EVALUATING AND EXPLOITING THE SUBSTRATE SELECTIVITY OF FARNESYLTRANSFERASE

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

GARRETT LEE SCHEY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISOR: PROF. MARK D. DISTEFANO

AUGUST 2023

Acknowledgments

With great thanks to my parents for supporting and encouraging me throughout the whole journey, friends both near and far for listening tireless, labmates who shared the stress of the journey and offered great advice, collaborators who helped expand the projects, my undergrads who were a joy to mentor, and Dr. Distefano, for the opportunity to research and helping me grow as a scientist. Thank you all!

Dedication

For Dr. Richard H. Smith, Jr.

For setting me on the path to becoming a scientist, exemplifying excellent scientific and mentoring skills, and always believing in and encouraging me.

“Work, Finish, Publish.” – Michael Faraday

Abstract

Protein prenylation is a post translational modification where an isoprenoid is attached to a CaaX sequence at the c terminus of a protein. Ultimately, after this modification and subsequent steps, prenylated proteins associate with the plasma membrane and are involved in many critical cellular functions. This dissertation describes an exploration of the amazing range of substrates of FTase, with both regards to the peptide and isoprenoid binding site, as well as how this selectivity can be expanded and leveraged for novel applications. The first chapter is a comprehensive literature review of the knowledge that got us to this point, discussing how the canonical rules of the composition of CaaX sequences was determined as how those rules were expanded, as well as the use of bioorthogonal isoprenoid probes and how they have been used to study prenylation. Chapter 2 describes the development of a MALDI-MS peptide library based assay to probe what amino acids are allowed in a pentapeptide CaaaX sequence, based on the recently discovered prenylatable sequence CMIIM. Upon success of this work, expansion to another sequence, CSLMQ, as well as a comparison between the specificity of rat and yeast FTase for these two sequences is discussed in chapter 3. Finally, chapter 4 focuses on the evaluation of a dual-bioorthogonal mutant FTase, an idea that combines two previous mutants, selectively transferring a bulky coumarin containing isoprenoid to an non-naturally recognized charged CaaX sequence. This dissertation expands our continuing understanding of potential substrates of FTase and their potential applications.

Table of Contents

List of Tables

List of Figures

Chapter 1: Defining and Utilizing the Substrate Selectivity of Farnesyltransferase

1.1.Introduction to Protein Prenylation

Protein prenylation is an post-translational modification in which either a 15 carbon farnesyl group or 20 carbon geranylgeranyl group is appended to a residue near the C-terminus of a protein from farnesyl diphosphate (FPP) or geranylgeranyl diphosphate (GGPP), respectively [1]. Three major prenyltransferase enzymes are responsible for these modifications. Farnesyltransferase (FTase) and geranylgeranyltransferase type 1 (GGTase-I) catalyze the attachment of a single farnesyl or geranylgeranyl isoprenoid group to the sulphur atom of a cysteine residue present in a shared recognition motif at the C terminus of a protein known as a "Ca_1a_2X box". In the canonical Ca_1a_2X model, C is Cys, a_1 and a_2 are generally aliphatic amino acids and X is normally the residue largely responsible for determining if the sequence is a substrate for FTase or GGTase-I. Geranylgeranyltransferase type 2 (GGTase-II) catalyzes the addition of two geranylgeranyl groups to two cysteine residues in sequences close to the C terminus of Rab proteins to recognition sequences CXCX or XCCX (Figure1.1) GGTase-II has a low affinity for Rab proteins, and a Rab Escort Protein (REP) is necessary for the prenylation of Rab proteins [2]. In addition, there is a more recently discovered GGTase-III which geranylgeranylates FBXL2, a ubiquitin ligase and Ykt6, a Golgi SNARE protein. FBXL2 terminates in the sequence CCVIL where the second Cys is geranylgeranylated, while Ykt6 terminates in CCAIM where the first Cys is geranylgeranylated and the second Cys is farnesylated [3,4]. Prenyltransferase enzymes are heterodimers, with FTase and GGTase-1 sharing an α

subunit but having different β subunits. GGTase-II has distinct α and β subunits. GGTase-III shares its β subunit with GGTase-II, but has a unique "orphaned" α subunit [3]. Monoprenylated proteins normally undergo further processing where the a_1a_2X motif is cleaved, followed by the carboxymethylation of the revealed prenylated Cys. Association to the plasma membrane occurs due to the synergistic effects of prenylation and either an upstream polybasic domain such as in K- Ras4B, or palmitoylation at an upstream Cys residues such as in H-Ras [5]. It has been shown that in yeast, some proteins are processed through what is known as a shunt pathway, in which they are fully functional after farnesylation without the need for further processing [6].

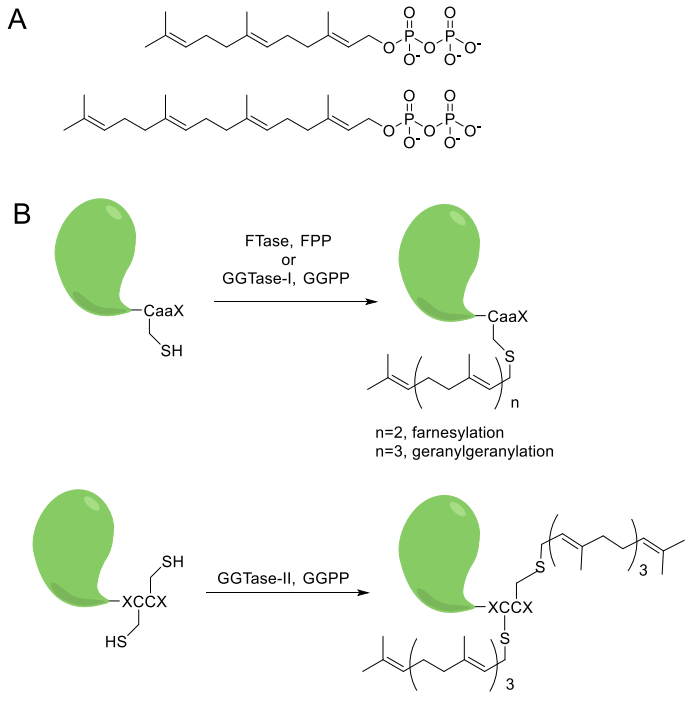


Figure1.1. Three major prenyltransferase reactions

A. Structures of FPP and GGPP. B. Reactions catalyzed by FTase, GGTase-I, and GGTase-II

Historically, prenylation has been implicated in many cancers and there has been substantial interest in developing inhibitors [7–10]. In recent years prenylation has also drawn more recent attention as a potential therapeutic target for many other diseases including Alzheimer's disease, Hutchinson−Gilford progeria syndrome, and more [11–15]. In 2020, the first inhibitor of FTase, lonafarnib, was approved for the treatment of progeria. [16,17]

The number of potential CaaX sequences is large and physiochemically diverse, and hence there has been substantial interest in understanding the rules of what sequences can or can't be prenylated [18]. There is also a wide variety of isoprenoid analogs with various functional groups that can be transferred [19]. Studies utilizing various isoprenoids indicate that the size and shape of the isoprenoid contribute more to the transfer to the peptide substrate than lipophilicity [20]. Analysis of these two flexibilities may be further complicated by the fact that different isoprenoid substrates can effect which amino acids are preferred in the CaaX sequence [21,22].

The mechanism of transfer of FPP to substrate peptide is well understood through both enzymological and crystallographic studies. The isoprenoid is bound in a pocket lined with aromatic residues. The peptide substrate is coordinated to Zn^{2+} via its thiol group and is recognized through several key hydrogen bonds with the C terminal carboxylate, both direct and water mediated [23]. A crystal structure of this binding site is shown in Figure 1.2. When both substrates are bound, a rotation of the first two isoprene units brings the C1 carbon toward the attacking Cys residue. This is consistent with previous kinetic models

[24]. Following this rotation, attack of the Cys to C1 occurs rapidly, followed by slow release of the product which is dependent on the binding of both a new FPP substrate and new peptide substrate [25]. While the mechanism of substrate binding is in theory random, it is in practicality ordered as the kinetics of the binding of the isoprenoid are more rapid than those of the peptide [26].

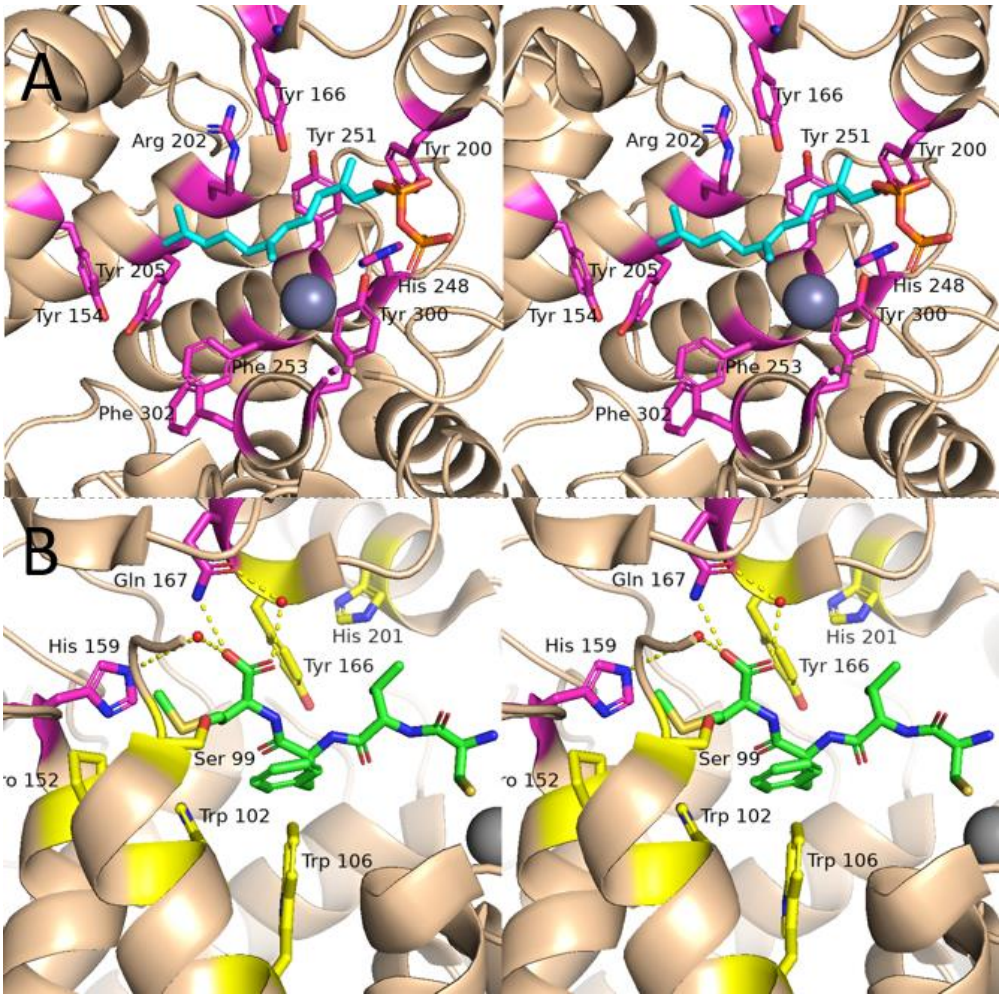


Figure 1.2 Active site of FTase

A. FPP binding site. FPP is shown in teal, and relevant residues in the binding site have their sidechains displayed in magenta. The catalytic Zn is displayed as a sphere. B. Peptide binding site. The peptide is shown in green, with magenta being used to highlight the direct

and water mediated hydrogen bonds responsible for the carboxyterminus to bind. Amino acids responsible for interacting with the peptide side chains are shown in yellow. The catalytic Zn is displayed as a sphere and can be seen coordinating Cys. Image based on pdb id 1JCR

Due to the fact that nearly any protein can become a substrate for prenyltransferase enzymes upon fusion of a CaaX-box sequence combined with isoprenoid flexibility, means there are a wide range of ways to leverage these enzymes for biotechnological functions including fluorescent labeling and targeted delivery of drug molecules [27,28].

1.2. Evaluation of the Selectivity of the Peptide Binding Site

1.2.1. Definition of CaaX requirements

Initial investigations of the sequence requirements for *Rattus norvegicus* FTase (rFTase) defined aliphatic amino acids in the a_1 and a_2 positions, with the X position being the most stringent [29]. Amino acids that were preferred in the X position were found to be Met, Ser, Phe, Gln, Ala, and Cys. Charged amino acids were somewhat tolerated in the a_1 position and far less tolerated in the a_2 position. Gly and Pro were generally found to make poor substrates at all positions. Further examination of the role of the X amino acid in both rFTase and rGGTase-I, indicated that peptides with Phe in this position could be modified by both enzymes. Peptides ending in branched amino acids such as Ile and Leu were preferred by rGGTase-I, and both enzymes were not tolerant of charged amino acids in the X position. It was also found that some sequences such as CVIM, while not a good substrate for geranylgeranylation, can be processed by rGGTase-I at a slow rate. The canonical

rGGTase-1 CaaX sequence CVLL cannot be processed by rFTase, leading to the idea that the farnesylation of peptide sequences terminating in Leu was not possible [30,31]. While such peptides are usually thought of as only rGGTase-I substrates the discovery that RhoB, terminating in CKVL, can both be farnesylated and geranylgeranylated challenged that view. In order to expand the understanding of CxxL peptides, the evaluation of a 41 peptide library was analyzed. These peptides were chosen based on predictions from a crystal structure and a database search. These proteins contained diverse amino acids in the a_1 position and approximately half were known to be geranylgeranylated while half had no known prenylation. Analysis of these peptides identified 10 sequences that were efficient rFTase substrates, four of which were not known to be prenylated *in vivo*. The best substrates were those in which the a_1 and a_2 amino acids were both smaller, uncharged amino acids such as Ile, Val, Leu, Cys, and Ser [32].

Further investigations that rFTase recognizes peptide substrates based on both hydrophobicity and side chain volume, and there is interdependence between the a_2 and X positions [33]; for example, large residues at the a_2 position preclude large residues at the X position. In general, the optimal volume at the a_2 position seemed to be Valine at 140 Angstroms when X is Ser or Ala, but it was found that changing the X position affected the reactivity when nonpolar amino acids were in the a_2 position. When the X amino acid was Met or Gln, reactivity did not substantially decrease when the a_2 amino acid was larger than Val, allowing for more variability in amino acids at this position. Additionally, bulky amino acids were found to be disfavored at the a_1 position, indicating a steric effect. Thus, the conclusion was made that rFTase substrate recognition is based on the sum of the total

features of the CaaX sequence rather than recognition of features from each individual amino acid. Another feature of CaaX sequences is that the upstream amino acids can affect the results of *in vitro* experiments, particularly when the Ras polybasic sequence is used [34].

Further work focused on large peptide libraries to further evaluate the interplay between the a_2 and X positions with both rat and yeast (*Saccharomyces cerevisiae)* FTase (yFTase). [35]. CaaX peptides with a photocleavable linker were synthesized on a membrane so that cyclization and cleavage allowed the peptides to be oriented in an N to C direction. The use of FTase along with an alkyne-containing analog aided in the search for substrates. Successful reactivity was evaluated by click chemistry with biotin azide followed by colorimetric detection based on enzyme-linked streptavidin. Using this method to evaluate amino acids preferred in the a_2 and X position using both the yeast and rat enzymes revealed many peptides that agreed with the canonical model and also a few outliers that are known to exist in viruses and bacteria, providing potentially new therapeutic targets [36]. Analysis with differing isoprenoid chain lengths revealed that analogs with 10 and 15 carbon lengths had similar reactivity, while a 5 carbon length displayed a different reactivity pattern. Comparing the reactivity of rFTase with the FTases of the yeasts *Saccharomyces cerevisiae* and *Candida albicans* showed that more substrates are shared between *R. norvegicus* and *S. cerevisiae* enzymes than *R. norvegicus* and *C. albicans*. A complete list of CaaX sequences identified by proteomics can be found in Table A1

1.2.2. Expanding the definition of CaaX sequences

While the canonical model was still being established, there was also interest in the use peptide libraries to define sequences that were outside of the canonical CaaX model that were bona fide substrates, finding that CKQQ and CKQM were efficient substrates of rFTase [37]. Subsequent work screened a library of over 200 CaaX sequences in the human genome and revealed many novel prenylated sequences that were observed under multiple turnovers conditions, as well as a variety of sequences that were only active under single turnover reaction conditions [38]. While the newly discovered peptide sequences that exhibited multiple turnovers were consistent with previous findings of allowed amino acids at the a_2 and X position, it was found that single turnover substrates had more variability at those positions. Following up on the expansion of allowed amino acids in a CaaX sequence, computational and *in vitro* assays determined that it is possible for CaaX sequences to terminate in a negatively charged residue, when the a_1 residue is usually Tyr or Phe, and the a_2 residue conforms to the canonical model [39].

While there have been a variety of *in vitro, in vivo,* and *in silico* experiments to define and expand the definition of a CaaX sequence, it appears that FTase is able to accept a wider variety of substrates in *in vitro* experiments vs *in vivo* [40]. To this end, assays utilizing high throughput yeast assays have greatly helped our understanding of prenylation in a model biological system. Early assays utilized a yeast-based assay, in which a-factor, a fully processed, prenylated peptide pheromone is required for yeast mating [41]. However, a more high throughput Ras-based reporter was later developed [42]. This assay utilizes a temperature sensitive strain of yeast that can only grow at higher temperatures when a constitutively active H-Ras isoform is correctly processed and transported to the plasma

membrane. This system supported the canonical model of the a_2 and X positions. Almost 8000 sequences were tested and it was discovered that only 6.2% of these possible sequences were farnesylated. The downside to such assays is that they rely on two additional processing steps, aaX cleavage and carboxymethylation, for a positive result. It is possible that a peptide can be prenylated but not further processed. Another assay based on the heat shock protein YDJ1 was later developed [40]. That protein must be prenylated for yeast to survive at higher temperatures, but it is part of the shunt pathway of prenylation and after farnesylation, no further processing is required for protein activity. They found that there was increased tolerability for polar uncharged amino acids at the a_2 position and smaller aliphatic amino acids at the X position. This study expanded the understanding of the complexity of the prenylation pathway and the necessity of varied reporter systems to assess prenylation.

Interestingly, the selectivity of rFTase can be additionally altered through mutagenesis in a rationally designed fashion [43]. It has been shown that by substituting a positively or negatively charged residue in the active site in concert with a second mutation to allow extra space, charged peptide sequences such as CVDS and CVKS can be efficient substrates, theoretically forming a strong charge-based interaction. These charged peptides have very poor initial velocities when reacted with wild type rFTase. Additionally, these sequences were shown to be functional in Hek293T cells by transfection with RFP modified with a charged CaaX sequence and associated mutant FTase and monitoring the localization of the prenylated protein. Another similar approach focused on having the X position in the CaaX sequence as a positively charged residue, with mutations adding a

negative charge in the active site [42]. They were able to show effective prenylation via the Ras recruitment assay noted above.

1.2.3. Prenylation of C-terminal sequences is not limited to four amino acids

While the canonical understanding of CaaX binding has long been thought to be limited to tetrapeptides, it has now been shown that extended pentapeptide CaaaX sequences, such as CMIIM, are capable of being farnesylated, but none of the sequences studied were able to be geranylgeranylated [44]. Evaluation of the kinetics of these sequences can be challenging, as many of them do not exhibit a fluorescence increase in a commonly used Dansylglycine-based continuous assay, despite the fact they are indeed farnesylated. However, these sequences have a k_{cat}/K_m value that is one to two orders of magnitude lower than a native CaaX sequence such as CVLS. Following up on these results, utilization of a MALDI based screening approach with yFTase evaluated peptide libraries based around the sequence CMIIM, individually changing each of the four terminal amino acids to be each of the canonical amino acids [45]. The greatest number of hits was found to be in the a_1 position, furthest away from the hydrogen bonds formed by the C terminal carboxylate that are necessary for binding. There was some variability in the a_2 and a_3 positions, while the X position displayed results that are consistent with the canonical CaaX model. These peptides were validated in an *in vivo* yeast model by examining prenylation of YDJ1 by Western blot. There was also an effort to search the human genome for CaaaX sequences with some "CaaX-like" nature. While some of the peptides examined indeed could be farnesylated, the sequence CSLMQ was found to be as efficiently prenylated as the native rFTase CaaX substrate CVLS. In addition to pentapeptide sequences, it was also found

that shorter tripeptide were also capable of being prenylated, with some, such as CFT, that were capable of being both farnesylated and geranylgeranylated [46]. It is noteworthy that both of these reactions were catalyzed by rFTase, an enzyme that is not normally capable of using GGPP as a substrate with tetrapeptides. These same tripeptide sequences were found to not be substrates for rGGTase-1. Taken together, these findings vastly expand the possible substrate space for FTase, and future efforts to define selectivity rules for these longer and shorter sequences should be evaluated. Because there are proteins in the human genome with pentapeptide and tripeptide sequences that can be prenylated, this also opens the possibility that such proteins could be prenylated *in vivo* and has significance in both biological and biotechnological dimensions.

1.2.4. Peptide and peptidomimetic inhibitors of farnesyltransferase

The flexibility of the peptide binding site has led to interest in the development of peptide and peptidomimetic inhibitors of FTase. Early work in this area found that tetrapeptides such as CVFM are not prenylated and act as competitive inhibitors of FTase, and that a positive charge on the amine group of the cysteine was necessary, as blocking or removing the amine resulted in peptides that could be prenylated [47,48]. Peptidomimetic inhibitors have utilized many strategies and isosteres such as methylation of the cysteine to block Zn coordination, replacing the aliphatic amino acids with a benzodiazepine moiety, use of non-canonical amino acids, pseudopeptide bonds, and beta amino acids [48–53]. There is interest in such molecules as a way to selectively target FTase in pathogenic organisms[54]. Structures of small molecule inhibitors of FTase are shown in Figure 1.3

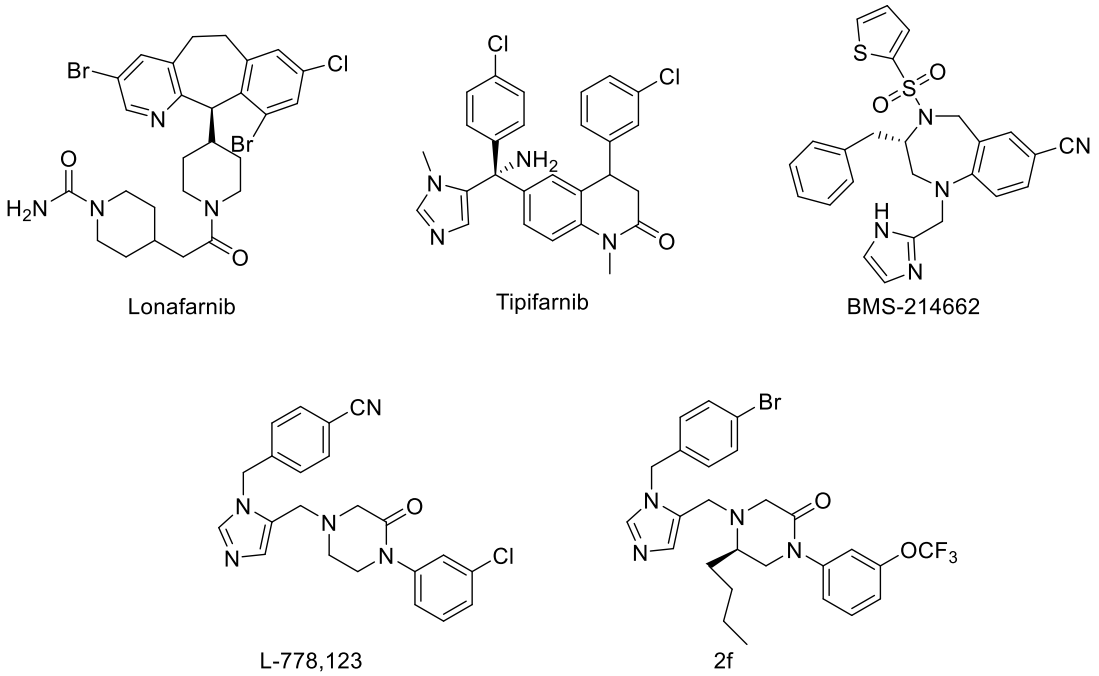


Figure 1.3 Small molecule inhibitors of FTase.

1.3. Leveraging the Selectivity of the Isoprenoid Binding Site

1.3.1 Fluorescent isoprenoid analogs

A wide variety of fluorescent analogs have been used to evaluate the kinetics of isoprenoid substrates and inhibitors for their respective prenyltransferase enzyme. Those highlighted in this work are shown in Figure 1.4.

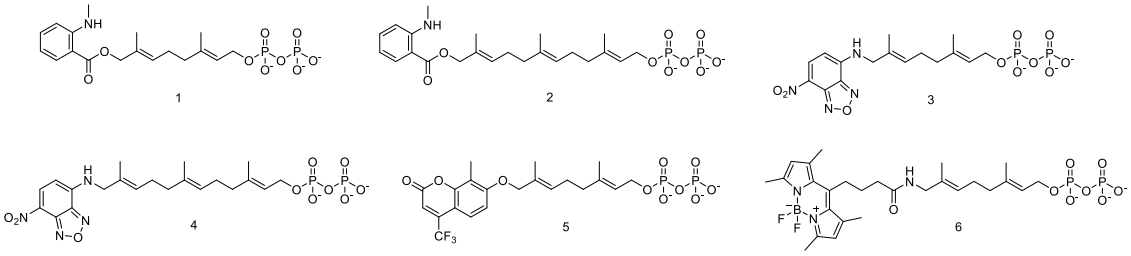


Figure 1.4. Structures of fluorescent isoprenoid analogs discussed in this work

Early work described the synthesis and use of both 10 carbon (**1**) and 15 carbon (**2**) N-methylanthranilate (NMA) isoprenoid analogs [55]. Both of these probes were efficient substrates of rGGTase-II and could be transferred to RAB7, with the longer probe binding better. Competition with GGPP reduced labeling by 70 percent, and these studies were not able to determine if lipid association was still possible after transfer. These probes were used to assess the binding kinetics of FPP and GGPP to rGGTase-II [56]. Direct and competitive titration of NMA analogs and FPP and GGPP allowed the calculation of binding constants, showing that the longer molecules were more tightly bound, with both native isoprenoids bunding tighter than their NMA analogs. FPP and GGPP had a K_d of 60 and 9 nM, respectively, with their NMA analogs having K_d values of 49 and 330 nM. Stop flow kinetic experiments revealed that while the association constants were all similar, dissociation constants differed, with less tightly bound isoprenoids dissociating faster. Kinetic parameters indicate that FPP binds in one step to rGGTase-II, but GGPP binds in a multi-step process that involves a conformational change to keep the isoprenoid tightly bound. NMA probes have use in studying other enzymes involved in prenyl pyrophosphate synthesis, with compound **1** being used to evaluate the kinetics of undecaprenyl pyrophosphate synthase (UPPS, an enzyme important for lipid construction of the bacterial cell wall) with its native substrate isopentyl pyrophosphate [57]. It was then subsequently utilized to explore the mechanism of inhibition of other pyrophosphate synthases such as farnesyl pyrophosphate synthase and geranylgeranyl pyrophosphate synthase [58].

A nitro-benzo[1,2,5]oxadiazol-4-ylamino (NBD) isoprenoid analog was developed with the purpose of profiling prenyltransferase inhibitors [59]. They synthesized both a 10 carbon (**3**) and 15 carbon (**4**) isoprenoid. They found that compound **3** was an efficient substrate for FTase (k_{cat}/K_m of $2x10^5$ M^{-1} s^{-1}), and compound **4** was an efficient substrate for both rGGTase-I and rGGTase-II (k_{cat}/K_m of $30x10^5$ M^{-1} s^{-1} and $0.1x10^5$ M^{-1} s^{-1} respectively). An on-bead fluorescence microscopy assay was utilized to evaluate peptidomimetic 3-mer sequences as inhibitors of prenyltransferase, using a library based on a previously discovered tripeptide. In this assay, substrates were attached to beads and incubated with the appropriate enzyme and probe, as well as a potential inhibitor. If the reaction occurs, there was a lack of inhibition and the NBD probe was transferred to the bead. Results were easily visualized by fluorescence microscopy. To further validate the usefulness of these probes, they showed that they were capable of entering epidermal carcinoma A431 cells in a manner independent of intracellular membrane transport. In addition, compound **4** was utilized in a continuous fluorescence assay with rGGTase-II, displaying a remarkable increase in fluorescence when transferred to Rab. Mutational studies indicated that the first Cys in the recognition sequence was the substrate for this probe [60].

A coumarin containing isoprenoid (5) was utilized to study the kinetics of UPPS [61]. It acts as a competitive inhibitor, undergoing one of the normal 8 elongation steps catalyzed by UPPS. Stop flow kinetic data revealed that while most prenyltransferase enzymes have a slow dissociation for FPP, the dissociation of FPP from UPPS is quite rapid by comparison. This probe was later used as a substrate for FTase, although it required a

mutation in the active site for the bulkier coumarin to be efficiently transferred [62]. This mutation increased the relative k_{cat}/K_M 300 fold vs FPP and could be efficiently transferred to a GFP-CaaX protein and used as an efficient FRET pair.

A 10 carbon isoprenoid analog containing a BODIPY fluorophore (**6**) was evaluated as a substrate for FTase, GGTase-I, and GGTase-II [63]. Interestingly, this molecule bound to all three enzymes, but was only able to be transferred by rGGTase-II. The K_d of **6** was the highest for rGGTase-II at 245 nM, while it displayed a 1 nM and 0.23 nM K_d for rFTase and rGGTase-I, respectively. It was found that compound **6** was well suited for *in vivo* biological experiments since it was able to easily enter baby hamster kidney cells. It was also used in animal studies where injection of the probe into 3 day old zebrafish embryos showed significant accumulation in the circulatory and digestive systems.

1.3.2 Photoaffinity labeling isoprenoid analogs

Photoaffinity labeling has been used to help determine how prenyltransferase enzymes bind isoprenoids, as well as evaluate their substrates. Those highlighted in this work are shown in Figure 1.5.

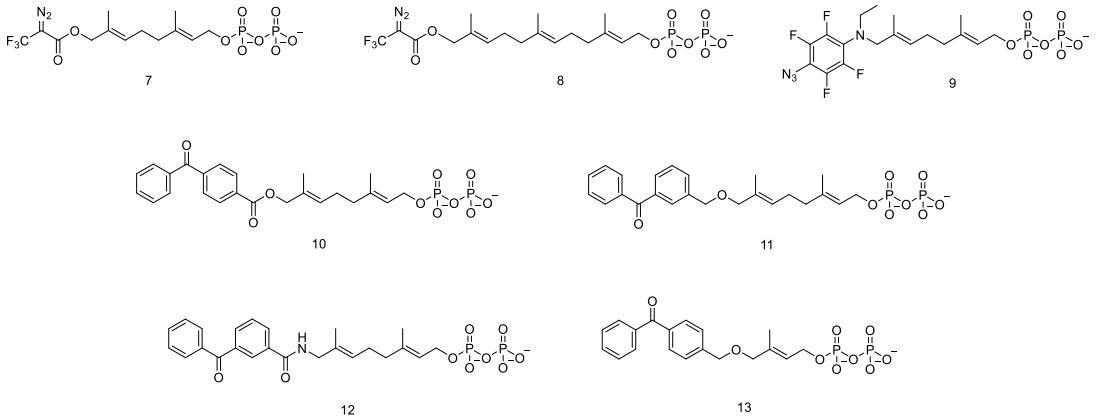


Figure 1.5. Structures of photoaffinity isoprenoid analogs discussed in this work

Early work involved the use of diazo-3-trifluoropropionyl (DAFTP) containing isoprenoids (**7**) as inhibitors of *Homo sapiens* FTase (hFTase) in both malignant and non-malignant lymphocytes [64]. By studying the inhibition of p21H-Ras in cell lysates by Western blot, it was shown that with 10 μM probe, greater than 70% of the prenylation was inhibited. Further studies using **7** in the characterization of hFTase showed that the β-subunit was labeled [65]. This was one of the earliest works to implicate the β-subunit as catalytically important in FPP transfer, as well as confirm both subunits need to be present in order for FPP to bind. A longer analog containing DAFTP (**8**) was used as in inhibitor for both hFTase and hGGTase-1 having a K_I of 100 nM against hFTase, and 18 nM against hGGTase-I, and a competitive inhibition profile [66]. Both probes **7** and **8** were used to confirm that yFTase had a similar labeling pattern to the human enzyme, and could also be enzymatically transferred to Ras proteins and peptides, making them valuable tools for learning what proteins interact with prenylated Ras proteins [67]. Probe **8** was also used to study GGTase-II mediated prenylation of Rab5 in rabbit reticulocyte lysate, confirming by Western blot that Rab5 can be doubly modified with the analog, and that the prenyl chain interacts with both Rep and GDI-1, as well as other unknown proteins [68].

An aryl azide containing probe (**9**) was completely transferred to peptide substrates by rFTase [69]. Photolabeling studies caused a 70% decrease in enzyme activity after irradiation with 1 μM probe. The presence of the probe with no irradiation led to a decrease

in 25% of the transfer of FPP, which indicated that the probe was also an alternative substrate for the enzyme.

Alternatively, benzophenone containing probes have been used to study prenyltransferase enzymes, with the idea that benzophenone compounds are more stable to handle, proceed with radical formation that does not get quenched by solvent, and absorb at higher wavelengths that are less damaging to the enzyme [70]. Use of an ester linked benzophenone (**10**) indicated that it was a competitive inhibitor compared to FPP, and showed that photocrosslinking preferentially labeled the β-subunit agreeing with the labeling identified by other probes [71]. An ether (**11**) and amide (**12**) linked benzophenone were also utilized. The ether linkage resulted in probes where both IC_{50} and K_I values were superior to the ester linked analogs [72]. A crystal structure showed that the benzophenone moiety occupies the same space as the C-terminal residue of the CaaX peptide, which explains why these substrates are inhibitors [73]. This means that while the C1 carbon and the diphosphate moiety are in the correct orientation, the peptide substrate cannot bind. It also predicted the importance of the diphosphate unit as a major contributor to the binding energy of the prenyl diphosphate analogues. The amide linkage resulted in a much lower affinity, perhaps due to the added rigidity of the amide bond [74]. These probes were utilized to study both yFTase and hGGTase-1, and results indicated that the β-subunit of both was preferentially labeled. A shorter ether linked probe (**13**) was shown to be a substrate of both yeast and human FTase, and an x-ray structure showed that the benzophenone does not adopt a coplanar configuration, which might explain poorer photolabeling than expected [75]. Also, the benzophenone is surrounded by tryptophan and

tyrosine residues, which are poorly labeled due to their sp2 hybridization and consequently less reactive C-H bonds.

1.3.3. Isoprenoid analogs used for bioconjugation

The use of isoprenoid probes with the ability to undergo conjugation reactions are of great interest in both defining substrates and creating biotechnologically useful tools. Those highlighted in this work are shown in Figure 1.6.

