INVESTIGATIONS OF AN EXPANDED PROTEIN ISOPRENYLATION PATHWAY

by

BRITTANY MICHELLE BERGER

(Under the Direction of WALTER K SCHMIDT)

ABSTRACT

The CaaX pathway is a set of three post-translational modifications that has long been associated with increasing hydrophobicity and membrane association of proteins. These modifications occur to proteins containing a CaaX motif at the COOH terminus, where “C” is a Cysteine, “a” is an aliphatic amino acid and “X” is one of several residues. The 3-step pathway begins with isoprenylation of the Cysteine, followed by proteolysis of the -aaX tripeptide, and concludes with carboxyl methylation of the lipidated Cysteine. However, in recent years, a new branch of CaaX pathway termed the “shunt” pathway has been characterized. In this alternative outcome, proteins undergo prenylation and then are shunted out of the pathway before undergoing sequential proteolysis and methylation steps. Notably, these shunted sequences often lack the aliphatic amino acids associated with traditional CaaX motifs, which suggests that a broader range of sequences may be able to undergo prenylation than previously expected. In this study, we used a genetic screen utilizing Ydj1, a known shunt protein in yeast, to identify nearly 140 shunted sequences. These newly identified sequences, together with previously published data, were then used to train the machine learning algorithm, Support Vector Machine (SVM). Using SVM, we were able to predict

prenylation for all 8000 possible Cxxx sequences. We then selected a subset of these predicted sequences to test experimentally, where we observed that the machine learning-based prediction method outperformed previously published methods. Lastly, we investigated the histone chaperone, Nap1, as a predicted shunt protein, and not only show that it is indeed prenylated, but also demonstrate that the prenylation status appears to impact Nap1 nuclear levels.

INDEX WORDS: Prenylation, CaaX, machine learning, FTase, shunt pathway, Nap1, Ydj1, Rce1.

INVESTIGATIONS OF AN EXPANDED PROTEIN ISOPRENYLATION PATHWAY

by

BRITTANY MICHELLE BERGER

BS, Eastern Michigan University, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

INVESTIGATIONS OF AN EXPANDED PROTEIN ISOPRENYLATION PATHWAY

by

BRITTANY MICHELLE BERGER

Major Professor:	Walter K. Schmidt
Committee:	Zachary Wood
	Robert Haltiwanger
	Vasant Muralidharan

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The CaaX Pathway

The CaaX pathway is a three-step, post-translational modification (PTM) pathway that was historically established to occur to proteins that have the “CaaX” motif as the final four amino acids at the COOH terminus, where “C” is a Cysteine, “a” is an aliphatic amino acid, and “X” can be one of several residues. The pathway is initiated by the addition of a C15 farnesyl or C20 geranylgeranyl isoprenyl lipid to the Cysteine by either Farnesyltransferase (FTase) or Geranylgeranyltransferase I (GGTase-I), respectively. After lipidation, proteolysis of the -aaX tripeptide is catalyzed by Rce1 or Ste24, and the pathway concludes with carboxyl methylation of the lipidated cysteine by isoprenylcysteine carboxyl methyltransferase (ICMT) or Ste14 in yeast (**Figure 1.1**) (1-3). Since their discovery, these CaaX modifications have long been thought to increase protein hydrophobicity necessary for membrane association. This is especially true in the case of the Ras GTPase isoforms, as early studies noted that by inhibiting CaaX modifications, Ras is mislocalized to the cytosol, resulting in reduced signaling and function (4, 5). As mutated Ras isoforms are estimated to be involved in approximately 30% of all human cancers and up to 90% of certain tumor types, this mislocalization and inhibition of function became an interest for anti-cancer therapies (4, 6-8). In this review, all five enzymes involved in the CaaX pathway, including the two prenyltransferases,

two proteases, and the carboxyl methyltransferase will be covered. Additionally, this review will provide information on the alternative CaaX PTMs associated with the yeast proteins, Ydj1 and Nap1.

Prenylation

The first step of the CaaX pathway, prenylation, was discovered in the late 1970s by observing a C15 lipid attached to the yeast mating pheromone **a**-factor (9). Prenylation is catalyzed by the addition of a C15 farnesyl group or a C20 geranylgeranyl group from the isoprenyl precursors farnesyl diphosphate (FPP) or geranylgeranyl diphosphate (GGPP), for FTase and GGTase-I respectively. Both these prenyltransferases add their respective isoprenyl lipids by a covalent thioether bond to the Cysteine 3 residues from the COOH terminus (10). Since their initial discovery in the 1990s, the FTase and GGTase-I enzymes both have been identified across eukaryotes, and in many ways the enzymes are very similar (11-14). Both FTase and GGTase-I are cytosolic enzymes composed of an α/β subunit heterodimer. In yeast, FTase and GGTase-I share the same α subunit, encoded by the *RAM2* gene. The β subunits, meanwhile, differ (*RAM1* for FTase, *CDC43* for GGTase-I) but share a similar overall structure of alpha helices (12, 15, 16). Both enzymes have a central cavity within the β subunit to bind the isoprenoid precursors, FPP and GGPP, with the charged diphosphate binding near the α/β interface. Although similar, the presence of a Tryptophan residue (W102β) in the FTase lipid binding pocket serves as the main determinant in lipid specificity. This Tryptophan residue in FTase appears to take up the space necessary for the fourth isoprene unit of GGPP. Indeed, mutating this Tryptophan to Threonine, which is present

in the GGTase-I binding pocket (T49β), results in FTase utilizing GGPP instead of FPP (17-19). Additionally, both enzymes are metalloenzymes, requiring a Zinc ion for coordination of the Cysteine residue. The residues for zinc coordination are highly conserved for both enzymes, with residues D307β, C309β, and H363β for yeast FTase, and 286β, 288β, 341β for yeast GGTase-I (D269β, C271β, and H362β for rat GGTase-I) (17, 20) . Interestingly, in addition to zinc, FTase alone requires a magnesium ion to position and stabilize the diphosphate leaving group from FPP. This magnesium ion has been shown to increase FTase activity by several hundred-fold, with the residue D352β shown by mutagenesis to be responsible for coordinating the Mg^{2+} ion (21-23).

After their initial discovery, prenyltransferases became a major area of interest for anti-cancer therapeutics, with the main focus being on FTase as the human Ras isoforms are farnesylated. (24, 25). FTase inhibition for Ras therapies was supported by studies showing that these COOH terminal modifications were crucial for proper localization and signaling of Ras and provided preliminary data that inhibition of Ras farnesylation may slow oncogenic growth (4, 5, 26). While several FTase inhibitors (FTIs) were developed throughout the 1990's and entered clinical trials, tipifarnib, developed by Janssen, ended up making it all the way to phase III clinical trials before being deemed unsuccessful (27, 28). While this FTI was mostly well tolerated in patients, the main reason for its ineffectiveness was due to the ability of GGTase-I to prenylate FTase substrates when FTase is inhibited (29-31). This was shown to be true for the Ras isoforms, where GGTase-I was shown to be able to modify K-Ras and N-Ras, effectively abolishing the purpose of FTI treatment (30-32). To circumvent this effect, it was hypothesized that a combination therapy of FTIs and GGTase-I inhibitors

(GGTIs) would be useful, however, this combination proved to be lethal in mice in preclinical studies (33). Together, this data led to most FTIs being abandoned for cancer treatments, leading to interest in inhibiting steps further down the CaaX pathway for Ras targeted cancer treatments (7, 34-38). Meanwhile, in recent years, FTIs have been shown to be effective in treating CaaX protein-related conditions outside of cancer, as prenylation was revealed to be important for growth in both *Plasmodium falciparum* and *Trypanosoma brucei*. Studies with previously developed FTIs have shown that these species are often more sensitive to treatment. Additionally, differences in specificity between these parasitic prenyltransferases and human may allow for development of parasite specific inhibitors (39-44). FTIs have also shown promise in treating hepatitis delta virus, where prenylation is necessary for viral particle assembly (45). Recently, an FTI became the first approved treatment for progeria, a rare disease caused by a mutation in the gene encoding the known CaaX protein, Prelamin A, leading to rapid aging and premature death (42-44, 46).

Although the prenyltransferases may be the most characterized enzymes within the CaaX pathway, the full scope of their specificity has yet to be determined. One explanation for this ambiguity is the fact that early studies examining prenylation specificity were done with reporters *in vivo*, where downstream processing may be limited to target sequences considered canonical CaaX motifs (47, 48). In recent years, *in vitro* and *in silico* approaches, as well as isoprenyl analogs using click-chemistry have been able to identify novel, non-canonical targets of prenylation (31, 49-55). Now, the new challenge is the large number of possible prenylation substrates. To date, the only known rules for prenylation were the aliphatic enrichment at the a_1 and a_2 positions.

However, it's now evident that this rule describes the specificity of CaaX proteases rather than that of prenyltransferases. Thus, this new finding abolishes the current CaaX motif to simply Cxxx for prenylation. With 20 possible amino acids in each "x" position, this allows for 8000 possible prenylatable sequences. While not all 8000 motifs are present *in vivo*, many standard approaches to identify the full scope of prenylatable sequences such as peptide libraries for *in vitro* screening would be cost-prohibitive, while using genetic reporters may be too labor intensive. An additional challenge when looking at prenylation specificity is determining whether the motif is modified by FTase versus GGTase-I. It was suggested in the early 1990's that the X position of CaaX motifs can be the determining factor for FTase or GGTase-I, with sequences ending in Leucine, Phenylalanine or Methionine being targets of geranylgeranylation (47, 56, 57). However, sequences with Methionine or Phenylalanine have also been shown to be farnesylated. Although the GGTase-I β subunit, *CDC43*, is essential for growth in yeast, it is possible to disrupt the FTase α subunit to observe if sequences are targets of geranylgeranylation, but this approach would raise two problems. The first being the natural occurrence of farnesylation versus geranylgeranylation *in vivo* for sequences that may be targets for both enzymes. The second problem is again this ability of GGTase-I to cross-prenylate FTase targets, as observed on Ras.

The Shunt Pathway

In 2016 a new outcome for the CaaX pathway was characterized (58). Identified using the yeast system and the Hsp40 chaperone, Ydj1, this alternative outcome yields proteins that undergo prenylation but are resistant to proteolysis and subsequent

methylation. While occasional examples of prenylated only proteins have been evident throughout literature, including PHK in rabbits (CAMQ), and Rab38 (CAKS) and Lbk1 (CKQQ) in humans (**Table 1.1**), Ydj1 was the first example of full CaaX processing having a negative impact on the protein's function (59-61). The impact of shunting on two yeast shunt proteins, Ydj1 and Nap1, will be discussed further later in this review.

Proteolysis

Following prenylation aliphatic-enriched CaaX motifs, often referred to as canonical CaaX motifs, undergo proteolysis of the -aaX tripeptide neighboring the prenylated Cysteine, followed by carboxyl methylation of the terminal Cysteine. Both enzymes involved in proteolysis are integral membrane proteins located at the endoplasmic reticulum, with multiple transmembrane passes (62-64). This has created major challenges for structural characterization, requiring detergent solubilization for isolation and structural studies. Additionally, although protease inhibitors were thought to hold promise as Ras therapeutics after the FTIs were deemed unsuccessful, there has been limited development of such inhibitors for therapeutic use.

As it was known that both the yeast mating pheromone, **a**-factor, and Ras require proteolysis and carboxymethylation for proper function, the two CaaX proteases were identified by genetic screens looking at **a**-factor and Ras defects. These phenotypes led to the discovery of Ras Converting Enzyme 1 (Rce1) and Ste24, initially named **a**-factor Converting Enzyme 1 (AFC1) (65-68). Although Rce1 and Ste24 are both multi-spanning membrane proteases, their similarities end there.

Rce1, often considered the main protease involved in the CaaX pathway, is a 35 kDa intramembrane protease hypothesized to have seven to eight transmembrane helices, depending on the species (69, 70). Rce1 structural studies in particular have remained a challenge, with purification attempts at isolating Rce1 from the membrane dating back to the 1990s remaining largely unsuccessful. Rce1 appears to be extremely unstable upon removal from the membrane, with the enzyme becoming inactive in the presence of detergents (64, 71). To date, only one structure of a distant Rce1 homolog has been solved from *Methanococcus maripaludis*, a prokaryotic methanogen, which was co-crystallized with an antibody for stability (72). The *Mm*Rce1 structure had eight transmembrane domains and shed insight on Rce1's mechanism for the first time. Rce1's mechanism was initially hypothesized to be a metalloprotease due to its conserved glutamate and histidine residues, or a Cysteine protease due to Rce1's sensitivity to Cysteine protease inhibitors. However, no metal ion was detected within Rce1, and the protease remained active in the presence of metal chelators, ruling out a zinc metalloenzyme mechanism. Meanwhile, a Rce1 mutant lacking all Cysteines was still active *in vivo*, eliminating Cysteine protease as the mechanism. Instead, it was implied that Rce1 may have a novel mechanism, involving the 3 conserved residues shown to be essential for Rce1 activity: E156, H194 and H248 in *S. cerevisiae* (E140, H173, and H227 of *Mm*Rce1) (64, 69, 73, 74). Manolaridis *et. al.* proposed that *Mm*Rce1 does indeed have a novel mechanism as a glutamyl protease, with a water molecule being activated by E140. Meanwhile, H173 was proposed to play a role stabilizing the water molecule and H227 forming an oxyanion hole with a conserved Asparagine, stabilizing the intermediate (H194, H248 for *Sc*Rce1, respectively). While

*Mm*Rce1 provides mechanistic and structural information, Rce1's presence in this organism is a mystery. As *M. maripaludis* is a prokaryote, there is no CaaX processing, suggesting that *Mm*Rce1 has no prenylated substrates *in vivo*. Additionally, there are a few caveats from the *Mm*Rce1 study, including being unable to cleave geranylgeranylated peptides, and cleaving at the P1 and P1´ (a_1 and a_2) positions instead of just P1 (a_1). As *Mm*Rce1 shares only 12% and 13% identity to yeast and human Rce1 orthologs, respectively, these differences in specificity perhaps provide rationale for the differences compared to eukaryotic orthologs (70).

