (2007) The growing applications of click chemistry. *Chem Soc Rev 36*, 1249-62.

(18) Hicks, K. A., Hartman, H. L., and Fierke, C. A. (2005) Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I. *Biochemistry 44*, 15325-33.

(19) Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R., 3rd, and Davis, N. G. (2006) Global analysis of protein palmitoylation in yeast. *Cell 125*, 1003-13.

(20) Taylor, J. S., Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2003) Structure of mammalian protein geranylgeranyltransferase type-I. *Embo J 22*, 5963-74.

(21) Ivanov, S. S., Charron, G., Hang, H. C., and Roy, C. R. (2010) Lipidation by the host prenyltransferase machinery facilitates membrane localization of Legionella pneumophila effector proteins. *J Biol Chem 285*, 34686-98.

(22) Hast, M. A., and Beese, L. S. (2008) Structure of protein geranylgeranyltransferase-I from the human pathogen Candida albicans complexed with a lipid substrate. *J Biol Chem 283*, 31933-40.

(23) Price, C. T., Al-Quadan, T., Santic, M., Jones, S. C., and Abu Kwaik, Y. (2010) Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of Legionella pneumophila. *J Exp Med 207*, 1713-26.

(24) Chung, J. A., Wollack, J. W., Hovlid, M. L., Okesli, A., Chen, Y., Mueller, J. D., Distefano, M. D., and Taton, T. A. (2009) Purification of prenylated proteins by affinity chromatography on cyclodextrin-modified agarose. *Anal Biochem 386*, 1-8.

(25) Nguyen, T., Joshi, N. S., and Francis, M. B. (2006) An affinity-based method for the purification of fluorescently-labeled biomolecules. *Bioconjug Chem 17*, 869-72.