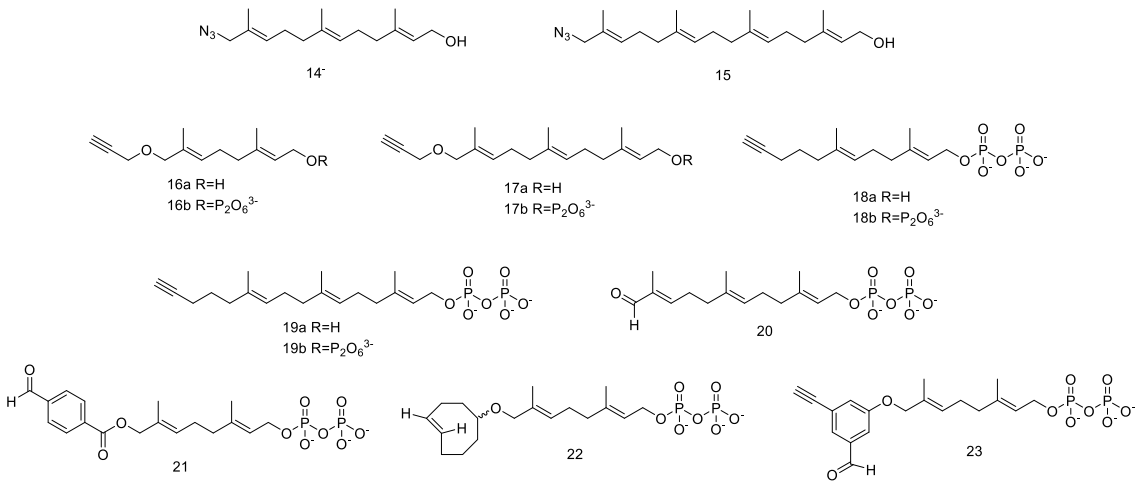


Figure 1.6. Structures of bioconjugatable isoprenoid analogs discussed in this work

Historically, isoprenoids modified with alkyne and azide moieties have been used to evaluate the prenylome in cells through a copper-catalyzed click reaction and subsequent in-gel analysis with a fluorophore or biotin enrichment [76–83]. These probes mimic FPP and GGPP with 10, 15, and 20 carbon chains respectively and can be utilized in an alcohol or diphosphate form, such as (**14**), (**15**), (**16a,b**) and (**17a,b**). The earliest work utilized azides, with (**14**) followed by a reaction with a biotinylated phosphine reagent to identify

farnesylated proteins by MS. Compound **15** was utilized in a similar manner to attach a fluorophore and identify geranylgeranylated proteins. Later, the azide was switched with an alkyne, and it was found that **16b** is an effective substrate for FTase and **17b** was efficiently transferred by both FTase and GGTase-I. Compound **17a** has been incorporated into mammalian cells, and following fixation and permeabilization, a fluorophore could be clicked on and used in confocal microscopy and flow cytometry experiments to demonstrate that different cell types have significant variation in their levels of prenylated proteins [84]. In addition, the level of prenylated proteins in a cell-based model of misregulated autophagy had a significant increase in the amount of prenylated protein compared with normal cells, indicating that there could be an implication for diseases in which autophagy is misregulated. This has been further combined with azide-containing mass cytometry probes to evaluate both levels of prenylation and autophagy biomarkers simultaneously [85]. Findings showed that statins, which are commonly used in metabolic incorporation experiments, are actually detrimental in these models as they activate autophagy related pathways.

Compound **17b** has also been utilized in probing the prenylome of the malaria-causing parasite, *Plasmodium falciparum* [13]. Fascinatingly, this probe may be modified endogenously in *P. falciparum* into a 20 carbon isoprenoid via the addition of one more isopentenyl diphosphate group, potentially providing a better picture of the organisms prenylome. It was found that *Plasmodium falciparum* has substantially fewer prenylated proteins compared to fungi or mammals and is mostly comprised of membrane transport and small GTPases. Additionally, **17b** has been used to study an Alzheimer's disease

mouse model [86]. Findings indicate that a lack of FTase or GGTase-I activity is detrimental to neurons, and a post-mortem *in vitro* proteomic analysis showed that there are very different groups of affected proteins between different knockout models of FTase or GGTase-I. This is made increasingly interesting by use of an *in vivo* model in which the probe was injected into the brains of living Alzheimer's' disease model mice, confirming that there are distinct difference in upregulated prenylated proteins in the disease model vs wild type [87]. Furthermore, there are both shared and unique prenylated proteins among neurons, glia, and astrocytes indicating that different individual proteins may play specific roles in different diseases [88]. Similar alkyne containing probes lacking an ether linkage (**18a,b** and **19a,b**) have been used to evaluate the dynamics of the prenylome in disease models and in combination with FTase inhibitors [89].

In addition to the versatile alkyne probes, other biorthogonal reactivities have been exploited, such as the use of aldehyde groups (**20** and **21**). These aldehyde probes were utilized in both oxime and hydrazone forming reactions [90]. GFP with a CaaX sequence can be modified with both probes, which can then be used for subsequent immobilization, fluorescent labeling, or PEGylation, all of interest for various biotechnological applications. Additionally, hydrazone formation was found to be reversible, and this was used to release fluorescently labeled proteins with amino-oxy-based reagents after the aldehyde-labeled protein had been immobilized via hydrozone formation. Compound **21** was additionally used to modify Designed Ankyrin Repeat Proteins (DARPins), a type of protein designed to tightly bind a specific target as an alternative to antibody scaffolds [27]. Modified DARPins could then be easily labeled with a fluorophore for localization and

binding studies or be conjugated to a toxic drug for directed cell killing. It was also shown that a bulky trans-cyclooctene analog (**22**) could be enzymatically transferred to both peptides and proteins, and had the advantage of being able to undergo tetrazine ligation which is nontoxic, unlike copper catalyzed click chemistry [91]. It was also possible to increase affinity for this probe by mutating FTase [62]. A triorthogonal probe (**23**) was shown to be capable of being added to a CaaX labeled protein, where it underwent both oxime formation and a separate click reaction to modify a protein of interest with both a fluorophore and a drug or PEG molecule to create multivalent conjugates [92]. This provided a facile way to create complex multiprotein assemblies in a predictable manner. Finally, it should be noted that the CaaX recognition differences in combination with isoprenoid preferences was used to synthesize specific diprotein constructs in one pot [93]. A GFP bearing the CaaX sequence CVLL will react only with rat GGTase-1 and a longer azide bearing isoprenoid, while an RFP with the CaaX sequence CVIA will selectively react with rat FTase and a shorter alkyne-bearing isoprenoid. The two could then be clicked together to produce a heterodimeric construct.

1.4.Conclusions

Prenyltransferase enzymes are able to catalyze the transfer of a wide variety of isoprenoid substrates to a large number of peptide and protein substrates. This flexibility has been the focus of numerous studies and has been particularly useful for the creation of probes to unravel the details of the biochemistry of these enzymes, as well as for the development of inhibitors that may be useful for biomedical applications. The application of this knowledge is as varied as the substrates themselves. These applications in protein-protein

constructs highlight the truly amazing versatility of prenyltransferase enzymes for their peptide and isoprenoid substrates, providing a unique and efficient approach to creating valuable tools for chemical biology and drug discovery.

Chapter 2: MALDI Analysis of Peptide Libraries Expands the Scope of Substrates for Farnesyltransferase

2.1. Introduction

Protein prenylation is a post-translational modification which involves the covalent attachment of a hydrophobic isoprenoid group to the thiol side chain of a cysteine residue located near the C-terminus of a protein. Farnesyltransferase (FTase) and Geranylgeranyl transferase (GGTase-I) transfer a 15-carbon farnesyl and 20-carbon geranylgeranyl group, respectively [94]. These enzymes recognize proteins with a C-terminal tetrapeptide consensus sequence known as a "CaaX box", where "C" is the cysteine residue that is covalently modified, "a" is usually an aliphatic amino acid, and the "X" is a residue that is largely responsible for determining whether the protein substrate is targeted by FTase or GGTase-I [1](Figure 1). Protein prenylation is essential for proper cellular localization and signaling activity, and misregulation of prenylated proteins is implicated in many diseases.[5,95] For this reason, prenylation has also drawn more recent attention as a potential target for treatment of Alzheimer's disease, Hutchinson−Gilford progeria

syndrome, and numerous other diseases.[11–15] In 2020, the first inhibitor of FTase, lonafarnib, was approved for the treatment of progeria. [16,17]

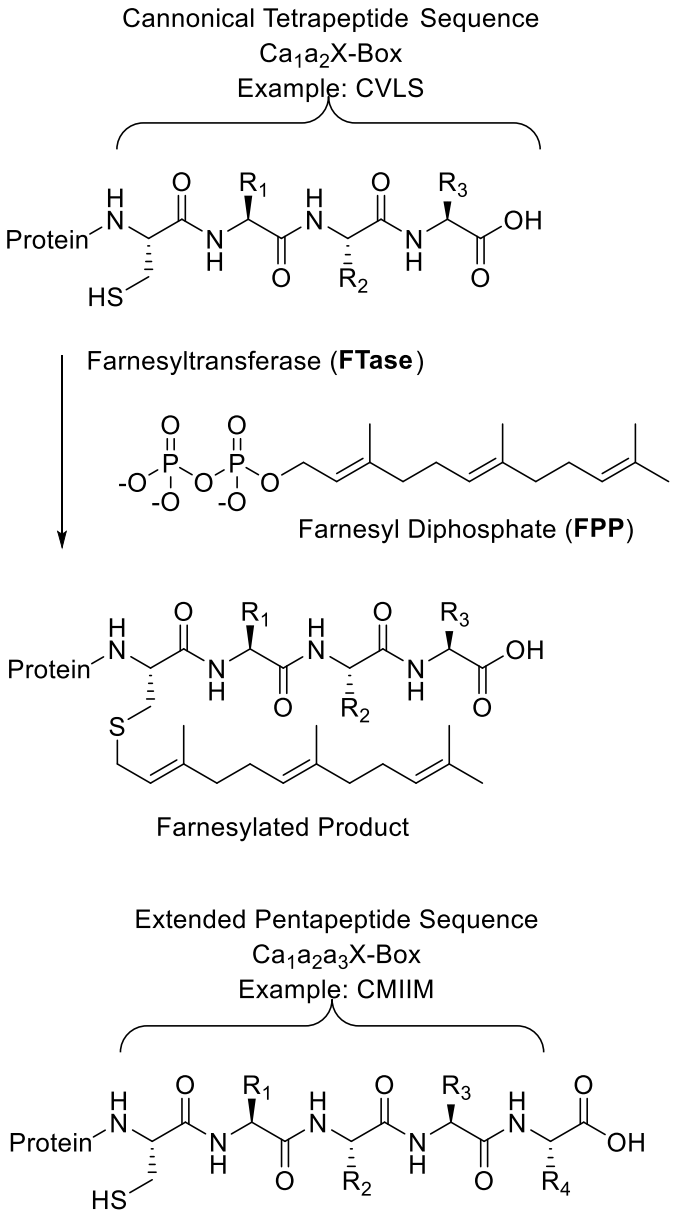


Figure 2.1. Diagram of the farnesylation of a C-terminal canonical tetrapeptide by FTase, as well as an example of an extended pentapeptide sequence

FTase manifests broad substrate specificity, catalyzing the transfer of a farnesyl group from Farnesyl Pyrophosphate (FPP) to a variety of polypeptide substrates, and many attempts have been made to define what amino acids are preferred or not preferred in the CaaX sequence.[18,29] This flexibility has even been leveraged in the design of novel mutant FTases for orthogonal labeling.[43] While the canonical model of the CaaX box is generally well understood, it has recently been found that certain sequences longer than the four-residue CaaX motif can also be farnesylated by both yeast and mammalian FTase orthologs.[44] These CaaaX motifs were first observed in yeast, and initial evaluation of CaaaX substrate space found the sequence CMIIM to be the prototype for the CaaaX sequence, with *in vitro* assays indicating that this peptide was a reasonable substrate ($k_{cat}/K_M = 1.9 \times 10^4\ M^{-1}s^{-1}$), although less efficient compared with the most efficient CaaX peptides such as CVLS ($k_{cat}/K_M = 2.0 \times 10^5\ M^{-1}s^{-1}$).[44]

Though the substrate space of the CaaX containing peptides has become increasingly well defined as a result of the past three decades of research, there are many questions remaining about how this information might apply to the extended CaaaX sequences. Work in our lab and others has relied on a library-based screening approach to probe peptide substrate space, which involves generating a large number of peptides that have a systematic combination of amino acids. Previous analysis of peptide libraries has involved the use of an alkyne-containing isoprenoid analogue to allow for biotin attachment by derivatizing with biotin-azide via copper-catalyzed azide−alkyne cycloaddition. [36] The attached biotin then allowed for visualization of farnesylated peptides via an enzyme-linked assay involving streptavidin-alkaline phosphatase to form a colored product. One disadvantage

of that approach is that it relies on the use of synthetic isoprenoid analogues that may perturb enzyme specificity. To complement that approach, we developed an alternative peptide library screening strategy to analyze the substrate space of pentapeptide CaaaX sequences that would employ MALDI-MS as the method of detection. Since peptide farnesylation results in a significant increase in mass, farnesylated products are easily separated from their unfarnesylated precursors. Moreover, MALDI typically generates singly charged species without fragmentation, allowing for rapid sample analysis, high sensitivity, and is amenable to complex mixtures, giving it numerous advantages.[96,97] Thus, it was hypothesized that MALDI would allow for libraries containing 10-20 members to be quickly and easily analyzed while utilizing the native substrate FPP without the need for a biorthogonal analog for subsequent visualization. Splitting libraries into 2 sets with 10 of the canonical amino acids allows for analysis without isotopic overlap of Leu/Ile and Lys/Gln.[98,99] To start this exploration of novel substrate space, we utilized peptide libraries based on randomization of the best characterized CaaaX sequence, CMIIM, to determine if this sequence's farnesylatability could be improved and if additional amino acid substitutions not normally considered canonically "CaaX-like" could be identified.

2.2. Research Objectives

In this work, we sought to evaluate the potential of FTase to discover novel extended CaaaX sequences that could be substrates of FTase. Our approach was to use a combinatorial peptide library approach coupled with MALDI analysis to probe for potential novel substrates of FTase. We created libraries based on a previously discovered

sequence, CMIIM, as well as searched the human genome for biologically relevant sequences. Sequences discovered through screening were subsequently validated through HPLC and *in vivo* yeast assays.

2.3. Results

2.3.1. Validation and optimization of MALDI method with known substrates

To determine whether the proposed MALDI method could be adapted to identify novel CaaaX sequences, initial efforts focused on peptides with the sequence DsGRAGCVa_2A (where Ds is a dansyl group). The canonical CaaX tetrapeptide CVIA is a native substrate for the yeast FTase, and libraries examining the sequence variability at the a_2 and X positions have been previously reported making this an excellent test case. A DsGRAG tag was appended onto the N-terminus to aid in purification and increase ionization efficiency.[100,101] Thus, a 17-membered DsGRAGCVa_2A library was synthesized, where X was varied to all 20 proteogenic amino acids except cysteine, leucine, and glutamine. Cysteine was omitted due to potential synthetic difficulties, while leucine and glutamine were omitted as they have nearly identical monoisotopic residue masses to isoleucine and lysine respectively, and thus, would be indistinguishable. In the unreacted library, all individual peptide peaks could be observed and resolved in the MALDI spectrum, including those with a difference of only one mass unit (Figure 2A). This library was farnesylated in the presence of different concentrations of FTase to determine the optimal conditions for farnesylating a complex mixture of substrates. Initial reactions containing 0.1 nM enzyme showed no appreciable product formation (Figure 2B). When the enzyme concentration was increased to 10 nM, the intensity of the unfarnesylated

peptides decreased in the MALDI spectrum, and several farnesylated peptides were observed with easily detectable intensity (Figure 2C).

Gratifyingly, most of these initial product peptides contain amino acids at the X position which were shown to be farnesylated effectively in previous studies.[18,35,36] Increasing the enzyme concentration to 100 nM yielded a similar reduction in the intensity of all unfarnesylated peptides and 10 farnesylated peptides were observed with remarkably high intensity (Figure 2D).

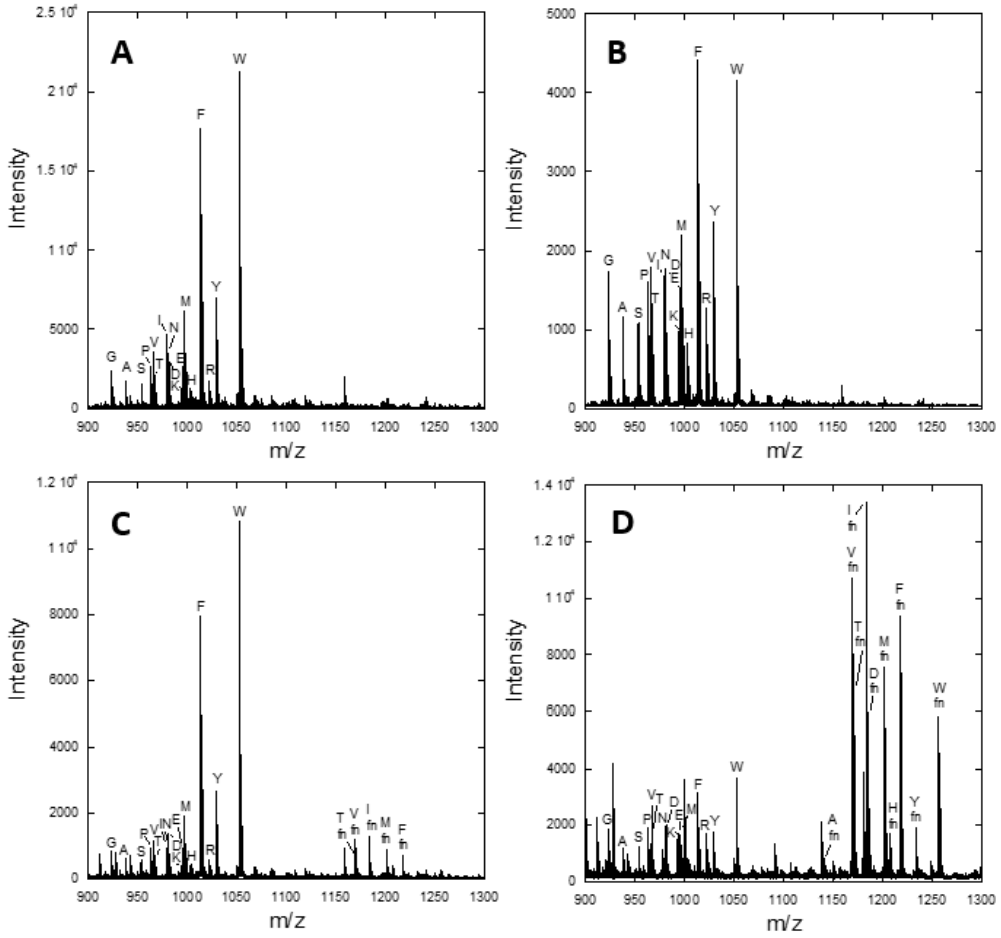


Figure 2.2. Farnesylation of a DsGRAGCVa_2A peptide library with varying yFTase concentrations.

Libraries reacted with **(A)** no enzyme **(B)** 0.1 nM enzyme **(C)** 10 nM enzyme **(D)** 100 nM enzyme. The identity of the residue in the X position is indicated with the letter above each peak. The farnesylated peptides are highlighted with the designator "fn".

2.3.2. Identification of novel substrates from the CMIIM motif using MALDI analysis

With the above validation complete for a simple CaaX library, several libraries were prepared based on the previously reported pentapeptide CaaaX box CMIIM, where the four positions following Cysteine were individually varied to all 20 proteogenic amino acids. This was done using two libraries of 10 peptides for each position, so that all possible amino acid substitutions could be evaluated without the overlap of amino acids with near identical molecular masses. Thus, eight libraries were created in total, two each for the sequences Ca_1IIM, CMa_2IM, CMIa_3M, and CMIIX. The predicted masses of the starting peptides and products along with their experimentally observed counterparts are provided in Table A2. While the ionization of the first unfarnesylated CXIIM library was sufficient, initial attempts to farnesylate this library with 100 nM yFTase proved unsuccessful. This was not surprising since CMIIM was reported to be a 10-fold poorer substrate relative to the native CVIA sequence. Increasing the enzyme concentration to 1 µM gave several hits (Figures 3B, 3D), while a reaction containing 10 µM enzyme did not greatly enhance the number of product peptides or their intensity (Figures A1-A4). Therefore, all farnesylation reactions studying the CaaaX libraries were performed using 1 µM yFTase. One potential cause of the need for increased enzyme for these pentapeptide libraries is that as well as

being poorer substrates, some may actually act as inhibitors of FTase. Peptides and peptidomimetic inhibitors of FTase have been known for many years.[53] Gratifyingly, the farnesylation of the CXIIM library showed farnesylation of the parent peptide, as well as several new amino acids. In total, 30 new potentially farnesylated CaaaX sequences were observed using a threshold of a signal to noise ratio of >12 across the eight different libraries (Table 1). The majority of the hits in the pentapeptide CMIIX library, including residues Q, A, M, S, were previously found to be good substrates when in the X position of the corresponding tetrapeptide CaaX box. Approximately half of known X position amino acids known to be substrates (in the context of the tetrapeptide CaaX box) were observed in our analysis of the CMIIX library. Overall, the X position libraries contained the lowest number of unexpected (or non-canonical) amino acids, leading us to hypothesize that recognition of the C-terminal residue in a pentapeptide in the enzyme active site may involve similar interactions as occurs with tetrapeptide CaaX box sequences. In contrast, some more unexpected amino acids, including Tyr or His in the a1 position were detected. Eight of the above peptides were selected for further evaluation, reflecting a mixture that represents a sampling of diverse substrate space.

Table 2.1. Summary of peptides observed in MALDI/MS libraries. Peptides chosen for further evaluation are bolded.

Library Sequence	Observed Amino Acid Hits
Ca_1IIM	**S**, C, M, F, **Y**, A, P, Q, E, **H**
CMa_2IM[a]	G, S, N, **K**, Q, E, H, R
$CMIa_3M$[a]	**G**, N, M, A, T, L, Q, E, H
CMIIX	**S**, C, **K**, A, **Q**, M

[a]Farnesylated isoleucine (I was present in the parent sequence CMIIM) was also observed in this library but with a signal to noise ratio of less than 10 and hence was not included in this tabulated data.

2.3.3. Evaluation of individual peptide hits by HPLC

To confirm the farnesylation of the hits in the library based MALDI analysis and to compare the farnesylation efficiencies of the various peptides with CMIIM, it was necessary to develop a secondary assay. This was readily accomplished by capitalizing on the presence of the dansyl fluorophore present in all of the peptide sequences, using High Pressure Liquid Chromatography (HPLC) to separate the starting peptides from their cognate farnesylated products. The extent of conversion was quantified by observing the loss of starting material over time via integration of the fluorescence signal. Since these reactions were performed with pure individual peptides, the possibility of inhibition by other nonsubstrate peptides was eliminated. Hence the concentration of enzyme was reduced from 1 μM to 25 nM. In each case, the structure of the farnesylated peptide was confirmed by Liquid Chromatography MS2 (LC-MS/MS) analysis. Of particular note, characteristic b4 and b5 ions were used to establish S-farnesylation on cysteine (Table A3 and Figure A14). The percent farnesylation of these peptides is summarized in Table 2. The CMIIM peptide displayed 56% conversion to product in 30 min with 25 nM yeast FTase (yFTase) (Figure 4A). Peptides considered to be similar to the related “CaaX-like” sequences, including CMIIQ, CMIIS, CSIIM, and CMKIM showed conversion similar or greater to that of the parent CMIIM (Figure 4B, Figures A5-A7). The sequences CMIGM and CMIIK both displayed little to no conversion under these reaction conditions; this was

not particularly surprising since both Gly and Lys are not normally observed in related CaaX sequences in the aliphatic or X position, respectively (Figure A8-A9); we speculate that the presence of the additional cationic lysine residue may boost ionization of that farnesylated peptide in the initial MALDI screening even if present in very low abundance. Interestingly, low conversion was observed for the sequence CHIIM, and high conversion for CYIIM, which contain residues that are also not commonly found in farnesylated CaaX sequences (Figures A10-A11). These initial findings indicate that while many of the canonical rules that define CaaX sequences also apply to the extended CaaaX sequence, the extra amino acid allows for some additional flexibility in allowable amino acids and expands the scope of these FTase substrates.

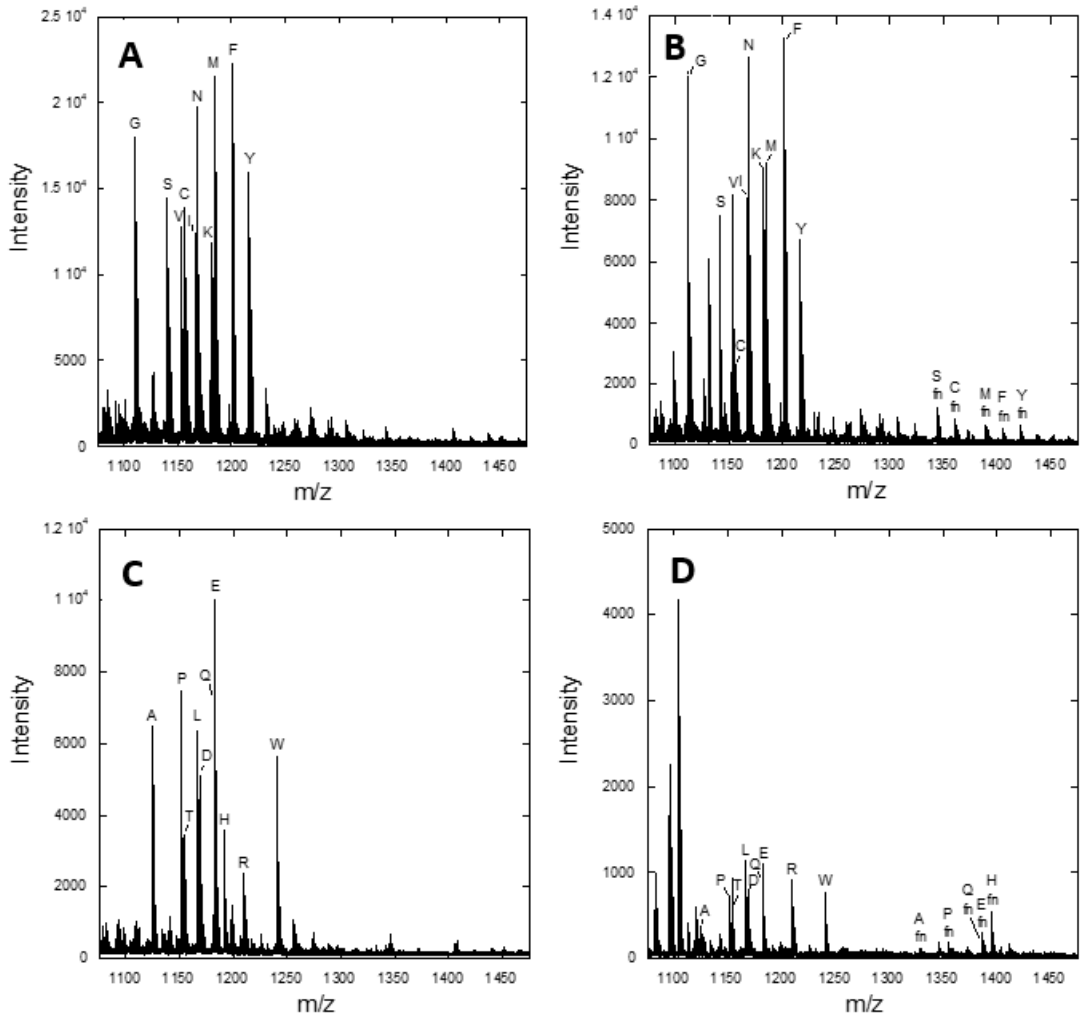


Figure 2.3. Farnesylation of DsGRAGCa$_1$IIM libraries.

Library 1 **(A)** before and **(B)** after farnesylation with 1 µM yFTase. Library 2 **(C)** before and (D) after farnesylation with 1 µM yFTase. The identity of the residue in the X position is indicated with the letter above each peak. The farnesylated peptides are highlighted with the designator "fn".

Overall, of the 8 peptide sequences that were selected for analysis in the secondary HPLC assay, 6 (75%) were confirmed as substrates. As a first step to querying the potential relevance of CaaaX sequences to a mammalian system, the above peptides from the yFTase

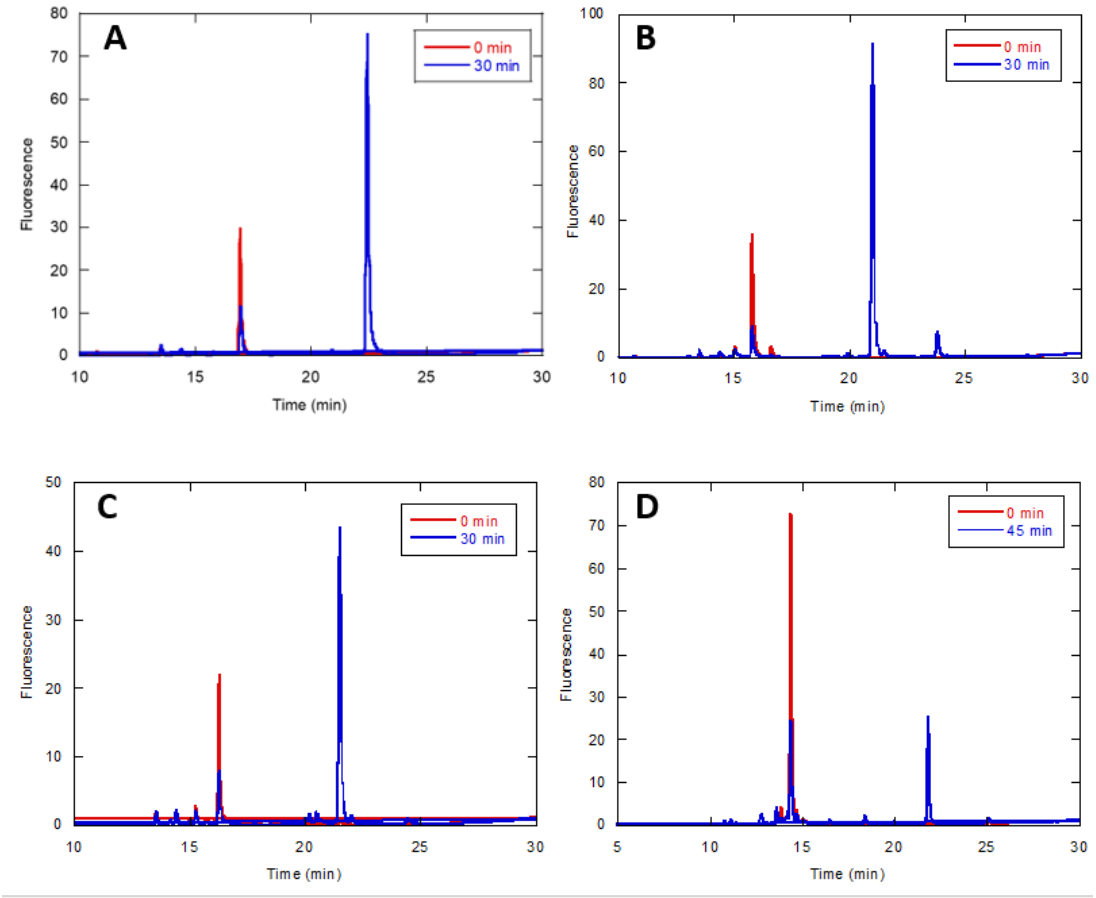


Figure 2.4. HPLC assays quantifying the conversion of selected peptides by the fluorescence of Dansylglycine (ex. 220/em. 495).

Reactions contained 2.4 µM peptide, 10 µM FPP and specified FTase concentration. **(A)** DsGRAGCMIIM (25 nM yFTase); **(B)** DsGRAGCMIIQ (25 nM yFTase); **(C)** DsGCMTSQ (100 nM yFTase); **(D)** DsGCSQAS (100 nM rFTase). Chromatograms of the reaction before (red) and after enzymatic reaction (blue) are shown for each peptide. The farnesylated peptides always elute later than their unfarnesylated counterparts. It should be noted that CMIIM and CMIIQ are sequences obtained from initial screening of peptide libraries via MALDI whereas CMTSQ and CSQAS are sequences identified from bioinformatic analysis of human genome sequences, guided by the results from the library screening.

screen were evaluated with rat FTase (rFTase) (Table 2). The only peptides that showed significant conversion with 200 nM rFTase were the CMIIS and CMIIQ sequences. This

is not particularly surprising, as the rat enzyme is generally more stringent in what substrates it accepts compared with the yeast enzyme.[35,36]

2.3.4. CaaaX hits in the mammalian genome

When searching the human genome for CaaaX sequences of the type CSXXX, CXXXQ, or CXXXS, 138 C-terminal CaaaX sequences were found that exist on known proteins. In order to narrow down this list, those which had unlikely motifs such as multiple charged residues or Gly and Pro amino acids were omitted. That left 28 sequences that might serve as prenyltransferase substrates. (Table A4). Of those sequences, 10 were selected for further evaluation. The results of HPLC assays with those 10 are shown in Table 3. While many of these sequences showed very limited activity with 100 nM rFTase, CMTSQ (Figure 4C) and CASQS (figure 4D) showed substantial conversion, with CSLMQ showing excellent conversion (Figure 5A). CSLMQ still showed high conversion with as little as 25 nM rFTase, and when this peptide was compared to the native CaaX sequence CVLS at the same enzyme concentration, the results were almost identical, with both peptides achieving 85% conversion (compare Figures 5A and 5B). Thus, it is striking that an extended CaaaX sequence can be farnesylated as efficiently as a native CaaX sequence, since even the best previously described pentapeptide CaaaX sequence, CMIIM, is approximately an order of magnitude worse compared with the native tetrapeptide ones.[44] These mammalian sequences were also evaluated with yFTase, and the three aforementioned peptides manifested excellent conversion with that enzyme as well, with CSQAS and CSLMQ still being efficiently farnesylated with as little as 25 nM yFTase,

with CMTSQ displaying 60% conversion with yFTase, while it had substantially less activity with the rat enzyme. CASSQ also displayed good conversion with the yeast enzyme, which is perhaps unsurprising due to its similarity to the native CaaX sequence CASQ.

Table 2.2. Percent farnesylation of CaaaX peptides derived from MALDI libraries. Each value is the result of triplicate experiments.

Sequence	25 nM yFTase, rt	200 nM rFTase, 35 C
CMIIM	56 ± 10	50 ± 3
CMIIS	61 ± 14	25 ± 3
CMIIQ	76 ± 3	80 ± 2
CSIIM	64 ± 5	< 1
CMKIM	54 ± 10	< 1
CYIIM	95 ± 1	49 ± 1
CHIIM	15 ± 6	< 1
CMIGM	< 1	< 1
CMIIK	< 1	< 1

2.3.5. Farnesylation of CaaaX sequences can occur efficiently in cells

While many of the above peptides were shown to be farnesylated *in vitro*, an important question concerns their ability to be farnesylated under cellular conditions. Accordingly, to determine whether these pentapeptide CaaaX sequences could be farnesylated *in vivo*, they were analyzed in the context of the yeast HSP40 protein Ydj1p, which has proven to be a useful reporter system for studying the specificity of farnesyltransferase in yeast.[6,40] Farnesylation of Ydj1p alters its mobility in SDS-PAGE such that farnesylated wild-type Ydj1p (CASQ) has increased mobility (i.e. smaller apparent kDa) relative to unfarnesylated Ydj1p-SASQ.[6] This mobility shift is entirely attributable to farnesylation because the

shift is eliminated in the absence of FTase activity as determined using a yeast knockout strain (Supplemental Figure A12).

Table 2.3. Percent farnesylation of CaaaX peptides derived from mammalian genome. Each value is the result of triplicate experiments, ND = not measured.