In terms of specificity, a substitution analysis using the yeast pheromone **a**-factor provided evidence that Rce1's specificity is mainly dependent on the a_2 position, largely preferring aliphatic (L, I, V) amino acids at this position. The a_1 position also plays a minor role in specificity, preferring aliphatic residues, as well as Alanine, Cysteine, Serine, and Threonine (48). Rce1 cleavage has also been shown to be prenylation dependent, unlike Ste24 (75). While Rce1 can clearly cleave farnesylated (C15) and geranylgeranylated (C20) substrates, it's unclear if it is able to cleave substrates with shorter or longer lipid groups. However, evidence has shown that Rce1 is partially able to cleave prenylated CaaX sequences that have one extra amino acid (5mer, Cxxxx) or one less (3mer, Cxx) (76, 77)

While several inhibitors of Rce1 have been developed, including a peptidomimetic Rce1 protease inhibitor (RPI) with an IC_{50} in the nanomolar range, there has been limited development of these inhibitors for therapeutics for several reasons. Although these peptidomimetic inhibitors exhibited better efficacy, they did have poorer cell permeability and were more susceptible to enzymatic degradation (70, 73, 78).

Additionally, the ambiguity of effects from Rce1 knockout studies may be a reason for a lack of interest in developing therapeutics. *rce1*$^{-/-}$ is embryonic lethal in mice, leading to the use of conditional knockouts and cell lines for further Rce1 studies.(79). While the absence of *Rce1* was shown to slow growth in a melanoma cell line, in hematopoietic cells, it led to progression of myeloproliferative disease (80, 81). Additionally, using a conditional knockout of Rce1 in heart tissue of mice lead to premature death (82). While it's unclear how Rce1 is causing these ill effects, the likely result is by disrupting processing of various CaaX proteins (83). All in all, preliminary data suggests that Rce1 may not be a good therapeutic target.

In addition to Rce1, Ste24 was also initially identified as a CaaX protease. Ste24 is a 52 kDa, zinc metalloprotease membrane protease with 7 transmembrane helices that create a large, inner membrane cavity. It also has short loops and helices forming a cap to this chamber. Thus, Ste24 has sometimes referred to as an "α-barrel" (84-86). This cavity, with an estimated volume of nearly 10,000 Å^3, has been proposed to be able to accommodate a 10 kDa protein or 450 water molecules. Although structures of Ste24 have been solved, including a homolog from the *Saccharomyces mikatae* yeast strain and human Ste24 (ZMPSTE24) with and without substrates, there are many questions regarding the function of this large chamber within Ste24's structure. It has been shown that Ste24's active site lies within this inner cavity, indicating that substrates must somehow enter this space for cleavage, however the dynamics of this substrate acquisition still remain unclear.

As described above, Ste24 was first identified as a CaaX protease by observing a reduction of **a**-factor mating in *ste24Δ* strains, yet to date **a**-factor remains the only

known substrate of Ste24 in yeast (75). In humans, Prelamin A is the only known substrate of the human ortholog, ZMPSTE24. For both **a**-factor and Prelamin A, Ste24 is responsible for an additional N-terminal cleavage. Meanwhile, recent studies have shown that at the COOH terminus, for both **a**-factor and Prelamin A, Rce1 is able to cleave equally or better than Ste24, suggesting Ste24 has no enzyme specific CaaX substrates (68, 87) (Chapter 3). It has also been shown that Ste24 can cleave substrates both with and without isoprenyl groups *in vitro*, implying that Ste24 may not necessarily be a true CaaX protease, which necessitates prenylation for activity (75). While Ste24's true function remains unclear, more recently it has been proposed that Ste24 acts as a translocon unclogger, cleaving misfolded proteins that block the ER translocon pore (88).

Carboxyl Methylation

After proteolysis, CaaX proteins are carboxyl methylated at the lipidated cysteine by Ste14 in yeast or Isoprenylcysteine Carboxyl Methyltransferase (ICMT) in humans. Ste14 is a 26 kDa integral membrane protein, with a proposed 6 transmembrane helices (89). Interestingly, Ste14 and the ICMT family are the only known membrane associated methyltransferases to date, with this membrane association and lack of characteristic methyltransferase motifs forming a new class of methyltransferases, known as class VI (90). Two structures from the ICMT family have been solved: a distantly related ortholog from the archaea *Methanosarcina acetivorans* (*Ma*ICMT), and a eukaryotic ortholog from the beetle *Tribolium casteaneum* (*Tc*ICMT) (90, 91). Similar to the *Mm*Rce1 structure, while this prokaryotic ICMT, *Ma*ICMT, did provide some structural information,

the lack of *in vivo* substrates and low homology left several questions to still be answered. In 2018, a more closely related ICMT structure was solved utilizing a monobody, a synthetic binding protein based on a randomized fibronectin protein domain, to stabilize ICMT during crystallization (91).

From mutagenesis studies, it was shown that the *Sc*Ste24 has 4 residues that are conserved and critical for function: G31, G132, P173, and E213. From the ICMT structures and further mutagenesis, putative roles could be assigned to these yeast residues, with G31 being a part of the GxxxG motif important for dimerization. G132, P173, and E213 were proposed to stabilize the transition state and bind the cofactor, S-adenosyl-L-methionine (SAM) with E213 having direct contact with SAM (91, 92). Additionally, the *Tc*ICMT structure allowed for the prediction of a lipid binding cavity by modeling. This cavity is lined with aromatic residues and partially exposed to the lipid bilayer, with mutagenesis showing a significant reduction or loss of activity when residues in this region were mutated. Although residues have been identified in cofactor binding and stabilizing transition states, as the *Tc*ICMT ortholog required a monobody, effectively inhibiting the enzyme and prevented co-crystallization studies, the exact mechanism of ICMT remains unclear. This is especially true as the ICMT family is suspected to be zinc metalloenzymes, but a zinc ion was not modeled in the solved *Tc*ICMT structure. The residues responsible for binding this zinc ion are still unknown, as the Cysteine residues proposed to be involved with zinc coordination resulted in little to no effect on ICMT function when mutated (91, 93, 94).

In terms of inhibition, as ICMT is the only enzyme responsible for methylation in the CaaX pathway and differs from cytosolic methyltransferases, it appeared to be the

best target for inhibiting the Ras modifications. Indeed, studies done in mice had shown that ICMT reduces growth in oncogenic K-Ras cells (95). These findings were further supported by work done in human cancer cell lines, where ICMT inhibitors were able to slow growth (96, 97). Opposite to Rce1, ICMT inhibitors did not seem to increase myeloproliferative disease, making ICMT an even more appealing target for Ras inhibition (98). However, a conflicting report suggested that ICMT may have a tumor suppressor role in pancreatic cancer in mice, with accelerated tumor growth following ICMT deletion (99). Interestingly, one ICMT inhibitor, Salirasib, entered clinical trials for several types of cancer and appeared to be well tolerated in Phase I, yet only one Phase II study was done in the US over 20 years ago (100, 101). Although there appears to be a renewed interest in Salirasib, it's unclear why this treatment did not progress further through the drug development pathway (102, 103).

Ydj1

Ydj1, a yeast homolog of the *E. coli* chaperone DnaJ, was first identified as prenylated in the early 1990s, with this prenylation shown to be important for proper growth of Ydj1 at higher temperatures (104, 105). Although there are 3 other DnaJ homologs in yeast, Ydj1 is the only one that is prenylated. While it has long been established that Ydj1 is prenylated, there are a few features of Ydj1 that do not fit a canonical CaaX protein. The first indicator is Ydj1's CaaX motif, CASQ, which does not contain the traditional aliphatic enriched definition at the a_1 and a_2 positions. Secondly, the yeast pheromone **a-**factor, which requires prenylation, proteolysis, methylation for proper mating, showed very little mating when the wildtype sequence was substituted

by CASQ (74). Lastly, while CaaX processing has long been associated with membrane association, Ydj1 was noted by fractionation and immunofluorescence studies to be cytosolic (58, 104). Mass spectrometry of native *Sc*Ydj1 purified from yeast showed that at the COOH-terminus, a farnesyl group is present but the -aaX peptide remains, indicating a lack of proteolysis and methylation. Interestingly, by introducing canonically processed CaaX motifs onto Ydj1, an intermediate growth phenotype was observed at high temperatures (58). While these sequences are able to grow better than the unfarnesylated Ydj1, there is a clear decrease compared to the wildtype, farnesylation only Ydj1 (**Figure 1.2**). This data shows that this alternative outcome is functionally separate than the canonical, 3-step CaaX pathway. While the exact reason for Ydj1's shunted (farnesylation only) outcome is unknown, there have been examples of prenyl groups playing a role structurally or in protein-protein interactions for Pex19 and Spindly in humans (106-108). Additionally, this identification of Ydj1 as the first shunted protein suggests the likelihood that more proteins may undergo this shunted CaaX prenylation.

Nap1

Another possible shunted protein is the histone chaperone, Nucleosome Assembly Protein 1 (Nap1). Nap1 was first identified in mammalian cell extracts by its involvement in nucleosome assembly (109). Since then, Nap1 has been implicated in many different roles and functions *in vivo*, many of which to date are still not well understood. More recently, emerging data for several orthologs from different species suggest that Nap1 may be a prenylprotein.

Nap1 is a 48 kDa, nucleocytoplasmic shuttle protein with several highly acidic regions thought to be important for binding basic histones. Nap1 also contains both nuclear localization and export signals (NLS/NES), and a long α helix acting as a dimerization domain. The NH_2 and COOH termini have disordered domains of approximately 60 residues, with the importance of these regions under considerable debate following several *in vitro* histone binding studies (110-112). Nap1 has several confirmed post-translational modifications, including several phosphorylation, acetylation, and glutamylation sites (113-116). While the impact of these modifications is unclear, both the phosphorylation and acetylation have been hypothesized to be cell cycle dependent, noting that Nap1 is a target of casein kinase II (CKII) (113, 116). In terms of acetylation, Nap1 appears to be a target of p300, with an increase in Nap1-p300 complexes seen during S-phase (115). Interestingly, the Nap1 family also possesses a conserved CKQ[Q/S] motif as the last 4 amino acids at the COOH terminus, suggesting possible isoprenylation. Indeed, prenylation has been confirmed *in vivo* for Nap1 in *Arabidopsis thaliana*, and studies utilizing metabolic labeling followed by mass spectroscopy identified two mammalian Nap1 homologs, Nap1L1 and Nap1L4, to be farnesylated (31, 117). While no such data exists for *Sc*Nap1, it is suspected that the yeast homolog will also be a prenylated due to its CKQS motif. Additionally, as this Nap1 family CKQ[Q/S] motif does not fit the canonical, aliphatic enriched definition of CaaX motifs, we predict that Nap1 undergoes the prenylation only outcome of the CaaX pathway. Supporting this hypothesis, the Nap1L1 and Nap1L4 homologs that were identified using metabolic labeling were found to be farnesylated and unprocessed, suggesting they are not cleaved or carboxyl methylated (31).

Although Nap1's functions are largely unknown, it does have several known protein interactors. These include several cyclins involved in cell cycle regulation, such as Clb2, Cln2 and Cln3, as well as Gin4, which is involved in septin assembly (118-122). Nap1 is also known to bind to histones H2A/H2B in the cytosol, which increases the affinity for the histone-Nap1 complex for their nuclear transporter, Kap114 (123, 124). Within the nucleus, Nap1 has been shown to be involved in transcription initiation and elongation, and chromatin assembly (125-130). *In vitro*, Nap1 is capable of forming nucleosomes by itself, depositing histones onto supercoiled DNA (109, 131-133). Nucleosomes, which are composed of an octamer of histones (a tetramer composed of 2 each of histone H3/H4 and two histones H2A/H2B dimers) were shown to be assembled by Nap1 first depositing the H3/H4 tetramer onto supercoiled DNA, followed by the two H2A/H2B dimers one at a time (131). Interestingly, while Nap1 has been shown to bind all 4 histones *in vitro*, Nap1 prefers H2A/H2B *in vivo* (123, 124). Additionally, while *in vitro* studies have shown Nap1 in a variety of different oligomeric states, including homodimers and octamers, it's unclear what Nap1's native state is *in vivo* (134-136).

Summary

While the CaaX pathway has long been regarded as a three-step process, the characterization of the alternative shunt pathway suggests the possibility of more sequences and proteins being prenylated than previously expected. Additionally, the identification of the shunt pathway allows for the first time the differentiation of shunted (prenylation only) and canonical sequences. Previously, these prenylation only motifs

were often considered anomalies or processing intermediates and broadly categorized as canonical CaaX proteins. Indeed, many tools used for studying the CaaX pathway, from prediction methods to genetic reporters, have an inherent bias towards canonical CaaX motifs, creating a challenge to identify shunted sequences. Altogether, this study looks to build upon the identification of the shunt pathway by determining additional sequences and substrates that can undergo the shunted, or prenylation only, outcome. In Chapter 2, we use the Ydj1 thermotolerant phenotype (**Figure 1.2**) as an *in vivo* reporter of the shunt pathway. From this study, we were able to identify approximately 140 other CaaX sequences that appear to undergo this prenylation only outcome. We then utilize this data, combined with a previously published dataset based on Ras, in Chapter 3, using machine learning trained on these newly identified sequences to create a predictive model for all 8000 possible Cxxx sequences (137). We were able to test a subset of our predictions to determine the accuracy of our machine learning method compared to previously published methods. Lastly, in Chapter 4 we present evidence for Nap1, a conserved histone chaperone, being prenylated and shunted, as well as explore a functional impact for this modification.

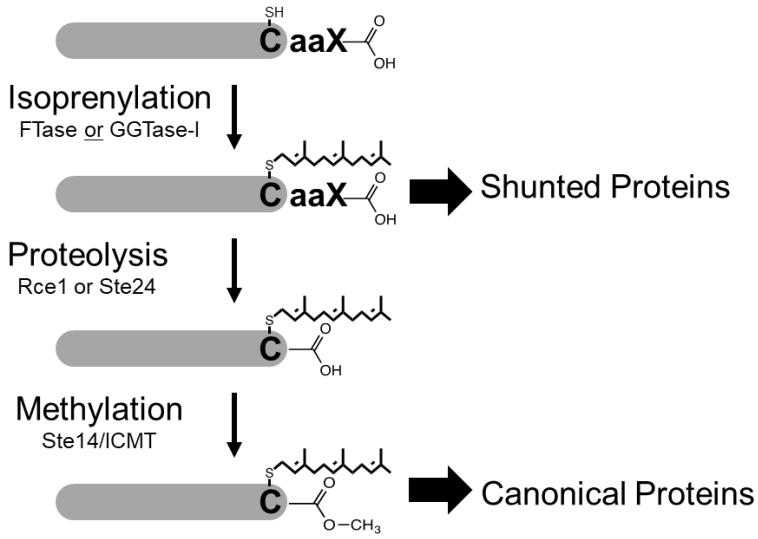
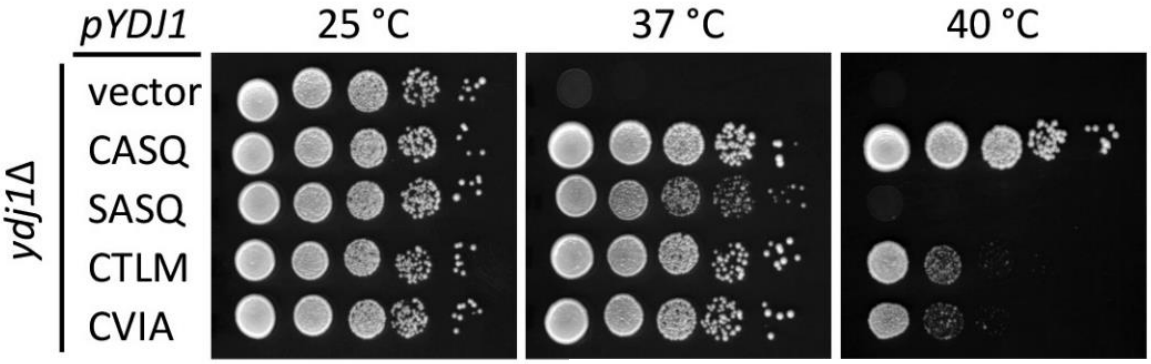


Figure 1.1: The CaaX Pathway. Proteins containing a CaaX motif (where “C” is a Cysteine, “a” are aliphatic amino acids, and “X” can be one of several amino acids) are subject to a set of post-translational modifications. This begins with isoprenylation (addition of C15 farnesyl or C20 geranylgeranyl) of the Cysteine residue by FTase or GGTase. Next, the -aaX tripeptide is cleaved by the CaaX protease Rce1 or Ste24, depending on the specific CaaX sequence. Lastly, the lipidated Cysteine is carboxyl methylated by an isoprenylcysteine carboxyl methyltransferase, Ste14 (yeast) or ICMT (humans). The three steps together are considered canonical CaaX processing and thought to increase membrane association. Additionally, an alternative “shunted” outcome of prenylation only has been characterized, where proteins retain their -aaX after prenylation, avoiding proteolysis and carboxyl methylation. The impact of this lipid only modification has only been investigated for a select number of shunted proteins.

Table 1.1: Proteins likely to be shunted

Species	Protein	CaaX	Reference
Saccharomyces cerevisiae	Ydj1	CASQ	(58, 104)
	Nap1	CKQS	Chapter 3,4 (138)
	Pex19	CKQQ*	(138, 139)
Homo sapiens	DNAJA1	CQTS*	(31, 54)
	DNAJA2	CAHQ*	(31, 54)
	Nap1L1/L4	CKQQ	(31, 54)
	Lkb1/Stk11	CKQQ	(140)
	Gγ5	CSFL	(60)
	Spindly	CNQQ*	(107)
	CENP-E	CKTQ*	(141)
	Rab38	CAKS	(59)
Oryctolagus cuniculus	PHK	CAMQ	(61)
Hepatitis Delta Virus	Delta Antigen	CRPQ*	(45)

*Indicates CaaX motif predicted to be shunted by SVM (Chapter 3) but lacks protein-specific data for shunting



Adapted from Hildebrandt *et. al.* (2016)

Figure 1.2: Ydj1 requires prenylation only for optimal growth. Saturated cultures of yeast strains lacking chromosomally encoded *YDJ1* (*ydj1*Δ) but expressing various Ydj1-Cxxx were pinned in 10-fold serial dilutions and incubated at the temperatures indicated. The thermotolerance of the various Ydj1-Cxxx altering Ydj1's COOH terminal modifications (SASQ: non-prenylated; CTLM, CVIA: cleaved and carboxyl methylated) were then compared to the wildtype Ydj1 motif (CASQ: prenylation only). The wildtype, prenylation only motif supports optimal growth at 40 °C compared to the non-prenylated or cleaved and carboxyl methylated Ydj1 mutants.

References

1. Zhang FL, Casey PJ. Protein prenylation: Molecular mechanisms and functional consequences. Annual review of biochemistry. 1996;65:241-69.

2. Wright LP, Philips MR. Thematic review series: Lipid posttranslational modifications. Caax modification and membrane targeting of ras. Journal of lipid research. 2006;47(5):883-91.

3. Wang M, Casey PJ. Protein prenylation: Unique fats make their mark on biology. Nature reviews Molecular cell biology. 2016;17(2):110-22.

4. Der JHJ, Cochrane CG, Bourne JR, Solski PA, Buss JE, C J. Farnesol modification of kirsten-ras exon 4b protein is essential for transformation. 1990.

5. Willumsen BM, Christensen A, Hubbert NL, Papageorge AG, Lowy DR. The p21 ras c-terminus is required for transformation and membrane association. Nature. 1984;310(5978):583-6.

6. The ras initiative: National Cancer Institute; [Available from: https://www.cancer.gov/research/key-initiatives/ras.

7. Winter-Vann AM, Casey PJ. Post-prenylation-processing enzymes as new targets in oncogenesis. Nature reviews Cancer. 2005;5(5):405-12.

8. Reuter CW, Morgan MA, Bergmann L. Targeting the ras signaling pathway: A rational, mechanism-based treatment for hematologic malignancies? Blood. 2000;96(5):1655-69.

9. Kamiya Y, Sakurai A, Tamura S, Takahashi N. Structure of rhodotorucine a, a novel lipopeptide, inducing mating tube formation in rhodosporidium toruloides. Biochemical and biophysical research communications. 1978;83(3):1077-83.

10. Clarke S. Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. Annual review of biochemistry. 1992;61:355-86.

11. Reiss Y, Goldstein JL, Seabra MC, Casey PJ, Brown MS. Inhibition of purified p21ras farnesyl:Protein transferase by cys-aax tetrapeptides. Cell. 1990;62(1):81-8.

12. Seabra MC, Reiss Y, Casey PJ, Brown MS, Goldstein JL. Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit. Cell. 1991;65(3):429-34.

13. Yokoyama K, Goodwin GW, Ghomashchi F, Glomset JA, Gelb MH. A protein geranylgeranyltransferase from bovine brain: Implications for protein prenylation specificity. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(12):5302-6.

14. Omer CA, Kral AM, Diehl RE, Prendergast GC, Powers S, Allen CM, et al. Characterization of recombinant human farnesyl-protein transferase: Cloning, expression, farnesyl diphosphate binding, and functional homology with yeast prenyl-protein transferases. Biochemistry. 1993;32(19):5167-76.

15. He B, Chen P, Chen SY, Vancura KL, Michaelis S, Powers S. Ram2, an essential gene of yeast, and ram1 encode the two polypeptide components of the farnesyltransferase that prenylates a-factor and ras proteins. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(24):11373-7.

16. Jin L, Orvell C, Myers R, Rota PA, Nakayama T, Forcic D, et al. Genomic diversity of mumps virus and global distribution of the 12 genotypes. Reviews in medical virology. 2015;25(2):85-101.

17. Taylor JS, Reid TS, Terry KL, Casey PJ, Beese LS. Structure of mammalian protein geranylgeranyltransferase type-i. Embo j. 2003;22(22):5963-74.

18. Long SB, Casey PJ, Beese LS. Cocrystal structure of protein farnesyltransferase complexed with a farnesyl diphosphate substrate. Biochemistry. 1998;37(27):9612-8.

19. Terry KL, Casey PJ, Beese LS. Conversion of protein farnesyltransferase to a geranylgeranyltransferase. Biochemistry. 2006;45(32):9746-55.

20. Harris CM, Derdowski AM, Poulter CD. Modulation of the zinc(ii) center in protein farnesyltransferase by mutagenesis of the zinc(ii) ligands. Biochemistry. 2002;41(33):10554-62.

21. Reiss Y, Brown MS, Goldstein JL. Divalent cation and prenyl pyrophosphate specificities of the protein farnesyltransferase from rat brain, a zinc metalloenzyme. The Journal of biological chemistry. 1992;267(9):6403-8.

22. Zhang FL, Casey PJ. Influence of metal ions on substrate binding and catalytic activity of mammalian protein geranylgeranyltransferase type-i. The Biochemical journal. 1996;320 (Pt 3)(Pt 3):925-32.

23. Pickett JS, Bowers KE, Fierke CA. Mutagenesis studies of protein farnesyltransferase implicate aspartate beta 352 as a magnesium ligand. The Journal of biological chemistry. 2003;278(51):51243-50.

24. Rowinsky EK, Windle JJ, Von Hoff DD. Ras protein farnesyltransferase: A strategic target for anticancer therapeutic development. Journal of Clinical Oncology. 1999;17(11):3631-52.

25. Casey PJ. Lipid modifications of g proteins. Current opinion in cell biology. 1994;6(2):219-25.

26. Willumsen BM, Norris K, Papageorge AG, Hubbert NL, Lowy DR. Harvey murine sarcoma virus p21 ras protein: Biological and biochemical significance of the cysteine nearest the carboxy terminus. The EMBO journal. 1984;3(11):2581-5.

27. Alsina M, Fonseca R, Wilson EF, Belle AN, Gerbino E, Price-Troska T, et al. Farnesyltransferase inhibitor tipifarnib is well tolerated, induces stabilization of disease, and inhibits farnesylation and oncogenic/tumor survival pathways in patients with advanced multiple myeloma. Blood. 2004;103(9):3271-7.

28. Sebti SM, Hamilton AD. Farnesyltransferase and geranylgeranyltransferase i inhibitors and cancer therapy: Lessons from mechanism and bench-to-bedside translational studies. Oncogene. 2000;19(56):6584-93.

29. James G, Goldstein JL, Brown MS. Resistance of k-rasbv12 proteins to farnesyltransferase inhibitors in rat1 cells. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(9):4454-8.

30. Pai DBW, Paul K, Tish NH, Irma N-O, Linda J, Joseph JC, et al. K- and n-ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors. 1997.

31. Storck EM, Morales-Sanfrutos J, Serwa RA, Panyain N, Lanyon-Hogg T, Tolmachova T, et al. Dual chemical probes enable quantitative system-wide analysis of protein prenylation and prenylation dynamics. Nature Chemistry. 2019.

32. Lerner EC, Zhang TT, Knowles DB, Qian Y, Hamilton AD, Sebti SM. Inhibition of the prenylation of k-ras, but not h- or n-ras, is highly resistant to caax peptidomimetics and requires both a farnesyltransferase and a geranylgeranyltransferase i inhibitor in human tumor cell lines. Oncogene. 1997;15(11):1283-8.

33. Lobell RB, Omer CA, Abrams MT, Bhimnathwala HG, Brucker MJ, Buser CA, et al. Evaluation of farnesyl:Protein transferase and geranylgeranyl:Protein transferase inhibitor combinations in preclinical models. Cancer research. 2001;61(24):8758-68.

34. Manandhar SP, Hildebrandt ER, Jacobsen WH, Santangelo GM, Schmidt WK. Chemical inhibition of caax protease activity disrupts yeast ras localization. Yeast (Chichester, England). 2010;27(6):327-43.

35. Manandhar SP, Hildebrandt ER, Schmidt WK. Small-molecule inhibitors of the rce1p caax protease. Journal of biomolecular screening. 2007;12(7):983-93.

36. Mohammed I, Hampton SE, Ashall L, Hildebrandt ER, Kutlik RA, Manandhar SP, et al. 8-hydroxyquinoline-based inhibitors of the rce1 protease disrupt ras membrane localization in human cells. Bioorganic & medicinal chemistry. 2016;24(2):160-78.

37. Rotblat B, Ehrlich M, Haklai R, Kloog Y. The ras inhibitor farnesylthiosalicylic acid (salirasib) disrupts the spatiotemporal localization of active ras: A potential treatment for cancer. Methods in enzymology. 2008;439:467-89.

38. Lau HY, Ramanujulu PM, Guo D, Yang T, Wirawan M, Casey PJ, et al. An improved isoprenylcysteine carboxylmethyltransferase inhibitor induces cancer cell death and attenuates tumor growth in vivo. Cancer biology & therapy. 2014;15(9):1280-91.

39. Yokoyama K, Lin Y, Stuart KD, Gelb MH. Prenylation of proteins in trypanosoma brucei. Molecular and biochemical parasitology. 1997;87(1):61-9.

40. Yokoyama K, Trobridge P, Buckner FS, Scholten J, Stuart KD, Van Voorhis WC, et al. The effects of protein farnesyltransferase inhibitors on trypanosomatids: Inhibition of protein farnesylation and cell growth. Molecular and biochemical parasitology. 1998;94(1):87-97.

41. Chakrabarti D, Da Silva T, Barger J, Paquette S, Patel H, Patterson S, et al. Protein farnesyltransferase and protein prenylation inplasmodium falciparum. Journal of Biological Chemistry. 2002;277(44):42066-73.

42. Nallan L, Bauer KD, Bendale P, Rivas K, Yokoyama K, Hornéy CP, et al. Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. Journal of medicinal chemistry. 2005;48(11):3704-13.

43. Eastman RT, Buckner FS, Yokoyama K, Gelb MH, Van Voorhis WC. Thematic review series: Lipid posttranslational modifications. Fighting parasitic disease by blocking protein farnesylation. Journal of lipid research. 2006;47(2):233-40.

44. Ochocki JD, Distefano MD. Prenyltransferase inhibitors: Treating human ailments from cancer to parasitic infections. Medchemcomm. 2013;4(3):476-92.

45. Otto JC, Casey PJ. The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells (∗). Journal of Biological Chemistry. 1996;271(9):4569-72.

46. Koh C, Canini L, Dahari H, Zhao X, Uprichard SL, Haynes-Williams V, et al. Oral prenylation inhibition with lonafarnib in chronic hepatitis d infection: A proof-of-concept randomised, double-blind, placebo-controlled phase 2a trial. The Lancet Infectious diseases. 2015;15(10):1167-74.

47. Moores SL, Schaber MD, Mosser SD, Rands E, O'Hara MB, Garsky VM, et al. Sequence dependence of protein isoprenylation. The Journal of biological chemistry. 1991;266(22):14603-10.

48. Trueblood CE, Boyartchuk VL, Picologlou EA, Rozema D, Poulter CD, Rine J. The caax proteases, afc1p and rce1p, have overlapping but distinct substrate specificities. Molecular and cellular biology. 2000;20(12):4381-92.

49. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA. Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities. Journal of molecular biology. 2010;395(1):176-90.

50. London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O. Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS computational biology. 2011;7(10):e1002170.

51. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA. Context-dependent substrate recognition by protein farnesyltransferase. Biochemistry. 2009;48(8):1691-701.