Sequence	25 nM yFTase	100 nM yFTase	25 nM rFTase	100 nM rFTase
CSLMQ	95 ± 4	> 99	79 ± 2	> 99
CSQAS	43 ± 3	> 99	ND	66 ± 1
CMTSQ	ND	62 ± 1	ND	9 ± 1
CASSQ	ND	33 ± 3	ND	< 1
CQYNS	ND	< 1	ND	< 1
CLACS	ND	< 1	ND	< 1
CVQTS	ND	< 1	ND	< 1
CASLS	ND	< 1	ND	< 1
CSKLN	ND	< 1	ND	< 1
CLLFS	ND	< 1	ND	< 1

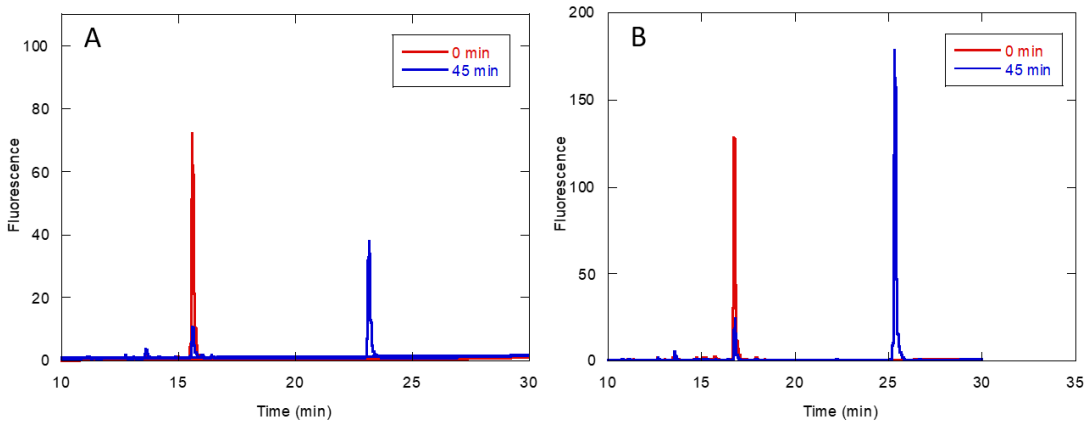


Figure 2.5. Comparison of the prenylation of CSLMQ and CVLS.

HPLC assays displaying that the best peptide from the bioinformatic screen, CSLMQ, was able to show near identical conversion to a native tetrapeptide sequence CVLS with conversion of 85% vs 84%, respectively. The quantification was performed as in Figure 4. Reactions contained of 2.4 µM peptide, 10 µM FPP and 25 nM enzyme. **(A)** DsGRAGCSLMQ with 25 nM rFTase; **(B)** DsGCVLS with 25 nM yFTase.

A similar FTase-dependent mobility shift was previously used to demonstrate the farnesylation of the reporter protein Ydj1p-CMIIM which bears a C-terminal pentapeptide CaaaX sequence.[44] By comparison to Ydj1p-SMIIM (a non-farnesylated protein), all of Ydj1p-CMIIM appears to be shifted to increased mobility, indicating that this and possibly other non-canonical length CaaaX sequences are able to undergo near complete farnesylation in cells. The CaaaX sequences obtained from the initial library screening described here were similarly individually transformed into yeast and evaluated using this mobility shift assay and were determined to be farnesylated to varying degrees (Figure 6). Quantification of the farnesylated and unfarnesylated species in each lane indicated that Ydj1p-CMIIQ, -CMKIM, -CSIIM, and -CYIIM appeared to be extensively farnesylated (100%, 98%, 100% and 100%, respectively), -CHIIM and -CMIIS were moderately farnesylated (55% and 77%, respectively), and -CMIGM was modestly farnesylated (20%). In addition, several of the CaaaX sequences observed in the mammalian genome were effectively farnesylated with -CSLMQ showing extensive farnesylation(91%) and -CLLFS and -CSKLN showing more limited farnesylation (18% and 11%, respectively) (Figure 7).This data with standard deviation and information on biological and technical replicates is shown in Table A5 Based on these *in vivo* results, it appears that the scope of farnesylatable sequences appears larger than previously believed.

2.4. Discussion

This work describes a new systematic approach to discover novel peptide substrates for FTase by combining the combinatorial power of Solid phase peptide synthesis (SPPS) with the ease of MALDI-MS. The workflow consists of synthesizing focused libraries

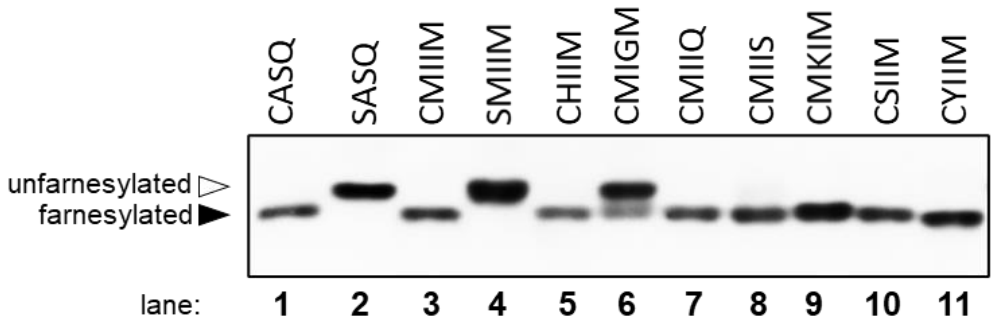


Figure 2.6. Mobility shift analysis of Ydj1p-CaaaX variants identified from peptide libraries.

Whole cell lysates prepared from yeast expressing the indicated Ydj1p-CaaaX variant were evaluated by SDS-PAGE and anti-Ydj1p immunoblot. The indicated Ydj1p variants were expressed in yWS2544 (ydj1::KANR) to eliminate any contribution from naturally encoded Ydj1p. Farnesylated Ydj1p (CASQ) exhibits a smaller apparent molecular mass relative to unmodified Ydj1p (SASQ). Farnesylation profiles for the indicated Ydj1p-CaaaX variants were determined across multiple biological and technical replicates, from which the percent of farnesylated species relative to the total signal for a sample was determined (see Figures A12-A13 and Table A6).

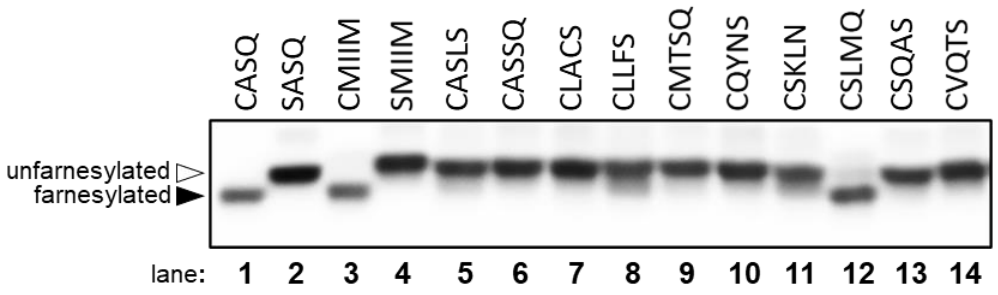


Figure 2.7. Mobility shift analysis of Ydj1p-CaaaX variants identified from analysis of mammalian genome.

Whole cell lysates prepared from yeast expressing the indicated Ydj1p-CaaaX variant were evaluated by SDS-PAGE and anti-Ydj1p immunoblot. The indicated Ydj1p variants were expressed in yWS2544 (ydj1::KANR) to eliminate any contribution from naturally encoded Ydj1p. Farnesylated Ydj1p (i.e., CASQ and CMIIM; closed triangle) exhibits a smaller apparent molecular mass relative to unmodified Ydj1p (i.e., SASQ and SMIIM; open triangle). Farnesylation profiles for the indicated Ydj1p-CaaaX variants were determined across multiple technical replicates, from which the percent of farnesylated species relative to the total signal for a sample was determined (see Figures A12-A13 and Table A6).

containing 10-20 sequences obtained by randomizing a synthetic peptide at a single position. As SPPS relies on iterative coupling reactions to extend a peptide from a solid support, the reaction of complex libraries is quite facile. Incubation of the library with FTase and FPP followed by MALDI analysis allows the enzymatic products to be clearly resolved from starting peptides due to the increase in mass that occurs upon farnesylation (a 204.1 Dalton increase). Positive hits are then confirmed using a secondary HPLC based assay with purified individual peptides functionalized with a dansyl fluorophore to facilitate quantification. Finally, expression of the yeast protein Ydj1p containing specific C-terminal CaaaX sequences in yeast followed by western blot analysis of cellular lysates allows the extent of farnesylation within live cells to be ascertained.

In analyzing the results from the extended CaaaX libraries, it was found that the a_1 position manifested the greatest number of hits (10), likely due to the fact that it is furthest away from the C-terminal X residue that plays a key role in substrate recognition. Numerous hits were also observed in the a_2 and a_3 positions with the a_3 position showing more canonical (canonical in the context of the tetrapeptide CaaX box) hydrophobic residues (Ala, Leu, and Met). The X position, considered the most important for peptide recognition, showed canonical hits (Ala, Cys, Ser, Gln, and Met) with one exception, Lys, which was not a bona fide substrate of the enzyme when evaluated in HPLC assays.

The results presented here illustrate that a large number of pentapeptide sequences are substrates for FTase. The list of those sequences includes motifs present on the C-termini of bonafide mammalian proteins including Transcription elongation factor A

protein 3 (CSLMQ), Xaa-Pro aminopeptidase 3 (CSQAS), and Beta-1,4 N-acetylgalactosaminyltransferase 1 (CMTSQ). While the data reported herein does not establish that those proteins are in fact farnesylated *in vivo*, the results imply that they could be. This in turn suggests that the number of farnesylated proteins within cells may be larger than previously thought. As a final point, it should be noted that the workflow presented here could be applied to other post-translational modifications including lipidation, glycosylation and others where a significant mass increase occurs. It is likely possible to achieve a similar observation by ESI, however the presence of multiply charged species and various salt adducts complicates the analysis of these complex mixtures. While MALDI-based screening has its limitations as a quantitative tool, it is particularly well suited to the type of experiments reported here where post-translational modification is accompanied by a significant mass shift.

2.5. Conclusions

Our work expands the understanding of what sequences can be prenylated by FTase, specifically the recently discovered pentapeptide CaaaX sequences. We have validated a MALDI-based peptide library approach as an excellent way to screen peptide chemical space for novel enzyme substrates. We have identified novel sequences based on libraries from a known substrate, as well as some from the human genome. This work has potentially important implications for the prenylation of these extended sequences in human health.

2.6. Methods

2.6.1 Library Synthesis

Peptide libraries were synthesized using Fmoc based solid-phase peptide synthesis (SPPS) on a Gyros Protein Technologies PS3® peptide synthesizer using four equivalents of Fmoc-protected amino acids from Aldrich®, Novabiochem® and P3 Biosystems and Fmoc-AA-Wang resins from P3 Biosystems. A one-pot synthesis of 10 amino acids per library was performed at the "X" position, varying the ratio of amino acids to account for coupling efficiencies.[102] Dansylglycine (DsG) was then coupled manually, using a twofold molar excess, reacting for 4-6 h. Following synthesis, peptides were cleaved from resin for 2 h with 5 mL reagent K (82.5% trifluoroacetic acid (TFA), 5% thioanisole, 5% phenol, 2.5% 1,2-ethanedithiol, and 5% H_2O, v/v) cleavage cocktail per 0.1 mmol of resin. Peptides were precipitated by draining the cleavage cocktail into 40 mL Et_2O cooled in an isopropanol / dry ice bath for ~10 min, centrifuging until pelleted, decanting the Et_2O layer, and repeating once to wash the residual cleavage cocktail from the crude peptide mixture. The peptides were then dissolved in 50:50 CH_3CN/H_2O containing 0.1% TFA. The total peptide concentration was determined by diluting in 1,4-dioxane and measuring the absorptivity at 338 nm using a Cary 50 Bio UV-Visible spectrophotometer, using beer-lambert's law in conjunction with dansylglycine's molar extinction coefficient (4300 cm^{-1} M^{-1}). Individual peptide hits were synthesized using similar conditions.

2.6.2. Enzymatic Farnesylation of Peptide Libraries

Enzymatic farnesylation of peptide libraries was performed by incubating FTase from *S. cerevisiae* (yFTase) in a reaction buffer that contained 20 μM total peptide, 40 μM FPP, 50 mM Tris-HCl pH 7.5, 10 μM ZnCl2, 5 mM MgCl2, 1 mM DTT in H_2O. Reactions were allowed to proceed for 5 h at 37° C. Upon completion, the samples were desalted using a Plus long Sep-Pak reverse-phase C18 environmental cartridge from Waters Corporation (WAT023635, length 3cm, diameter 1 cm). Sep-Paks were primed by washing with 3 mL Buffer B (CH_3CN with 0.1% TFA) and equilibrated with 3 mL Buffer A (H_2O with 0.1% TFA). The sample was then loaded, washed with 2 mL each of 100% Buffer A, 10% Buffer B in Buffer A, and 20% Buffer B in Buffer A, and then eluted with 2 mL Buffer B. Samples were immediately spotted on a MALDI plate or stored at -80 °C. Control libraries were prepared under identical conditions, without the addition of FTase.

2.6.3. MALDI-TOF MS of Farnesylated Peptide libraries

Samples eluted from the Sep-Pak columns (0.5 μL) were cospotted with 0.5 μL of 10 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA) matrix in 50:50 Buffer A: Buffer B on an AB Sciex 384 Opti TOF plate. The typical spotting procedure involved spotting the matrix first, then immediately spotting the sample on top of the matrix, rapidly pipetting up and down to mix. Samples were then analyzed with an AB-Sciex 5800 13 MALDI/TOF mass spectrometer using the reflector positive mode. A laser intensity of ~4000-5000 was applied, with a pulse rate of 400 Hz. Laser intensity was increased in increments of 200 if signal was not readily apparent. 4000 laser shots were applied per spectrum, and the entire spot surface was sampled.

2.6.4. HPLC based Enzymatic Farnesylation Assay

Enzymatic farnesylation reactions with purified peptides were performed by incubating FTase from *S. cerevisiae* (yFTase) or *R. norvegicus* (rFTase) in a reaction buffer that contained 2.4 μM peptide, 10 μM FPP, 50 mM Tris-HCl pH 7.5, 10 μM ZnCl2, 5 mM MgCl2, 1 mM DTT in H_2O.[103,104] Reactions with yFTase were at rt for 30 min, reactions with rFTase were run at 35 C for 45 min. The reactions were flash frozen to stop enzymatic activity, and 200 uL aliquots were injected onto an Agilent 1100 HPLC instrument equipped with an FLD detector and a Phenomenex Luna 5-micron C18 100 Å pore size 250 x 4.60 mm 5 micron analytical column. Fluorescence of the dansylated peptides was monitored with an excitation of 220 nm and an emission of 495 nm with a PMT gain of 12. All reactions were run in triplicate. The extent of peptide farnesylation was quantified by integration of the starting material peak from the HPLC chromatogram. The identity of the starting peptides and farnesylated products were confirmed by LC-MS/MS analysis using a ThermoFischer LTQ Orbitrap Velos instrument.

2.6.5. Peptide search of the human proteome

The scanProsite tool of Expasy was used to scan the uniprotKB for known protein sequences that contain a potential pentapeptide CaaaX sequence (https://prosite.expasy.org/scanprosite/). The search was limited to C-terminal sequences representative of some of our best peptide hits; the queries searched were CSXXX,

CXXXQ, and CXXXS, where any amino acids were allowed in the varied X positions. The scan was performed as a motif search against the UniProtKB, including isoforms. Results were then filtered to only show sequences specific to the human proteome (*H. sapiens*).

2.6.6. Yeast strains and plasmids

The yeast strains used in this study have been previously described (Table A5).[46] Yeast strains were propagated at rt in either YPD or appropriate selective media when plasmid transformed. Introduction of plasmids into yeast strains was accomplished via a lithium acetate-based transformation procedure.[105] Several of the plasmids used in this study have also been previously described (Table A6). Others were created *in vivo* by recombinational cloning using similar methods. In brief, yeast cells were co-transformed by the lithium acetate-based procedure with DNA fragments derived from a NheI and AflII restriction digest of parent plasmid pWS1132 and a DNA fragment encoding the desired CaaaX sequence that was created by PCR using a high-fidelity polymerase. The PCR product encoding the CaaaX sequence was flanked by 5´ and 3´ sequences that were identical to regions of the parent plasmid to facilitate homologous recombination to repair the gapped parent plasmid. Candidate plasmids were recovered from yeast, transformed into and amplified in E. coli, and evaluated by restriction digest and commercial DNA sequencing to confirm the presence of the desired YDJ1-CaaaX open reading frame.

2.6.7. Mobility shift analysis of Ydj1p farnesylation

Whole cell lysates of late-log yeast were prepared as previously described, separated by large format (19.5x16 mm) SDS-PAGE (9.5%), transferred onto nitrocellulose, and blots incubated with rabbit anti-Ydj1p primary antibody (courtesy of Dr. Avrom Caplan) and HRP-conjugated goat anti-rabbit secondary antibody (Kindle Biosciences, Greenwich, CT).[106] After development of blots with the WesternBright TM ECL-spray (Advansta, San Jose, CA), protein bands were detected using a KwikQuant Imager at multiple exposure times. Levels of farnesylation were quantified from KwikQuant images using ImageJ software for at least 2 technical replicates.

Acknowledgments

This work was funded by MARK DISTEFANO, National Institutes of Health grant numbers GM084152 and GM141853, and WALTER SCHMIDT, National Institutes of Health grant number GM132606. GARRETT SCHEY was funded in part by an NIH Training Grant (GM008347). The authors thank Dr. Avrom Caplan (City College of New York, New York, NY) for the anti-Ydj1p primary antibody. Garrett Schey performed peptide synthesis, MALDI analysis, and HPLC assays as well as prepared figures and tables. Peter Buttery performed peptide synthesis and MALDI analysis. Elizabeth Hildebrandt performed yeast transformation and gel shift assays. Mark Distefano, James Hougland, and Walter Schmidt provided guidance in the design of the experiments.

Chapter 3: MALDI Analysis of Peptide Libraries for Farnesyltransferase: Pushing Toward Prenylation of Pentapeptide Sequences in the Human Genome

3.1. Introduction

Protein prenylation is a post-translational modification in which a hydrophobic isoprenoid group is covalently attached to the thiol side chain of a cysteine residue located near the C-terminus of a protein. Farnesyltransferase (FTase) transfers a 15-carbon farnesyl group using FPP while Geranylgeranyltransferase (GGTase-I) employs a 20-carbon GGPP prenyl donor [94]. These enzymes recognize proteins with a C-terminal tetrapeptide consensus sequence known as a "CaaX box". In the canonical view of prenyltransferase selectivity, "C" is the cysteine residue that is covalently modified, "a" is usually an aliphatic amino acid, and the "X" is a residue that is largely responsible for determining whether the protein substrate is targeted by FTase or GGTase-I [1]. Protein prenylation is essential for proper cellular localization, protein-protein interactions and signaling activity, and misregulation of prenylation is implicated in many diseases [5,95]. In addition to cancer, prenylation and the prenylation pathway have drawn considerable attention as potential targets for treatment of Alzheimer's disease, Hutchinson−Gilford progeria syndrome, and numerous other diseases [11–15]. In 2020, the first FDA approved inhibitor of FTase, lonafarnib, initially designed as a potential cancer drug, was approved for the treatment of progeria since it prevents the prenylation of prelamin A [16,17]. Another inhibitor, Tipifarnib, has long been under investigation for cancer therapy and is still under investigation in clinical trials for head and neck squamous cell cancers driven by H-RAS mutations [107,108].

FTase manifests broad substrate specificity, catalyzing the transfer of a farnesyl group from Farnesyl Pyrophosphate (FPP) to a wide variety of protein and peptide substrates, and many attempts have been made to define what amino acids are allowed or not allowed in the CaaX sequence [18,29]. Interestingly, it has been found that while CaaX binding largely occurs through the binding of the X amino acid, there is synergy between the amino acids that affects their efficiency as substrates [23,33]. The flexibility of the peptide binding site has been used in the rational design of novel mutant FTases for orthogonal peptide reactivity and to enhance the ability to accept fluorescent isoprenoid analogs [42,43,62]. While the canonical model of the CaaX box is generally well understood, it has recently been found that certain sequences other than the four-residue CaaX motif can also be farnesylated by both yeast and mammalian FTase orthologs, specifically both tripeptides and pentapeptides [44,46]. The prenylation of pentapeptide CaaaX sequences was first observed in yeast, and initial evaluation of CaaaX substrate space found the sequence CMIIM to be the most efficiently farnesylated for the CaaaX sequence, with *in vitro* assays indicating that this peptide was a reasonable substrate ($k_{cat}/K_M = 1.9 \times 10^4\ M^{-1}s^{-1}$), although less efficient compared with the most efficient CaaX peptides such as CVLS ($k_{cat}/K_M = 2.0 \times 10^5\ M^{-1}s^{-1}$) [44]. None of these peptides were found to be substrates for GGTase-I.

Due to the past three decades of research, the selectivity rules for CaaX containing peptides have become increasingly well defined. However, there are many questions remaining about how this information might apply to pentapeptide CaaaX sequences, since it is

known that the FTase from different organisms display different preferences for peptide substrates [36]. Previous work in our lab and others have relied on the analysis of peptide libraries through the use of an alkyne-containing isoprenoid analogue to allow for biotin attachment by derivatizing with biotin-azide via copper-catalyzed azide−alkyne cycloaddition. [36] Since that approach relies on synthetic isoprenoid analogues that may perturb enzyme specificity, it may provide misleading results concerning the specificity with the native substrate FPP. This may be particularly problematic since the isoprenoid and peptide substrates are known to extensively interact over a broad surface as evidenced from the structure of the FTase ternary complex. In an effort to study peptide libraries with the native FPP substrate, we have recently utilized a MALDI-MS based detection assay. Peptide libraries were synthesized with a DansylGlycine (DsG) and RAG sequence upstream of the variable CaaaX sequence to aid in solubility and ionization [45]. This method relied on the 204 Dalton mass shift due to the addition of the farnesyl group to easily observe the formation of prenylated product peptides. The gentle ionization of MALDI resulted in clear results by producing singly charged ions without fragmentation, making it highly amenable in for the analysis of peptide libraries [96,97]. We utilized this peptide library approach to study variations of the most efficiently prenylated CaaaX sequence, CMIIM, using yFTase. Over 30 new prenylated sequences were observed using that approach and validation of eight of these sequences in a yeast-based reporter assay showed that they were prenylated *in vivo*. In addition, a curated search of the human genome revealed that there may be many potentially prenylatable CaaaX sequences. Analysis of one such sequence, CSLMQ, showed that it was prenylated as efficiently as the sequence CVLS, the CaaX box present at the C-terminus of H-Ras, and was also

prenylated *in vivo*. To expand our understanding of the recognition of CaaaX sequences, we have analyzed CMIIM libraries with rFTase, as well as generated and tested a new set of libraries based on CSLMQ with both yeast and rat FTase. In addition, we have looked more broadly by evaluating approximately xxx CaaaX sequences that occur naturally in the human genome.

3.2. Research Objectives

In this work, we sought to expand our understanding of extended pentapeptide CaaaX sequence prenylation by analyzing previously described libraries based on the sequence CMIIM with a mammalian enzyme, rat FTase. Libraries based on a previously discovered efficient sequence, CSLMQ, were also synthesized and analyzed with both yeast and rat FTase to explore the recognition diversity within that different class sequences in comparison to previous CMIIM libraries. A search of the human genome revealed that there are over 1000 pentapeptide CaaaX sequences.

3.3. Results

3.3.1. Identification of novel substrates from the CMIIM motif using MALDI analysis

Utilizing our previously described method of evaluating peptide libraries by MALDI/MS, libraries based on CMIIM were analyzed with rFTase as a comparison to yFTase [45]. Using 2 μM enzyme, it was possible to observe a number of hits for the library CMXIM where X represents the variable position. Positive hits in this position included **Gly**, Met, Cys, **Asn**, **Ser**, **Gln**, **His**, Arg, and Tyr, with those in bold being shared between libraries analyzed using yFTase or rFTase (Figure 3.1). Of the shared hits, the data was normalized

by dividing the ionization of the prenylated peak by the combined ionization of the parent ion and product (Table 3.1).

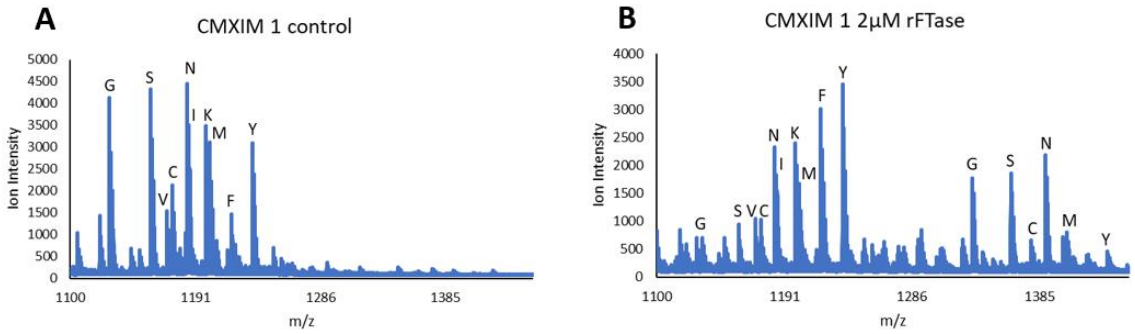


Figure 3.1. Farnesylation of DsGRAGCMXIM library 1.

Library 1 **(A)** before and **(B)** after farnesylation with 2 µM rFTase at 35° C for eight hours. The identity of the residue in the X position is indicated with the letter above each peak. X = G, S, V, C, N, I, K, M, F, and Y.

Differences were observed in the prenylation levels of the peptides present in these libraries, with some amino acids such as Gly being preferred by rFTase and His being preferred by yFTase. Amino acids such as Gln appear to be prenylated equally by both enzymes. While this quantitative analysis should be interpreted with caution due to differences in the ionization efficiency of the product versus reactant, it can provide a useful metric to estimate efficiency of substrates between different enzymes.

Table 3.1. Normalization of hits from DsGRAGCMa$_2$IM 1 library shared between yFTase and rFTase

Amino Acid	rFTase	yFTase	rFTase:yFTase
Gly	0.72	0.29	2.5
Ser	0.66	0.46	1.4
Asn	0.55	0.38	1.4
Gln	0.49	0.48	1.0
His	0.31	0.63	0.49

Interestingly, analysis of the other three positions in the CMIIM-based libraries yielded no prenylated peptides at this enzyme concentration. While the rat enzyme is more stringent in its substrate specificity than the yeast enzyme, this is surprising as it was possible to find hits at all positions using yFTase[36]. However, increasing the concentration of enzyme to 10 μM rFTase afforded some prenylated peaks, although the peak intensities were quite low compared to earlier results with the DsGRAGCMa_2IM libraries (Table 3.2). This is in contrast to libraries analyzed using yFTase, where increasing the concentration of enzyme did not significantly increase peak intensity. It is somewhat difficult to interpret these latter results since those experiments required higher enzyme concentrations, but it is clear that rFTase has the most flexibility for the a_2 position. While canonical hits such as Gln, Met, and Ser were observed at the X position, it is interesting that none were observed at lower enzyme concentrations, as this is the residue considered most important for substrate binding[33]. Finally, it was surprising that screening of the a_1 position yielded only Val as a hit. In general, tetrapeptide sequences have shown a similar number of hits for the aliphatic positions, and previous analysis showed this trend continued to pentapeptides when analyzed with yFTase (Table 3.3). Therefore, it is striking that such differences in reactivity between the aliphatic positions are observed when using rFTase. While substrate inhibition is certainly a possibility, it is difficult to use this as the reasoning behind our results.

Next several of the peptide hits were validated using a previously described HPLC-based fluorescence assay. Prenylation of peptides containing CMGIM, CMNIM, and CMSIM

sequences was observed in 45 min reactions with 200 nM rFTase giving conversions of 50%, 41%, and 31% respectively. This agrees with the conversion levels previously found with rFTase, although nothing stood out as a more efficient substrate than CMIIM for rFTase.

Table 3.2. Summary of peptides observed in CMIIM MALDI/MS libraries using rFTase at 10μM.

Library Sequence	Observed Amino Acid Hits
Ca_1IIM	V
CMa_2IM[a]	G, S, N, Q, H, C, M, Y, R
$CMIa_3M$	S, V, C, N, M, A, L, Q, E, H
CMIIX	Q, H, S, C, N, M, F, A, E

[a]These libraries showed conversion at 2uM, while all others had no prenylation at this concentration. Reactions were run at 35° C for 8 hours. Peptides were synthesized and tested with a DsGRAG sequence upstream of the CaaaX box.

Table 3.3. Summary of peptides observed in CMIIM MALDI/MS libraries using yFTase at 2 μM.

Library Sequence	Observed Amino Acid Hits
Ca_1IIM	S, C, M, F, Y, A, P, Q, E, H
CMa_2IM[a]	G, S, N, K, Q, E, H, R
$CMIa_3M$[a]	G, N, M, A, T, L, Q, E, H
CMIIX	S, C, K, A, Q, M

[a]Farnesylated isoleucine (I was present in the parent sequence CMIIM) was also observed in this library but with a signal to noise ratio of less than 10 and hence was not included in this tabulated data. Reactions were run at 35° C for 8 hours. Peptides were synthesized and tested with a DsGRAG sequence upstream of the CaaaX box.

3.3.2. Identification of novel substrates from the CSLMQ motif using MALDI analysis

Based on previous results showing that CSLMQ appeared to be a more efficient substrate than CMIIM, we next explored libraries based on that sequence [45]. Initially, examination of the libraries using 1 µM yFTase showed different trends than were observed with the CMIIM libraries (Table 3.4). In the case of CMIIM-based libraries, they contained a variety of hits at all positions (a1, a2, a3 and X), In contrast, with CSLMQ-based libraries, only two hits at the a_1 and X positions and four at the a3 position were obtained. This tighter selectivity is interesting, as it might be expected that the generally more flexible yFTase might reveal more hits. The position with the most allowed amino acids was the a_2 position, which is also what was observed with the CMIIM libraries with rFTase. Normally, bulky aromatic amino acids in the aliphatic region are not favored, but Tyr and Phe have been consistently observed in our libraries, with even the conformationally restrained Pro appearing as a hit (Figures A15 and A16).

Table 3.4. Summary of peptides observed in CSLMQ MALDI/MS libraries using yFTase at 1 µM. [a]

Library Sequence	Observed Amino Acid Hits
Ca_1LMQ	S, V
CSa_2MQ	G, S, N, Q, H, C, M, Y, R, F, L, D, E, P, A
CSLa_3Q	A, T, Q, M
CSLMX	N, Q

[a]Reactions were run at 35° C for 8 hours. Peptides were synthesized and tested with a DsGRAG sequence upstream of the CaaaX box.

Reacting CSLMQ libraries in the presence of rFTase yielded results indicating a surprising number of hits at all positions except a_1, which only showed Ile as a hit (Table 3.5). Given that the yFTase is more lenient in its selectivity for tetrapeptide CaaX sequences compared with the rFTase, the greater number of hits obtained using the rat enzyme with these pentapeptide libraries is surprising. Positive hits included a variety of canonical amino acids and more unusual aromatic residues, including Trp in the a_3 position. The X position also showed hits for charged amino acids including Asp and Glu, which are typically associated with single turnover reactions, although that is unlikely here given the enzyme to peptide ratio employed and the large number of peptides hits observed in that library reaction Figures (A17-A23).

Table 3.5. Summary of peptides observed in CSLMQ MALDI/MS libraries using rFTase at 1 µM. [a]

Library Sequence	Observed Amino Acid Hits
Ca_1LMQ	I
CSa_2MQ	G, S, N, C, M, Y, F, V
$CSLa_3Q$	G, S, V, C, I, M, F, Y, A, T, L, D, Q, H, W
CSLMX	S, C, D, M, F, Y, A, T, N, Q, E

[a]Reactions were run at 35° C for 8 hours. Peptides were synthesized and tested with a DsGRAG sequence upstream of the CaaaX box.

3.3.3. CaaaX hits in the mammalian genome

To identify additional pentapeptide sequences that might be FTase substrates, a search of the human genome was performed using the search motif CXXXX where any amino acids were allowed in the varied X positions, finding over 1000 potentially prenylated peptides as previously described [44]. From that list, 192 of those peptides were chosen by removing sequences that had multiple charged residues, multiple conformationally challenging residues (Gly and Pro), or one of each from the aforementioned categories. This parallel synthesis effort employed an Intavis Multipep RS instrument, allowing for the synthesis of 24 peptides at a time on a 10 μmol resin scale using standard HCTU coupling chemistry. This is in contrast to standard SPPS scale which is usually on the order of 0.1-0.2 mM. These peptides were cleaved and reacted in crude form with 2 μM rFTase. Analysis of these peptides revealed a small number of hits. Of the 192 tested, only seven positive sequences (CITTL, CVHAL, CQTLI, CRFVT, CHSIA, CTSEI, CYLVK) were obtained. Efforts to validate these sequences in HPLC assays are underway. Interestingly, one sequence, CSLQQ, identified from a library based on CSLMQ that was found to be a substrate using both the rat and yeast FTase enzymes, appeared in the human genome search as well.

3.4. Discussion

This work describes the continued evaluation of pentapeptide CaaaX sequences using a previously utilized workflow employing focused libraries based on CaaaX sequences of interest. Libraries containing 10 sequences, chosen to eliminate isobaric overlap, obtained by randomizing a synthetic peptide at a single position were enzymatically farnesylated and evaluated by MALDI/MS. Libraries based on CMIIM, a sequence previously discovered and analyzed with the yeast enzyme and in whole yeast cell assays, were further investigated utilizing rFTase, a homolog more relevant to human health (and less flexible in terms of substrate specificity) to evaluate differences between the two. Libraries based on CSLMQ, a pentapeptide that exists on the C terminus of the human protein Transcription elongation factor A protein 3, were also analyzed with both yFTase and rFTase. Although it is unknown whether this protein is prenylated *in vivo*, our previous work and additional results presented here suggest it may be possible. To expand the repertoire of possible pentapeptide prenylation substrates, parallel synthesis coupled with MALDI-MS analysis allowed the evaluation of 192 CXXXX sequences from the human genome.

In analyzing the results from the CMIIM libraries reacted with rFTase at 2 μM enzyme, only variants at the a_2 position showed prenylated products. This is in contrast to our previous findings using yFTase where there were hits at all four variable positions with the most at the a_1 position, likely due to the fact that it is furthest away from the C-terminal X residue which is involved in the critical binding interactions. Increasing the concentration to 10 μM resulted in additional reactivity. Interestingly, the a1 position was still the least permissive to change with only one hit, Val. While this is a canonical amino acid for this

position (in a tetrapeptide CaaX context), it is noteworthy that this position was found to have the most hits when CMIIM-based libraries were analyzed with yFTase. The X position showed a wider variety of substitutions using rFTase compared with yFTase, although a much higher concentration of enzyme was required. Canonical hits including Gln, Ser, Cys, Met, and Ala, and more unusual ones such as His, Asn, Phe, and Glu were observed in that case. These additional hits indicate rFTase has more flexibility for these extended sequences at the X position.