52. Sousa SF, Coimbra JT, Paramos D, Pinto R, Guimarães RS, Teixeira V, et al. Molecular dynamics analysis of a series of 22 potential farnesyltransferase substrates containing a caax-motif. Journal of molecular modeling. 2013;19(2):673-88.

53. DeGraw AJ, Palsuledesai C, Ochocki JD, Dozier JK, Lenevich S, Rashidian M, et al. Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. Chemical biology & drug design. 2010;76(6):460-71.

54. Onono FO, Morgan MA, Spielmann HP, Andres DA, Subramanian T, Ganser A, et al. A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with western blotting. Molecular & cellular proteomics : MCP. 2010;9(4):742-51.

55. Palsuledesai CC, Ochocki JD, Kuhns MM, Wang YC, Warmka JK, Chernick DS, et al. Metabolic labeling with an alkyne-modified isoprenoid analog facilitates imaging and quantification of the prenylome in cells. ACS chemical biology. 2016;11(10):2820-8.

56. Finegold AA, Johnson DI, Farnsworth CC, Gelb MH, Judd SR, Glomset JA, et al. Protein geranylgeranyltransferase of saccharomyces cerevisiae is specific for cys-xaa-xaa-leu motif proteins and requires the cdc43 gene product but not the dpr1 gene product. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(10):4448-52.

57. Yokoyama K, McGeady P, Gelb MH. Mammalian protein geranylgeranyltransferase-i: Substrate specificity, kinetic mechanism, metal requirements, and affinity labeling. Biochemistry. 1995;34(4):1344-54.

58. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

59. Leung KF, Baron R, Ali BR, Magee AI, Seabra MC. Rab gtpases containing a caax motif are processed post-geranylgeranylation by proteolysis and methylation. The Journal of biological chemistry. 2007;282(2):1487-97.

60. Kilpatrick EL, Hildebrandt JD. Sequence dependence and differential expression of ggamma5 subunit isoforms of the heterotrimeric g proteins variably processed after prenylation in mammalian cells. The Journal of biological chemistry. 2007;282(19):14038-47.

61. Heilmeyer LM, Jr., Serwe M, Weber C, Metzger J, Hoffmann-Posorske E, Meyer HE. Farnesylcysteine, a constituent of the alpha and beta subunits of rabbit skeletal muscle phosphorylase kinase: Localization by conversion to s-ethylcysteine and by tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(20):9554-8.

62. Schmidt WK, Tam A, Fujimura-Kamada K, Michaelis S. Endoplasmic reticulum membrane localization of rce1p and ste24p, yeast proteases involved in carboxyl-terminal caax protein processing and amino-terminal a-factor cleavage. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(19):11175-80.

63. Tam A, Schmidt WK, Michaelis S. The multispanning membrane protein ste24p catalyzes caax proteolysis and nh2-terminal processing of the yeast a-factor precursor. The Journal of biological chemistry. 2001;276(50):46798-806.

64. Chen Y, Ma YT, Rando RR. Solubilization, partial purification, and affinity labeling of the membrane-bound isoprenylated protein endoprotease. Biochemistry. 1996;35(10):3227-37.

65. Boyartchuk VL, Ashby M, N. , Jasper R. Modulation of ras and a-factor function by carboxyl-terminal proteolysis. 1997.

66. Fujimura-Kamada K, Nouvet FJ, Michaelis S. A novel membrane-associated metalloprotease, ste24p, is required for the first step of nh2-terminal processing of the yeast a-factor precursor. The Journal of cell biology. 1997;136(2):271-85.

67. Boyartchuk VL, Rine J. Roles of prenyl protein proteases in maturation of saccharomyces cerevisiae a-factor. Genetics. 1998;150(1):95-101.

68. Tam A, Nouvet FJ, Fujimura-Kamada K, Slunt H, Sisodia SS, Michaelis S. Dual roles for ste24p in yeast a-factor maturation: Nh2-terminal proteolysis and cooh-terminal caax processing. The Journal of cell biology. 1998;142(3):635-49.

69. Hildebrandt ER, Davis DM, Deaton J, Krishnankutty RK, Lilla E, Schmidt WK. Topology of the yeast ras converting enzyme as inferred from cysteine accessibility studies. Biochemistry. 2013;52(38):6601-14.

70. Hampton SE, Dore TM, Schmidt WK. Rce1: Mechanism and inhibition. Critical reviews in biochemistry and molecular biology. 2018:1-18.

71. Nishii W, Muramatsu T, Kuchino Y, Yokoyama S, Takahashi K. Partial purification and characterization of a caax-motif-specific protease from bovine brain using a novel fluorometric assay. Journal of biochemistry. 1997;122(2):402-8.

72. Manolaridis I, Kulkarni K, Dodd RB, Ogasawara S, Zhang Z, Bineva G, et al. Mechanism of farnesylated caax protein processing by the intramembrane protease rce1. Nature. 2013;504(7479):301-5.

73. Ma YT, Gilbert BA, Rando RR. Inhibitors of the isoprenylated protein endoprotease. Biochemistry. 1993;32(9):2386-93.

74. Plummer LJ, Hildebrandt ER, Porter SB, Rogers VA, McCracken J, Schmidt WK. Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues. The Journal of biological chemistry. 2006;281(8):4596-605.

75. Hildebrandt ER, Arachea BT, Wiener MC, Schmidt WK. Ste24p mediates proteolysis of both isoprenylated and non-prenylated oligopeptides. The Journal of biological chemistry. 2016;291(27):14185-98.

76. Blanden MJ, Suazo KF, Hildebrandt ER, Hardgrove DS, Patel M, Saunders WP, et al. Efficient farnesylation of an extended c-terminal c(x)3x sequence motif expands the scope of the prenylated proteome. The Journal of biological chemistry. 2018;293(8):2770-85.

77. Ashok S, Hildebrandt ER, Ruiz CS, Hardgrove DS, Coreno DW, Schmidt WK, et al. Protein farnesyltransferase catalyzes unanticipated farnesylation and geranylgeranylation of shortened target sequences. Biochemistry. 2020;59(11):1149-62.

78. Otto JC, Kim E, Young SG, Casey PJ. Cloning and characterization of a mammalian prenyl protein-specific protease. The Journal of biological chemistry. 1999;274(13):8379-82.

79. Kim E, Ambroziak P, Otto JC, Taylor B, Ashby M, Shannon K, et al. Disruption of the mouse rce1 gene results in defective ras processing and mislocalization of ras within cells. The Journal of biological chemistry. 1999;274(13):8383-90.

80. Bergo MO, Ambroziak P, Gregory C, George A, Otto JC, Kim E, et al. Absence of the caax endoprotease rce1: Effects on cell growth and transformation. Molecular and cellular biology. 2002;22(1):171-81.

81. Wahlstrom AM, Cutts BA, Karlsson C, Andersson KM, Liu M, Sjogren AK, et al. Rce1 deficiency accelerates the development of k-ras-induced myeloproliferative disease. Blood. 2007;109(2):763-8.

82. Bergo MO, Lieu HD, Gavino BJ, Ambroziak P, Otto JC, Casey PJ, et al. On the physiological importance of endoproteolysis of caax proteins: Heart-specific rce1 knockout mice develop a lethal cardiomyopathy. The Journal of biological chemistry. 2004;279(6):4729-36.

83. Karlsson C, Akula MK, Staffas A, Cisowski J, Sayin VI, Ibrahim MX, et al. Knockout of the ras endoprotease rce1 accelerates myeloid leukemia by downregulating gadd45b. Leukemia. 2021;35(2):606-9.

84. Pryor EE, Horanyi PS, Clark KM, Fedoriw N, Connelly SM, Koszelak-Rosenblum M, et al. Structure of the integral membrane protein caax protease ste24p. Science. 2013;339(6127):1600-4.

85. Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, et al. The structural basis of zmpste24-dependent laminopathies. Science. 2013;339(6127):1604-7.

86. Goblirsch BR, Wiener MC. Ste24: An integral membrane protein zinc metalloprotease with provocative structure and emergent biology. Journal of molecular biology. 2020;432(18):5079-90.

87. Nie L, Spear E, Babatz TD, Quigley A, Dong YY, Chu A, et al. A new paradigm for prelamin a proteolytic processing by zmpste24: The upstream sy^ll cleavage occurs first and there is no caax processing by zmpste24. bioRxiv. 2020:2020.05.13.093849.

88. Ast T, Michaelis S, Schuldiner M. The protease ste24 clears clogged translocons. Cell. 2016;164(1-2):103-14.

89. Romano JD, Schmidt WK, Michaelis S. The saccharomyces cerevisiae prenylcysteine carboxyl methyltransferase ste14p is in the endoplasmic reticulum membrane. Molecular biology of the cell. 1998;9(8):2231-47.

90. Yang J, Kulkarni K, Manolaridis I, Zhang Z, Dodd RB, Mas-Droux C, et al. Mechanism of isoprenylcysteine carboxyl methylation from the crystal structure of the integral membrane methyltransferase icmt. Mol Cell. 2011;44(6):997-1004.

91. Diver MM, Pedi L, Koide A, Koide S, Long SB. Atomic structure of the eukaryotic intramembrane ras methyltransferase icmt. Nature. 2018;553(7689):526-9.

92. Romano JD, Michaelis S. Topological and mutational analysis of saccharomyces cerevisiae ste14p, founding member of the isoprenylcysteine carboxyl methyltransferase family. Molecular biology of the cell. 2001;12(7):1957-71.

93. Hodges HB, Zhou M, Haldar S, Anderson JL, Thompson DH, Hrycyna CA. Inhibition of membrane-associated methyltransferases by a cholesterol-based metal chelator. Bioconjugate chemistry. 2005;16(3):490-3.

94. Desrosiers RR, Nguyen QT, Béliveau R. The carboxyl methyltransferase modifying g proteins is a metalloenzyme. Biochemical and biophysical research communications. 1999;261(3):790-7.

95. Bergo MO, Leung GK, Ambroziak P, Otto JC, Casey PJ, Gomes AQ, et al. Isoprenylcysteine carboxyl methyltransferase deficiency in mice. Journal of Biological Chemistry. 2001;276(8):5841-5.

96. Bergo MO, Gavino BJ, Hong C, Beigneux AP, McMahon M, Casey PJ, et al. Inactivation of icmt inhibits transformation by oncogenic k-ras and b-raf. The Journal of clinical investigation. 2004;113(4):539-50.

97. Lau H, Tang J, Casey P, Wang M. Isoprenylcysteine carboxylmethyltransferase is critical for malignant transformation and tumor maintenance by all ras isoforms. Oncogene. 2017;36(27):3934-42.

98. Wahlstrom AM, Cutts BA, Liu M, Lindskog A, Karlsson C, Sjogren A-KM, et al. Inactivating icmt ameliorates k-ras–induced myeloproliferative disease. Blood, The Journal of the American Society of Hematology. 2008;112(4):1357-65.

99. Court H, Amoyel M, Hackman M, Lee KE, Xu R, Miller G, et al. Isoprenylcysteine carboxylmethyltransferase deficiency exacerbates kras-driven pancreatic neoplasia via notch suppression. The Journal of clinical investigation. 2013;123(11):4681-94.

100. Badar T, Cortes JE, Ravandi F, O'Brien S, Verstovsek S, Garcia-Manero G, et al. Phase i study of s-trans, trans-farnesylthiosalicylic acid (salirasib), a novel oral ras inhibitor in patients with refractory hematologic malignancies. Clinical Lymphoma Myeloma and Leukemia. 2015;15(7):433-8.e2.

101. Tsimberidou AM, Rudek MA, Hong D, Ng CS, Blair J, Goldsweig H, et al. Phase 1 first-in-human clinical study of s-trans, trans-farnesylthiosalicylic acid (salirasib) in patients with solid tumors. Cancer chemotherapy and pharmacology. 2009;65(2):235.

102. Furuse J, Kurata T, Okano N, Fujisaka Y, Naruge D, Shimizu T, et al. An early clinical trial of salirasib, an oral ras inhibitor, in japanese patients with relapsed/refractory solid tumors. Cancer chemotherapy and pharmacology. 2018;82(3):511-9.

103. Sekerdag E, Lüle S, Bozdağ Pehlivan S, Öztürk N, Kara A, Kaffashi A, et al. A potential non-invasive glioblastoma treatment: Nose-to-brain delivery of farnesylthiosalicylic acid incorporated hybrid nanoparticles. Journal of Controlled Release. 2017;261:187-98.

104. Caplan AJ, Douglas MG. Characterization of ydj1: A yeast homologue of the bacterial dnaj protein. The Journal of cell biology. 1991;114(4):609-21.

105. Caplan AJ, Tsai J, Casey PJ, Douglas MG. Farnesylation of ydj1p is required for function at elevated growth temperatures in saccharomyces cerevisiae. The Journal of biological chemistry. 1992;267(26):18890-5.

106. Emmanouilidis L, Schutz U, Tripsianes K, Madl T, Radke J, Rucktaschel R, et al. Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor pex19. Nature communications. 2017;8:14635.

107. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, et al. A novel role of farnesylation in targeting a mitotic checkpoint protein, human spindly, to kinetochores. The Journal of cell biology. 2015;208(7):881-96.

108. Mosalaganti S, Keller J, Altenfeld A, Winzker M, Rombaut P, Saur M, et al. Structure of the rzz complex and molecular basis of its interaction with spindly. The Journal of cell biology. 2017;216(4):961-81.

109. Ishimi Y, Yasuda H, Hirosumi J, Hanaoka F, Yamada M. A protein which facilitates assembly of nucleosome-like structures in vitro in mammalian cells. Journal of biochemistry. 1983;94(3):735-44.

110. Park YJ, Luger K. The structure of nucleosome assembly protein 1. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(5):1248-53.

111. Zlatanova J, Seebart C, Tomschik M. Nap1: Taking a closer look at a juggler protein of extraordinary skills. The FASEB Journal. 2007;21(7):1294-310.

112. Fujii-Nakata T, Ishimi Y, Okuda A, Kikuchi A. Functional analysis of nucleosome assembly protein, nap-1. The negatively charged cooh-terminal region is not necessary for the intrinsic assembly activity. The Journal of biological chemistry. 1992;267(29):20980-6.

113. Calvert MEK, Keck KM, Ptak C, Shabanowitz J, Hunt DF, Pemberton LF. Phosphorylation by casein kinase 2 regulates nap1 localization and function. Molecular and cellular biology. 2008;28(4):1313-25.

114. Regnard C, Desbruyères E, Huet JC, Beauvallet C, Pernollet JC, Eddé B. Polyglutamylation of nucleosome assembly proteins. The Journal of biological chemistry. 2000;275(21):15969-76.

115. Asahara H, Tartare-Deckert S, Nakagawa T, Ikehara T, Hirose F, Hunter T, et al. Dual roles of p300 in chromatin assembly and transcriptional activation in cooperation with nucleosome assembly protein 1 in vitro. Molecular and cellular biology. 2002;22(9):2974-83.

116. Huang Z-X, Zhao P, Zeng G-S, Wang Y-M, Sudbery I, Wang Y. Phosphoregulation of nap1 plays a role in septin ring dynamics and morphogenesis in candida albicans. mBio. 2014;5(1):e00915-13.

117. Galichet A, Gruissem W. Developmentally controlled farnesylation modulates atnap1;1 function in cell proliferation and cell expansion during arabidopsis leaf development. Plant Physiol. 2006;142(4):1412-26.

118. Kellogg DR, Murray AW. Nap1 acts with clb1 to perform mitotic functions and to suppress polar bud growth in budding yeast. The Journal of cell biology. 1995;130(3):675.

119. Kellogg DR, Kikuchi A, Fujii-Nakata T, Turck CW, Murray AW. Members of the nap/set family of proteins interact specifically with b-type cyclins. The Journal of cell biology. 1995;130(3):661-73.

120. Altman R, Kellogg D. Control of mitotic events by nap1 and the gin4 kinase. The Journal of cell biology. 1997;138(1):119-30.

121. Mortensen EM, McDonald H, Yates J, 3rd, Kellogg DR. Cell cycle-dependent assembly of a gin4-septin complex. Molecular biology of the cell. 2002;13(6):2091-105.

122. Ito T, Bulger M, Kobayashi R, Kadonaga JT. Drosophila nap-1 is a core histone chaperone that functions in atp-facilitated assembly of regularly spaced nucleosomal arrays. Molecular and cellular biology. 1996;16(6):3112-24.

123. Mosammaparast N, Jackson KR, Guo Y, Brame CJ, Shabanowitz J, Hunt DF, et al. Nuclear import of histone h2a and h2b is mediated by a network of karyopherins. The Journal of cell biology. 2001;153(2):251-62.

124. Mosammaparast N, Ewart CS, Pemberton LF. A role for nucleosome assembly protein 1 in the nuclear transport of histones h2a and h2b. The EMBO Journal. 2002;21(23):6527-38.

125. Ohkuni K, Shirahige K, Kikuchi A. Genome-wide expression analysis of nap1 in saccharomyces cerevisiae. Biochemical and biophysical research communications. 2003;306(1):5-9.

126. Mosammaparast N, Del Rosario BC, Pemberton LF. Modulation of histone deposition by the karyopherin kap114. Molecular and cellular biology. 2005;25(5):1764-78.

127. Del Rosario BC, Pemberton LF. Nap1 links transcription elongation, chromatin assembly, and messenger rnp complex biogenesis. Molecular and cellular biology. 2008;28(7):2113-24.

128. Xue YM, Kowalska AK, Grabowska K, Przybyt K, Cichewicz MA, Del Rosario BC, et al. Histone chaperones nap1 and vps75 regulate histone acetylation during transcription elongation. Molecular and cellular biology. 2013;33(8):1645-56.

129. D'Arcy S, Martin KW, Panchenko T, Chen X, Bergeron S, Stargell LA, et al. Chaperone nap1 shields histone surfaces used in a nucleosome and can put h2a-h2b in an unconventional tetrameric form. Mol Cell. 2013;51(5):662-77.

130. Aguilar-Gurrieri C, Larabi A, Vinayachandran V, Patel NA, Yen K, Reja R, et al. Structural evidence for nap1-dependent h2a-h2b deposition and nucleosome assembly. Embo j. 2016;35(13):1465-82.

131. Nakagawa T, Bulger M, Muramatsu M, Ito T. Multistep chromatin assembly on supercoiled plasmid DNA by nucleosome assembly protein-1 and atp-utilizing chromatin assembly and remodeling factor. The Journal of biological chemistry. 2001;276(29):27384-91.

132. Ishimi Y, Hirosumi J, Sato W, Sugasawa K, Yokota S, Hanaoka F, et al. Purification and initial characterization of a protein which facilitates assembly of nucleosome-like structure from mammalian cells. European journal of biochemistry. 1984;142(3):431-9.

133. Wagner G, Bancaud A, Quivy JP, Clapier C, Almouzni G, Viovy JL. Compaction kinetics on single dnas: Purified nucleosome reconstitution systems versus crude extract. Biophysical journal. 2005;89(5):3647-59.

134. Tóth KF, Mazurkiewicz J, Rippe K. Association states of nucleosome assembly protein 1 and its complexes with histones. The Journal of biological chemistry. 2005;280(16):15690-9.

135. McBryant SJ, Park YJ, Abernathy SM, Laybourn PJ, Nyborg JK, Luger K. Preferential binding of the histone (h3-h4)2 tetramer by nap1 is mediated by the amino-terminal histone tails. The Journal of biological chemistry. 2003;278(45):44574-83.

136. McBryant SJ, Peersen OB. Self-association of the yeast nucleosome assembly protein 1. Biochemistry. 2004;43(32):10592-9.

137. Stein V, Kubala MH, Steen J, Grimmond SM, Alexandrov K. Towards the systematic mapping and engineering of the protein prenylation machinery in saccharomyces cerevisiae. PloS one. 2015;10(3):e0120716.

138. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS computational biology. 2007;3(4):e66.

139. Jones JM, Morrell JC, Gould SJ. Pex19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. The Journal of cell biology. 2004;164(1):57-67.

140. Collins SP, Reoma JL, Gamm DM, Uhler MD. Lkb1, a novel serine/threonine protein kinase and potential tumour suppressor, is phosphorylated by camp-dependent protein kinase (pka) and prenylated in vivo. The Biochemical journal. 2000;345 Pt 3(Pt 3):673-80.

141. Ashar HR, James L, Gray K, Carr D, Black S, Armstrong L, et al. Farnesyl transferase inhibitors block the farnesylation of cenp-e and cenp-f and alter the association of cenp-e with the microtubules. The Journal of biological chemistry. 2000;275(39):30451-7.

CHAPTER 2

Protein isoprenylation in yeast targets COOH-terminal sequences not adhering to the CaaX consensus

Berger, B.M.*, Kim, J.H.*, Hildebrandt, E.R., Davis, I.C., Morgan, M.C., Hougland, J.L., Schmidt, W.K. 2018. Genetics, 210 1301-1316. Reprinted with permission of publisher.

*These authors contributed equally to this work.

ABSTRACT

Protein isoprenylation targets a subset of COOH-terminal Cxxx tetrapeptide sequences that has been operationally defined as a CaaX motif. The specificity of the farnesyl transferase toward each of the possible 8000 combinations of Cxxx sequences, however, remains largely unresolved. In part, it has been difficult to consolidate results stemming from *in vitro* and *in silico* approaches that yield a wider array of prenylatable sequences relative to those known *in vivo*. We have investigated whether this disconnect results from the multi-step complexity of post-translational modification that occurs *in vivo* to CaaX proteins. For example, the Ras GTPases undergo isoprenylation followed by additional proteolysis and carboxyl methylation events at the COOH-terminus. By contrast, *Saccharomyces cerevisiae* Hsp40 Ydj1 is isoprenylated but not subject to additional modification. In fact, additional modifications are detrimental to Ydj1 activity *in vivo*. We have taken advantage of the properties of Ydj1 and a Ydj1-dependent growth assay to identify sequences that permit Ydj1 isoprenylation *in vivo* while simultaneously selecting against non-prenylatable and more extensively modified sequences. The recovered sequences are largely non-overlapping with those previously identified using an *in vivo* Ras-based yeast reporter. Moreover, most of the sequences are not readily predicted as isoprenylation targets by existing prediction algorithms. Our results reveal that the yeast CaaX-type prenyltransferases can utilize a range of sequence combinations that extend beyond the traditional constraints for CaaX proteins, which implies that more proteins may be isoprenylated than previously considered.

Introduction

CaaX proteins are abundant eukaryotic proteins with diverse biological functions. They are operationally defined by a COOH-terminal CaaX motif that is subject to an ordered series of post-translational modifications involving covalent attachment of a C15 (farnesyl) or C20 (geranylgeranyl) isoprenoid to cysteine, endoproteolysis to remove aaX, and carboxyl methylation **(Figure 2.1A**) (1, 2). More complexity can occur in the form of additional modification to certain CaaX proteins (e.g., palmitoylation). The multi-step canonical modification of CaaX proteins has been extensively studied in vivo using relatively few CaaX proteins, with Ras-related GTPases being the most often cited. Modifications modulate CaaX protein function and/or localization, and there is intense interest on developing therapeutic inhibitors for all steps of the pathway (e.g., prenyl transferase inhibitors, protease inhibitors, methyltransferase inhibitors) (1, 3-7).

The isoprenylation step of the canonical modification pathway has received the most investigative attention. The two isoprenoid transferases targeting CaaX sequences are the farnesyl transferase (FTase) and geranylgeranyl transferase (GGTase-I). Their specificities toward CaaX motifs clearly involve sequence determinants, but these are not yet fully resolved despite intensive investigations using *in vivo, in vitro* and *in silico* methods. Historically, the determinants have been defined as a cysteine (required), followed by two aliphatic amino acids, and one of several amino acids at the last position. While CaaX sequences can be recognized by both FTase and GGTase-I, geranylgeranylation is reportedly enhanced for sequences ending Leu or Phe (8-13). The aliphatic requirement at a_1 and a_2 positions of the CaaX motif should not be viewed

as rigid, however, because prenylatable sequences clearly fall outside the traditional consensus. Examples include yeast Ydj1 (CASQ) and human Stk11/Lkb1 (CKQQ), among many others (14, 15). Increasing evidence indicates that non-canonical sequences are modified by isoprenylation but are not cleaved and carboxymethylated (i.e. shunted products) (16).

Comparisons of known prenylated sequences and subsequent systematic amino acid substitution analysis of associated CaaX sequences initially contributed to the development of rules for prenyltransferase selectivity (9, 17-24). This *in vivo* work, much of it originally performed using the yeast system, suggested enrichment of aliphatic residues at a_1 and a_2, more so at a_2, and significant variation at X, leading to the canonical definition now widely accepted. A recent study utilizing a high-throughput yeast genetic approach and a Ras reporter revealed a similar bias for aliphatic residues at a_1 and a_2, again more so at a_2 (25). Additional rules have derived from studies involving mammalian FTase mutants with altered selectivity (26), computational modeling of substrates within the mammalian FTase active site (27, 28), and reactivity of mammalian FTase against arrayed peptide sets (13, 29-31). Collectively, these *ex vivo* approaches suggest that reactive sequences can significantly deviate from the historical CaaX definition. The rules proposed to govern FTase selectivity have been incorporated into predictive algorithms (12, 28). These algorithms often fail to predict, however, non-canonical motifs found on well-documented prenylated CaaX proteins (e.g., CKQQ present on human Stk11/Lkb1, human Nap1L1, and yeast Pex19; CASQ present on yeast Ydj1).

A striking difference between *in vivo* and *ex vivo* studies of FTase specificity is the strong enrichment of branched chain residues (BCAs) (i.e., Ile, Leu, Val) at the a_2 position in many *in vivo* studies. We hypothesize that the results of *in vivo* studies are inherently biased because reporters used in those studies (e.g., **a**-factor; Ras and Ras-related GTPases) use the canonical prenylation pathway and are subject to multiple post-translational modifications (i.e., Ras-like modifications). In such situations, outputs are governed not only by isoprenylation efficiency but also proteolysis and carboxyl methylation efficiencies. At least two *in vivo* approaches are being developed that minimize the impact of proteolysis and methylation on specificity studies. One involves the combined use of prenylation probes (e.g., alkyne farnesyl analogs) and mass spectrometry methods to identify prenylated proteins (i.e., the prenylome) (32-35). Application of such technology has begun to confirm known and identify novel prenylated proteins, but direct detection of the prenyl group itself on these proteins remains a key challenge. Moreover, there may not be exact equivalency of chemical probes and farnesyl diphosphate when used by the FTase (36). A second approach that we describe in this study involves the use of the yeast Hsp40 Ydj1 chaperone as a genetic reporter (14, 37). An advantage of Ydj1 over previously used *in vivo* reporters is that it is an uncleaved CaaX protein (16). It is thus useful for identifying prenylatable sequences without concern for proteolysis and methylation, which are actually detrimental to Ydj1 activity (16). We report the use of Ydj1-based screening to recover sequences that support prenylation of Ydj1, which upon evaluation largely fail to match the operationally defined CaaX consensus, supporting a broader specificity than anticipated for the yeast CaaX-type prenyltransferases.

Materials and Methods

Yeast strains

Strains used in this study are listed in **Table S2.1**. Most have been previously described, several were isolated from a commercial *MAT***a** haploid genomic deletion library, and a couple were created for this study (16, 38-41). Plasmids were introduced into strains via a lithium acetate-based transformation procedure (42).

yWS2542 (*MAT***a** *his3 leu2 met15 ura3 ydj1::NAT^R ram1::KAN^R*) was created by replacing the *YDJ1* open reading frame with the nourseothricin resistance cassette (NAT^R) in yWS1632 (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ram1::KAN^R*). This was accomplished by transformation of the strain with an extensive digestion of pWS1623 (*BamHI*, *HindIII*, *PvuII*) and selection on YPD containing 100 µg/ml nourseothricin. yWS2544 (*MAT***a** *his3 leu2 met15 ura3 ydj1::NAT^R*) was made in similar fashion using BY4741 as the parent strain. The gene replacements were confirmed by PCR and Western blot against Ydj1.

yWS304 (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ydj1::KAN^R*) and yWS1635 (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ydj1::KAN^R ste14::KAN^R*) were used for unbiased selection screens to identify Ydj1-Cxxx mutants that could support thermotolerance. Unless otherwise noted, strains were routinely propagated at 30 °C or room temperature if temperature sensitive (23-25 °C) on either YPD or selective media as appropriate.