When analyzing the results with the CSLMQ libraries, we again found numerous differences occurring when using yeast versus rat FTase. To our surprise, the normally more flexible yFTase yielded fewer hits than rFTase, most notably in the X position. The yFTase results showed that only Asn and Gln were tolerated in the X position, while rFTase gave a mix of canonical and noncanonical hits, similar to the results from the CMIIM libraries. Interestingly, both enzymes displayed very low tolerance to amino acid substitutions in the a_1 position, in stark contrast to how yFTase behaves with CMIIM libraries. This highlights the importance of studying enzymes from different organisms and the use of multiple libraries, as a high number of hits at the a_1 position appears to be an aberration instead of the rule. The a_2 position is the most variable across all libraries with both enzymes showing a wide range of hits in each experiment. The a_3 position for CSLMQ libraries with yFTase showed several canonical hits, while it had the highest number of hits (15) for rFTase.

Analysis of the human genome for sequences of the type CXXXX identified over 1000 results. We were able to test 192 of them by utilizing parallel synthesis and evaluating crude peptide material via MALDI-MS analysis. Of the 192, only CITTL, CVHAL, CQTLI, CRFVT, CHSIA, CTSEI, CYLVK displayed any reactivity. In analyzing the human genome for specific sequences of the types CXLMQ, CSXMQ, CSLXQ, and CSLMX, it was found that one hit beyond the parent (CSLMQ), CSLQQ, exists on C-terminus of the protein Ribosome biogenesis protein C1orf109. Excitingly, this sequence was one of the many hits discovered using both the yeast and rat enzymes. When CSLMQ was first analyzed, it was found to be an efficient substrate for both rat and yeast FTase, as well as was the only sequence from the human genome we previously tested to be efficiently prenylated in whole yeast. While the libraries based on CSLMQ gave a large number of results, it is possible this is another potentially prenylated CaaaX sequence in humans.

3.5. Conclusions

The work reported here further expands the understanding of what sequences can be prenylated by FTase and continues to elucidate the differences in specificity between yFTase and rFTase with regards to these extended sequences. The use of peptide libraries based on a previously validated pentapeptide sequence has led to the discovery of numerous, previously unidentified, peptide substrates. Analysis of the human genome reveals that there are a large number of potentially extended sequences that may have the

capacity to be prenylated. This work has potentially important implications for the prenylation of these extended sequences in a biological setting.

3.6. Methods

3.6.1 Library Synthesis

Peptide libraries were synthesized using Fmoc-based solid-phase peptide synthesis (SPPS) employing a Gyros Protein Technologies PS3® peptide synthesizer using four equivalents of Fmoc-protected amino acids from Aldrich®, Novabiochem® and P3 Biosystems and Fmoc-AA-Wang resins from P3 Biosystems. HCTU was used as the coupling reagent (0.4 mM) and 0.8 M DIEA was used as the base. A one-pot synthesis of 10 amino acids per library was performed at the "X" position. Dansylglycine (DsG) was then coupled manually, using a twofold molar excess, reacting for 4-6 h. Following synthesis, peptides were cleaved from resin for 2 h with 5 mL reagent K (82.5% trifluoroacetic acid (TFA), 5% thioanisole, 5% phenol, 2.5% 1,2-ethanedithiol, and 5% H_2O, v/v) cleavage cocktail per 0.1 mmol of resin. Peptides were precipitated by draining the cleavage cocktail into 40 mL Et_2O cooled in an isopropanol / dry ice bath for ~10 min, centrifuging until pelleted, decanting the Et_2O layer, and repeating once to remove any residual cleavage cocktail from the crude peptide mixture. The peptides were then dissolved in 50:50 CH_3CN/H_2O containing 0.1% TFA. The total peptide concentration was determined by diluting in 1,4-dioxane and measuring the absorptivity at 338 nm using a Cary 50 Bio UV-Visible spectrophotometer, using the Beer-Lambert law in conjunction with dansylglycine's molar extinction coefficient (4300 cm^{-1} M^{-1}). Individual peptide hits were synthesized using similar conditions.

3.6.2. Enzymatic Farnesylation of Peptides

Enzymatic farnesylation of peptide libraries was performed by incubating FTase from *S. cerevisiae* (yFTase) or *R. norvegicus* (rFTase) in a reaction buffer that contained 20 μM total peptide, 40 μM FPP, 50 mM Tris-HCl pH 7.5, 10 μM $ZnCl_2$, 5 mM MgCl2, 1 mM DTT in H_2O. Reactions were allowed to proceed for 5 h at 37° C. Upon completion, the samples were desalted using a Plus long Sep-Pak reverse-phase C18 environmental cartridge from Waters Corporation (WAT023635, length 3 cm, 1 cm diameter). Sep-Paks were primed by washing with 3 mL Buffer B (CH_3CN with 0.1% TFA) and equilibrated with 3 mL Buffer A (H_2O with 0.1% TFA). The sample was then loaded, washed with 2 mL each of 100% Buffer A, 10% Buffer B in Buffer A, and 20% Buffer B in Buffer A, and then eluted with 2 mL Buffer B. Samples were immediately spotted on a MALDI plate or stored at -80 °C. Control libraries were prepared under identical conditions, without the addition of FTase.

3.6.3. MALDI-TOF MS of Farnesylated Peptide libraries

Samples eluted from the Sep-Pak columns (0.5 μL) were cospotted with 0.5 μL of 10 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA) matrix dissolved in 50:50 Buffer A: Buffer B on an AB Sciex 384 Opti TOF plate. The typical spotting procedure involved spotting the matrix first, then immediately spotting the sample on top of the matrix, rapidly pipetting up and down to mix. Samples were then analyzed with an AB-Sciex 5800 13

MALDI/TOF mass spectrometer using the reflector positive mode. A laser intensity of ~4000-5000 was applied, with a pulse rate of 400 Hz. Laser intensity was increased in increments of 200 if signal was not readily apparent. 4000 laser shots were applied per spectrum, and the entire spot surface was sampled.

3.6.4. HPLC based Enzymatic Farnesylation Assay

Enzymatic farnesylation reactions with purified peptides were performed by incubating FTase from *S. cerevisiae* (yFTase) or *R. norvegicus* (rFTase) in a reaction buffer that contained 2.4 μM peptide, 10 μM FPP, 50 mM Tris-HCl pH 7.5, 10 μM $ZnCl_2$, 5 mM $MgCl_2$, 1 mM DTT in H_2O.[103,104] Reactions with yFTase were at rt for 30 min, reactions with rFTase were run at 35 C °C for 45 min. The reactions were flash frozen to stop enzymatic activity, and 200 uL aliquots were injected onto an Agilent 1100 HPLC instrument equipped with an FLD detector and a Phenomenex Luna 5-micron C18 100 Å pore size 250 x 4.60 mm, 5 μm analytical column. Fluorescence of the dansylated peptides was monitored with an excitation of 220 nm and an emission of 495 nm with a PMT gain of 12. All reactions were run in triplicate. The extent of peptide farnesylation was quantified by integration of the starting material peak from the HPLC chromatogram.

3.6.5. Peptide search of the human proteome

The scanProsite tool of Expasy was used to scan the uniprotKB for known protein sequences that contain a potential pentapeptide CaaaX sequence

(https://prosite.expasy.org/scanprosite/). The search was limited to C-terminal sequences CXXXX, where any amino acids were allowed in the varied X positions. The scan was performed as a motif search against the UniProtKB, including isoforms. Results were then filtered to only show sequences specific to the human proteome (*H. sapiens*). Further filtering was performed to remove sequences that had multiple charged residues, multiple conformationally challenging residues (Gly and Pro), or one of each from the aforementioned categories.

3.6.6 Synthesis of Peptide Hits from the Human Genome

Peptides were synthesized using Fmoc based solid-phase peptide synthesis (SPPS) using an Intavis MultiPep RS® peptide synthesizer using four equivalents of Fmoc-protected amino acids from Aldrich®, Novabiochem® and P3 Biosystems and Fmoc-AA-Wang resins from P3 Biosystems. Coupling was performed with standard HCTU coupling procedures, (0.4 mM) and 0.8 M DIEA was used as the base. with Dansylglycine being allowed to react for an extended 4 h time. Following synthesis, peptides were cleaved from resin for 2 h with 5 mL reagent K (82.5% trifluoroacetic acid (TFA), 5% thioanisole, 5% phenol, 2.5% 1,2-ethanedithiol, and 5% H_2O, v/v) cleavage cocktail per 0.1 mmol of resin. Peptides were precipitated by draining the cleavage cocktail into 40 mL Et_2O cooled in an isopropanol / dry ice bath for ~10 min, centrifuging until pelleted, decanting the Et_2O layer, and repeating once to wash the residual cleavage cocktail from the crude peptide mixture. The crude peptides were then dissolved in 50:50 CH_3CN/H_2O containing 0.1% TFA and evaluated by enzymatic reaction and MALDI analysis as described above.

Chapter 4: A Farnesyltransferase Mutant Displaying Dual Substrate Orthogonality with Application for Enzymatic Labeling

4.1 Introduction

The site-specific modification of proteins is of great interest in many areas, including therapeutics and biotechnology. The enzymatic labeling of proteins is known to increase their pharmacokinetic properties, and has been found useful for proteins ranging in size from interferon to antibodies[27,109–112]. In biotechnology, labeling applications include biosensors, support materials, and microarrays[113–115]. Despite the wide range of applications, this is still a challenging and growing area of interest[116,117]. Enzymatic labeling involves the attachment of a small functional tag, which can then be used as a fluorescent reporter or for a variety of reactions[1]. Such an approach is particularly useful since potentially any synthetic chemical moiety can be attached. This is in contrast to a genetic labeling strategy such as incorporating GFP or another fluorescent protein, to the protein of interest [118]. There have been many proteins used for enzymatic labeling, including Formylglycine generating enzyme, Sortase A, Biotin ligase, and Farnesyltransferase[119,120].

Farnesyltransferase (FTase), a member of the prenyltransferase family, catalyzes farnesylation, a post-translational modification which involves the covalent attachment of a hydrophobic 15-carbon isoprenoid group (FPP) to the thiol side chain of a cysteine residue located near the C-terminus of a protein [94]. This enzyme recognize proteins with a C-terminal tetrapeptide consensus sequence known as a “CaaX box”, where “C” is the cysteine residue that is covalently modified, “a” is usually an aliphatic amino acid, and the

"X" is a residue that is largely responsible for determining which prenyltransferase the protein is a substrate for.[1] FTase is somewhat unique in that it has high substrate flexibility for both its peptide substrates, as well as the isoprenoids it can transfer. There has been great interest in determining which amino acids are preferred or not preferred in the CaaX sequence, specifically at the a_2 and X positions.[18,29] In addition, the flexibility of the isoprenoid binding site has been leveraged to transfer a variety of isoprenoids with many different handles for click chemistry, as well as biotin and fluorophores[1].

While enzymes are highly efficient catalysts, there has been great interest over the past 40 years in altering the specificity and reactivity of enzymes through mutagenesis. This has been accomplished through a variety of techniques including site-directed mutagenesis as well as directed evolution, and has yielded many enzymes with altered substrate specificity, elevated catalytic activity, and reversed stereochemistry with applications in biosynthesis and pharmaceutical production[121,122]. These enzymes have been used in a variety of applications including biosynthesis of amino acids, production of antimicrobial drugs, and environmental remediation[123–125].

The inherent flexibility of FTase has led our group and others to investigate what rational design can be done to alter the substrate specificity of FTase through site directed mutagenesis. While the isoprenoid binding pocket can accept a variety of isoprenoids, there was interest in creating a mutant that could transfer a bulky coumarin containing isoprenoid (CoumarinOPP) that had been previously used as a substrate for another enzyme in the isoprenoid biosynthesis pathway[61]. The isoprenoid binding site is filled with aromatic

residues, and it was found that the mutation Y205A allowed for efficient transfer of CoumarinOPP to a model protein with a CaaX box[62].

The flexibility of the peptide binding site is in part due to the fact that the key interactions are the thiol chelating a zinc atom in the active site, as well as the carboxyterminus making a key hydrogen bond[53]. Variability in substrate binding is then affected by the hydrophobic nature of the amino acids. Rational design led to the creation of a mutant with a charged amino acid that could preferentially accept a charged residue at the a_2 position, something that normally is not preferred. It was found that the double mutant W102R/W106L was capable of efficiently transferring FPP to the charged CaaX box CVDS. We aspired to combine these mutants into a single enzyme (RLA mutant, named after the one letter amino acid code for the changed amino acids) which would retain the orthogonal chemistry of each individual mutant (Figure 4.1). A model of the coumarin probe docked in the RLA mutant active site can be seen in Figure 4.2

4.2. Research Objectives

In this work, we sought to evaluate the potential of combining two separate FTase mutants to create a mutant that had orthogonality in both the peptide and isoprenoid binding site. Our approach was to use fluorescence based HPLC assays to evaluate the activity and selectivity of the RLA mutant for the designed substrates of its two different mutation sets. We also employed a MALDI approach to search for an optimal substrate. We evaluated the enzyme for its selectivity for both native and charged substrates as well as FPP and CoumarinOPP in a variety of competition experiments.

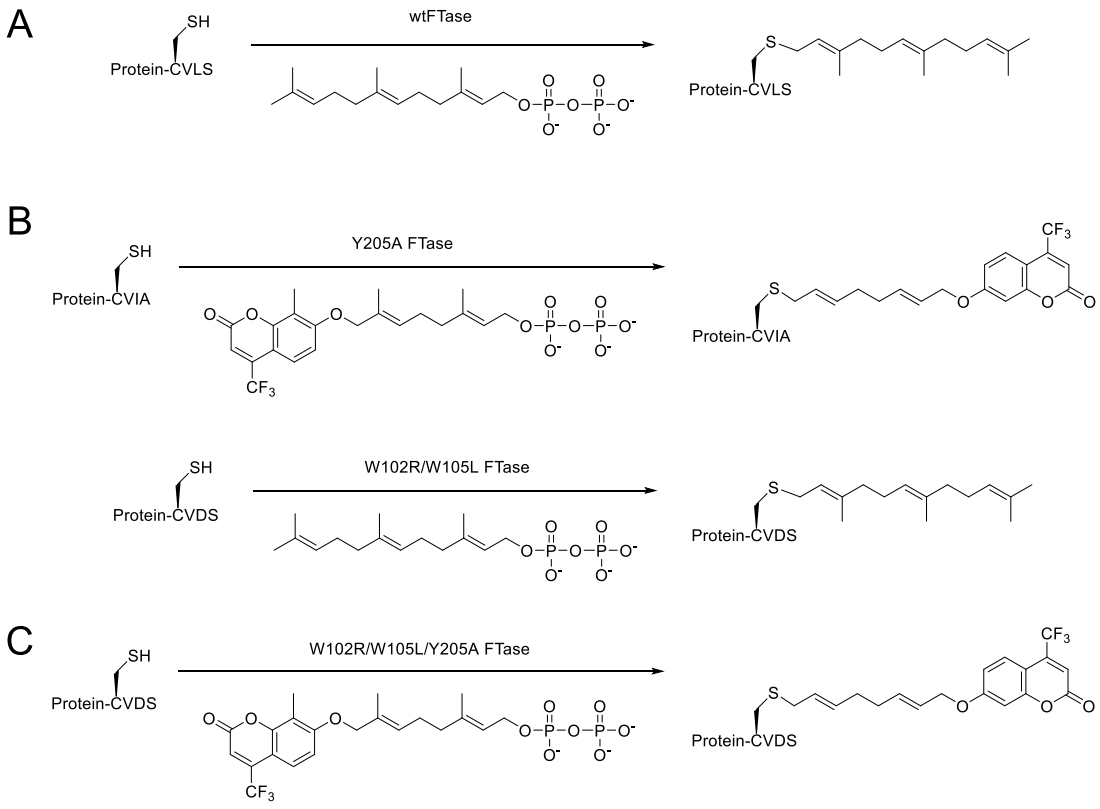


Figure 4.1. Native prenylation and mutants of interest

(A) The native prenylation reaction of CVLS with FPP and wtFTase. (B) Previously described mutants showing that Y205A allows for the transfer of CoumarinOPP and that W102R/W105L allows for FPP to be transferred to the charged peptide CVDS. (C) This work, which combines the mutations from B.

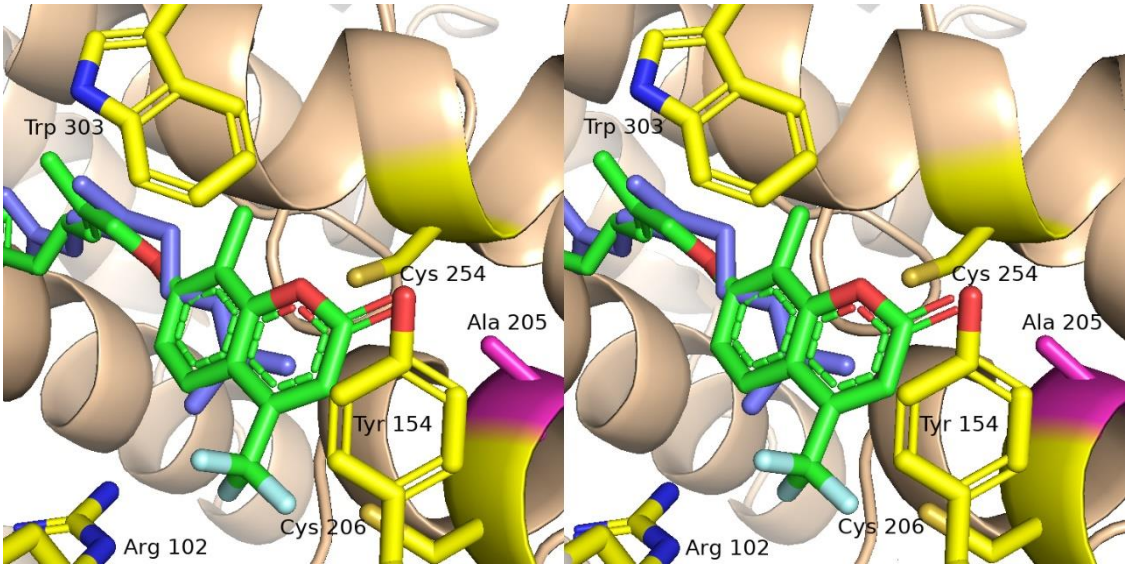


Figure 4.2. Docking of CoumarinOPP in FTase

Model of the CoumarinOPP docked in the RLA mutant. The coumarin probe (green) is shown overlaid with the native FPP substrate (blue). The Y205A mutation that allows the extra space for the Coumarin OPP to bind is shown in pink. Residues near the coumarin are shown in yellow. Peptide substrates are omitted for clarity.

4.3. Results

4.3.1 Combining previous mutants produces an active enzyme with dual substrate orthogonality

Initially, we sought to determine if combining the previous mutants yielded an enzyme that was both active and specific for the orthogonal reactions that each individual mutant catalyzed. Using a previously described continuous florescence assay, we were gratified to

find that the RLA mutant was able to transfer FPP to both a native and charged substrate. While the larger isoprenoid binding site allows for the transfer of CoumarinOPP, it still retains the ability to bind and transfer FPP. The enzyme retained a faster initial rate for the charged CVDS peptide (0.30μM/min) compared to the native CVLS peptide (.092μM/min) (Figure 4.3). We then looked to ascertain if the RLA mutant retained the ability to accept and transfer the CoumarinOPP, specifically to the charged CVDS peptide. Analysis of the enzymatic reaction of CoumarinOPP with CVDS indicates that within 45 minutes, the transfer of probe to the charged peptide had reached completion (Figure 4.4). These data allowed us to conclude that the combination of our two previous mutants was successful and retained the individual orthogonal reactions each was designed to catalyze.

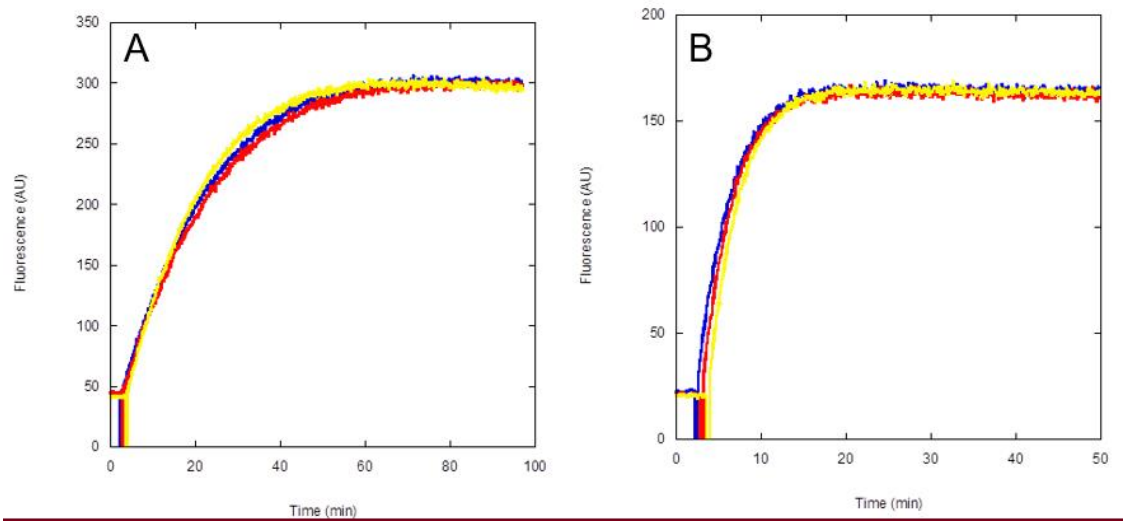


Figure 4.3. Fluorescence activity assay of triple mutant FTase with native and charged substrates

Fluorescence trace of the enzymatic reaction of the RLA mutant with the peptide (A) CVLS or (B) CVDS with FPP. The initial rates are 0.092μM/min 0.30μM/min, respectively, indicating that there is faster turnover favoring the charged substrate.

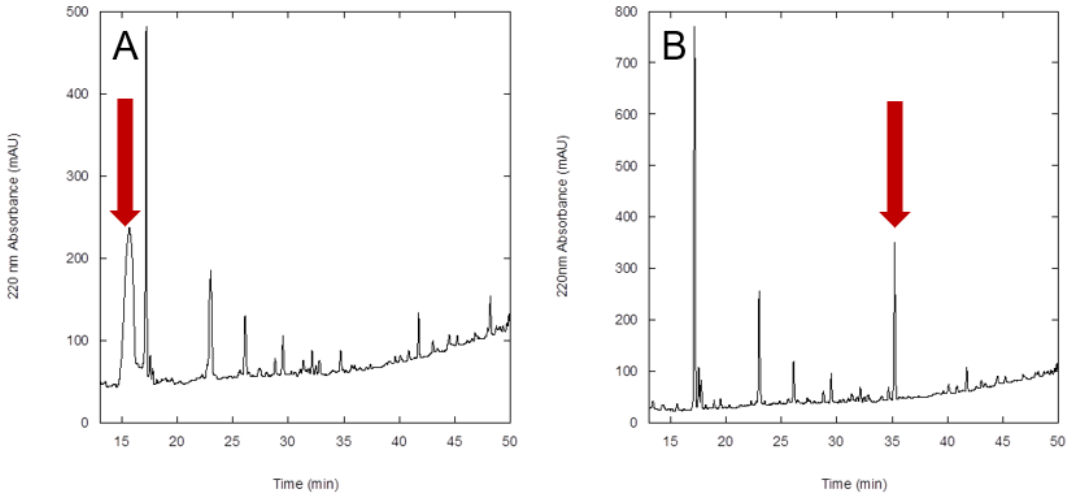


Figure 4.4. CoumarinOPP can be effectively transferred to the peptide CVDS

Enzymatic reaction buffer monitored at 220 nm (A) without enzyme and (B) with enzyme. Peptide is indicated by red arrow and confirmed by MS. The peptide has been completely converted to product.

4.3.2 The triple mutant is selective for a charged peptide but still preferentially transfers FPP over CoumarinOPP

For this enzyme to have applications in chemical biology and biotechnology, it would need to have some level of selectivity for its designed substrates. Analysis of a competition reaction of equal amounts of CVDS and CVLS with CoumarinOPP revealed a 3:1 preference for the charged substrate (Figure 4.5). The data indicate that the RLA mutant is selective for the charged peptide when either FPP or CoumarinOPP is the isoprenoid substrate, as we would expect based on the enzyme's design. Interestingly, when comparing the competition with a different CaaX sequence, CVIA, we found that the coumarinylated CVDS peptide was only observed in a trace amount, and there was no observed coumarinylated CVDA product, indicating that CVDA inhibits the enzyme (Figure 4.6). This is quite surprising, since although there are known peptides that inhibit

WT FTase, CVIA is a one of the most efficiently prenylated CaaX sequences in yeast[35]. While CVIA is one of the preferred substrates for yeast FTase, it can still be prenylated efficiently by wild type rat FTase[36]. Having determined that the mutant enzyme had retained a preference for the charged substrate, we investigated a competition between CoumarinOPP and FPP at a 10:1 ratio. We believe this to be an obtainable level in potential cell experiments by incubating cells with CoumarinOPP and reducing the native FPP concentrations through the use of statins[84]. Unfortunately, HPLC assays indicated that even at this ratio of the two substrates, the enzyme preferred to transfer FPP to CVDS by 2:1, causing us to rethink our peptide substrate choice (Figure 4.7).

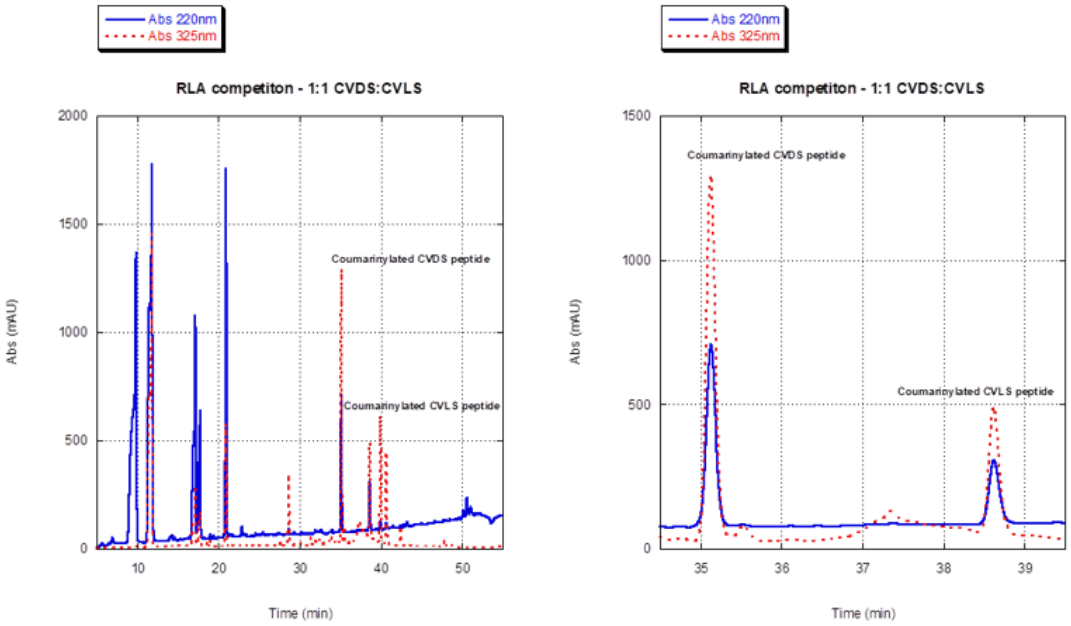


Figure 4.5. Competition between charged CVDS peptide and native CVLS peptide

A competition between equal amounts of CVDS and CVLS peptide shows that transfer of the CoumarinOPP is selective for CVDS. Integration of the 220 nm signal indicates a ratio of approximately 3:1 CVDS vs CVLS. (A) HPLC trace of competition experiment at 220 and 325 nm. 220 nm is a general absorption wavelength, while absorption at 325 is specific to the Dansyl and Coumarin moieties (B) zoom in of relevant product peaks.

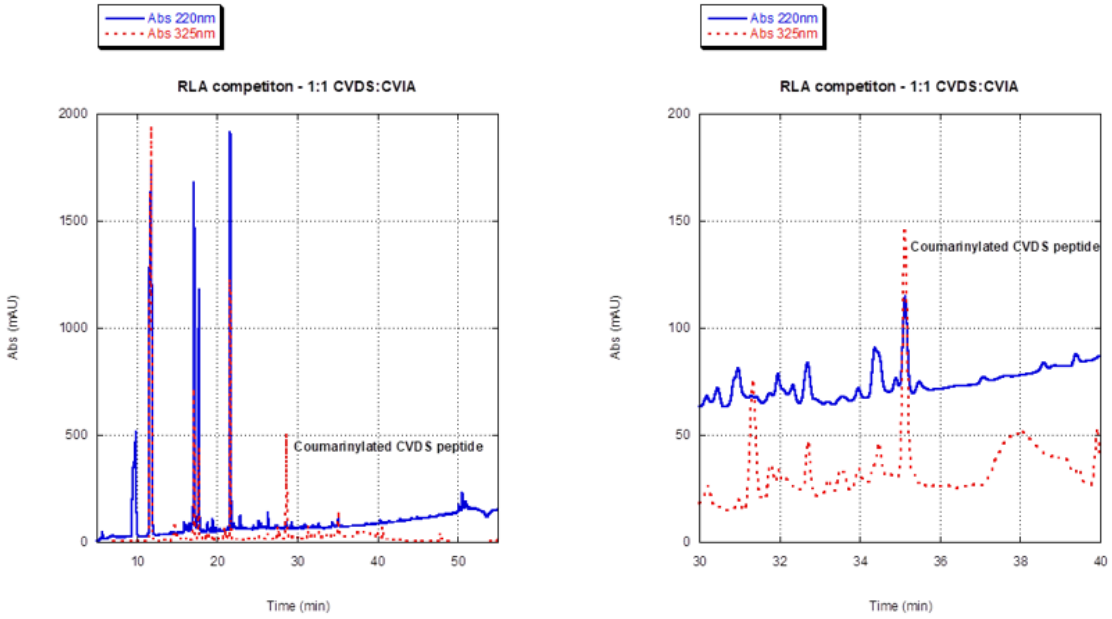


Figure 4.6. Competition between charged CVDS peptide and native CVIA peptide

A competition between equal amounts of CVDS and CVIA peptides Indicates that there is no attachment of the CoumarinOPP probe to CVIA, and a highly limited transfer of CoumarinOPP to CVDS indicates inhibition of the enzyme. (A) HPLC trace of competition experiment at 220 and 325 nm (B) zoom in of relevant product peak.

4.3.3 Optimization of the peptide substrate for CoumarinOPP

We decided to utilize a MALDI-MS approach previously described by our lab to attempt to discover a charged peptide substrate to which the mutant enzyme would preferentially transfer CoumarinOPP over FPP[45]. We created a library of peptides of the sequence DsGRAGCVDX, where X is all 20 proteogenic amino acids, to determine an optimal substrate for CoumarinOPP. We focused our search on the C-terminal amino acid, as that position makes the key hydrogen bond in the WT protein-peptide binding, whereas the

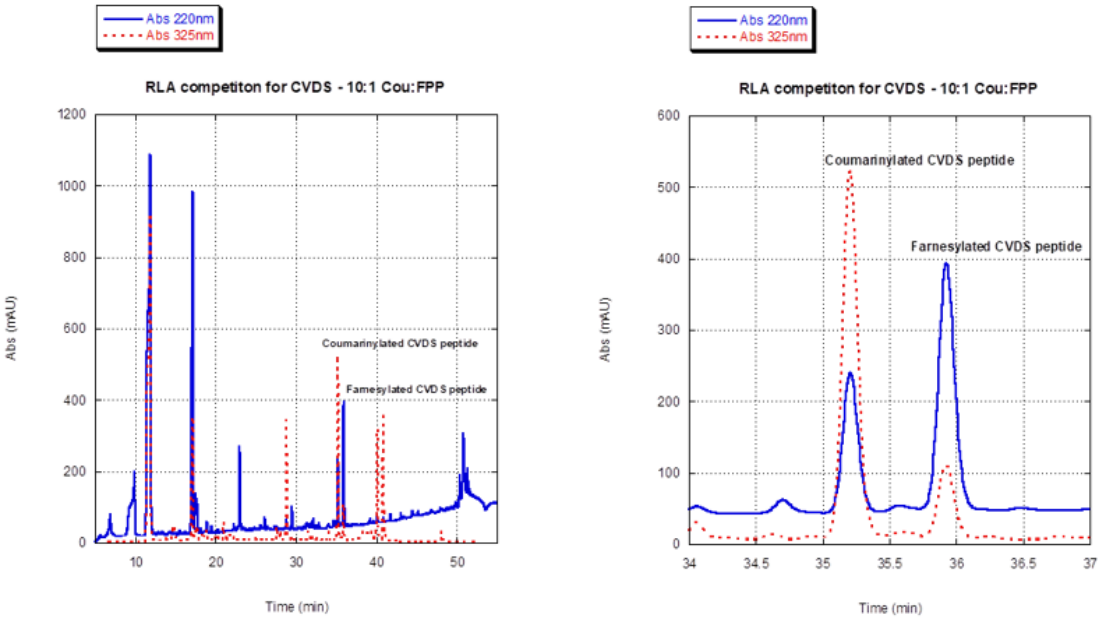


Figure 4.7. Competition between CoumarinOPP and FPP for charged CVDS peptide

A competition between 10:1 CoumarinOPP and FPP for the charged CVDS peptide. Integration of the 220 nm signal indicates that FPP is the preferred isoprenoid substrate by a ratio of 2:1 (A) HPLC trace of competition experiment at 220 and 325 nm (B) zoom in of relevant product peaks.

had the greatest ionization, and Ser was in the middle of all observed products (Figure 4.8). When the same reaction was run with CoumarinOPP, the ionization pattern of the products was altered, as we would expect based on our CoumarinOPP vs FPP experiments. We observed that Ala in the X position displayed the greatest ionization (Figure 4.9). This was slightly surprising given that we had observed CVIA inhibit the transfer of CoumarinOPP, although in general alanine is known to create efficient CaaX boxes when in the X position. While the ionization of these modified peptides provides a qualitative readout of product formation, we decided to synthesize the peptide DsGRAGCVDA to determine if single

peptide *in vitro* assays would match the MALDI results. We were gratified to find that by changing from S to A, we were able to alter the selectivity of the mutant enzyme to prefer CoumarinOPP over FPP at a ratio of 2:1, further validating our MALDI assays and optimizing the peptide substrate for the mutant enzyme (Figure 4.10).