Plasmids

The plasmids used in this study, other than those recovered by screening, are listed in **Table S2.2**. Plasmids were either previously reported or constructed by standard molecular methods. All new plasmids created for this study were analyzed by restriction digest and DNA sequencing (Genewiz, South Plainfield, NJ or Eurofins Genomics, Louisville, KY) to verify proper sequence of the entire open reading frame (*MFA1*) or the 3´-end encoding the Cxxx sequence (*YDJ1;* approximately 900 bp 5´ from stop codon). Plasmids encoding Ydj1 and **a**-factor with specific Cxxx sequences were derived by the same approach using pWS1132 (*CEN URA3 YDJ1-SASQ*) and pWS610 (*CEN LEU2 MFA1*), respectively, as the parent plasmids. Mutagenic oligonucleotides were designed to encode the desired Cxxx sequences and used to produce PCR products compatible with PCR-directed, plasmid-based recombination methods (43) (**Table S2.3**). The PCR products have homology to the appropriate parent plasmid in regions flanking the intended mutation site. The parent plasmids were readied for recombination-based gap repair by digestion with *NheI* (pWS1132) or *MluI* and *SphI* (pWS610). Following co-transformation of digested plasmid and PCR product into yeast and appropriate selection (i.e., SC-uracil or SC-leucine), plasmids were recovered from individual yeast colonies, evaluated by restriction enzyme mapping, and sequenced to confirm the identity of recovered plasmids. pWS1623 was made in similar fashion using an *NheI* and *BsaBI* digestion of pWS1132 and a PCR product encoding NAT^R to replace the entirety of the *YDJ1* ORF.

Plasmids encoding Ydj1 with random Cxxx sequences for the purpose of thermotolerance selection were also created by PCR-directed, plasmid-based recombination. In this case, the mutagenic oligonucleotide used for PCR was synthesized to contain random nucleotides for the xxx codons (oWS986). The COOH-terminal sequences associated with Ydj1 mutants recovered by the selection scheme are listed in **Table S2.4**.

Thermotolerance selection

Yeast deficient for *YDJ1* (*ydj1*Δ) were cultured to late log phase, harvested, and co-transformed with *NheI*-linearized pWS1132 and PCR product; control transformations with each DNA component alone were also prepared. For the co-transformed condition, multiple replicates were prepared. A portion of one transformation mix (10%) was plated onto SC-uracil and incubated at room temperature (23-25 °C). This allowed for an estimation of the number of recombinant plasmids created by the procedure. The remaining portion of the transformation mix (90%) and replicate transformation mixes (100%) were plated onto YPD and incubated at 40 °C for 48 hours followed by growth at room temperature (24-36 hours) to facilitate better visual identification of colonies. Surviving colonies were amplified as liquid cultures in SC-uracil, and cell pellets used for isolation of plasmids via sequential yeast and *E. coli* DNA miniprep protocols. Several selection rounds were performed over the course of the study to accumulate the plasmids. In these experiments, both yWS304 and yWS1635 were used as the *ydj1*Δ background.

We used the above selection method to directly assess the impact of high temperature on transformation and recombination efficiencies by plating out equal portions of a single transformation mix in replicate onto SC-uracil (23-25 °C) and YPD (40 °C incubation). The transformation mix was prepared using yWS304, *NheI*-linearized pWS1132, and a PCR product designed to encode a thermotolerant Ydj1 mutant (AQCASQ). The fraction of colonies observed at 40 °C over 23-25 °C was calculated and applied as a correction factor to the initial estimate of colonies screened.

Temperature sensitivity assay and scoring of Ydj1 Cxxx variants

Thermotolerance assays were performed as previously described with minor modifications (16, 44). In brief, plasmid-transformed strains expressing Ydj1 Cxxx variants were cultured to saturation (25 °C, 24-30 hours) in SC-uracil liquid media, serially diluted into H_2O (10-fold dilutions), and replica pinned in duplicate onto YPD solid media. For assays involving temperature and time optimizations, saturated cultures were diluted at fixed dilutions into YPD and spotted using a multi-channel pipettor; the dilutions are specified in the associated figure legend. Strains expressing unmodified Ydj1 (SASQ), cleaved and carboxymethylated Ydj1 (CVIA), and shunted Ydj1 (CASQ) were typically included as controls, although the combination of controls pinned/spotted onto YPD plates varied between sets of mutants evaluated; shunted refers to CaaX motifs that are isoprenylated but not cleaved and carboxymethylated (16). Plates were typically incubated at various temperatures after an appropriate time of growth: 25 °C for 72 hours; 37 °C for 48 hours; 40 °C and 41 °C for 72 hours plus 24 hours at non-restrictive temperature to allow better visualization of micro-colonies;

alterations to these temperature and time schedules are noted in appropriate figure legends.

On average, 4.25 replicates were evaluated for each strain expressing a Ydj1 mutant under the conditions described above; the range was 3-8 replicates. Four independent observers scored the qualitative growth phenotype of replicates using a 1-5 score range in single-blind fashion. The scores were determined relative to reference strains expressing Ydj1 (SASQ), Ydj1 (CVIA) and Ydj1 (CASQ), which were standardly assigned scores of 1 (no growth at 40 °C and 41 °C), 3 (weak growth at 40 °C and 41°C), and 5 (strong growth at 40 °C and 41°C), respectively. Intermediate phenotypic scores of 2 and 4 were allowed for Ydj1 mutants. An average thermotolerance (T) score and standard deviation were calculated for each reference and mutant from the indicated number of replicates (*n*) reported with each T score. Because reference strains were typically included in each experiment, the number of replicates for these is much higher than that of mutants examined.

Yeast lysate preparations for SDS-PAGE analysis and immunoblot

Yeast were cultured to log phase (A_{600} 0.75-1.0) in selective SC-uracil at 25 °C unless otherwise noted. Cell pellets of equal mass were harvested by centrifugation, washed with water, and processed by alkaline hydrolysis and TCA precipitation (45). Total cell precipitates were resuspended in urea-containing Sample Buffer (250 mM Tris, 6 M urea, 5% β-mercaptoethanol, 4% SDS, 0.01% bromophenol blue, pH 8), and analyzed by SDS-PAGE and immunoblot. Blots were processed according to standard protocols

using appropriate dilutions of rabbit anti-Ydj1 (courtesy of Dr. Avrom Caplan) and HRP-conjugated donkey or goat anti-rabbit antibodies (GE Healthcare, Little Chalfont, UK; Kindle Biosciences, Greenwich, CT). Antibody dilutions were prepared using TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween-20, pH 7.5) containing 1% milk (w/v). Immune complexes on blots were detected using X-ray film after treatment with HyGLO development solution (Denville Scientific, South Plainfield, NJ) or using a KwikQuant Imager at multiple exposure times after treatment with the KwikQuant Western Blot Detection Kit (Kindle Biosciences).

Yeast mating assay

Qualitative and quantitative yeast mating assays were performed as previously described except that mating mixtures were pinned instead of spotted with a multi-channel pipettor (16, 45, 46). In brief, *MAT***a** strains expressing various **a**-factor mutants were cultured to saturation SC-leucine liquid media; the *MAT*α (IH1793) was cultured in parallel in YPD liquid media. Saturated cultures were normalized to an A_{600} value 1.0 ± 0.05 using appropriate fresh media, and normalized cell suspensions were mixed 1:9 (*MAT***a**: *MAT*α) in individual wells of a 96-well plate. The mating mixtures were further subject to 10-fold serial dilution within the 96-well plate using the normalized *MAT*α cell suspension as the diluent. For qualitative analysis, each diluted series was pinned in duplicate onto SC-lysine and minimal SD solid media. For quantitative analysis, equivalent volumes of empirically identified mixtures from the dilution series were spread onto SC-lysine and SD plates in duplicate, such that individual colony density was projected to be 50-200 colonies per plate the dilution mixture used varied between

all samples. The SC-lysine cell count reports on the total number of mating competent cells (i.e., *MAT***a** haploid cells), while the SD cell count reports on the number of mating competent cells that underwent mating events (i.e., diploids). To cross-compare values obtained, the colony counts were mathematically corrected for the dilution evaluated in order to estimate the number of colony forming units (CFUs) in each undiluted sample. The CFU values were used to determine mating efficiency (CFU_{SD} / $CFU_{SC\text{-}lysine}$) relative to a positive control expressing wildtype **a**-factor that was operational defined as having 100% mating efficiency (i.e., SM2331 transformed with pWS610; Tables S2.1 and S2.2). For both qualitative and quantitative analyses, plates were typically incubated 3 days at 30 °C.

Digital imaging of yeast plates and immunoblots

A Cannon flat-bed scanner was used to image plates and X-ray films (300 dpi; grayscale; TIFF format). Plates were scanned face down without lids using a black background; films were scanned using a white background. Some immunoblot images were captured using a KwikQuant Imager system (TIFF format). Digitized images of plates and immunoblots were imported into Photoshop for minor adjustments (i.e., image rotation, contrast, cropping, etc.) then copied to PowerPoint for final figure assembly. Contrast settings were adjusted within Photoshop to be identical for all plate images and to maximize dynamic range of signal; contrast settings for film-based immunoblot images were subject to Photoshop's auto-contrast function; the contrast settings for KwikQuant-based images were unaltered.

Amino acid frequency analysis

For our analyses, the entire set of Ydj1-based sequences was always evaluated (n=153). For the Ras-based sequences, the high probability sequences (i.e. enrichment score >3) were culled to eliminate low confidence sequences as suggested by the authors of the original study, which created a reduced set of high probability sequences (n=369) (25). For Weblogo-based analyses, appropriate groupings of Ydj1 and Ras-based sequences were uploaded to the WebLogo server (http://weblogo.berkeley.edu/logo.cgi) and analyzed for amino acid frequency using a custom color scheme (47). Cys was set to blue; polar charged amino acids were set to green (Asp, Arg, Glu, His and Lys); polar uncharged residues were set to black (Asn, Gln, Ser, Thr and Tyr); branched chain amino acids (BCAs) were set to red (Ile, Leu and Val); all other residues were set to purple (Ala, Gly, Met, Phe, Pro and Trp). For bar graphs, the number of occurrences of a specific amino acid at each position of the Ydj1-based sequences was normalized to the number of codons for that particular amino acid. Normalization was not applied to Ras-based sequences because this issue was addressed by the study design. High and low frequency amino acids for both sets of sequences were defined as those with normalized frequencies outside a 95% confidence level relative to the mean frequency for all amino acids at each position.

Prenylation Predictions

For analyses using the Prenylation Prediction Suite algorithm (PrePS; http://mendel.imp.ac.at/PrePS), each Cxxx sequence was evaluated in the context of a 26 amino acid window representing the COOH terminus of a protein (12). While PrePS

requires a minimal length of 15 amino acids for analysis, we used 26 to be somewhat consistent with our previous amino acid frequency analysis of the COOH-terminal region of Ydj1 homologs where a window of 25 amino acids was evaluated (16). Ydj1-derived sequences were evaluated in the context of Ydj1 (i.e., RASRGGANYDSDEEEQGGEGVQCxxx). Ras-derived sequences and the full set of 8000 Cxxx sequences were evaluated in the context of human H-Ras (i.e., RQHKLRKLNPPDESGPGCMSCKCxxx). The Cxxx sequences associated with the set of 89 yeast Cxxx proteins identified in the *Saccharomyces* Genome database (SGD; https://www.yeastgenome.org) were evaluated in the context of the parental protein. Sequences were binned into groups based on their PrePS values for predicted probability of farnesylation: highly probable (scores greater than 0); ambiguous (-2 to 0); weakly predicted (less than -2) (12). For predictions using FlexPepBind scores, each Ydj1 and Ras-based sequence was associated with its FlexPepBind score, and sequences binned into groups based on predicted probability of farnesylation: highly probable (scores less than -1.1); ambiguous (-1.1 to -0.4); weakly predicted (greater than -0.4) (28).

Our in-house prenylation and cleavage prediction rules were based on a simple point system involving assessment of the amino acid at each position of the Cxxx sequence. Prenylation potential was scored using a negative point scale where disfavored amino acids were counted. One negative point was assigned when the amino acid was low frequency in both Ydj1 and Ras-based sets of sequences (e.g., Phe at x_1); one extra negative point was assigned when it was absent in both (e.g., Lys at x_2) (see **Table**

2.2). The potential range of scores with this method was 0 to -5. Sequences with scores of 0 were categorized as having a strong probability of prenylation, those with scores of -1 were categorized as having ambiguous potential, and those scoring -2 points or less were considered to have weak prenylation potential. Cleavage potential was scored using a mixture of positive and negative point scales where only the Ras-based data set was considered. One positive point was assigned when the amino acid was high frequency (e.g., Ala at x_1). Negative points were assigned as for the prenylation rule (e.g., Phe at x_1), also including the extra count for absent residues (e.g., Lys at x_2). The potential range of scores with this method was 3 to -5. Sequences with scores of 1 or more were categorized as having a strong probability of cleavage, and all other sequences were categorized as having weak probability.

Data Availability

Strains and plasmids are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Supplemental Material is deposited at figshare portal: https://doi.org/10.25386/genetics.7075025

Results

Multiple Cxxx sequences can sustain Ydj1-dependent thermotolerance

Farnesylation of Ydj1 is required for high temperature growth of yeast (i.e. thermotolerance) and mitochondrial import (48-51). Replacing the normally uncleaved prenylation motif of Ydj1 (CASQ) with cleavable motifs (CTLM or CVIA) alters yeast

thermotolerance, the ability to over-express Ydj1, and Ydj1 subcellular localization (16). These effects are correlated with the COOH-terminal cleavage state of the reporter rather than altered protein expression, stability or farnesylation. It remains unclear why optimal Ydj1 thermotolerance function requires an uncleaved COOH-terminus. Nevertheless, Ydj1 serves as a unique reporter to investigate protein prenylation because it does not require subsequent proteolytic and methylation modifications associated with the canonical modification pathway (i.e., Ras-like modifications).

The last six residues of Ydj1 are conserved across species (16) (**Figure 2.1B**; **Figure S2.1**). The last four amino acids of the sequence form the CaaX motif that we operationally refer to as a Cxxx motif in this study. To initially determine whether any of the residues in this region, besides cysteine, contributed to the thermotolerance function of Ydj1, substitution mutations within the conserved COOH-terminal region were created and thermotolerance profiles examined after plasmid-based reintroduction of the mutants into *ydj1*Δ yeast. For one set of mutants, alanine was substituted at every conserved position, except at x_1 of the Cxxx sequence where glutamine was used to replace the naturally occurring alanine at that position. In a second set of mutants, the x_2 position was varied with aliphatic amino acids isoleucine, leucine, methionine, and valine in an effort to make the sequence more canonical (i.e., aliphatic residues at both x_1 and x_2). The substitution mutants all supported thermotolerance behavior at 40 °C, although subtle growth pattern differences were observed at 41 °C (**Figure 2.1C**). Importantly, the mutants were all more thermotolerant at higher temperatures than non-farnesylated Ydj1 (i.e., SASQ) or Ydj1 that was fully modified in a manner typically

associated with traditional CaaX proteins (i.e., CVIA). For the latter, colonies were typically smaller at higher temperatures and colony growth was less robust and reproducible at 41 °C between experiments.