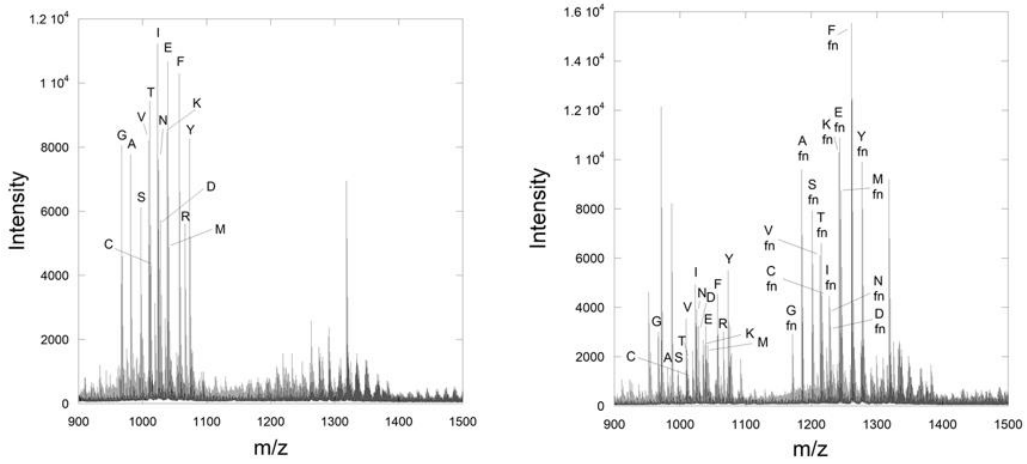


Figure 4.8. MALDI-MS of a CVDX peptide library with FPP

CVDX library before enzymatic reaction (A) and after reaction with FPP and triple mutant (B). Reactions were run with 2.4 µM peptide, 100 nM enzyme, and 10 µM FPP

4.4. Discussion

This work focuses on the analysis of an enzyme with dual orthogonality created by combining mutations from two previously rationally designed enzymes. The first of these sets of mutations alters the specificity of the peptide binding site by inserting an Arginine residue, bestowing preferential reactivity for a charged CaaX box, CVDS, over a native one such as CVLS. The second mutation increases the size of the isoprenoid binding site, allowing for the binding of a coumarin containing isoprenoid that is much bulkier than the

native FPP substrate. We were able to utilize an HPLC based assay to evaluate the retention of orthogonal reactivity for these combined mutations, utilizing 220 nm absorbance or

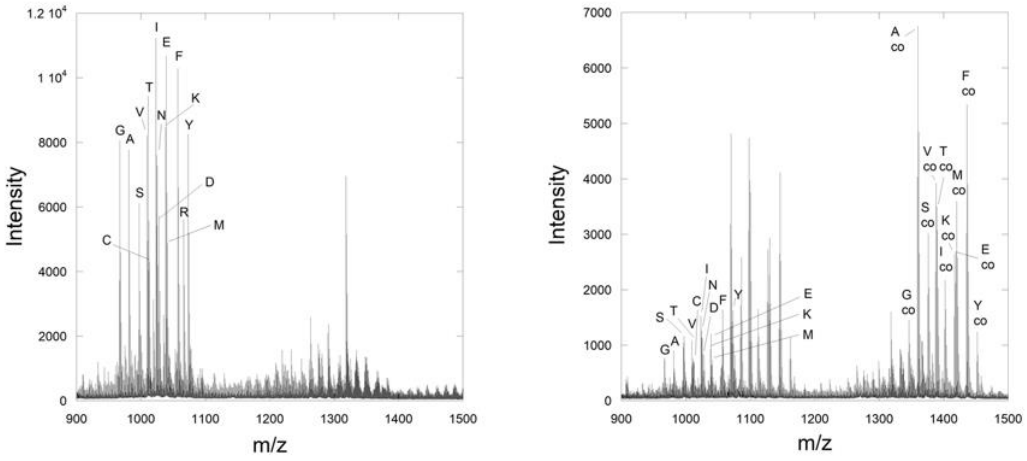


Figure 4.9. MALDI-MS of a CVDX peptide library with CoumarinOPP

CVDX library before enzymatic reaction (A) and after reaction with CoumarinOPP and triple mutant (B). The product ionization indicates that Ala in the X position might be a suitable substrate for CoumarinOPP. Reactions were run with 2.4 µM peptide, 100nM enzyme, and 10 µM CoumarinOPP.

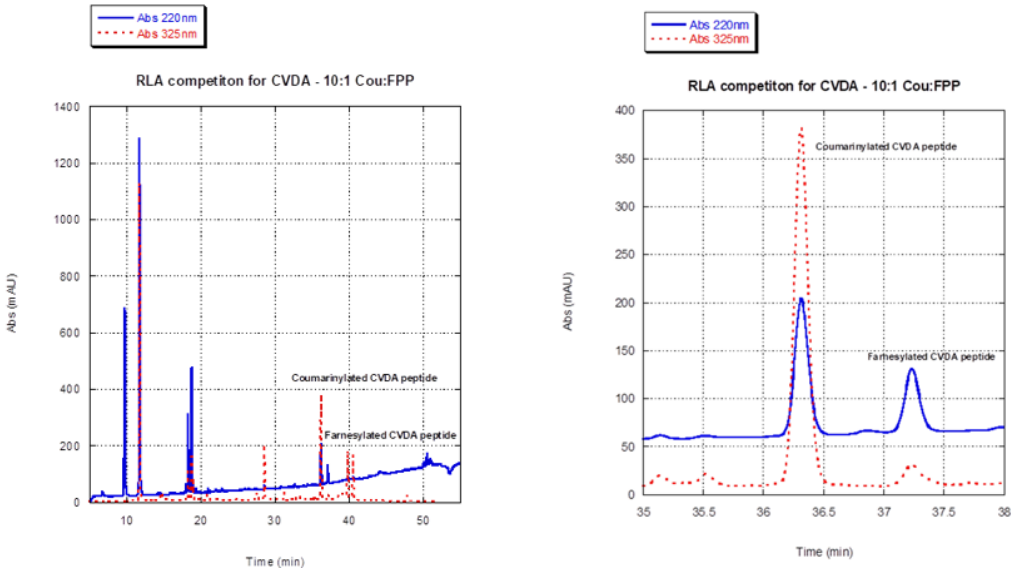


Figure 4.10. Competition between CoumarinOPP and FPP for charged CVDA peptide

A competition between 10:1 CoumarinOPP and FPP for the charged CVDA peptide. Integration of the 220 nm signal indicates that CoumarinOPP is now the preferred isoprenoid substrate by a ratio of 2:1 (A) HPLC trace of competition experiment at 220 and 325 nm (B) zoom in of relevant product peaks.

Coumarin fluorescence, as necessary. We found that while the peptides analyzed were very similar, they could still be separated in a reasonably short gradient. The prenylated products displayed a large shift in retention time, as expected, and the farnesylated and coumarinylated products were also separable.

We were able to determine that the combination of previous mutations yielded an active enzyme that could transfer FPP to both a native and charged peptide. While the charged arginine in the active site provides a preference for a charged substrate, it does not occlude the binding of the native peptide. When analyzing the activity for the charged and native peptides, we found that the initial rate of the reaction was three times faster for the charged substrate over the native one. In addition, the reaction for the charged peptide reaches completion four times faster than its uncharged counterpart. This indicates that the addition of a larger isoprenoid binding site does not remove the enzyme's preference for the charged substrate. We were also able to show that the CoumarinOPP could be efficiently transferred to the charged substrate, validating that the combination of individual mutations retained their desired orthogonal reactivities.

In further analysis of the active mutant, we evaluated its reactivity when there was competition between native and charged substrates. Gratifyingly, the enzyme was able to

selectively transfer CoumarinOPP to the charged substrate, indicating that the preference for the charged substrate was retained regardless of the isoprenoid used. Unfortunately, while the larger isoprenoid binding site allows for the binding of the bulky CoumarinOPP, FPP is still bound very well and is selectively transferred, even when the amount of CoumarinOPP is in a 10:1 ratio to FPP. However, we were able to utilize a MALDI peptide library screen to analyze how a change in the C terminal amino acid might alter the transfer of isoprenoid probes. Changing the C-terminal serine to an alanine was enough to reverse the selectivity to favor CoumarinOPP over FPP, highlighting the importance and intricacies of peptide substrates of FTase for enzymatic labeling applications. We have validated this enzyme as a potential tool for future enzymatic labeling applications, as the dual substrate orthogonality would provide a double layer of selectivity.

4.5. Conclusions

Our work evaluates the possibility of combining two previous FTase mutants into a single enzyme that displays the orthogonal activities of both individual mutants. We have validated that combining mutations that allow both peptide binding site and isoprenoid binding site orthogonality produce an enzyme that is both active and retains the unique reactivity of all mutations. We have also employed a MALDI assay to optimize the peptide substrate for this new enzyme. In combination with cellular engineering, this work could be used as an important tool for the specific tracking of proteins of interest in cells. This combination of genetic engineering and peptide library screening could be a valuable strategy applicable to many other systems.

4.6. Methods

4.6.1 Peptide Synthesis

Peptide libraries were synthesized using Fmoc based solid-phase peptide synthesis (SPPS) using a Gyros Protein Technologies PS3® peptide synthesizer using four equivalents of Fmoc-protected amino acids from Aldrich®, Novabiochem® and P3 Biosystems and Fmoc-AA-Wang resins from P3 Biosystems. Dansylglycine (DsG) was then coupled manually, using a twofold molar excess, reacting for 4-6 h. Following synthesis, peptides were cleaved from resin for 2 h with 5 mL reagent K (82.5% trifluoroacetic acid (TFA), 5% thioanisole, 5% phenol, 2.5% 1,2-ethanedithiol, and 5% H_2O, v/v) cleavage cocktail per 0.1 mmol of resin. Peptides were precipitated by draining the cleavage cocktail into 40 mL Et_2O cooled in an isopropanol / dry ice bath for ~10 min, centrifuging until pelleted, decanting the Et_2O layer, and repeating once to wash the residual cleavage cocktail from the crude peptide mixture. The peptides were then dissolved in 50:50 CH_3CN/H_2O containing 0.1% TFA. The total peptide concentration was determined by diluting in 1,4-dioxane and measuring the absorptivity at 338 nm using a Cary 50 Bio UV-Visible spectrophotometer, using Beer-Lambert's law in conjunction with dansylglycine's molar extinction coefficient (4300 cm^{-1} M^{-1}).

4.6.2. Continuous Fluorescence Assay

Monitoring of activity and initial rates of peptide substrates was performed as previously described[126].

4.6.3. Synthesis of CoumarinOPP

Synthesis of this molecule was performed as previously described[61].

4.6.4. Enzymatic Farnesylation of Peptides

Enzymatic farnesylation of peptide libraries was performed by incubating mutant FTase in a reaction buffer that contained 2.4 μM total peptide, 10 μM FPP or CoumarinOPP, 50 mM Tris-HCl, pH 7.5, 10 μM $ZnCl_2$, 5 mM $MgCl_2$, 1 mM DTT in H_2O. Reactions were allowed to proceed for 5 h at 37° C. The reactions were flash frozen to stop enzymatic activity. Upon completion, the samples were desalted using a Plus long Sep-Pak reverse-phase C18 environmental cartridge from Waters Corporation (WAT023635, length 3 cm, diameter 1 cm). Sep-Paks were primed by washing with 3 mL Buffer B (CH_3CN with 0.1% TFA) and equilibrated with 3 mL Buffer A (H_2O with 0.1% TFA). The sample was then loaded, washed with 2 mL each of 100% Buffer A, 10% Buffer B in Buffer A, and 20% Buffer B in Buffer A, and then eluted with 2 mL Buffer B. All reactions were run in triplicate. Control experiments were performed under identical conditions, without the addition of FTase.

4.6.5. HPLC and MALDI analysis of Prenylated Peptides

For HPLC analysis, 200 uL aliquots of enzymatically reacted peptides were injected onto an Agilent 1100 HPLC instrument equipped with an FLD detector and a Phenomenex Luna 5-micron C18 100 Å pore size 250 x 4.60 mm 5 micron analytical column. Fluorescence of the dansylated peptides was monitored with an excitation of 220 nm and an emission of 495 nm with a PMT gain of 12. The gradient began with a 5 min hold at 1% buffer A, followed by a 25 min gradient from 1 to 100% buffer B. For MALDI analysis, 0.5 μL of enzymatic reactions were cospotted with 0.5 μL of 10 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA) matrix in 50:50 Buffer A: Buffer B on an AB Sciex 384 Opti TOF plate. The typical spotting procedure involved spotting the matrix first, then immediately spotting the sample on top of the matrix, rapidly pipetting up and down to mix. Samples were then analyzed with an AB-Sciex 5800 13 MALDI/TOF mass spectrometer using the reflector positive mode. A laser intensity of ~4000-5000 was applied, with a pulse rate of 400 Hz. Laser intensity was increased in increments of 200 if signal was not readily apparent. 4000 laser shots were applied per spectrum, and the entire spot surface was sampled.

Bibliography

[1] C.C. Palsuledesai, M.D. Distefano, Protein Prenylation: Enzymes, Therapeutics, and Biotechnology Applications, ACS Chem. Biol. 10 (2015) 51–62.

[2] L. Desnoyers, M.C. Seabra, Single prenyl-binding site on protein prenyl transferases, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12266–12270. https://doi.org/10.1073/pnas.95.21.12266.

[3] S. Kuchay, H. Wang, A. Marzio, K. Jain, H. Homer, N. Fehrenbacher, M.R. Philips, N. Zheng, M. Pagano, GGTase3 is a newly identified geranylgeranyltransferase targeting a ubiquitin ligase, Nat. Struct. Mol. Biol. 26 (2019) 628–636. https://doi.org/10.1038/s41594-019-0249-3.

[4] R. Shirakawa, S. Goto-Ito, K. Goto, S. Wakayama, H. Kubo, N. Sakata, D.A. Trinh, A. Yamagata, Y. Sato, H. Masumoto, J. Cheng, T. Fujimoto, S. Fukai, H. Horiuchi, A SNARE geranylgeranyltransferase essential for the organization of the Golgi apparatus , EMBO J. 39 (2020) 1–20. https://doi.org/10.15252/embj.2019104120.

[5] F. Ghomashchi, X. Zhang, L. Liu, M.H. Gelb, Binding of Prenylated and Polybasic Peptides to Membranes: Affinities and Intervesicle Exchange, Biochemistry. 34 (1995) 11910–11918. https://doi.org/10.1021/bi00037a032.

[6] E.R. Hildebrandt, M. Cheng, P. Zhao, J.H. Kim, L. Wells, W.K. Schmidt, A shunt pathway limits the CaaX processing of Hsp40 Ydj1p and regulates Ydj1p-dependent phenotypes, Elife. 5 (2016) 1–22. https://doi.org/10.7554/eLife.15899.

[7] N. Berndt, A.D. Hamilton, S.M. Sebti, Targeting protein prenylation for cancer therapy, Nat Rev Cancer. 11 (2011) 775–791. http://dx.doi.org/10.1038/nrc3151.

[8] S.M. Sebti, A.D. Hamilton, Inhibition of Ras prenylation: A novel approach to cancer chemotherapy, Pharmacol. Ther. 74 (1997) 103–114. https://doi.org/10.1016/S0163-7258(97)00014-4.

[9] N.E. Kohl, C.A. Omer, M.W. Conner, N.J. Anthony, J.P. Davide, S.J. Desolms, E.A. Giuliani, R.P. Gomez, S.L. Graham, K. Hamilton, L.K. Handt, G.D. Hartman, K.S. Koblan, A.M. Kral, P.J. Miller, S.D. Mosser, T.J. O'Neill, E. Rands, M.D. Schaber, J.B. Gibbs, A. Oliff, Inhibition of farnesyltransferase induces regression of mammary and salivary carcinomas in ras transgenic mice, Nat. Med. 1 (1995) 792–797. https://doi.org/10.1038/nm0895-792.

[10] R.B. Lobell, D. Liu, C.A. Buser, J.P. Davide, E. DePuy, K. Hamilton, K.S. Koblan, Y. Lee, S. Mosser, S.L. Motzel, J.L. Abbruzzese, C.S. Fuchs, E.K. Rowinsky, E.H. Rubin, S. Sharma, P.J. Deutsch, K.E. Mazina, B.W. Morrison, L. Wildonger, S.L. Yao, N.E. Kohl, Preclinical and clinical pharmacodynamic assessment of L-778,123, a dual inhibitor of farnesyl:protein transferase and geranylgeranyl:protein transferase type-I, Mol. Cancer Ther. 1 (2002) 747–758.

[11] D.A. Hottman, L. Li, Protein Prenylation and Synaptic Plasticity: Implications for Alzheimer's Disease, Mol. Neurobiol. 50 (2014) 177–185. https://doi.org/10.1007/s12035-013-8627-z.

[12] B.C. Capell, M.R. Erdos, J.P. Madigan, J.J. Fiordalisi, R. Varga, K.N. Conneely, L.B. Gordon, C.J. Der, A.D. Cox, F.S. Collins, Inhibiting farnesylation of progerin prevents the characteristic nuclear blebbing of Hutchinson-Gilford progeria syndrome, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 12879–12884. https://doi.org/10.1073/pnas.0506001102.

[13] K.F. Suazo, C. Schaber, C.C. Palsuledesai, A.R. Odom John, M.D. Distefano, Global proteomic analysis of prenylated proteins in Plasmodium falciparum using an alkyne-modified isoprenoid analogue, Sci. Rep. 6 (2016) 1–11. https://doi.org/10.1038/srep38615.

[14] B.B. Bordier, P.L. Marion, K. Ohashi, M.A. Kay, H.B. Greenberg, J.L. Casey, J.S. Glenn, A Prenylation Inhibitor Prevents Production of Infectious Hepatitis Delta Virus Particles, J. Virol. 76 (2002) 10465–10472. https://doi.org/10.1128/jvi.76.20.10465-10472.2002.

[15] M. Amaya, A. Baranova, M.L. Van Hoek, Protein prenylation: A new mode of host-pathogen interaction, Biochem. Biophys. Res. Commun. 416 (2011) 1–6. https://doi.org/10.1016/j.bbrc.2011.10.142.

[16] L.B. Gordon, M.E. Kleinman, J. Massaro, R.B. D'Agostino, H. Shappell, M. Gerhard-Herman, L.B. Smoot, C.M. Gordon, R.H. Cleveland, A. Nazarian, B.D. Snyder, N.J. Ullrich, V.M. Silvera, M.G. Liang, N. Quinn, D.T. Miller, S.Y. Huh, A.A. Dowton, K. Littlefield, M.M. Greer, M.W. Kieran, Clinical Trial of the Protein Farnesylation Inhibitors Lonafarnib, Pravastatin, and Zoledronic Acid in Children with Hutchinson-Gilford Progeria Syndrome, Circulation. 134 (2016) 114–125. https://doi.org/10.1161/CIRCULATIONAHA.116.022188.

[17] L.B. Gordon, H. Shappell, J. Massaro, R.B. D'Agostino, J. Brazier, S.E. Campbell, M.E. Kleinman, M.W. Kieran, Association of lonafarnib treatment vs no treatment with mortality rate in patients with Hutchinson-Gilford progeria syndrome, JAMA - J. Am. Med. Assoc. 319 (2018) 1687–1695. https://doi.org/10.1001/jama.2018.3264.

[18] Y.C. Wang, M.D. Distefano, Synthesis and screening of peptide libraries with free C-termini, Curr. Top. Pept. Protein Res. 15 (2014) 1–23.

[19] K.F. Suazo, K. Park, M.D. Distefano, A Not-So-Ancient Grease History: Click Chemistry and Protein Lipid Modifications, Chem. Rev. 121 (2021) 7178–7248. https://doi.org/10.1021/acs.chemrev.0c01108.

[20] T. Subramanian, S. Liu, J.M. Troutman, D.A. Andres, H.P. Spielmann, Protein farnesyltransferase-catalyzed isoprenoid transfer to peptide depends on lipid size and shape, not hydrophobicity., Chembiochem. 9 (2008) 2872–2882.

https://doi.org/10.1002/cbic.200800248.

[21] S.A. Reigard, T.J. Zahn, K.B. Haworth, K.A. Hicks, C.A. Fierke, R.A. Gibbs, Interplay of isoprenoid and peptide substrate specificity in protein farnesyltransferase, Biochemistry. 44 (2005) 11214–11223. https://doi.org/10.1021/bi050725l.

[22] B.C. Jennings, A.M. Danowitz, Y.C. Wang, R.A. Gibbs, M.D. Distefano, C.A. Fierke, Analogs of farnesyl diphosphate alter CaaX substrate specificity and reactions rates of protein farnesyltransferase, Bioorganic Med. Chem. Lett. 26 (2016) 1333–1336. https://doi.org/10.1016/j.bmcl.2015.12.079.

[23] C.L. Strickland, W.T. Windsor, R. Syto, L. Wang, R. Bond, Z. Wu, J. Schwartz, H. V. Le, L.S. Beese, P.C. Weber, Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue, Biochemistry. 37 (1998) 16601–16611. https://doi.org/10.1021/bi981197z.

[24] J.M. Dolence, D.B. Rozema, C.D. Poulter, Yeast protein farnesyltransferase. Site-directed mutagenesis of conserved residues in the β-subunit, Biochemistry. 36 (1997) 9246–9252. https://doi.org/10.1021/bi970039p.

[25] Stephan B. Long, Patrick J. Casey, Lorena S. Beese, Reaction path of protein farnesyltransferase at atomic resolution, Nature. 419 (2002) 645–650.

[26] J.M. Dolence, P.B. Cassidy, J.R. Mathis, C.D. Poulter, Yeast Protein Famesyltransferase: Steady-State Kinetic Studies of Substrate Binding, Biochemistry. 34 (1995) 16687–16694. https://doi.org/10.1021/bi00051a017.

[27] Y. Zhang, Y. Wang, S. Uslu, S. Venkatachalapathy, M. Rashidian, J. V. Schaefer, A. Plückthun, M.D. Distefano, Enzymatic Construction of DARPin-Based Targeted Delivery Systems Using Protein Farnesyltransferase and a Capture and Release Strategy, Int. J. Mol. Sci. 23 (2022). https://doi.org/10.3390/ijms231911537.

[28] Y. Wang, O. Kilic, C.M. Csizmar, S. Ashok, J.L. Hougland, M.D. Distefano, C.R. Wagner, Engineering reversible cell-cell interactions using enzymatically lipidated chemically self-assembled nanorings, Chem. Sci. 12 (2021) 331–340. https://doi.org/10.1039/d0sc03194a.

[29] Y. Reiss, S.J. Stradley, L.M. Gierasch, M.S. Brown, J.L. Goldstein, Sequence requirement for peptide recognition by rat brain p21(ras) protein farnesyltransferase, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 732–736. https://doi.org/10.1073/pnas.88.3.732.

[30] P.F. Lebowitz, P.J. Casey, G.C. Prendergast, J.A. Thissen, Farnesyltransferase inhibitors alter the prenylation and growth- stimulating function of RhoB, J. Biol. Chem. 272 (1997) 15591–15594. https://doi.org/10.1074/jbc.272.25.15591.

[31] R. Baron, E. Fourcade, I. Lajoie-Mazenc, C. Allal, B. Couderc, R. Barbaras, G.

Favre, J.C. Faye, A. Pradines, 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 (2000) 11626–11631. https://doi.org/10.1073/pnas.97.21.11626.

[32] A.J. Krzysiak, A. V. Aditya, J.L. Hougland, C.A. Fierke, R.A. Gibbs, Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates, Bioorganic Med. Chem. Lett. 20 (2010) 767–770. https://doi.org/10.1016/j.bmcl.2009.11.011.

[33] J.L. Hougland, C.L. Lamphear, S.A. Scott, R.A. Gibbs, C.A. Fierke, Context-dependent substrate recognition by protein farnesyltransferase, Biochemistry. 48 (2009) 1691–1701. https://doi.org/10.1021/bi801710g.

[34] K.A. Hicks, H.L. Hartman, C.A. Fierke, Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I, Biochemistry. 44 (2005) 15325–15333. https://doi.org/10.1021/bi050951v.

[35] Y.C. Wang, M.D. Distefano, Solid-phase synthesis of C-terminal peptide libraries for studying the specificity of enzymatic protein prenylation, Chem. Commun. 48 (2012) 8228–8230. https://doi.org/10.1039/c2cc31713c.

[36] Y.C. Wang, J.K. Dozier, L.S. Beese, M.D. Distefano, Rapid analysis of protein farnesyltransferase substrate specificity using peptide libraries and isoprenoid diphosphate analogues, ACS Chem. Biol. 9 (2014) 1726–1735. https://doi.org/10.1021/cb5002312.

[37] J.A. Boutin, W. Marande, L. Petit, A. Loynel, C. Desmet, E. Canet, J.L. Fauchère, Investigation of S-farnesyl transferase substrate specificity with combinatorial tetrapeptide libraries, Cell. Signal. 11 (1999) 59–69. https://doi.org/10.1016/S0898-6568(98)00032-1.

[38] J.L. Hougland, K.A. Hicks, H.L. Hartman, R.A. Kelly, T.J. Watt, C.A. Fierke, Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities, J. Mol. Biol. 395 (2010) 176–190. https://doi.org/10.1016/j.jmb.2009.10.038.

[39] N. London, C.L. Lamphear, J.L. Hougland, C.A. Fierke, O. Schueler-Furman, Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity, PLoS Comput. Biol. 7 (2011). https://doi.org/10.1371/journal.pcbi.1002170.

[40] B.M. Berger, J.H. Kim, E.R. Hildebrandt, I.C. Davis, M.C. Morgan, J.L. Hougland, W.K. Schmidt, Protein isoprenylation in yeast targets COOH-terminal sequences not adhering to the CaaX consensus, Genetics. 210 (2018) 1301–1316. https://doi.org/10.1534/genetics.118.301454.

[41] C.E. Trueblood, V.L. Boyartchuk, J. Rine, Substrate specificity determinants in the

farnesyltransferase β-subunit, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 10774–10779. https://doi.org/10.1073/pnas.94.20.10774.

[42] V. Stein, M.H. Kubala, J. Steen, S.M. Grimmond, K. Alexandrov, Towards the systematic mapping and engineering of the protein prenylation machinery in Saccharomyces cerevisiae, PLoS One. 10 (2015) e0120716. https://doi.org/10.1371/journal.pone.0120716.

[43] J.L. Hougland, S.A. Gangopadhyay, C.A. Fierke, Expansion of protein farnesyltransferase specificity using "tunable" active site interactions: Development of bioengineered prenylation pathways, J. Biol. Chem. 287 (2012) 38090–38100. https://doi.org/10.1074/jbc.M112.404954.

[44] M.J. Blanden, K.F. Suazo, E.R. Hildebrandt, D.S. Hardgrove, M. Patel, W.P. Saunders, M.D. Distefano, W.K. Schmidt, J.L. Hougland, Efficient farnesylation of an extended C-terminal C(x)3X sequence motif expands the scope of the prenylated proteome, J. Biol. Chem. 293 (2018) 2770–2785. https://doi.org/10.1074/jbc.M117.805770.

[45] G.L. Schey, P.H. Buttery, E.R. Hildebrandt, S.X. Novak, W.K. Schmidt, J.L. Hougland, M.D. Distefano, Maldi-ms analysis of peptide libraries expands the scope of substrates for farnesyltransferase, Int. J. Mol. Sci. 22 (2021). https://doi.org/10.3390/ijms222112042.

[46] S. Ashok, E.R. Hildebrandt, C.S. Ruiz, D.S. Hardgrove, D.W. Coreno, W.K. Schmidt, J.L. Hougland, Protein Farnesyltransferase Catalyzes Unanticipated Farnesylation and Geranylgeranylation of Shortened Target Sequences, Biochemistry. 59 (2020) 1149–1162. https://doi.org/10.1021/acs.biochem.0c00081.

[47] J.L. Goldstein, M.S. Brown, S.J. Stradley, Y. Reiss, L.M. Gierasch, Nonfarnesylated tetrapeptide inhibitors of protein farnesyltransferase, J. Biol. Chem. 266 (1991) 15575–15578. https://doi.org/10.1016/s0021-9258(18)98441-3.

[48] M.S. Brown, J.L. Goldstein, K.J. Paris, J.P. Burnier, J.C. Marsters, Tetrapeptide inhibitors of protein farnesyltransferase: Ammo-terminal substitution in phenylalanine-containing tetrapeptides restores farnesylation, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 8313–8316. https://doi.org/10.1073/pnas.89.17.8313.

[49] J.C. Marsters, R.S. McDowell, M.E. Reynolds, D.A. Oare, T.C. Somers, M.S. Stanley, T.E. Rawson, M.E. Struble, D.J. Burdick, K.S. Chan, C.M. Duarte, K.J. Paris, J.Y.K. Tom, D.T. Wan, Y. Xue, J.P. Bumier, Benzodiazepine peptidomimetic inhibitors of farnesyltransferase, Bioorganic Med. Chem. 2 (1994) 949–957. https://doi.org/10.1016/S0968-0896(00)82044-1.

[50] Y. Qian, A. Vogt, S.M. Sebti, A.D. Hamilton, Design and synthesis of non-peptide Ras CAAX mimetics as potent farnesyltransferase inhibitors, J. Med. Chem. 39 (1996) 217–223. https://doi.org/10.1021/jm950414g.

[51] A. Wallace, K.S. Koblan, K. Hamilton, D.J. Marquis-Omer, P.J. Miller, S.D. Mosser, C.A. Omer, M.D. Schaber, R. Cortese, A. Oliff, J.B. Gibbs, A. Pessi, Selection of potent inhibitors of farnesyl-protein transferase from a synthetic tetrapeptide combinatorial library, J. Biol. Chem. 271 (1996) 31306–31311. https://doi.org/10.1074/jbc.271.49.31306.

[52] S.L. Graham, J. DeSolms, E.A. Giuliani, N.E. Kohl, S.D. Mosser, A.I. Oliff, D.L. Pompliano, E. Rands, Pseudopeptide Inhibitors of Ras Farnesyl-Protein Transferase, J. Med. Chem. 26237 (1994) 725–732.

[53] S.B. Long, P.J. Hancock, A.M. Kral, H.W. Hellinga, L.S. Beese, 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 (2001) 12948–12953. https://doi.org/10.1073/pnas.241407898.

[54] D. Chakrabarti, T. Da Silva, J. Barger, S. Paquette, H. Patel, S. Patterson, C.M. Allen, Protein farnesyltransferase and protein prenylation in Plasmodium falciparum, J. Biol. Chem. 277 (2002) 42066–42073. https://doi.org/10.1074/jbc.M202860200.

[55] D.J. Owen, K. Alexandrov, E. Rostkova, A.J. Scheidig, R.S. Goody, H. Waldmann, Chemo-enzymatic synthesis of fluorescent Rab 7 proteins: Tools to study vesicular trafficking in cells, Angew. Chemie - Int. Ed. 38 (1999) 509–512. https://doi.org/10.1002/(sici)1521-3773(19990215)38:4<509::aid-anie509>3.0.co;2-3.

[56] N.H. Thoma, A. Iakovenko, D. Owen, A.S. Scheidig, H. Waldmann, R.S. Goody, K. Alexandrov, Phosphoisoprenoid binding specificity of geranylgeranyltransferase type II, Biochemistry. 39 (2000) 12043–12052. https://doi.org/10.1021/bi000835m.

[57] K.H. Teng, A.P.C. Chen, C.J. Kuo, Y.C. Li, H.G. Liu, C.T. Chen, P.H. Liang, Fluorescent substrate analog for monitoring chain elongation by undecaprenyl pyrophosphate synthase in real time, Anal. Biochem. 417 (2011) 136–141. https://doi.org/10.1016/j.ab.2011.05.043.

[58] K.H. Teng, E.T. Hsu, Y.H. Chang, S.W. Lin, P.H. Liang, Fluorescent Farnesyl Diphosphate Analogue: A Probe To Validate trans-Prenyltransferase Inhibitors, Biochemistry. 55 (2016) 4366–4374. https://doi.org/10.1021/acs.biochem.6b00486.

[59] B. Dursina, R. Reents, C. Delon, Y. Wu, M. Kulharia, M. Thutewohl, A. Veligodsky, A. Kalinin, V. Evstifeev, D. Ciobanu, S.E. Szedlacsek, H. Waldmann, R.S. Goody, K. Alexandrov, Identification and specificity profiling of protein prenyltransferase inhibitors using new fluorescent phosphoisoprenoids, J. Am. Chem. Soc. 128 (2006) 2822–2835. https://doi.org/10.1021/ja052196e.

[60] Y.W. Wu, H. Waldmann, R. Reents, F.H. Ebetino, R.S. Goody, K. Alexandrov, A

protein fluorescence amplifier: Continuous fluorometric assay for rab geranylgeranyltransferase, ChemBioChem. 7 (2006) 1859–1861. https://doi.org/10.1002/cbic.200600377.

[61] A.P.C. Chen, Y.H. Chen, H.P. Liu, Y.C. Li, C.T. Chen, P.H. Liang, Synthesis and application of A fluorescent substrate analogue to study ligand interactions for undecaprenyl pyrophosphate synthase, J. Am. Chem. Soc. 124 (2002) 15217–15224. https://doi.org/10.1021/ja020937v.

[62] J.K. Dozier, S.L. Khatwani, J.W. Wollack, Y.-C. Wang, C. Schmidt-Dannert, M.D. Distefano, Engineering Protein Farnesyltransferase for Enzymatic Protein Labeling Applications, Bioconjug. Chem. 25 (2014) 1203–1212. https://doi.org/10.1021/bc500240p.

[63] D. Das, Z. Tnimov, U.T.T. Nguyen, G. Thimmaiah, H. Lo, D. Abankwa, Y. Wu, R.S. Goody, H. Waldmann, K. Alexandrov, Flexible and General Synthesis of Functionalized Phosphoisoprenoids for the Study of Prenylation in vivo and in vitro, ChemBioChem. 13 (2012) 674–683. https://doi.org/10.1002/cbic.201100733.

[64] C.M. Das, Nagaratnam P. and Allen, INHIBITION OF FARNESYL TRANSFERASES FROM MALIGNANT AND NON-MALIGNANT CULTURED HUMAN LYMPHOCYTES BY PRENYL SUBSTRATE ANALOGUES, Biochem. Biophys. Res. Commun. 181 (1991) 729–735.

[65] C.A. Omer, A.M. Kral, R.E. Diehl, G.C. Prendergast, J.B. Gibbs, N.E. Kohl, S. Powers, C.M. Allen, Characterization of Recombinant Human Farnesyl-Protein Transferase: Cloning, Expression, Farnesyl Diphosphate Binding, and Functional Homology with Yeast Prenyl-Protein Transferases, Biochemistry. 32 (1993) 5167–5176. https://doi.org/10.1021/bi00070a028.

[66] Y.E. Bukhtiyarov, C.A. Omer, C.M. Allen, Photoreactive analogues of prenyl diphosphates as inhibitors and probes of human protein farnesyltransferase and geranylgeranyltransferase type I, J. Biol. Chem. 270 (1995) 19035–19040. https://doi.org/10.1074/jbc.270.32.19035.

[67] R.L. Edelstein, M.D. Distefano, Photoaffinity labeling of yeast farnesyl protein transferase and enzymatic synthesis of a Ras protein incorporating a photoactive isoprenoid, Biochem. Biophys. Res. Commun. 235 (1997) 377–382. https://doi.org/10.1006/bbrc.1997.6792.

[68] G.J. Quellhorst, C.M. Allen, M. Wessling-Resnick, Modification of Rab5 with a Photoactivatable Analog of Geranylgeranyl Diphosphate, J. Biol. Chem. 276 (2001) 40727–40733. https://doi.org/10.1074/jbc.M104398200.

[69] K.A.H. Chehade, K. Kiegiel, R.J. Isaacs, J.S. Pickett, K.E. Bowers, C.A. Fierke, D.A. Andres, H.P. Spielmann, Photoaffinity analogues of farnesyl pyrophosphate transferable by protein farnesyl transferase, J. Am. Chem. Soc. 124 (2002) 8206–

8219. https://doi.org/10.1021/ja0124717.

[70] T.C. Turek, I. Gaon, M.D. Distefano, Analogs of farnesyl pyrophosphate incorporating internal benzoylbenzoate esters: Synthesis, inhibition kinetics and photoinactivation of yeast protein farnesyltransferase, Tetrahedron Lett. 37 (1996) 4845–4848. https://doi.org/10.1016/0040-4039(96)00972-0.

[71] I. Gaon, T.C. Turek, V.A. Weller, R.L. Edelstein, S.K. Singh, M.D. Distefano, Photoactive analogs of farnesyl pyrophosphate containing benzoylbenzoate esters: Synthesis and application to photoaffinity labeling of yeast protein farnesyltransferase, J. Org. Chem. 61 (1996) 7738–7745. https://doi.org/10.1021/jo9602736.

[72] I. Gaon, T.C. Turek, M.D. Distefano, Farnesyl and geranylgeranyl pyrophosphate analogs incorporating benzoylbenzyl ethers: Synthesis and inhibition of yeast protein farnesyltransferase, Tetrahedron Lett. 37 (1996) 8833–8836. https://doi.org/10.1016/S0040-4039(96)02066-7.

[73] T.C. Turek-Etienne, C.L. Strickland, M.D. Distefano, Biochemical and structural studies with prenyl diphosphate analogues provide insights into isoprenoid recognition by protein farnesyl transferase, Biochemistry. 42 (2003) 3716–3724. https://doi.org/10.1021/bi0266838.

[74] T.C. Turek, I. Gaon, D. Gamache, M.D. Distefano, Synthesis and evaluation of benzophenone-based photoaffinity labeling analogs of prenyl pyrophosphates containing stable amide linkages, Bioorganic Med. Chem. Lett. 7 (1997) 2125–2130. https://doi.org/10.1016/S0960-894X(97)00373-9.

[75] T.C. Turek, I. Gaon, M.D. Distefano, C.L. Strickland, Synthesis of farnesyl diphosphate analogues containing ether-linked photoactive benzophenones and their application in studies of protein prenyltransferases, J. Org. Chem. 66 (2001) 3253–3264. https://doi.org/10.1021/jo991130x.

[76] Y. Kho, S.C. Kim, C. Jiang, D. Barma, S.W. Kwon, J. Cheng, J. Jaunbergs, C. Weinbaum, F. Tamanoi, J. Falck, Y. Zhao, A tagging-via-substrate technology for detection and proteomics of farnesylated proteins, Proc. Natl. Acad. Sci. United States Am. . 101 (2004) 12479–12484. https://doi.org/10.1073/pnas.0403413101.

[77] A.F.H. Berry, W.P. Heal, A.K. Tarafder, T. Tolmachova, R.A. Baron, M.C. Seabra, E.W. Tate, Rapid multilabel detection of geranylgeranylated proteins by using bioorthogonal ligation chemistry, ChemBioChem. 11 (2010) 771–773. https://doi.org/10.1002/cbic.201000087.

[78] A.J. DeGraw, C. Palsuledesai, J.D. Ochocki, J.K. Dozier, S. Lenevich, M. Rashidian, M.D. Distefano, Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation, Chem. Biol. Drug Des. 76 (2010) 460–471. https://doi.org/10.1111/j.1747-0285.2010.01037.x.

[79] G. Charron, L.K. Tsou, W. Maguire, J.S. Yount, H.C. Hang, Alkynyl-farnesol

reporters for detection of protein S-prenylation in cells, Mol. Biosyst. 7 (2011) 67–73. https://doi.org/10.1039/c0mb00183j.

[80] B.P. Duckworth, Z. Zhang, A. Hosokawa, M.D. Distefano, Selective labeling of proteins by using protein farnesyltransferase, ChemBioChem. 8 (2007) 98–105. https://doi.org/10.1002/cbic.200600340.

[81] A. Hosokawa, J.W. Wollack, Z. Zhang, L. Chen, G. Barany, M.D. Distefano, Evaluation of an alkyne-containing analogue of farnesyl diphosphate as a dual substrate for protein-prenyltransferases, Int. J. Pept. Res. Ther. 13 (2007) 345–354. https://doi.org/10.1007/s10989-007-9090-3.

[82] M.W. Rose, N.D. Rose, J. Boggs, S. Lenevich, J. Xu, G. Barany, M.D. Distefano, Evaluation of geranylazide and farnesylazide diphosphate for incorporation of prenylazides into a CAAX box-containing peptide using protein farnesyltransferase, J. Pept. Res. 65 (2005) 529–537. https://doi.org/10.1111/j.1399-3011.2005.00261.x.

[83] L.N. Chan, C. Hart, L. Guo, T. Nyberg, B.S.J. Davies, L.G. Fong, S.G. Young, B.J. Agnew, F. Tamanoi, A novel approach to tag and identify geranylgeranylated proteins, Electrophoresis. 30 (2009) 3598–3606. https://doi.org/10.1002/elps.200900259.

[84] C.C. Palsuledesai, J.D. Ochocki, M.M. Kuhns, Y.C. Wang, J.K. Warmka, D.S. Chernick, E. V. Wattenberg, L. Li, E.A. Arriaga, M.D. Distefano, Metabolic Labeling with an Alkyne-modified Isoprenoid Analog Facilitates Imaging and Quantification of the Prenylome in Cells, ACS Chem. Biol. 11 (2016) 2820–2828. https://doi.org/10.1021/acschembio.6b00421.

[85] Z.A. Maxwell, K.F. Suazo, H.M.G. Brown, M.D. Distefano, E.A. Arriaga, Combining Isoprenoid Probes with Antibody Markers for Mass Cytometric Analysis of Prenylation in Single Cells, Anal. Chem. 94 (2022) 11521–11528. https://doi.org/10.1021/acs.analchem.2c01509.

[86] W. Qu, K.F. Suazo, W. Liu, S. Cheng, A. Jeong, D. Hottman, L.L. Yuan, M.D. Distefano, L. Li, Neuronal Protein Farnesylation Regulates Hippocampal Synaptic Plasticity and Cognitive Function, Mol. Neurobiol. 58 (2021) 1128–1144. https://doi.org/10.1007/s12035-020-02169-w.

[87] A. Jeong, S.A. Auger, S. Maity, K. Fredriksen, R. Zhong, L. Li, M.D. Distefano, In Vivo Prenylomic Profiling in the Brain of a Transgenic Mouse Model of Alzheimer's Disease Reveals Increased Prenylation of a Key Set of Proteins, ACS Chem. Biol. 17 (2022) 2863–2876. https://doi.org/10.1021/acschembio.2c00486.

[88] K.F. Suazo, A. Jeong, M. Ahmadi, C. Brown, W. Qu, L. Li, M.D. Distefano, Metabolic labeling with an alkyne probe reveals similarities and differences in the prenylomes of several brain-derived cell lines and primary cells, Sci. Rep. 11 (2021) 1–17. https://doi.org/10.1038/s41598-021-83666-3.

[89] E.M. Storck, J. Morales-Sanfrutos, R.A. Serwa, N. Panyain, T. Lanyon-Hogg, T. Tolmachova, L.N. Ventimiglia, J. Martin-Serrano, M.C. Seabra, B. Wojciak-Stothard, E.W. Tate, Dual chemical probes enable quantitative system-wide analysis of protein prenylation and prenylation dynamics, Nat. Chem. 11 (2019) 552–561. https://doi.org/10.1038/s41557-019-0237-6.

[90] M. Rashidian, J.M. Song, R.E. Pricer, M.D. Distefano, Chemoenzymatic reversible immobilization and labeling of proteins without prior purification, J. Am. Chem. Soc. 134 (2012) 8455–8467. https://doi.org/10.1021/ja211308s.

[91] J.W. Wollack, B.J. Monson, J.K. Dozier, J.J. Dalluge, K. Poss, S.A. Hilderbrand, M.D. Distefano, Site-specific labeling of proteins and peptides with trans-cyclooctene containing handles capable of tetrazine ligation, Chem. Biol. Drug Des. 84 (2014) 140–147. https://doi.org/10.1111/cbdd.12303.

[92] M. Rashidian, S.C. Kumarapperuma, K. Gabrielse, A. Fegan, C.R. Wagner, M.D. Distefano, Simultaneous dual protein labeling using a triorthogonal reagent, J. Am. Chem. Soc. 135 (2013) 16388–16396. https://doi.org/10.1021/ja403813b.

[93] Y. Zhang, M.J. Blanden, C. Sudheer, S.A. Gangopadhyay, M. Rashidian, J.L. Hougland, M.D. Distefano, Simultaneous Site-Specific Dual Protein Labeling Using Protein Prenyltransferases, Bioconjug. Chem. 26 (2015) 2542–2553. https://doi.org/10.1021/acs.bioconjchem.5b00553.

[94] F.L. Zhang, P.J. Casey, Protein Prenylation: Molecular Mechanisms and Functional Consequences, Annu. Rev. Biochem. 65 (1996) 241–269. https://doi.org/10.1146/annurev.bi.65.070196.001325.

[95] M.D. Resh, Trafficking and signaling by fatty-acylated and prenylated proteins, Nat Chem Biol. 2 (2006) 584–590. http://dx.doi.org/10.1038/nchembio834.

[96] W.M. Nadler, D. Waidelich, A. Kerner, S. Hanke, R. Berg, A. Trumpp, C. Rösli, MALDI versus ESI: The Impact of the Ion Source on Peptide Identification, J. Proteome Res. 16 (2017) 1207–1215. https://doi.org/10.1021/acs.jproteome.6b00805.

[97] W. Sandoval, Matrix-assisted laser desorption/ionization time-of-flight mass analysis of peptides, Curr. Protoc. Protein Sci. 2014 (2014) 16.2.1-16.2.11. https://doi.org/10.1002/0471140864.ps1602s77.

[98] R. Scott Youngquist, G.R. Fuentes, M.P. Lacey, T. Keough, T.A. Baillie, Matrix-assisted laser desorption ionization for rapid determination of the sequences of biologically active peptides isolated from support-bound combinatorial peptide libraries, Rapid Commun. Mass Spectrom. 8 (1994) 77–81. https://doi.org/10.1002/rcm.1290080115.

[99] G. Schlosser, G. Pocsfalvi, E. Huszár, A. Malorni, F. Hudecz, MALDI-TOF mass spectrometry of a combinatorial peptide library: Effect of matrix composition on signal suppression, J. Mass Spectrom. 40 (2005) 1590–1594.

https://doi.org/10.1002/jms.937.

[100] S.J. Park, J.S. Song, H.J. Kim, Dansylation of tryptic peptides for increased sequence coverage in protein identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometric peptide mass fingerprinting, Rapid Commun. Mass Spectrom. 19 (2005) 3089–3096. https://doi.org/10.1002/rcm.2166.

[101] E. Krause, H. Wenschuh, P.R. Jungblut, The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins, Anal. Chem. 71 (1999) 4160–4165. https://doi.org/10.1021/ac990298f.

[102] W. Lee Herman, G. Tarr, S.A. Kates, Optimization of the synthesis of peptide combinatorial libraries using a one-pot method, Mol. Divers. 2 (1997) 147–155. https://doi.org/10.1007/bf01682202.

[103] S.L. Khatwani, J.S. Kang, D.G. Mullen, M.A. Hast, L.S. Beese, M.D. Distefano, T.A. Taton, Covalent protein-oligonucleotide conjugates by copper-free click reaction, Bioorganic Med. Chem. 20 (2012) 4532–4539. https://doi.org/10.1016/j.bmc.2012.05.017.

[104] A.J. DeGraw, M.A. Hast, J. Xu, D. Mullen, L.S. Beese, G. Barany, M.D. Distefano, Caged protein prenyltransferase substrates: Tools for understanding protein prenylation, Chem. Biol. Drug Des. 72 (2008) 171–181. https://doi.org/10.1111/j.1747-0285.2008.00698.x.

[105] R.C. Elble, A simple and efficient procedure for transformation of yeasts, Biotechniques. 13 (1992) 18–20. https://doi.org/10.1002/yea.1104.

[106] S. Kim, A.N. Lapham, C.G.K. Freedman, T.L. Reed, W.K. Schmidt, Yeast as a tractable genetic system for functional studies of the insulin-degrading enzyme, J. Biol. Chem. 280 (2005) 27481–27490. https://doi.org/10.1074/jbc.M414192200.

[107] W.R. Bishop, R. Doll, P. Kirschmeier, Farnesyl Transferase Inhibitors: From Targeted Cancer Therapeutic to a Potential Treatment for Progeria, 1st ed., Elsevier Inc., 2011. https://doi.org/10.1016/B978-0-12-381339-8.00015-9.

[108] A.L. Ho, I. Brana, R. Haddad, J. Bauman, K. Bible, S. Oosting, D.J. Wong, M.J. Ahn, V. Boni, C. Even, J. Fayette, M.J. Flor, K. Harrington, S.B. Kim, L. Licitra, I. Nixon, N.F. Saba, S. Hackenberg, P. Specenier, F. Worden, B. Balsara, M. Leoni, B. Martell, C. Scholz, A. Gualberto, Tipifarnib in head and neck squamous cell carcinoma with HRAS mutations, J. Clin. Oncol. 39 (2021) 1856–1864. https://doi.org/10.1200/JCO.20.02903.

[109] K. Rajender Reddy, M.W. Modi, S. Pedder, Use of peginterferon alfa-2a (40 KD) (Pegasys®) for the treatment of hepatitis C, Adv. Drug Deliv. Rev. 54 (2002) 571–586. https://doi.org/10.1016/S0169-409X(02)00028-5.

[110] J. Hu, G. Wang, W. Zhao, X. Liu, L. Zhang, W. Gao, Site-specific in situ growth

of an interferon-polymer conjugate that outperforms PEGASYS in cancer therapy, Biomaterials. 96 (2016) 84–92. https://doi.org/10.1016/j.biomaterials.2016.04.035.

[111] T. Pleiner, M. Bates, S. Trakhanov, C.T. Lee, J.E. Schliep, H. Chug, M. Böhning, H. Stark, H. Urlaub, D. Görlich, Nanobodies: Site-specific labeling for super-resolution imaging, rapid epitope- mapping and native protein complex isolation, Elife. 4 (2015) 1–21. https://doi.org/10.7554/eLife.11349.

[112] G. Falck, K.M. Müller, Enzyme-based labeling strategies for antibody–drug conjugates and antibody mimetics, Antibodies. 7 (2018). https://doi.org/10.3390/antib7010004.

[113] A. Ibraheem, R.E. Campbell, Designs and applications of fluorescent protein-based biosensors, Curr. Opin. Chem. Biol. 14 (2010) 30–36. https://doi.org/10.1016/j.cbpa.2009.09.033.

[114] S. Datta, L.R. Christena, Y.R.S. Rajaram, Enzyme immobilization: an overview on techniques and support materials, 3 Biotech. 3 (2013) 1–9. https://doi.org/10.1007/s13205-012-0071-7.

[115] A. Syahir, K. Usui, K. Tomizaki, K. Kajikawa, H. Mihara, Label and Label-Free Detection Techniques for Protein Microarrays, Microarrays. 4 (2015) 228–244. https://doi.org/10.3390/microarrays4020228.

[116] I.S. Carrico, Chemoselective modification of proteins: Hitting the target, Chem. Soc. Rev. 37 (2008) 1423–1431. https://doi.org/10.1039/b703364h.

[117] S. Jevševar, M. Kunstelj, V.G. Porekar, PEGylation of therapeutic proteins, Biotechnol. J. 5 (2010) 113–128. https://doi.org/10.1002/biot.200900218.

[118] E.C. Jensen, Use of Fluorescent Probes: Their Effect on Cell Biology and Limitations, Anat. Rec. 295 (2012) 2031–2036. https://doi.org/10.1002/ar.22602.

[119] M. Rashidian, J.K. Dozier, M.D. Distefano, Enzymatic labeling of proteins: Techniques and approaches, Bioconjug. Chem. 24 (2013) 1277–1294. https://doi.org/10.1021/bc400102w.

[120] Y. Zhang, K.-Y. Park, K.F. Suazo, M.D. Distefano, Recent progress in enzymatic protein labelling techniques and their applications, Chem. Soc. Rev. 47 (2018) 9106–9136. https://doi.org/10.1039/C8CS00537K.

[121] S. Lutz, S.M. Iamurri, Protein Engineering: Past, Present, and Future, in: Protein Eng. Methods Protoc., Springer, 2018: pp. 1–12. https://doi.org/10.1007/978-1-4939-7366-8_1.

[122] J.A. Brannigan, A.J. Wilkinson, Protein engineering 20 years on, Nat. Rev. Mol. Cell Biol. 3 (2002) 964–970. https://doi.org/10.1038/nrm975.

[123] E. Watkins-Dulaney, S. Straathof, F. Arnold, Tryptophan Synthase: Biocatalyst Extraordinaire, ChemBioChem. 22 (2021) 5–16.

https://doi.org/10.1002/cbic.202000379.

[124] G.K. Meghwanshi, N. Kaur, S. Verma, N.K. Dabi, A. Vashishtha, P.D. Charan, P. Purohit, H.S. Bhandari, N. Bhojak, R. Kumar, Enzymes for pharmaceutical and therapeutic applications, Biotechnol. Appl. Biochem. 67 (2020) 586–601. https://doi.org/10.1002/bab.1919.

[125] V. Tournier, C.M. Topham, A. Gilles, B. David, C. Folgoas, E. Moya-Leclair, E. Kamionka, M.L. Desrousseaux, H. Texier, S. Gavalda, M. Cot, E. Guémard, M. Dalibey, J. Nomme, G. Cioci, S. Barbe, M. Chateau, I. André, S. Duquesne, A. Marty, An engineered PET depolymerase to break down and recycle plastic bottles, Nature. 580 (2020) 216–219. https://doi.org/10.1038/s41586-020-2149-4.

[126] D.L. Pompliano, R.P. Gomez, N.J. Anthony, Continuous Assay of Ras FamesyhProtein Transferase Kedarcidin , a New Chromoprotein Antitumor Antibiotic : Structure Elucidation of Kedarcidin Chromophore, J. Am. Chem. Soc. (1992) 7945–7946.

Appendix

Gene Name	Enzyme	CAAX
BROX	FTase	CYI S
CDC42	GGTase I	CAVR
DNAJA1	FTase	CNVL
DNAJA2	FTase	CAHQ
GNG12	GGTase I	CIIL
GNG5	GGTase I	CSFL
LMNB1	FTase	CAIM
NAP1L1	FTase	CKQ Q
PEX19	FTase	CLIM
RAB11B	GGTase II	CCQNL
RAB13	GGTase II	CSLG
RAB14	GGTase II	GCGC
RAB15	GGTase II	TCWC
RAB16	GGTase II	TCWC
RAB18	GGTase II	CSVL
RAB1A	GGTase II	GGCC
RAB1B	GGTase II	GGCC
RAB23	GGTase II	CSIP
RAB2B	GGTase II	SGCC
RAB31	GGTase II	RRCC
RAB3B	GGTase II	NCSC
RAB4A	GGTase II	ECGC
RAB5A	GGTase II	CCSN
RAB5B	GGTase II	CCSN
RAB5C	GGTase II	CCSN
RAB6B	GGTase II	GCSC
RAB7A	GGTase II	SCSC
Rac2	GGTase I	CS LL
RALA	FTase	CCIL
RAP1B	GGTase I	CQLL
RHEB	FTase	CSVM
RHOA	GGTase I	CLVL
RHOB	GGTase I	CCKVL
RHOG	GGTase I	CILL
RRAS	FTase	CVIF
RRAS2	FTase	CVIF
YKT6	FTase	CAIM

Gng2	GGTase I	CAIL
PALM	FTase	CCSIM
PSIP1	FTase	CNLQ
RAB22	GGTase II	RSCC
RAP1A	GGTase I	CLLL
CNP	GGTase I	CTII
KRAS	FTase	CIIM
NRAS	FTase	CVVM
PTP4A1	FTase	CCIQ
RAB1C	GGTase II	GGCC
RAP2B	GGTase I	CVIL
GBP2	GGTase I	CNIL
RAB8A	GGTase II	CVLL
Dpcd	FTase?	Prickle1
Mien1	GGTase I	CVIL
NAP1L4	FTase	CKQQ
RAB10	GGTase II	SKCC
RAB21	GGTase II	CCSSG
RAB24	GGTase II	CCHH
RAB28	FTase?	CAVQ
RAB2A	GGTase II	GGCC
RAB33B	GGTase II	TCWC
RAB35	GGTase II	KRCC
RAB3A	GGTase II	DCAC
RAB3D	GGTase II	SCSC
RAB6A	GGTase II	GCSC
RAB11A	GGTase II	CCQNL
INF2	FTase	CVIQ
Lrrfip1	FTase?	CTMS
Phkb	FTase	CLVS
Rab12	GGTase II	VRCC
Rab30	GGTase II	CCNFN
RALB	GGTase I	CCLL
Rhobtb3	GGTase I	CLVM
Rhoc	GGTase I	CPIL
Rhoj	FTase	CAII
Rnd3	FTase	CTVM
ZC3HAV1	FTase	CVIS
ALDH9A1	SELF	
Cenpe	FTase	CKTQ
DNAJA4	FTase	CQTA

Fam63a	GGTase I	CVLL
Fam63b	GGTase I	CVI L
Fbxl20	GGTase III?	CCIIL
INPP5A	FTase	CCV VQ
MRAS	GGTase I	CVIL
Rab27b	GGTase II	KCAC
RAB3C	GGTase II	NCAC
RAP2A	GGTase I	CVIL
Rpgr	GGTase I	CTIL
STON1	FTase	CITQ
Ston2	FTase?	CGVQ
UBL3	GGTase I	CVIL
CEP85	FTase	CVTQ
Ptp4a2	FTase	CCVQ
Eras	GGTase I	CVIM
Palm3	FTase	CAVM
RAB22A	GGTase II	RSCC
RAB32	GGTase II	SQCC
RAB8B	GGTase II	CVLL
C16orf45	FTase	CNIM
CENPF	FTase	CKVQ
DCAF8	FTase	CMPS
DNAJB2	GGTase I	CLIL
EHBP1	FTase	Dnaja2
EHBP1L1	FTase	Uchl1
FBXL2	GGTase I	CVIL
GBP1	GGTase I	CTIS
GNG10	GGTase I	CALL
LMNA	FTase	CSIM
LMNB2	FTase	CYVM
MAPKAPK3	FTase	CN NQ
RAB27A	GGTase II	ACGC
Rab29	GGTase II	WSCC
RAB39A	GGTase II	ECFC
RAB9B	GGTase II	SSCC
RAP2C	GGTase I	CVIL
RHOF	GGTase I	CLLL
SPDL1	FTase	CPQQ
ULK3	FTase	CTLQ
WDTC1	FTase	CRPS
4930544G11Rik	GGTase I	CFVF

ALDH3B1	GGTase I	CTLL
ALDH3B2	GGTase I	CTLL
CAMK1G	unknown	CLIM
CLN3	FTase	CQLS
CPLX3	FTase	CHVM
CPLX4	FTase	CSVM
CRACR2A	GGTase I	SCCG
DHX32	FTase	C TLQ
DIRAS1	GGTase I	CTLM
DIRAS2	GGTase I	CVIM
DIRAS3	GGTase I	CIIM
GBP5	GGTase I	CVLL
GNG11	FTase	CVIS
GNG13	GGTase I	CTIL
GNG3	GGTase I	CALL
GNG4	GGTase I	CTIL
GNG7	GGTase I	CIIL
GNG8	GGTase I	CVLL
GNGT1	FTase	CVIS
GNGT2	FTase	CLIS
GRK1	FTase	CLVS
GRK7	GGTase I	CLLL
HRAS	FTase	CVLS
INPP5B	FTase	CNPL
INPP5E	FTase	CSVS
PALM2	FTase	CVVM
PDE6A	FTase	CCIQ
PDE6B	GGTase I	CCIL
PDE6C	FTase	CCIQ
PHKA1	FTase	CAMQ
PHKA2	FTase	CQMQ
PLA2G4C	FTase	CCL A
PPP1R16A	FTase	CLLM
PPP1R16B	FTase	CRIS
PPP1R32	MAY	CVAHS
PRICKLE1	FTase	CIIS
PRICKLE2	FTase	CIIS
PTGIR	FTase	CSLC
PTP4A3	FTase	CCVM
RAB17	GGTase II	CCAH
RAB19	GGTase II	HCTC

RAB20	GGTase II	GCCA
RAB25	GGTase II	CCISL
RAB26	GGTase II	CCRP
RAB33A	GGTase II	SCPC
RAB34	GGTase II	TCCP
RAB36	GGTase II	LGCC
RAB37	GGTase II	CCSFM
Rab38	GGTase II	CAKS
RAB39B	GGTase II	RCLC
RAB40A	GGTase I	CKIS
RAB40AL	GGTase I	CKIS
RAB40B	GGTase I	CKIS
RAB40C	GGTase I	CKIS
RAB41	GGTase II	RSYC
RAB42	GGTase II	PCQC
Rab43	GGTase II	GCGC
Rab44	GGTase II	GCCH
RAB4B	GGTase II	PCGC
Rab6c	NO	PVSWR
RAB7B	GGTase II	SRCC
RAB9A	GGTase II	SSCC
Rac1	GGTase I	CLLL
RAC3	GGTase I	CTVF
RAP1blike	GGTase I	CQLL
RASD1	FTase	CVIS
RASD2	FTase	CTIQ
RASL10A	FTase	CSLM
RASL10B	GGTase I	CAIM
RHEBL1	FTase	CHLM
RHOD	GGTase I	CVVT
RHOH	GGTase I	CKIF
RHOQ	FTase	CLIT
RND1	GGTase I	CSIM
RND2	GGTase I	CNLM
RTL8C	FTase	CVLA
STK11	FTase	CKQQ
UCHL1	FTase	CKAA
USP32	FTase	CVLQ
ZFAND2B	GGTase I	CSLC

Table A1. CaaX sequences discovered in proteomics experiments, along with associated gene and enzyme they are a substrate for.

Library	Peptide Sequence	Expected Mass	Observed Mass
Ca_1IIM 1	DsGRAGCGIIM	1110.4	1110.2
Ca_1IIM 1	DsGRAGCSIIM	1140.4	1140.2
Ca_1IIM 1	DsGRAGC(fn)SIIM	1344.8	1344.4
Ca_1IIM 1	DsGRAGCVIIM	1152.5	1152.3
Ca_1IIM 1	DsGRAGCCIIM	1156.5	1156.2
Ca_1IIM 1	DsGRAGC(fn)CIIM	1360.8	1360.4
Ca_1IIM 1	DsGRAGCIIIM	1166.5	1166.3
Ca_1IIM 1	DsGRAGCNIIM	1167.4	1167.3
Ca_1IIM 1	DsGRAGC(fn)NIIM	1371.8	1371.4
Ca_1IIM 1	DsGRAGCKIIM	1181.5	1181.3
Ca_1IIM 1	DsGRAGCMIIM	1184.5	1184.2
Ca_1IIM 1	DsGRAGC(fn)MIIM	1388.9	1388.4
Ca_1IIM 1	DsGRAGCFIIM	1200.5	1200.3
Ca_1IIM 1	DsGRAGC(fn)FIIM	1404.9	1404.4
Ca_1IIM 1	DsGRAGCYIIM	1216.5	1216.2
Ca_1IIM 1	DsGRAGC(fn)YIIM	1420.9	1420.4
Ca_1IIM 2	DsGRAGCAIIM	1124.4	1124.3
Ca_1IIM 2	DsGRAGC(fn)AIIM	1328.8	1328.4
Ca_1IIM 2	DsGRAGCPIIM	1150.4	1150.3
Ca_1IIM 2	DsGRAGC(fn)PIIM	1354.8	1354.4
Ca_1IIM 2	DsGRAGCTIIM	1154.4	1154.2
Ca_1IIM 2	DsGRAGCLIIM	1166.5	1166.3
Ca_1IIM 2	DsGRAGCDIIM	1168.4	1168.2
Ca_1IIM 2	DsGRAGCQIIM	1181.5	1181.2
Ca_1IIM 2	DsGRAGC(fn)QIIM	1385.8	1385.4
Ca_1IIM 2	DsGRAGCEIIM	1182.4	1182.3
Ca_1IIM 2	DsGRAGC(fn)EIIM	1386.8	1386.4
Ca_1IIM 2	DsGRAGCHIIM	1190.5	1190.3
Ca_1IIM 2	DsGRAGC(fn)HIIM	1394.8	1394.4
Ca_1IIM 2	DsGRAGCRIIM	1209.5	1209.3
Ca_1IIM 2	DsGRAGCWIIM	1239.5	1239.3
CMa_2IM 1	DsGRAGCMGIM	1128.4	1128.1
CMa_2IM 1	DsGRAGC(fn)MGIM	1332.8	1332.2
CMa_2IM 1	DsGRAGCMSIM	1158.4	1158.1
CMa_2IM 1	DsGRAGC(fn)MSIM	1362.8	1362.3
CMa_2IM 1	DsGRAGCMVIM	1170.5	1170.1
CMa_2IM 1	DsGRAGCMCIM	1174.5	1174.1
CMa_2IM 1	DsGRAGCMIIM	1184.5	1184.2
CMa_2IM 1	DsGRAGCMNIM	1185.5	1185.1

CMa_2IM 1	DsGRAGC(fn)MNIM	1389.2	1389.3
CMa_2IM 1	DsGRAGCMKIM	1199.5	1199.2
CMa_2IM 1	DsGRAGC(fn)MKIM	1403.9	1303.4
CMa_2IM 1	DsGRAGCMMIM	1202.6	1202.1
CMa_2IM 1	DsGRAGCMFIM	1218.5	1218.1
CMa_2IM 1	DsGRAGCMYIM	1234.5	1234.1
CMa_2IM 2	DsGRAGCMAIM	1142.4	1142.1
CMa_2IM 2	DsGRAGCMPIM	1168.5	1168.1
CMa_2IM 2	DsGRAGCMTIM	1172.5	1172.1
CMa_2IM 2	DsGRAGCMLIM	1184.5	1184.1
CMa_2IM 2	DsGRAGCMDIM	1186.5	1186.1
CMa_2IM 2	DsGRAGCMQIM	1199.5	1199.1
CMa_2IM 2	DsGRAGC(fn)MQIM	1403.9	1403.3
CMa_2IM 2	DsGRAGCMEIM	1200.5	1200.1
CMa_2IM 2	DsGRAGC(fn)MEIM	1404.8	1404.3
CMa_2IM 2	DsGRAGCMHIM	1208.5	1208.1
CMa_2IM 2	DsGRAGC(fn)MHIM	1412.9	1412.3
CMa_2IM 2	DsGRAGCMRIM	1227.6	1227.1
CMa_2IM 2	DsGRAGC(fn)MRIM	1431.9	1431.3
CMa_2IM 2	DsGRAGCMWIM	1257.6	1257.1
$CMIa_3M$ 1	DsGRAGCMIGM	1128.4	1128.2
$CMIa_3M$ 1	DsGRAGC(fn)MIGM	1332.8	1332.3
$CMIa_3M$ 1	DsGRAGCMISM	1158.4	1158.2
$CMIa_3M$ 1	DsGRAGC(fn)MISM	1362.8	1362.3
$CMIa_3M$ 1	DsGRAGCMIVM	1170.5	1170.3
$CMIa_3M$ 1	DsGRAGCMICM	1174.5	1174.2
$CMIa_3M$ 1	DsGRAGC(fn)MICM	1378.9	1378.3
$CMIa_3M$ 1	DsGRAGCMIIM	1184.5	1184.3
$CMIa_3M$ 1	DsGRAGCMINM	1185.5	1185.2
$CMIa_3M$ 1	DsGRAGC(fn)MINM	1389.2	1389.3
$CMIa_3M$ 1	DsGRAGCMIKM	1199.5	1199.3
$CMIa_3M$ 1	DsGRAGCMIMM	1202.6	1202.2
$CMIa_3M$ 1	DsGRAGC(fn)MIMM	1406.9	1406.3
$CMIa_3M$ 1	DsGRAGCMIFM	1218.5	1218.2
$CMIa_3M$ 1	DsGRAGC(fn)MIFM	1422.9	1422.3
$CMIa_3M$ 1	DsGRAGCMIYM	1234.5	1234.2
$CMIa_3M$ 2	DsGRAGCMIAM	1142.4	1142.3
$CMIa_3M$ 2	DsGRAGC(fn)MIAM	1346.8	1346.4
$CMIa_3M$ 2	DsGRAGCMIPM	1168.5	1168.3
$CMIa_3M$ 2	DsGRAGCMITM	1172.5	1172.4
$CMIa_3M$ 2	DsGRAGC(fn)MITM	1376.8	1376.5
$CMIa_3M$ 2	DsGRAGCMILM	1184.5	1184.3
$CMIa_3M$ 2	DsGRAGC(fn)MILM	1388.9	1388.5
$CMIa_3M$ 2	DsGRAGCMIDM	1186.5	1186.4
$CMIa_3M$ 2	DsGRAGCMIQM	1199.5	1199.4
$CMIa_3M$ 2	DsGRAGC(fn)MIQM	1403.9	1403.5
$CMIa_3M$ 2	DsGRAGCMIEM	1200.5	1200.3

CMIa$_3$M 2	DsGRAGC(fn)MIEM	1404.8	1404.5
CMIa$_3$M 2	DsGRAGCMIHM	1208.5	1208.4
CMIa$_3$M 2	DsGRAGC(fn)MIHM	1412.9	1412.5
CMIa$_3$M 2	DsGRAGCMIRM	1227.6	1227.4
CMIa$_3$M 2	DsGRAGCMIWM	1257.6	1257.4
CMIIX 1	DsGRAGCMIIG	1110.4	1110.3
CMIIX 1	DsGRAGCMIIS	1140.4	1140.3
CMIIX 1	DsGRAGC(fn)MIIS	1344.8	1344.4
CMIIX 1	DsGRAGCMIIV	1152.5	1152.3
CMIIX 1	DsGRAGCMIIC	1156.5	1156.3
CMIIX 1	DsGRAGC(fn)MIIC	1360.8	1360.4
CMIIX 1	DsGRAGCMIII	1166.5	1166.3
CMIIX 1	DsGRAGCMIIN	1167.4	1167.2
CMIIX 1	DsGRAGCMIIK	1181.5	1181.3
CMIIX 1	DsGRAGC(fn)MIIK	1385.9	1385.5
CMIIX 2	DsGRAGCMIIA	1124.4	1124.2
CMIIX 2	DsGRAGC(fn)MIIA	1328.8	1328.4
CMIIX 2	DsGRAGCMIIP	1150.4	1150.3
CMIIX 2	DsGRAGCMIIT	1154.4	1154.2
CMIIX 2	DsGRAGCMIIL	1166.5	1166.3
CMIIX 2	DsGRAGCMIID	1168.4	1168.2
CMIIX 2	DsGRAGCMIIQ	1181.5	1181.3
CMIIX 2	DsGRAGC(fn)MIIQ	1385.8	1385.4
CMIIX 2	DsGRAGCMIIM	1184.5	1184.2
CMIIX 2	DsGRAGC(fn)MIIM	1388.9	1388.4
CMIIX 2	DsGRAGCMIIR	1209.5	1209.3
CMIIX 2	DsGRAGCMIIW	1239.5	1239.3

Table A2. Mass of CaaaX library peptides. Expected and observed mass for all pentapeptide starting material from MALDI libraries, as well as the prenylated products.