With the exception of the non-farnesylated Ydj1 mutant (i.e., SASQ), Ydj1 substitution mutants appeared fully isoprenylated as judged by a gel-shift assay (**Figure 2.1D**). In this gel-shift assay, farnesylated Ydj1 migrates faster than non-prenylated Ydj1 generated through either mutation of the Cxxx motif (i.e., SASQ) or expression of Ydj1 in a farnesylation-defective yeast background (i.e., *ram1*Δ). These results were interpreted to indicate that multiple Cxxx motifs can promote Ydj1 farnesylation and support Ydj1-dependent thermotolerance. It remains unclear why Ydj1 family members have a conserved COOH-terminal sequence, but we speculate that these residues provide optimal functionality for at least one of the many roles attributed to Ydj1 (e.g., protein translocation, protein folding, prion clearance, etc.) (37, 50, 52-54).

Unbiased identification of sequences that support Ydj1-based thermotolerance

The ability of Ydj1 to support thermotolerance and be isoprenylated in the context of different Cxxx sequences suggested that many sequences might be able to promote this behavior. This led us to hypothesize that the thermotolerance profile of Ydj1 could be used as a genetic reporter to identify the breadth of Cxxx sequences capable of supporting thermotolerance. Moreover, we predicted that carefully selecting thermotolerance conditions would allow for genetic enrichment of shunted Cxxx motifs (i.e., prenylated but not cleaved and carboxyl methylated) over non-prenylated and fully

modified sequences (e.g., SASQ and CVIA, respectively). We therefore designed a high temperature selection strategy to enrich for shunted Ydj1 Cxxx mutants. At the highest temperatures applied, the selection strategy prevented growth of non-prenylated Ydj1 (i.e., SASQ) and forced slow growth of fully modified Ydj1 (i.e., CVIA and CTLM) relative to shunted Ydj1 (i.e., CASQ) (**Figure 2.2A-B**).

We next devised a strategy to create and evaluate thermotolerance of a library of Ydj1 Cxxx mutants *in vivo*. In brief, *ydj1*Δ yeast was co-transformed with a linearized yeast expression vector encoding non-prenylated Ydj1 (i.e., SASQ) and a library of PCR products encoding Cxxx sequences. This resulted in recombination events between the DNA fragments, leading to the formation of a library of Ydj1 mutants having the potential to encode all 8000 possible COOH-terminal tetrapeptide combinations of the form Cxxx, where C is Cys and "x" is any amino acid (**Figure 2.2C**) (43). Importantly, this strategy allowed for immediate selection of colonies with thermotolerant properties. We initially estimated that we evaluated ~480,000 recombination events, which were enough to achieve near complete coverage of all Cxxx permutations (estimated 99.9% completeness; ~1% probability that all Cxxx sequences were sampled) (55). This value was revised to ~67,200 recombination events (93.5% completeness; ~0% probability of full coverage) when it was determined that the number of colonies capable of forming at higher temperature was ~14% that observed at room temperature, which we infer is due to reductions in transformation and/or recombination efficiency.

Our selection strategy yielded 172 thermotolerant colonies from which plasmids were recovered and subject to DNA sequencing. One of the plasmids encoded wildtype Ydj1 (i.e., CASQ) thus validating the design of the screen. A set of 153 plasmids encoding unique Ydj1 Cxxx sequences, inclusive of the wildtype sequence, was defined by eliminating a small number of sequences recovered multiple times. The unique set of plasmids was re-introduced into *ydj1*Δ yeast for more detailed analyses of thermotolerance (**Figures 2.2D, S2.2**). This analysis yielded a thermotolerance (T) score for each Ydj1 Cxxx variant, where higher scores were associated with better thermotolerant behavior (see *Methods* for details on the scoring rubric). We binned the Cxxx variants based on their T scores (**Figure 2.2E, Table 2.1**). The CASQ-like group formed the largest cohort and displayed phenotypic growth similar to that of wildtype Ydj1 (CASQ) (54% of hits; T score 4.5-5.0). These Cxxx variants supported strong thermotolerance (i.e., growth). The CASQ-like (weak) group was next largest (37%; T score 3.5-4.5.). These Cxxx variants were phenotypically less thermotolerant than wildtype Ydj1 but more tolerant than cleaved Ydj1 (i.e., CVIA). The CVIA-like group accounted for a small percentage of total hits (9%; T score 2.5-3.5). These Cxxx variants behaved much like cleaved Ydj1 (i.e., CVIA). The CVIA-like (weak) group was formed by a single Cxxx variant (0.6%; T-score 2.47). Overall, our scoring analysis indicated that 90% of Cxxx variants were categorized as having thermotolerance profiles better than that of cleaved Ydj1 (i.e., CVIA).

Weaker thermotolerance phenotypes are associated with prenylation defects and enhanced cleavage propensity

To investigate the extent of prenylation associated with sequences conferring the strongest thermotolerance phenotype, we randomly identified 10% of the Ydj1 Cxxx variants in each of the CASQ-like and CASQ-like (weak) phenotypic categories and assessed prenylation by gel-shift assay (**Figure 2.3A**). An online algorithm (Research Randomizer; https://www.randomizer.org/) was used to randomly identify the 10% subsets (8 and 6 sequences per category, respectively) (56). Most of these sequences (9 of 14 evaluated) presented as a single band with the same mobility as farnesylated Ydj1. Several sequences presented as doublet bands, indicative of incomplete prenylation. The prenylated species (i.e., lower band) appeared qualitatively stronger than that of the non-prenylated species (i.e., upper band) in most instances, with the exception of CDFI where the bands were qualitatively assessed to be about equal intensity. Thus, partial prenylation of 50% or greater appears to be sufficient to support Ydj1-dependent thermotolerance. We also performed gel-shift studies on the 15 sequences conferring the weakest thermotolerance phenotypes – CVIA-like and CVIA-like (weak) (**Figure 2.3B**). The CWGG mutant was the only mutant in this set to present with a doublet pattern, with the non-prenylated species being the major band. Thus, incomplete prenylation most likely explains the weak thermotolerance profile of this sequence. Considering all our data, however, it appears that the completeness of isoprenylation cannot account for the range of thermotolerance behaviors observed for mutant sequences.

We determined that the majority of the 15 CVIA-like and CVIA-like (weak) mutants were fully prenylated because they presented as a single band with mobility similar to prenylated Ydj1. We thus hypothesized that the reduced thermotolerance associated with these mutants was likely due to the presence of cleavable sequences (i.e., Ras-like). To assess cleavage status, we first evaluated the thermotolerant behavior of *ydj1*Δ *ste14*Δ yeast expressing the 15 Ydj1 Cxxx variants. This genetic background lacks the isoprenylcysteine carboxylmethyl transferase and improves the thermotolerance profiles of yeast expressing cleavable Ydj1 Cxxx variants (16). Indeed, thermotolerance profiles improved for all but Ydj1 CWGG (**Figure S2.3**). To examine cleavage status more directly, we next evaluated the sequences in the context of the yeast **a**-factor mating pheromone (**Figure 2.3C**). The biological activity of **a**-factor *in vivo* requires it to be farnesylated, cleaved, and carboxylmethylated (39, 57-59). Defects in any one of these steps reduce bioactivity as measured through a pheromone-based biological mating assay. Through both qualitative and quantitative mating assays, we observed that the bioactivities of the **a**-factor Cxxx mutants could be categorized into two groups. Category I mutants formed the largest group and had substantial bioactivity relative to wildtype **a**-factor (i.e., 10% or greater). This observation indicates that these sequences are susceptible to cleavage. We suspect that the varied bioactivities of category I mutants reflect differences in their cleavage efficiencies or perhaps alternative geranylgeranylation of these sequences. Four-fold reduced bioactivity has been reported for synthetic geranylgeranylated **a**-factor, which can also be produced *in vivo* in the context of the CVIL sequence (60). Category II mutants had very limited or no bioactivity (<1%) and were represented by CALL, CWGG, CAGF, and CSFN. Whereas

the absence of bioactivity associated with CWGG is most likely due to a prenylation defect, this does not readily explain the loss of bioactivity observed for the other category II sequences. These sequences appear to be fully prenylated and cleaved in the context of Ydj1, thus we speculate that they may have a prenylation or cleavage defect in the context of **a**-factor. Such defects would negatively impact multiple downstream steps required for pheromone bioactivity that depend on efficient initial farnesylation and cleavage (e.g., Ste14-mediated carboxylmethylation; Ste6-dependent export; Ste3 receptor interaction).

Ydj1-based thermotolerant behavior may involve alternative isoprenylation

Yeast thermotolerance depends on the farnesylation of Ydj1. It is unclear whether geranylgeranylation can substitute in this capacity. We were especially interested in sequences ending in Leu or Phe, which reportedly confer GGTase-I reactivity (8-13, 61). To address this potential, we investigated whether the 10 Ydj1 CxxL and CxxF variants recovered through screening could be geranylgeranylated and promote thermotolerance in the absence of FTase activity (**Figure 2.4A**).

The two CxxF sequences, CAGF and CIGF, both displayed weak thermotolerance. We also analyzed several sequences similar to CAGF that were recovered by screening (i.e., CAGx), and these did not demonstrate thermotolerance. More substantial thermotolerance was observed for CxxL sequences, including CRPL and CAPL that only differ at the x_1 position. By contrast to CAPL, the very similar CAPQ sequence recovered by screening did not support thermotolerance. The remaining CxxL

sequences were part of subsets CALx, CGLx and CVGx. Among these subsets, those with Leu or Tyr at x_3 generally displayed minor thermotolerance. Leu at the x_3 position, however, was not always a positive predictor of thermotolerance in the absence of FTase activity (e.g., CVCL). Collectively, these results indicate that Ydj1 Cxx(L/F/Y) sequences , where Leu is potentially more favorably than Phe or Tyr, are susceptible to prenylation in the absence of FTase, consistent with previous reports of GGTase-I selectivity being influenced by the x_3 position (11). But it is not clear under our screening conditions whether geranylgeranylation is competing with farnesylation for these sequences or only occurs when farnesylation is genetically disrupted (i.e., *ram1*Δ background-specific). Thus, we cannot exclude the possibility that some sequences recovered by thermotolerance screening are indeed geranylgeranylated.

We also examined the properties of potentially geranylgeranylated Ydj1 Cxxx variants by gel-shift assay using lysates prepared from yeast cultured at elevated temperature (**Figure 2.4B**). GGTase-I activity (Cdc43p/Ram2p) is essential, so we could not express variants in a strain background lacking this activity. Instead, we expressed variants in the FTase-defective background (i.e., *ram1*Δ) and reasoned that any shifted bands still present in this background were a consequence of GGTase-I activity. We focused on the subset of four Ydj1 Cxxx variants that supported robust growth of the *ydj1*Δ *ram1*Δ strain on solid media because yeast expressing other Ydj1 Cxxx variants did not grow well at elevated temperature. The four Ydj1 Cxxx variants were also evaluated in a background with normal FTase activity (i.e., *RAM1*). In the presence of FTase, a shifted protein band indicative of prenylated Ydj1 was present in each case. This species was

estimated to be at least 50% or more of the total Ydj1 present in each of the lysates. In the absence of FTase, the shifted protein band was present but represented a reduced amount of the total Ydj1 in the sample. We interpret these observations to indicate that low levels of prenylation can occur to Ydj1 Cxxx variants in the absence of FTase. We also evaluated several other Ydj1 Cxxx variants, including CASQ, SASQ and CAPQ. Yeast expressing these variants were cultured at room temperature when expressed in the *ram1*Δ background because of their inability to grow at elevated temperature. None of these variants displayed a shifted band in the absence of FTase.

Frequency analyses of recovered sequences

The Ydj1-based selection strategy recovered numerous sequence combinations. WebLogo frequency analysis of all the identified sequences revealed no obvious enrichment of any particular amino acid at the x_1, x_2, or x_3 position (**Figure 2.5A**). Analysis of sequences grouped by thermotolerance scores also failed to reveal an enrichment pattern for the variants that performed phenotypically most similarly to shunted Ydj1 (i.e., CASQ-like) (**Figure 2.5B**). While the small subset of variants that performed similarly to cleaved Ydj1 (i.e., CVIA-like) also lacked obvious enrichment of any type of amino acid at x_1 or x_3, they were enriched for branched chain amino acids (BCAs) at x_2 (**Figure 2.5C**). The presence of a BCAs at x_2 alone, however, was not a good predictor of weaker thermotolerance. Frequency analysis of CASQ-like sequences with x_2 BCAs revealed that charged residues (i.e., Arg, Asp, Glu, His, or Lys) were present at x_1 or x_3 in seven of ten instances (**Figure 2.5D**); the outliers were CAVQ,

CGLL, and CGVQ. Similar analysis of CASQ-like (weak) sequences revealed charged residues at x_1 or x_3 in ten of eleven instances; the outlier was CAVG.

We also conducted frequency analysis of sequences predicted to be prenylated in yeast that were obtained using a Ras-based *in vivo* reporter (Ras Recruitment System screen; RRS screen) (25). The Ras-based strategy employed an enrichment scoring system to identify sequences having a high probability of prenylation (n=496). For our analyses, we culled the high probability sequences to eliminate low confidence sequences as suggested by the authors, which created a reduced set of sequences (n=369). This reduced set was enriched for BCAs at x_2, while a wide range of amino acids were present x_1 and x_3 (**Figure 2.5E**). Interestingly, most of the sequences identified through Ydj1-based screening were scored as low probability sequences in the Ras-based screen (**Figure S2.4A**). Among the overlapping set of sequences identified in both screens, six of the eight sequences presented phenotypically as CVIA-like or CVIA-like (weak) in the Ydj1-based thermotolerance test; the two outliers were CAVQ (CASQ-like) and CVTS (CASQ-like (weak).