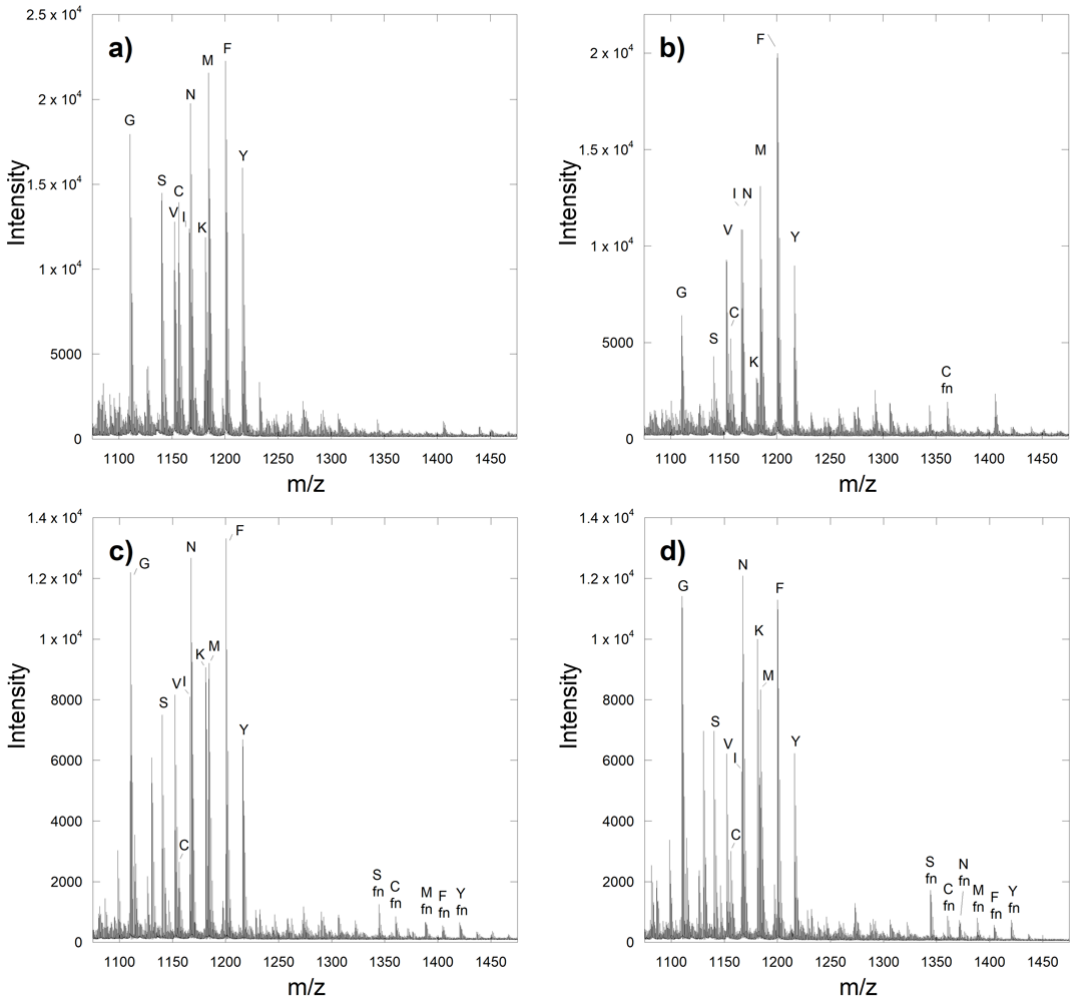


Figure A1. Optimization of enzyme concentration for CaaaX MALDI. DsGRAGCMIIX library 1 reacted with (a) no enzyme (b) 100 nM enzyme (c) 1 uM enzyme (d) 10 uM enzyme

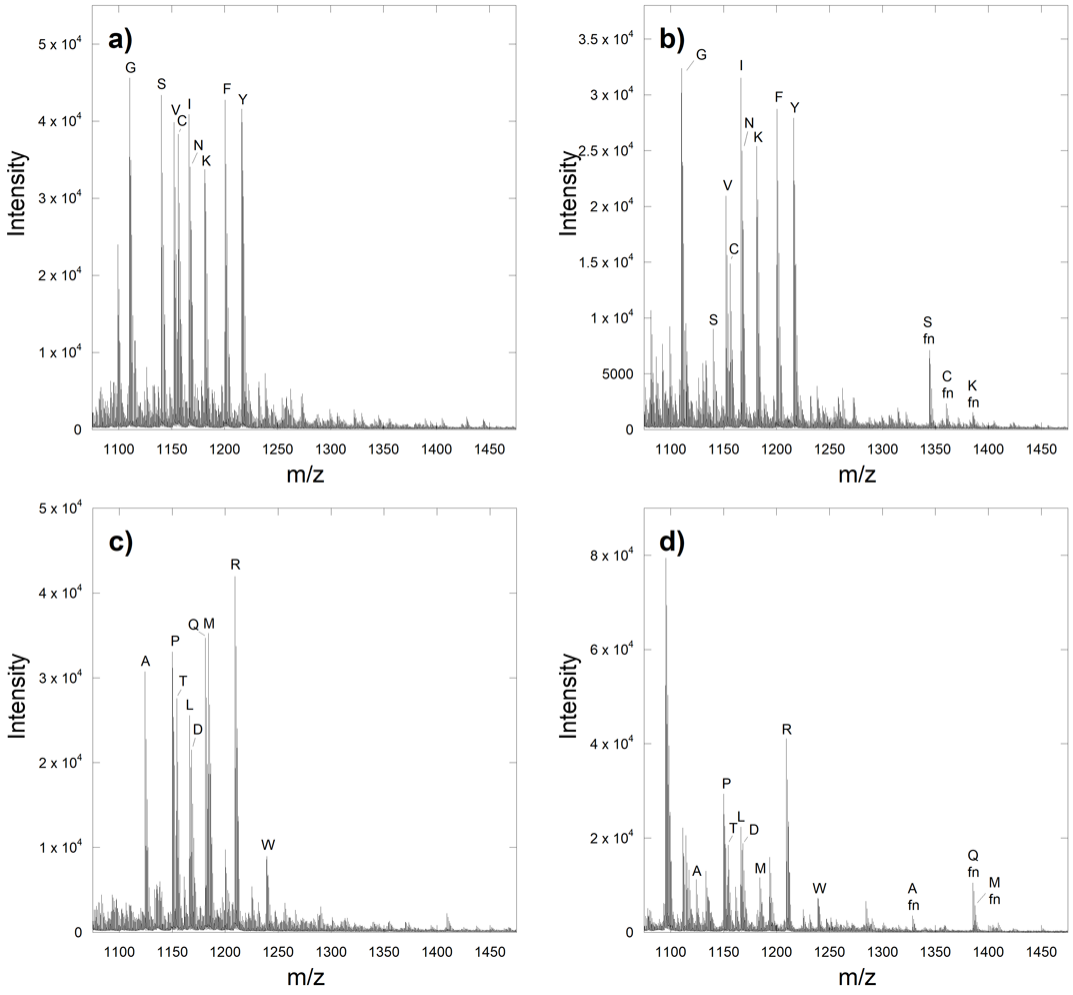


Figure A2. MALDI/MS of DsGRAGCMIIX libraries. Library 1 (a) before and (b) after prenylation with 1 uM yFTase. Library 2 (c) before and (d) after prenylation with 1 uM yFTase.

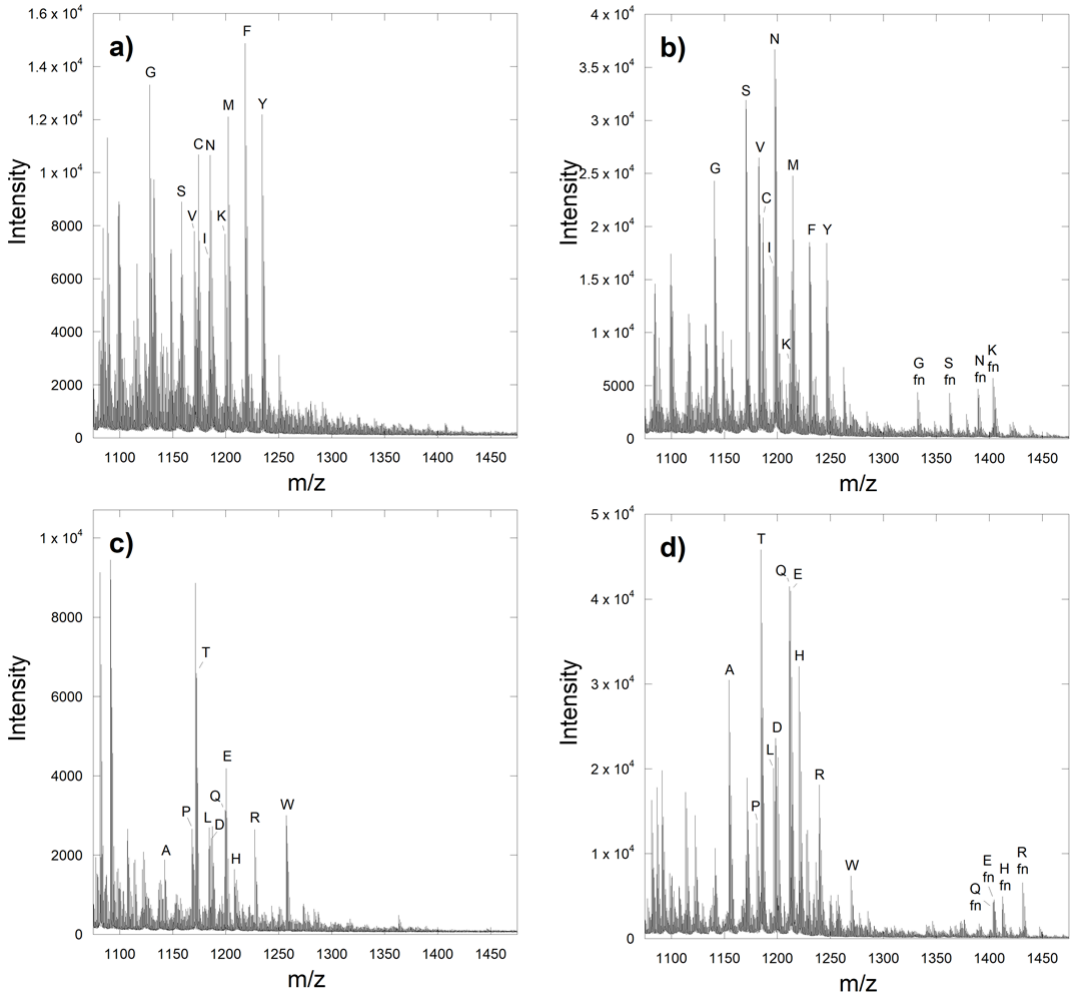


Figure A3. MALDI/MS of DsGRAGCMa$_2$IM libraries. Library 1 (a) before and (b) after prenylation with 1 uM yFTase. Library 2 (c) before and (d) after prenylation with 1 uM yFTase.

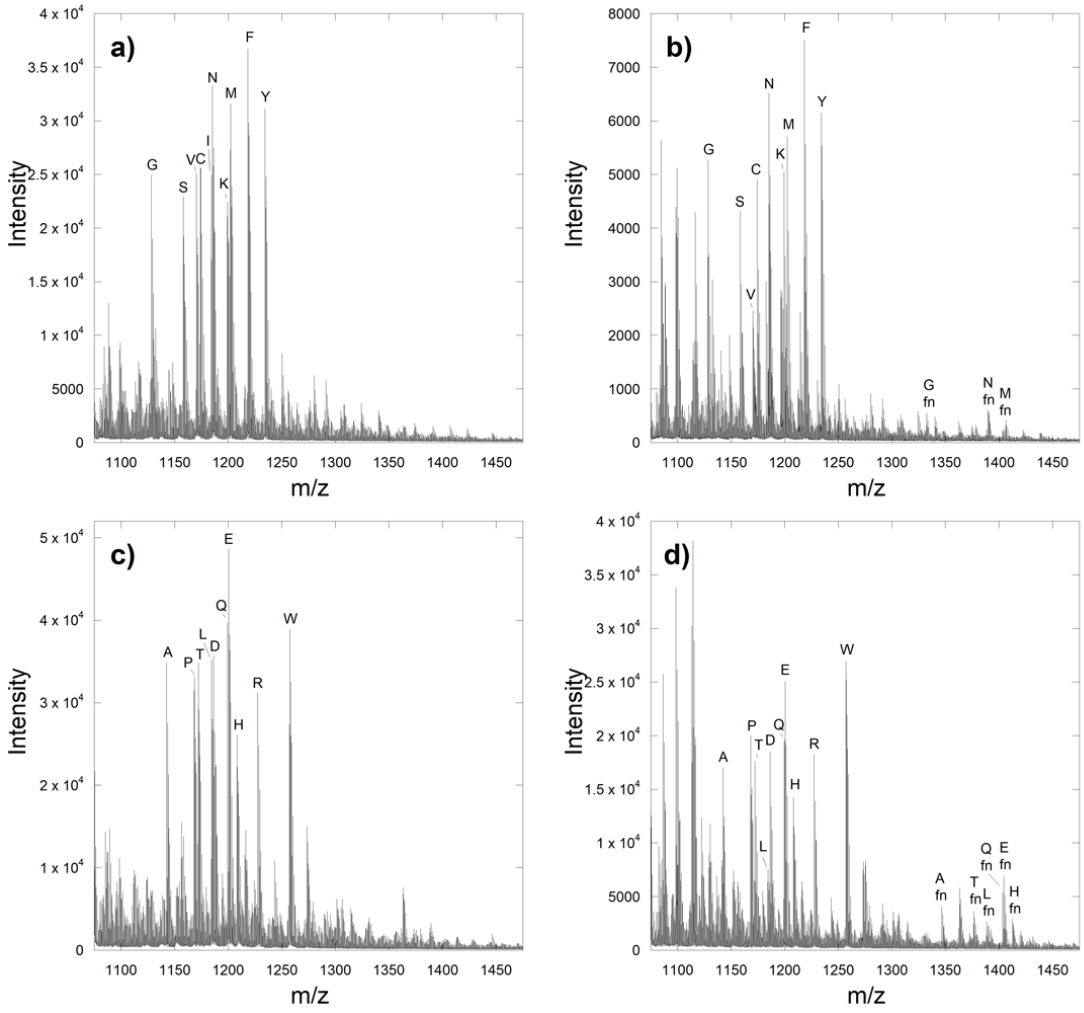


Figure A4. MALDI/MS of DsGRAGCMIa$_3$M libraries. Library 1 (a) before and (b) after prenylation with 1 uM yFTase. Library 2 (c) before and (d) after prenylation with 1 uM yFTase.

Sequence	Formula	+/- Fn	ion	Expected mass	Observed mass
DsGRAGCMIIS	$C_{63}H_{103}N_{13}O_{13}S_3$2+	+	M+2H	672.85	672.85
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43
	$C_{48}H_{73}N_{10}O_8S_3$+	+	b6	1013.48	1013.48
	$C_{54}H_{84}N_{11}O_9S_3$+	+	b7	1126.56	1126.56
	$C_{48}H_{78}N_{13}O_{13}S_3$+	-	M+H	1140.50	1140.50
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{33}H_{49}N_{10}O_8S_3$+	-	b6	809.29	809.29
DsGRAGCMIIQ	$C_{65}H_{106}N_{14}O_{13}S_3$2+	+	M+2H	693.36	693.36
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43
	$C_{48}H_{73}N_{10}O_8S_3$+	+	b6	1013.48	1013.47
	$C_{54}H_{84}N_{11}O_9S_3$+	+	b7	1126.56	1126.56
	$C_{60}H_{95}N_{12}O_{10}S_3$+	+	b8	1239.65	1239.64
	$C_{50}H_{81}N_{14}O_{13}S_3$+	-	M+H	1181.53	1181.53
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{33}H_{49}N_{10}O_8S_3$+	-	b6	809.29	809.29
DsGRAGCSIIM	$C_{63}H_{103}N_{13}O_{13}S_3$2+	+	M+2H	672.85	672.85
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43
	$C_{46}H_{69}N_{10}O_9S_2$+	+	b6	969.47	969.46
	$C_{52}H_{80}N_{11}O_{10}S_2$+	+	b7	1082.55	1082.54
	$C_{48}H_{78}N_{14}O_{14}S_3$+	-	M+H	1140.50	1140.50
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{31}H_{45}N_{10}O_9S_2$+	-	b6	765.28	765.28

DsGRAGCMKIM	$C_{65}H_{108}N_{14}O_{12}S_4$+	+	M+2H	702.36	702.36
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43
	$C_{48}H_{73}N_{10}O_8S_3$+	+	b6	1013.48	1013.48
	$C_{50}H_{83}N_{14}O_{12}S_4$+	-	M+H	1199.53	1199.53
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{33}H_{49}N_{10}O_8S_3$+	-	b6	809.29	809.29
	$C_{39}H_{61}N_{12}O_9S_3$+	-	b7	937.38	937.39
	$C_{45}H_{72}N_{13}O_{10}S_3$+	-	b8	1050.47	1050.47
DsGRAGCYIIM	$C_{69}H_{107}N_{13}O_{13}S_3$2+	+	M+2H	710.86	710.86
	$C_{23}H_{32}N_7O_5S$+	+	b3	518.22	518.22
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.44
	$C_{52}H_{73}N_{10}O_9S_2$+	+	b6	1045.50	1045.50
	$C_{58}H_{84}N_{11}O_{10}S_2$+	+	b7	1158.58	1158.58
	$C_{23}H_{33}N_7O_5S$+	-	b3	518.22	518.23
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{37}H_{49}N_{10}O_9S_2$+	-	b6	841.31	841.31
DsGRAGCHIIM	$C_{66}H_{105}N_{15}O_{12}S_3$2+	+	M+2H	697.86	697.86
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.24
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43
	$C_{49}H_{71}N_{12}O_8S_2$+	+	b6	1019.50	1019.49
	$C_{51}H_{80}N_{15}O_{12}S_3$+	-	M+H	1190.53	1190.53
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.24
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{34}H_{47}N_{12}O_8S_2$+	-	b6	815.31	815.32
DsGRAGCMIGM	$C_{61}H_{99}N_{13}O_{12}S_4$2+	+	M+2H	666.82	666.82
	$C_{25}H_{35}N_8O_6S$+	+	b4	575.24	575.23
	$C_{43}H_{64}N_9O_7S_2$+	+	b5	882.44	882.43

	$C_{48}H_{78}N_{10}O_8S_3$+	+	b6	1013.48	1013.48
	$C_{54}H_{84}N_{11}O_9S_3$+	+	b7	1026.56	1026.58
	$C_{56}H_{87}N_{12}O_{10}S_3$+	+	b8	1183.58	1183.58
	$C_{46}H_{74}N_{13}O_{12}S_4$+	-	M+H	1128.45	1128.45
	$C_{25}H_{35}N_8O_6S$+	-	b4	575.24	575.23
	$C_{28}H_{40}N_9O_7S_2$+	-	b5	678.25	678.25
	$C_{33}H_{49}N_{10}O_8S_3$+	-	b6	809.29	809.28

Table A3. MS/MS of prenylated CaaaX library hits. MS/MS ions of prenylated peptides, indicating the Cys is prenylated. b ions for the prenylated peptide are shown, as well as b ions that have also lost the prenyl group.

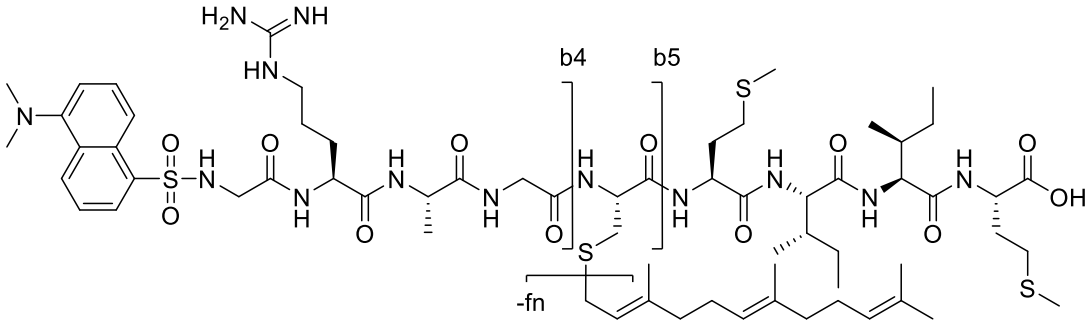


Figure A5. Structure of representative DsGRAGC(fn)MIIM peptide. The major ions used to confirm S-farnesylation are labeled

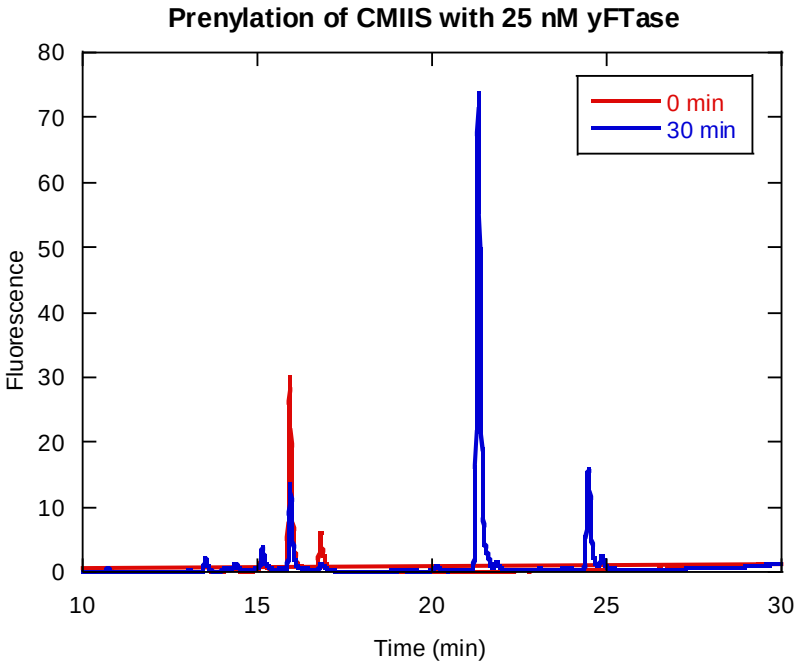


Figure A6. HPLC Fluorescence assay of CMIIS. HPLC assay quantifying the farnesylation of the peptide DsGRAGCMIIS by the fluorescence of the Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase

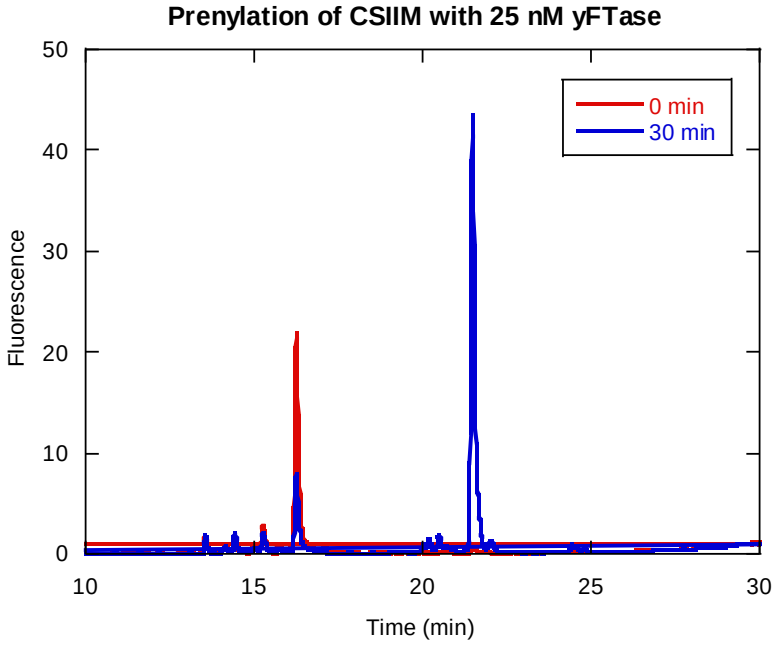


Figure A7. HPLC Fluorescence assay of CSIIM**.** HPLC assay quantifying the farnesylation of the peptide DsGRAGCSIIM by the fluorescence of th Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase.

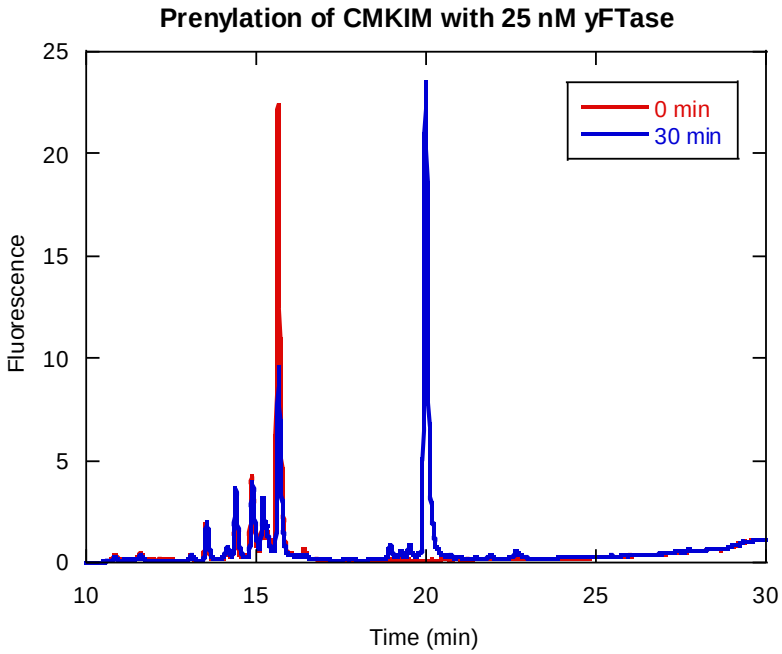


Figure A8. HPLC Fluorescence assay of CMKIM**.** HPLC assay quantifying the conversion the peptide DsGRAGCMKIM by the fluorescence of the Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase.

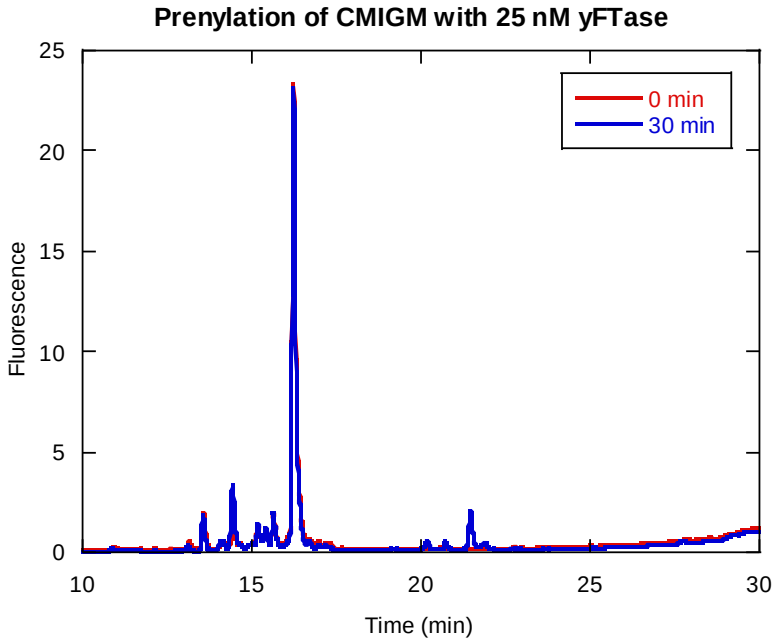


Figure A9. HPLC Fluorescence assay of CMGIM. HPLC assay quantifying the conversion the peptide DsGRAGCMIGM by the fluorescence of the Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase.

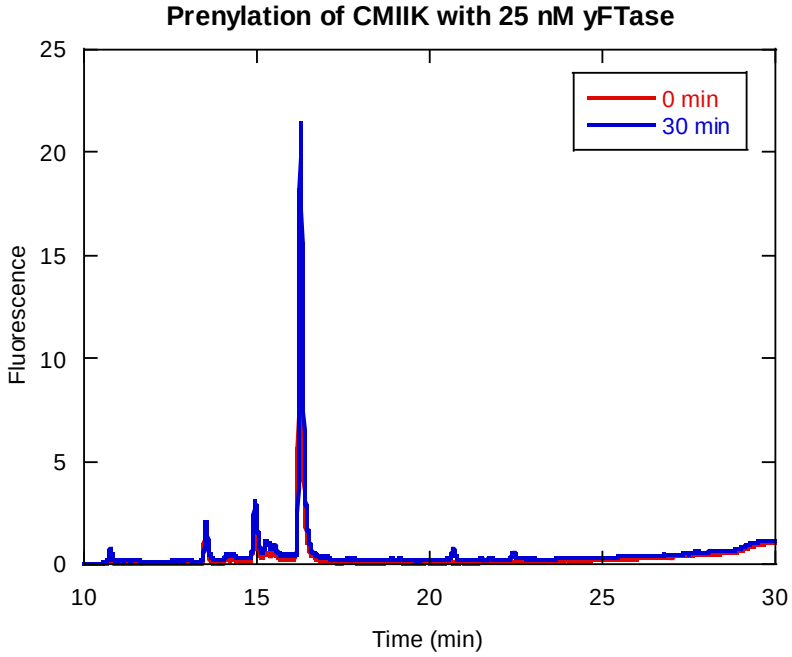


Figure A10. HPLC Fluorescence assay of CMIIK. HPLC assay quantifying the conversion the peptide DsGRAGCMIIK by the fluorescence of Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase.

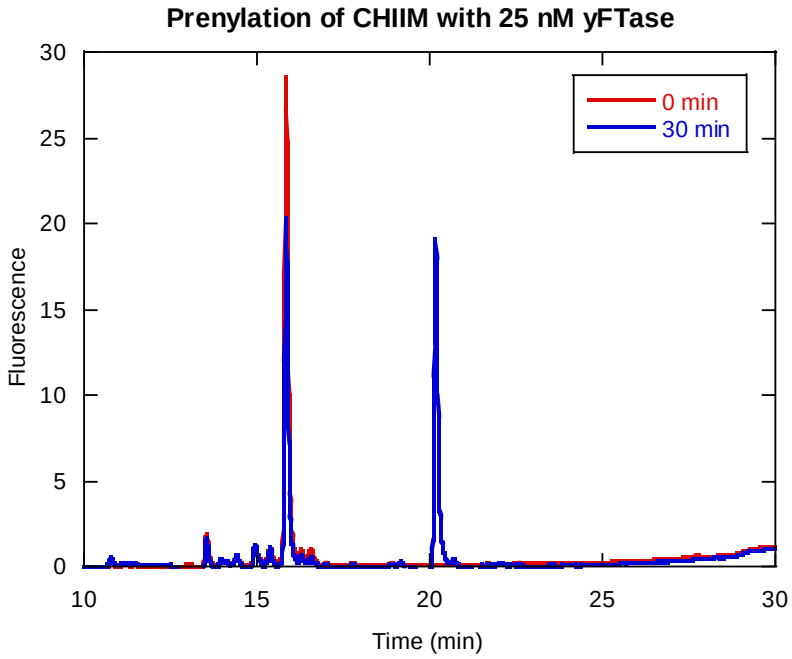


Figure A11. HPLC Fluorescence assay of CHIIM. HPLC assay quantifying the conversion the peptide DsGRAGCHIIM by the fluorescence of Dansylglycine (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase

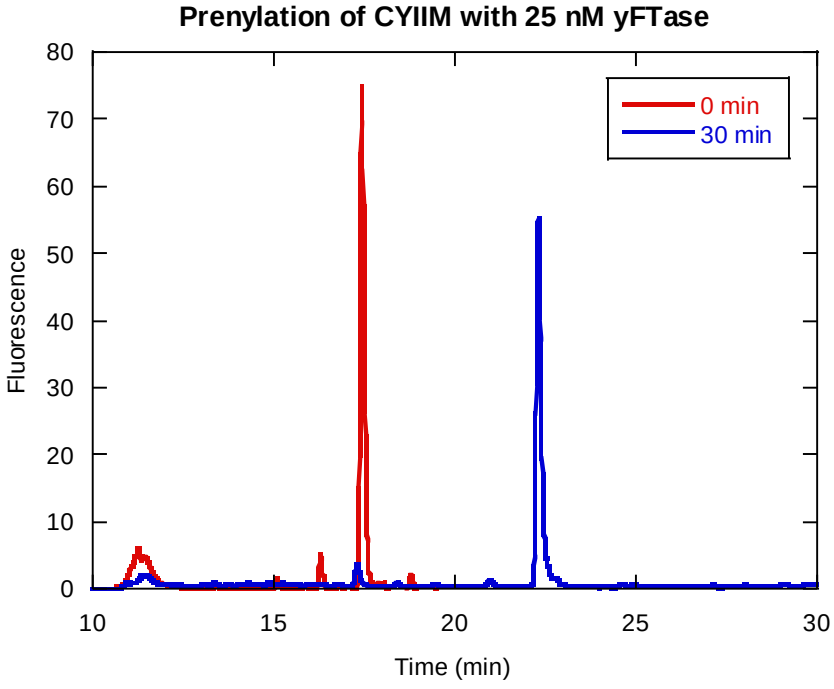


Figure A12. HPLC Fluorescence assay of CYIIM. HPLC assay quantifying the conversion the peptide DsGRAGCYIIM by the fluorescence of Dansyl group (ex. 220/em. 495). Reaction of 2.4 uM peptide with 25 nM yFTase.

Sequence	Protein name	Uniprot ID
CSQAS	Xaa-Pro aminopeptidase 3	https://www.uniprot.org/uniprot/Q9NQH7
CSLMQ	Transcription elongation factor A protein 3	https://www.uniprot.org/uniprot/O75764-2
CSTAN	Neuronal acetylcholine receptor subunit alpha-7	https://www.uniprot.org/uniprot/P36544-3
CASSQ	T cell receptor beta variable 14	https://www.uniprot.org/uniprot/A0A5B0
CMTSQ	Beta-1,4 N-acetylgalactosaminyltransferase 1	https://www.uniprot.org/uniprot/Q00973
CLTIQ	Acetyl-coenzyme A synthetase, cytoplasmic	https://www.uniprot.org/uniprot/Q9NR19
CVQTS	Thymosin beta-15A	https://www.uniprot.org/uniprot/P0CG34
CNVTS	Orphan sodium- and chloride-dependent neurotransmitter transporter NTT5	https://www.uniprot.org/uniprot/Q9GZN6
CASLS	Proline-rich protein 5-like	https://www.uniprot.org/uniprot/Q6MZQ0
CLISS	Regulator of G-protein signaling 7-binding protein	https://www.uniprot.org/uniprot/Q6MZT1
CQLNS	Dynein heavy chain 7	https://www.uniprot.org/uniprot/Q8WXX0
CTASS	Serine/threonine-protein kinase Nek7	https://www.uniprot.org/uniprot/Q8TDX7
CSKLN	NUAK family SNF1-like kinase 1	https://www.uniprot.org/uniprot/O60285
CSKLN	Olfactory receptor 4A8	https://www.uniprot.org/uniprot/P0C604
CSKVN	Thioredoxin domain-containing protein 16	https://www.uniprot.org/uniprot/Q9P2K2
CSLQQ	Uncharacterized protein C1orf109	https://www.uniprot.org/uniprot/Q9NX04-2
CSLLL	Arylsulfatase B	https://www.uniprot.org/uniprot/P15848-2
CSLFA	NF-kappa-B-activating kinase-associated protein 1	https://www.uniprot.org/uniprot/Q9H6S1-4
CSIFI	Nuclear receptor coactivator 7	https://www.uniprot.org/uniprot/Q8NI08-6
CSSAV	Upstream transcription factor 3	https://www.uniprot.org/uniprot/Q68DE3
CSNTF	Protein virilizer homolog	https://www.uniprot.org/uniprot/Q69YN4-4
CAFLS	Histone-lysine N-methyltransferase MECOM	https://www.uniprot.org/uniprot/Q03112-9
CFLSS	N-alpha-acetyltransferase 16, NatA auxiliary subunit	https://www.uniprot.org/uniprot/Q6N069-4
CLLFS	Nuclear mitotic apparatus protein 1	https://www.uniprot.org/uniprot/Q14980-3
CVSVS	Phosphatidylinositol 4-phosphate 5-kinase type-1 gamma	https://www.uniprot.org/uniprot/O60331-2
CQYNS	Pituitary homeobox 1	https://www.uniprot.org/uniprot/P78337
CLFLS	Regulation of nuclear pre-mRNA domain-containing protein 2	https://www.uniprot.org/uniprot/Q5VT52-4
CLACS	Slit homolog 3 protein	https://www.uniprot.org/uniprot/O75094

CASWQ	lamin subunit gamma 3	https://www.uniprot.org/uniprot/Q9Y6N6

Table A4. CaaaX sequences in the human genome. Selected CaaaX sequences in the human genome and their associated proteins. Peptides selected for evaluation are shaded in grey.

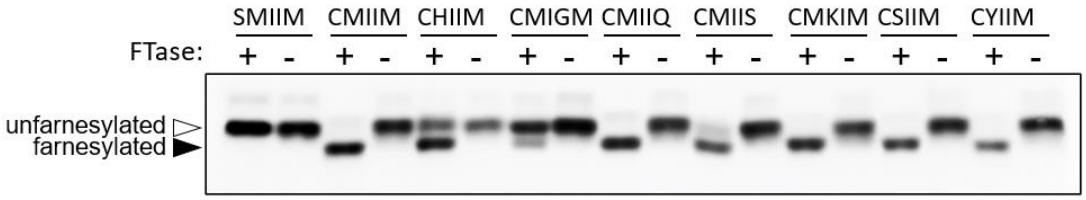


Figure A13. Mobility shift analysis of Ydj1p-CaaaX variants in the presence and absence of farnesyltransferase. Whole cell lysates were prepared and analyzed as described for Figure 5 except that each Ydj1p-CaaaX variant was expressed in two strains. One expresses FTase (yWS2544; *ydj1::KAN*^{R}) and the other lacks FTase (yWS2452; *ydj1::KAN*^{R} *ram1*Δ). The presence (+) or absence of FTase (-) correlates with the presence and absence, respectively, of the non-essential yeast gene RAM1 that encodes the beta subunit of farnesyltransferase.

Sample	Sequence	Biological Replicates	Technical Replicates	Average of Tech. Replicates[1]	Standard Deviation[2]
Controls	CASQ	4	5	100.0%	0.0%
	SASQ	4	5	0.0%	0.0%
	CMIIM	4	7	100.0%	0.0%
	SMIIM	4	9	0.0%	0.0%
Sequences from MALDI libraries	CHIIM	3	4	76.9%	25.9%
	CMIGM	3	5	39.5%	24.8%
	CMIIQ	2	4	100.0%	0.0%
	CMIIS	3	5	88.8%	10.9%
	CMKIM	3	5	99.4%	1.3%
	CSIIM	3	5	100.0%	0.0%
	CYIIM	3	5	100.0%	0.0%
Sequences from the mammalian genome	CASLS	1	3	2.1%	3.6%
	CASSQ	1	3	0.0%	0.0%
	CLACS	1	3	0.6%	1.0%
	CLLFS	1	3	16.4%	2.4%
	CMTSQ	1	3	0.7%	1.2%
	CQYNS	1	3	0.0%	0.0%
	CSKLN	1	3	13.0%	5.5%
	CSLMQ	1	4	96.7%	4.7%
	CSQAS	1	3	0.0%	0.0%

CVQTS	1	3	1.1%	2.0%

Table A5. Percent farnesylation for Ydj1p-CaaaX variants evaluated in this study.

[1]Values for band intensities of farnesylated and unfarnesylated species within gel lanes of Figures A12 and A13 were determined by the peak integration method using ImageJ [4]. reference; see below for options). These values were used to calculate percent farnesylation for individual samples that were averaged for multiple biological and/or technical replicates.

[2]Standard deviations were derived for all replicates using the Excel STDEV function.

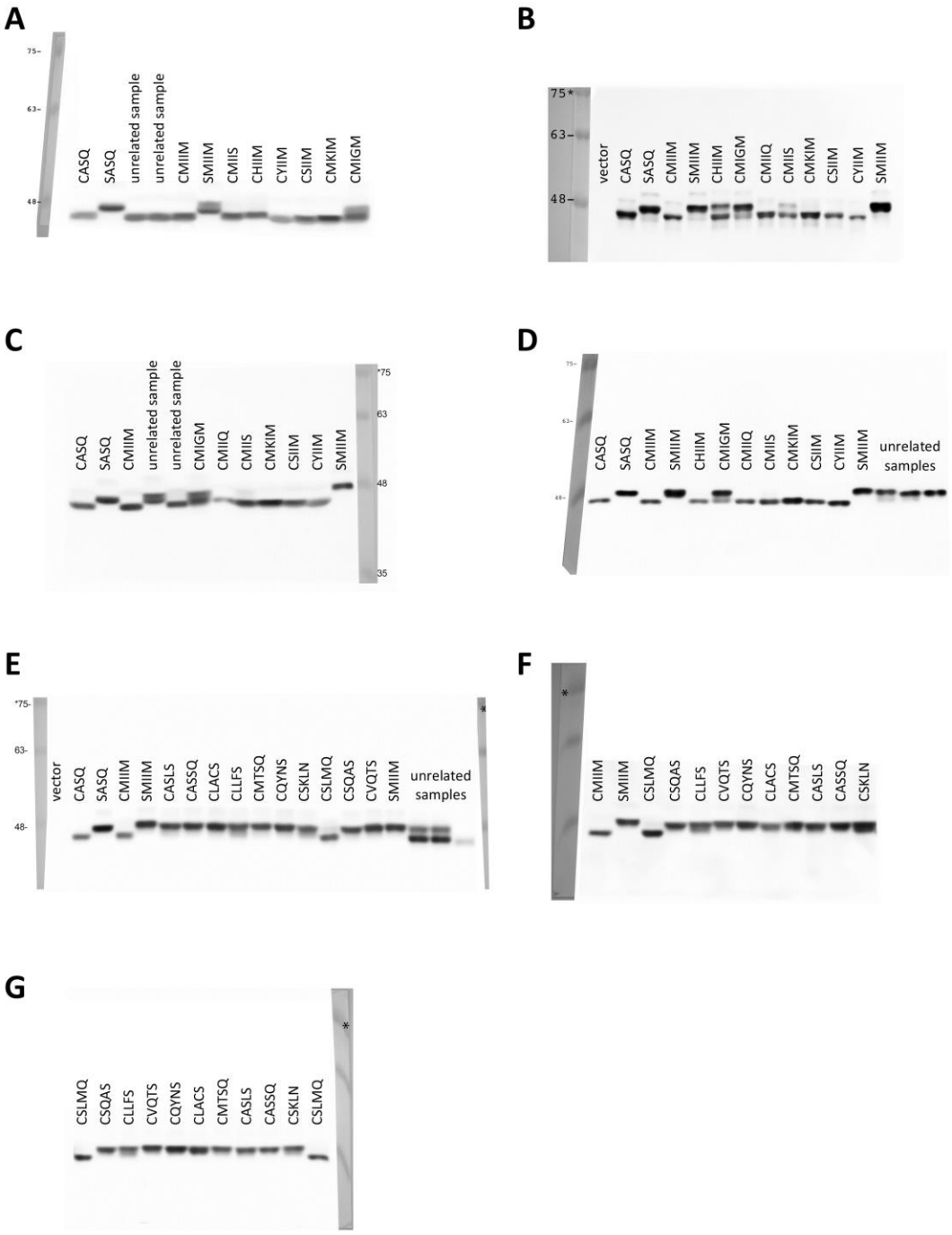


Figure A14. Mobility shift analysis of Ydj1p-CaaaX variants. A collage of anti-Ydj1p immunoblots representing the multiple biological and technical replicates of each Ydj1p-CaaaX variant that was analyzed in this study. Analysis was as described for Figures 5 and 6, with panels D and E representing uncropped versions of the blots used for Figures 5 and 6, respectively. These blots and the blot described for Figure A12 were used for determining the percent prenylation observed for each Ydj1p-CaaaX variant that is reported in Table A6. In each

panel, the digital layer containing the anti-Ydj1p chemiluminescence signal is juxtaposed with the layer containing the prestained molecular weight standards that was captured in parallel under ambient light.

Strain	Genotype	Reference
yWS2542	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ram1::KAN*^{R} *ydj1::NAT*^{R}	[1]
yWS2544	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ydj1::NAT*^{R}	[1]

Table A6. Yeast strains used in this study.

plasmid	genotype	Reference
pRS416	*CEN URA3* (vector)	[2]
pWS942	*CEN URA3 YDJ1*	[3]
pWS1132	*CEN URA3 YDJ1-SASQ*	[3]
pWS1488	*CEN URA3 YDJ1-CMIIM*	[4]
pWS1917	*CEN URA3 YDJ1-SMIIM*	this study
pWS1918	*CEN URA3 YDJ1-CMIIS*	this study
pWS1919	*CEN URA3 YDJ1-CMIIQ*	this study
pWS1920	*CEN URA3 YDJ1-CHIIM*	this study
pWS1921	*CEN URA3 YDJ1-CYIIM*	this study
pWS1922	*CEN URA3 YDJ1-CSIIM*	this study
pWS1923	*CEN URA3 YDJ1-CMKIM*	this study
pWS1924	*CEN URA3 YDJ1-CMIGM*	this study
pWS1981	*CEN URA3 YDJ1-CSLMQ*	this study
pWS1982	*CEN URA3 YDJ1-CSQAS*	this study
pWS1983	*CEN URA3 YDJ1-CLLFS*	this study
pWS1984	*CEN URA3 YDJ1-CVQTS*	this study
pWS1985	*CEN URA3 YDJ1-CQYNS*	this study
pWS1986	*CEN URA3 YDJ1-CLACS*	this study
pWS1987	*CEN URA3 YDJ1-CMTSQ*	this study
pWS1988	*CEN URA3 YDJ1-CASLS*	this study
pWS1989	*CEN URA3 YDJ1-CASSQ*	this study
pWS1990	*CEN URA3 YDJ1-CSKLN*	this study

Table A7. Plasmids used in this study.

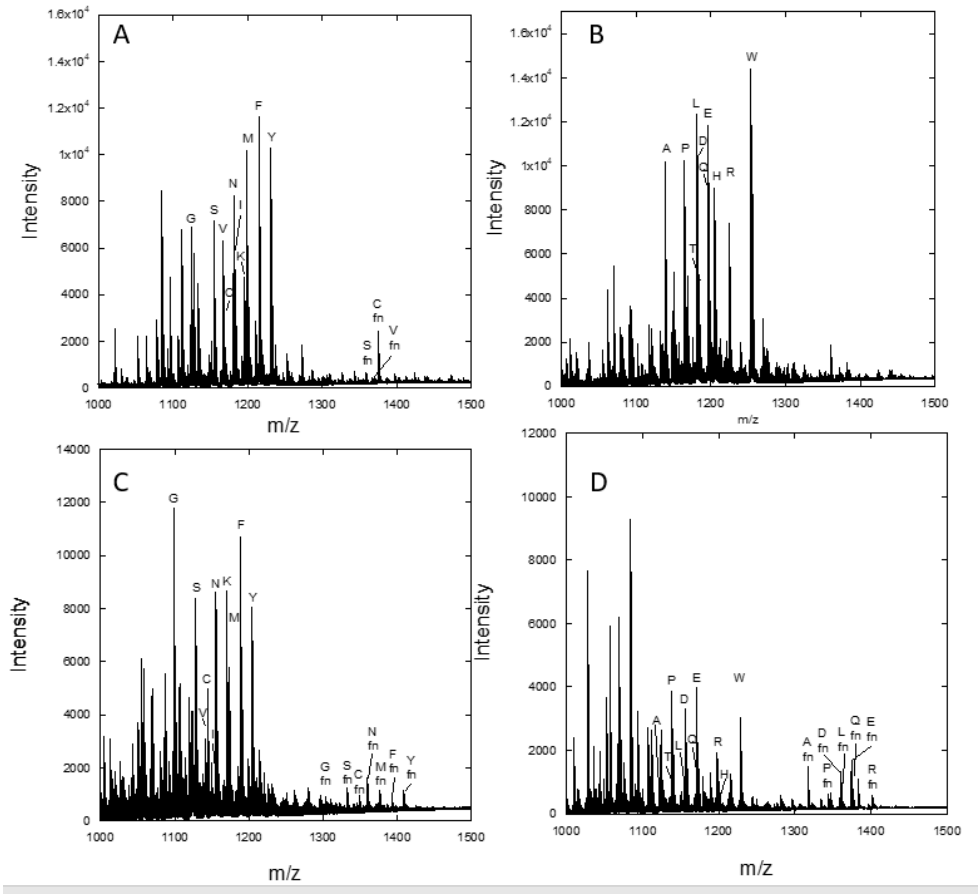


Figure A15 Prenylation of CSLMQ libraries with yFTase. A. CXLMQ 1, B. CXLMQ 2, C. CSXMQ 1, D. CSXMQ 2. Reactions performed with 2uM enzyme for eight hours.

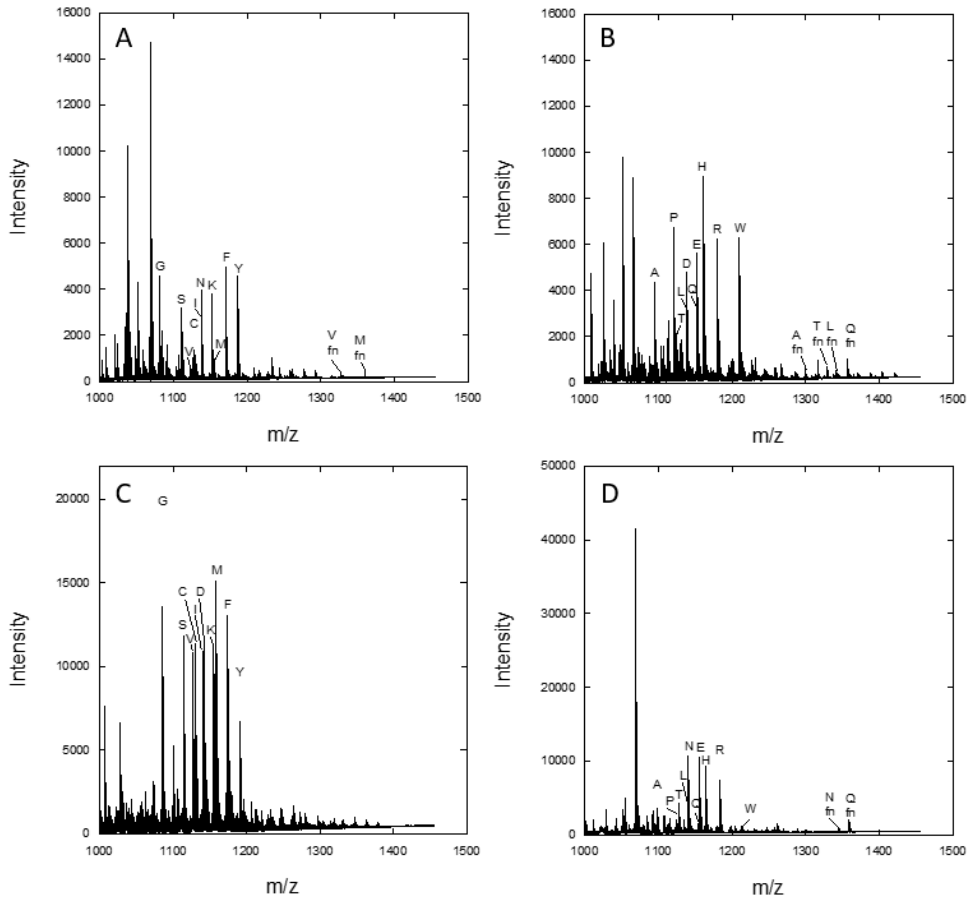


Figure A16 Prenylation of additional CSLMQ libraries with yFTase. A. CSLXQ 1, B. CSLXQ 2, C. CSLMX 1, D. CSLMX 2. Reactions performed with 2uM enzyme for eight hours.

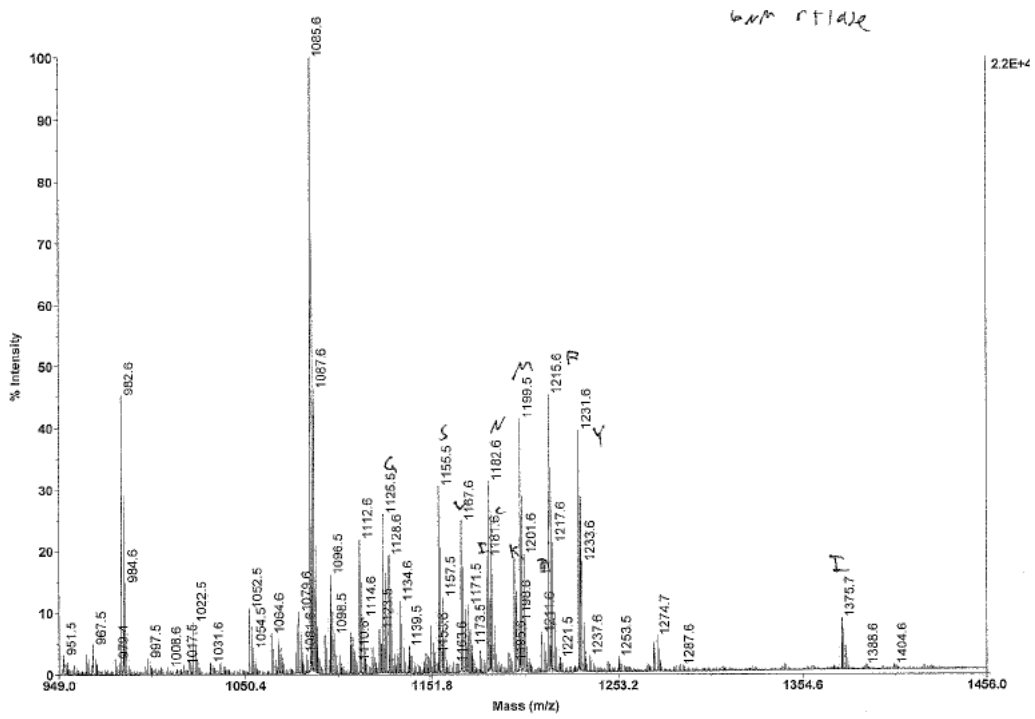


Figure A17. Prenylation of CXLMQ 1 with 3uM rFTase

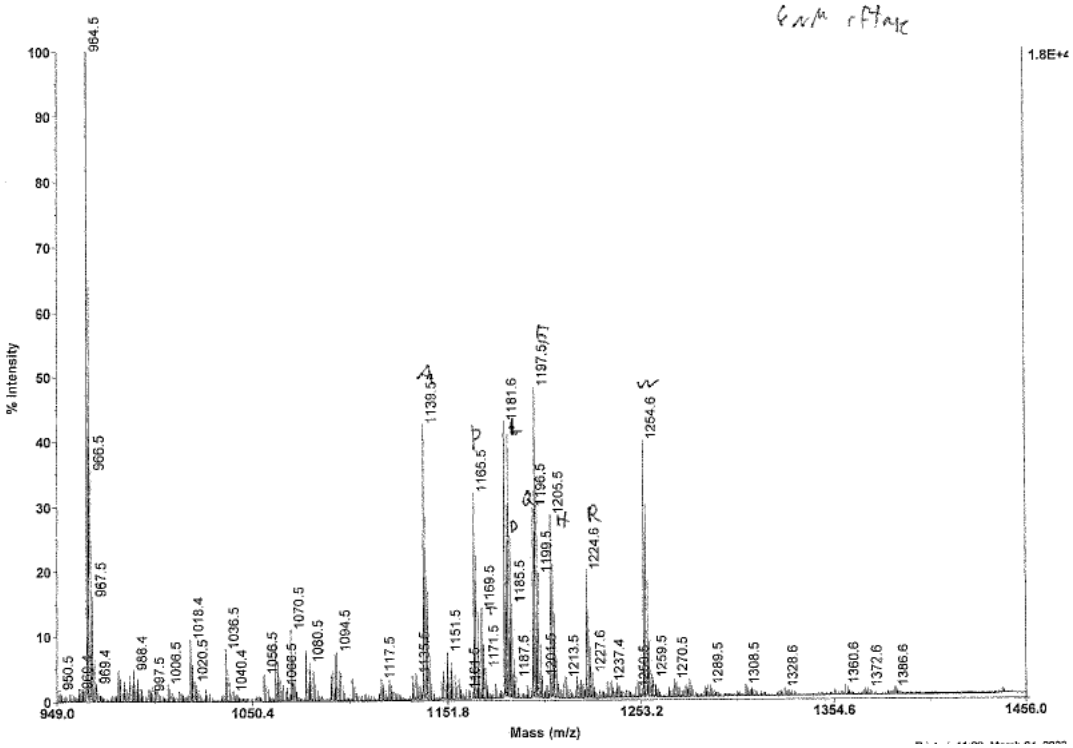


Figure A18. Prenylation of CXLMQ 2 with 3uM rFTase

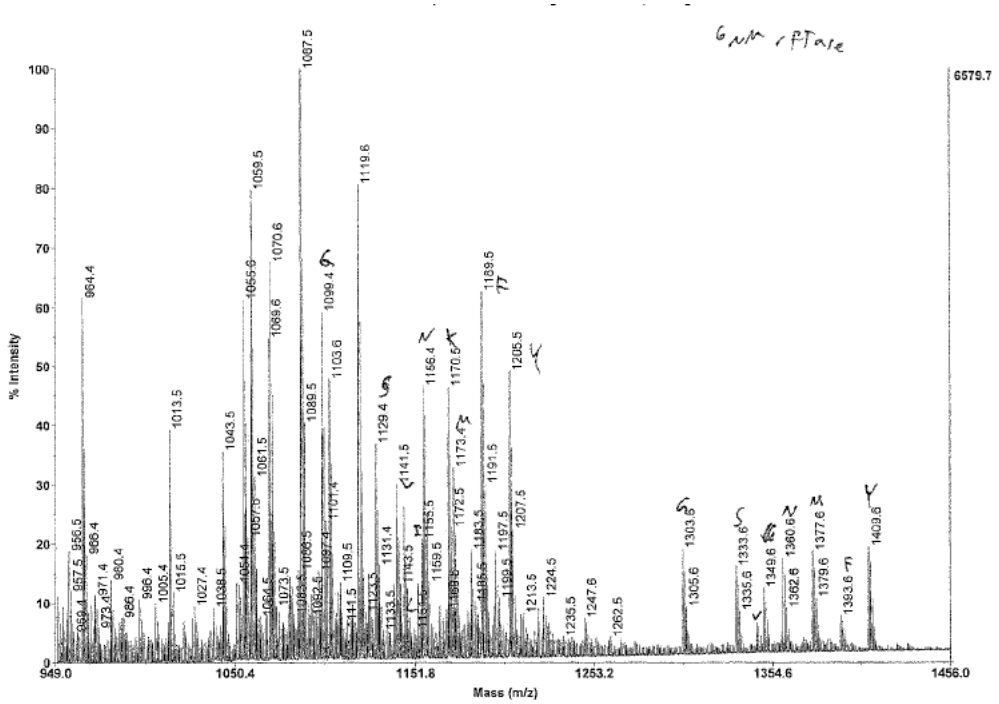
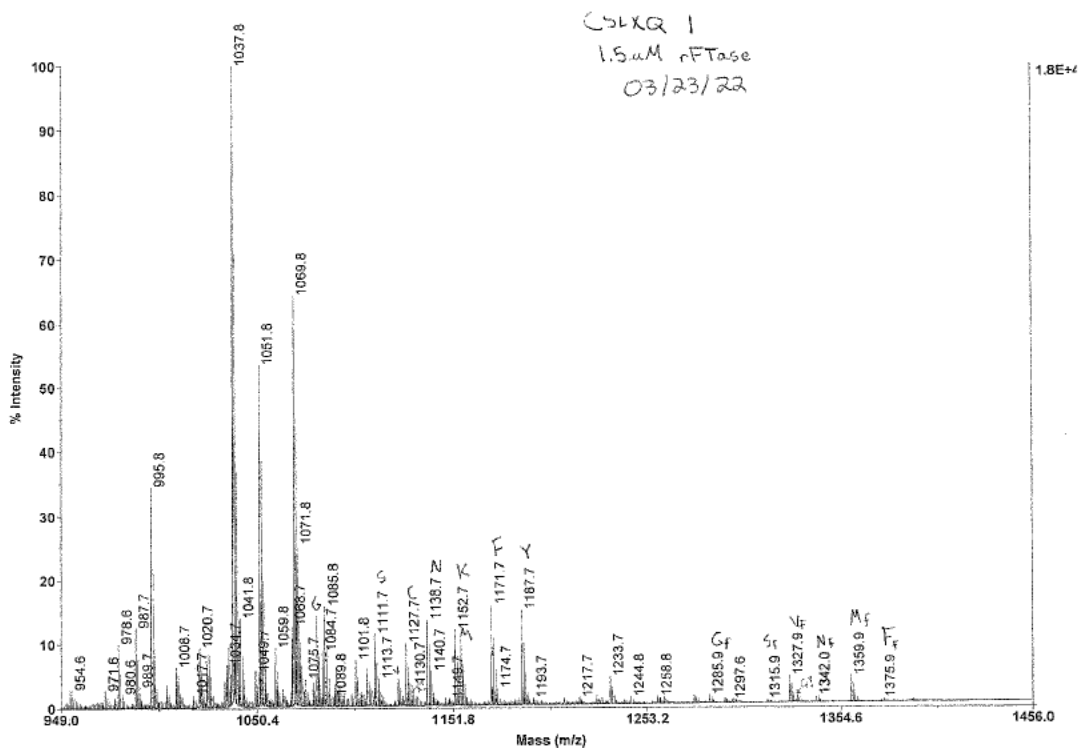
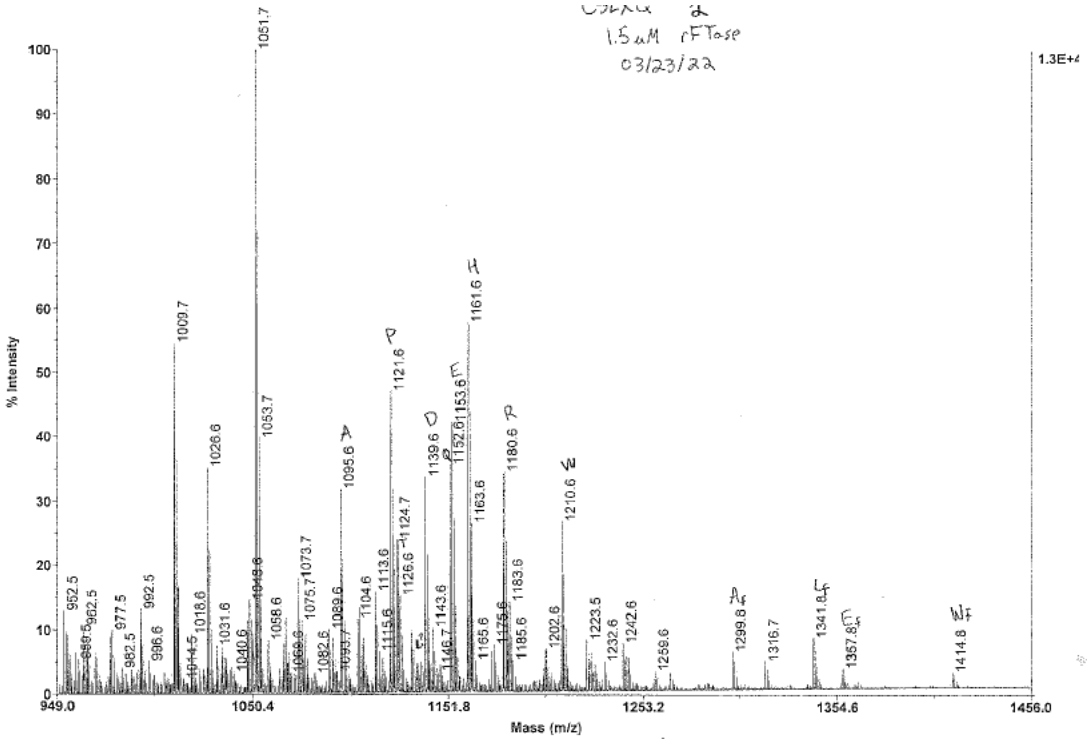


Figure A19. Prenylation of CSXMQ 1 with 3uM rFTase



A20. Prenylation of CSLXQ 1 with 3 uM rFTase



A21. Prenylation of CSLXQ 2 with 3 uM rFTase

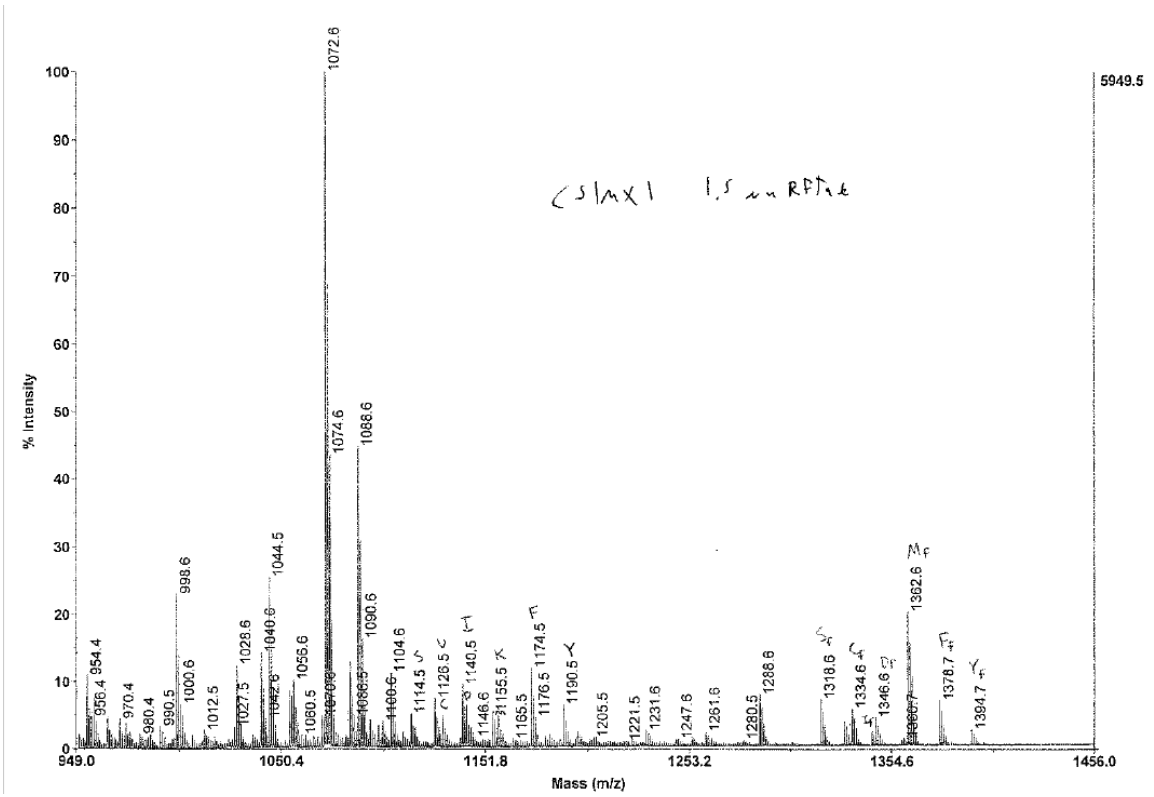


Figure A22. Prenylation of CSLMX 1 with 3uM rFTase

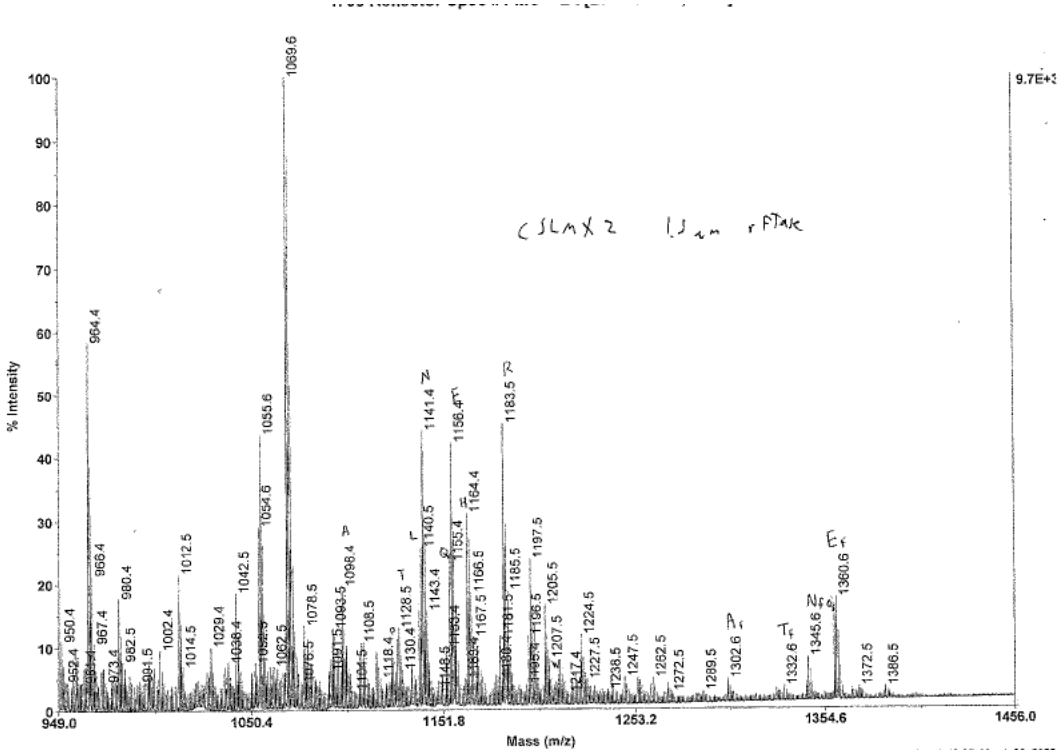


Figure A23. Prenylation of CSLMX 2 with 3 uM rFTase

References

[1] S. Ashok, E.R. Hildebrandt, C.S. Ruiz, D.S. Hardgrove, D.W. Coreno, W.K. Schmidt, J.L. Hougland, Protein Farnesyltransferase Catalyzes Unanticipated Farnesylation and Geranylgeranylation of Shortened Target Sequences, Biochemistry. 59 (2020) 1149–1162. https://doi.org/10.1021/acs.biochem.0c00081.

[2] R.S. Sikorski, P. Hieter, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae., Genetics. 122 (1989) 19–27. https://doi.org/10.1093/genetics/122.1.19.

[3] E.R. Hildebrandt, M. Cheng, P. Zhao, J.H. Kim, L. Wells, W.K. Schmidt, A shunt pathway limits the CaaX processing of Hsp40 Ydj1p and regulates Ydj1p-dependent phenotypes, Elife. 5 (2016) 1–22. https://doi.org/10.7554/eLife.15899.

[4] M.J. Blanden, K.F. Suazo, E.R. Hildebrandt, D.S. Hardgrove, M. Patel, W.P. Saunders, M.D. Distefano, W.K. Schmidt, J.L. Hougland, Efficient farnesylation of an extended C-terminal C(x)3X sequence motif expands the scope of the prenylated proteome, J. Biol. Chem. 293 (2018) 2770–2785. https://doi.org/10.1074/jbc.M117.805770.