Initial evaluation of the Ydj1 and Ras-based sequences revealed that some amino acids were absent or at low frequency in one, the other, or both sets. Some amino acids were also very common. For example, Ser was observed over 20 times at each position in Ydj1-based sequences. This over-representation likely reflects codon bias in the degenerate oligo used to create Cxxx permutations (i.e., Ser is encoded by six codons). To better understand the frequency occurrence of each amino acid in the Ydj1 set

without codon bias, we normalized the occurrence of each residue based on the number of potential codons for that amino acid (**Figure 2.6A**). We built high and low frequency groups by identifying amino acids with normalized frequencies that were outside a 95% confidence interval level relative to the average frequency of all amino acids (**Table 2.2**). We built the same groups for the reduced set of Ras-based sequences, but normalization was not needed due the nature of the experimental strategy used to identify those sequences (**Figure 2.6B, Table 2.2**) (25). Few amino acids were consistently present at high frequency independent of reporter: x_1 (Ala, Thr, Val); x_2 (Val); x_3 (Gln and Ser); these results are similar to those observed *in vitro* with mammalian FTase (11, 29, 30). We view these amino acids as extremely favorable for farnesylation. There were additional high frequency amino acids identifiable in the context of one or the other reporter, indicating that reporter-specific effects need to be considered when evaluating the data sets. Considering the combined set of high frequency amino acids, 8 were strongly favored at x_1, 11 at x_2, and 10 at x_3. We interpret this observation to indicate that the yeast FTase can tolerate many different amino acids at these individual positions, similar to the reported behavior of mammalian FTase *in vitro*. Of note, charged amino acids were generally excluded as high frequency amino acids from both data sets, with the exception of Asp (x_1) and His (x_2 and x_3) that were highly enriched in Ydj1 sequences. By contrast to high frequency amino acids, more amino acids were consistently present at low frequency independent of reporter: x_1 (His, Gln, Phe, Trp, Tyr); x_2 (Arg, Asp, Glu, Lys, Trp); x_3 (Arg, Glu, Lys, Pro, Tyr). We view these low frequency amino acids as incompatible with efficient farnesylation when present at the indicated positions, although their incompatibility may be context specific

(i.e., dependent on residues at other positions) (29, 30). Considering the combined set of low frequency amino acids, 13 were disfavored at x_1, 13 at x_2, and 12 at x_3. Some of the disfavored amino acids from the RRS set may be incompatible with proteolysis rather than farnesylation. Thus, the low frequency amino acids common to both the Ydj1 and RRS data sets are likely the extent of disfavored amino acids: x_1 (His); x_2 (Arg, Asp, Glu, Lys) and x_3 (Arg, Glu, Lys).

Overall, our frequency analyses established that membership in a frequency group was clearly influenced by the reporter utilized in the screen. Of additional note, the Ydj1-based selection recovered CASQ as an independent isolate whereas the RRS screen did not identify this sequence as significantly enriched. Collectively, our observations indicate that sequences recovered by Ydj1 and Ras-based approaches yield largely non-overlapping sets of sequences, which we hypothesize reflect shunted and cleaved sequences, respectively. Additively, the two sets of sequences likely represent a comprehensive spectrum of prenylatable sequences, although it is possible that additional Ydj1-based sequences have gone unidentified. For example, we did not recover the Pex19 CKQQ motif, whose prenylation is well documented (62). This sequence is associated with CASQ-like behavior in the context of Ydj1 (Hildebrandt and Schmidt, unpublished observation).

Evaluation of Ydj1-based sequences using prenylation prediction algorithms

We evaluated our Ydj1-based hits in the context of several prenylation prediction algorithms, as it was unclear how efficiently these non-canonical sequences would be

scored by methods trained using learning sets based on different measures of prenylation reactivity. These included recently developed FlexPepBind, which takes into account structure-based constraints derived from the conserved features in solved FTase structures, the web-hosted PrePS algorithm, and an in-house algorithm that used disfavored amino acids identified by our study as the guide to predict likelihood of farnesylation (**Table 2.3; Figure S2.4B-C**) (12, 28). It should be noted that FlexPepBind and PrePS were optimized for mammalian FTase, potentially limiting their utility for predicting modification of yeast Cxxx sequences.

Relatively few Ydj1-based sequences were identified as having a high probability of prenylation by FlexPepBind and PrePS (27% and 7%, respectively), with the majority of sequences judged to have a low probability (54% and 68%, respectively). By contrast, a majority of sequences were judged to have a high probability of prenylation by in-house rules (86%), and none were judged to have a low probability. Considering the Ras-based-sequences, in-house rules also identified more high probability sequences (86%) than FlexPepBind and PrePS (60% and 56%, respectively). We also evaluated the full set of 8000 Cxxx combinations using our in-house rules and the FlexPepBind and PrePS algorithms. In-house rules identified many high probability sequences (42%) whereas fewer were identified by FlexPepBind and PrePS (17%, and 5%, respectively). A similar pattern was observed for a set of Cxxx proteins Identified in the *Saccharomyces* Genome database (SGD). The initially identified set of sequences (n=108) was culled to a smaller set (n=89) to by eliminating those annotated as dubious open reading frames. In-house rules identified a majority of sequences as having a high

probability of prenylation (61%), whereas FlexPepBind and PrePS identified fewer (28% and 26%, respectively) (**Table 2.3**; **Table S2.5**).

Discussion

There are a limited number of genetic reporters suitable for investigations of prenyltransferase specificity. Past studies in yeast have relied on the Ras GTPase and **a**-factor mating pheromone, which both require full COOH-terminal modification for their optimal activities. By contrast, Ydj1 only requires isoprenylation for its optimal activity, and lack thereof or more extensive Ras-like modification of its COOH-terminus are actually detrimental to its activity in promoting yeast thermotolerance. The specific reason that shunted Ydj1 is required for optimal thermotolerance remains undefined. We hypothesize that Ydj1 uses its farnesylated COOH-terminal region for physical interactions with key client proteins required for an effective heat stress response. This is akin to the chaperone Pex19 where its farnesylation stabilizes conformations that promote association with cargo proteins (63). Full modification of Ydj1 would thus be expected to alter the biophysical properties of its COOH-terminus, potentially disrupting client interactions. A free charged COOH terminus may be needed for client interactions, which can be provided by any sequence that can be prenylated but resists cleavage (i.e., shunted). Alternatively, it may be that the canonical multi-step modification of Ydj1 enhances its hydrophobicity and promotes increased association with membranes, limiting its accessibility to cytosolic client proteins. Regardless of the reason that Ydj1-based thermotolerance requires shunting, we were able to successfully take advantage of this phenotype to develop Ydj1 as a novel reporter for

sampling the prenylation potential of sequence space. We did this to specifically test our prediction that the multi-step complexity of post-translational modification that occurs *in vivo* to Ras and **a**-factor may limit interpretation of prenyltransferase specificity.

Through a Ydj1-based genetic thermotolerance selection strategy, we recovered many sequences that did not conform to the CaaX motif typically associated with prenylproteins. Interestingly, we observed that the sequences recovered using Ydj1 were mostly non-overlapping with those obtained using Ras-based methods, suggesting that the combined set of sequences represents a more comprehensive set of prenylatable motifs that is larger than previously appreciated. The combined set suggests that a large combination of amino acids is tolerated by the prenyltransferases across all positions of the motif. Because early investigations into the specificity of the prenyltransferases were often limited to a small subset of prenylproteins (e.g. Ras and Ras-related GTPases), this led to the use of the traditional CaaX consensus often described in the literature (20-22, 24), although “Cxxx” is a more accurate description of the consensus in light of our findings and others (64-66).

With respect to yeast prenyltransferase specificity, frequency analysis of both Ydj1 and Ras data sets suggests strongly preferred and disfavored amino acids within the Cxxx motif. At x_1, eight amino acids were frequently observed. These included expected aliphatic amino acids (Ala, Gly, Ile, Leu, Val), small polar uncharged residues (Ser, Thr), and an unexpected polar charged residue (Asp). At x_2, eleven amino acids were frequently observed. These included expected aliphatic amino acids (Ala, Gly, Ile, Leu,

Met, Val), unexpected polar uncharged residues (Asn, Gln, Ser, Tyr), and a polar/weakly charged residue (His; the charged state of His depends on local environment; only 5-10% of free His is charged at physiological pH). At x_3, ten amino acids were frequently observed. These included aliphatic amino acids (Ala, Gly, Ile, Met, Val), polar uncharged residues (Asn, Cys, Gln, Ser), and a polar/weakly charged residue (His). The above reflects amino acids most highly enriched relative to other amino acids (i.e., above a 95% confidence interval), so the number of amino acids tolerated at each position is actually greater when considering amino acids with average to above average frequency occurrences. We expect that future studies involving modeling or co-crystallization of atypical peptide sequences within the prenyltransferase active sites will help establish how these enzymes can accommodate such a variety of side chains at each position. We also expect that multi-variate analysis will lead to a better understanding of the substrate features being recognized, but such studies will likely require more Ydj1-based sequences than we have presently collected.

Perhaps more telling for yeast prenyltransferase specificities are the residues that were infrequently recovered in the combined data sets, and in particular those that were reporter independent. While each position of the Cxxx motif seems to have a disfavored set of amino acids, poor prenylation outcomes were common when the x_1 position had a bulky amino acid, and x_2 and x_3 positions had either charged or bulky residues; notably, mammalian fTase also has steric restrictions at these positions (27, 29). The negative constraint for Pro at x_3 is interesting in that it suggests that the conformation restrictions

introduced by Pro at this position may eliminate critical interactions with the prenyltransferase and/or lead to unfavorable contacts within the substrate binding site.

We initiated our study expecting that Ydj1 could serve as a reporter for identifying shunted Cxxx sequences (i.e., CASQ-like). Indeed, the vast majority of recovered sequences behave phenotypically as if they are not cleaved. It remains to be determined, however, whether each identified sequence is actually shunted. Such an analysis would require either individual purification and mass-spec analysis of the COOH-terminus of each Ydj1 Cxxx variant or indirectly assessing cleavage through other reporters such as **a**-factor (i.e., mating assays) or Ras (i.e., localization assays). It also remains to be determined how well each sequence is prenylated and by which prenyl group. While we did not evaluate all Cxxx variants, our investigations with subsets of sequences indicate that many are substantially farnesylated, often to completion.

Besides information on yeast prenyltransferase specificity, our Ydj1-based data also hints at specificity determinants related to Cxxx cleavage. While our screen was not primarily intended to identify cleaved Cxxx sequences, we did recover a small set of sequences that behave phenotypically as if they are cleaved (i.e., CVIA-like). Not surprising, this phenotype is generally favored when aliphatic amino acids are present at x_2, except in instances where charged residues flank this position or prenylation is otherwise disfavored. The limited number of CVIA-like sequences in our Ydj1-based data set limits our ability to analyze cleavage specificity. We propose that sequences

recovered through Ras-based screening offer a better set for such analyses. Because the Ras-based sequences are largely non-overlapping with the sequences obtained by Ydj1-based screening, the sets of sequences clearly differ in some capacity. The most parsimonious explanation is that the Ras-based sequences are cleaved while the Ydj1-based sequences are shunted. Analysis of Ras-based sequences indicates that they are also enriched for aliphatic residues at x_2. This suggests that Cxax (a = aliphatic) may be a more precise consensus for cleavage by the yeast CaaX proteases. It remains to be determined whether cleavage in these instances is due to Rce1 or Ste24, the two proteases identified as cleaving CaaX proteins (59).

Overall, we interpret our analyses to indicate that many sequences have the potential to be prenylated in yeast (i.e., Cxxx), and that a subset have characteristics that make them susceptible to cleavage (i.e., Cxax). Using our in-house rules for prenylation, and additional rules to predict cleavage potential, we categorized the 89 Cxxx proteins identifiable in the *Saccharomyces* Genome Database (**Table S2.5**). Over one-third of the sequences (36%; n = 32) were predicted to be prenylated and cleaved (i.e., Ras-like). This group includes the Ras and Rho GTPases (*RAS2*, *RHO1*, *RHO2*, etc.), a Gγ subunit (*STE18*), and **a**-factor (*MFA1*, *MFA2*). Other members of this group have not been investigated with respect to their Cxxx modifications (e.g., ABC transporter Atr1; ureidoglycolate lyase Dal3; Hsp40 Xdj1) and are interesting candidates to investigate as part of future studies on protein prenylation. About one-quarter of sequences (24%; n = 22) were predicted to be prenylated and not cleaved (i.e., shunted). This group included Ydj1, as expected, and notables Pex19 and Nap1. The CKQQ motif associated with

Pex19 is farnesylated, and this motif is also present on the mammalian tumor suppressor STK11/Lkb1 for which there is evidence of farnesylation and shunting (15, 62, 67). A similar CKQx motif is present on Nap1 (yeast, human and plant orthologs). The farnesylation status of yeast Nap1 has not yet been specifically investigated, but both human and plant Nap1 are farnesylated (32, 33, 68); cleavage status is unknown. Of the remaining sequences, about one-fifth (16%; n = 14) had ambiguous prenylation predictions with varying cleavage potential, and about one-quarter (23%; n = 21) were not predicted to be farnesylated. None of the proteins in these sets are known to be farnesylated (e.g., *CUP1-1*, *HMG1*, etc.). The proper binning of many proteins into what are appropriate categories provides confidence that our in-house prediction methods have potential applicability. Our prediction methods can be improved in future studies by incorporating additional discrimination parameters. For example, the accessibility of Cxxx sequences to the cytosolic prenyltransferases was not considered, so some proteins may be false positives within the canonical or shunted sequence, groups.

The activities of the prenyltransferases and CaaX proteases have received much attention over the years, but their specificities have been hard to resolve despite a combination of *in vivo*, *in vitro*, and *in silico* methodologies. Here we used a genetic approach to identify a large set of prenylatable sequences in yeast that are not predicted to be modified by existing prediction methods. While this data set provides strong evidence for broader yeast prenyltransferase specificity, it has also allowed for re-interpretation of pre-existing data in a manner that informs on cleavage specificity. Beyond studies of the yeast enzymes, we envision that the specificities of other

prenyltransferases and CaaX proteases could be investigated using our methods by heterologous expression of desired enzymes (e.g., yeast expressing human Ftase instead of yeast enzyme). Intriguingly, considerable evidence suggests conserved specificity between yeast and human prenyltransferases, suggesting that the specificities observed in our study may ultimately hold true for the human enzymes (31, 69, 70).

Acknowledgments

The authors thank Dr. Avrom Caplan (City College of New York) for anti-Ydj1 primary antibody, Dr. Ora Furman-Schueler (Hebrew University of Jerusalem) for sharing FlexPepBind scores, and Wayland Yeung (University of Georgia) for help with PrePS scores. We also thank members of the Schmidt lab for critical discussions and comments on this manuscript, and the following lab members for technical assistance: Alona Botnar, Manuel Fierro, and Ryan Peppenhorst, and Rajani Ravishankar. This work was funded by a grant from the National Institutes of Health to WKS (GM117148).

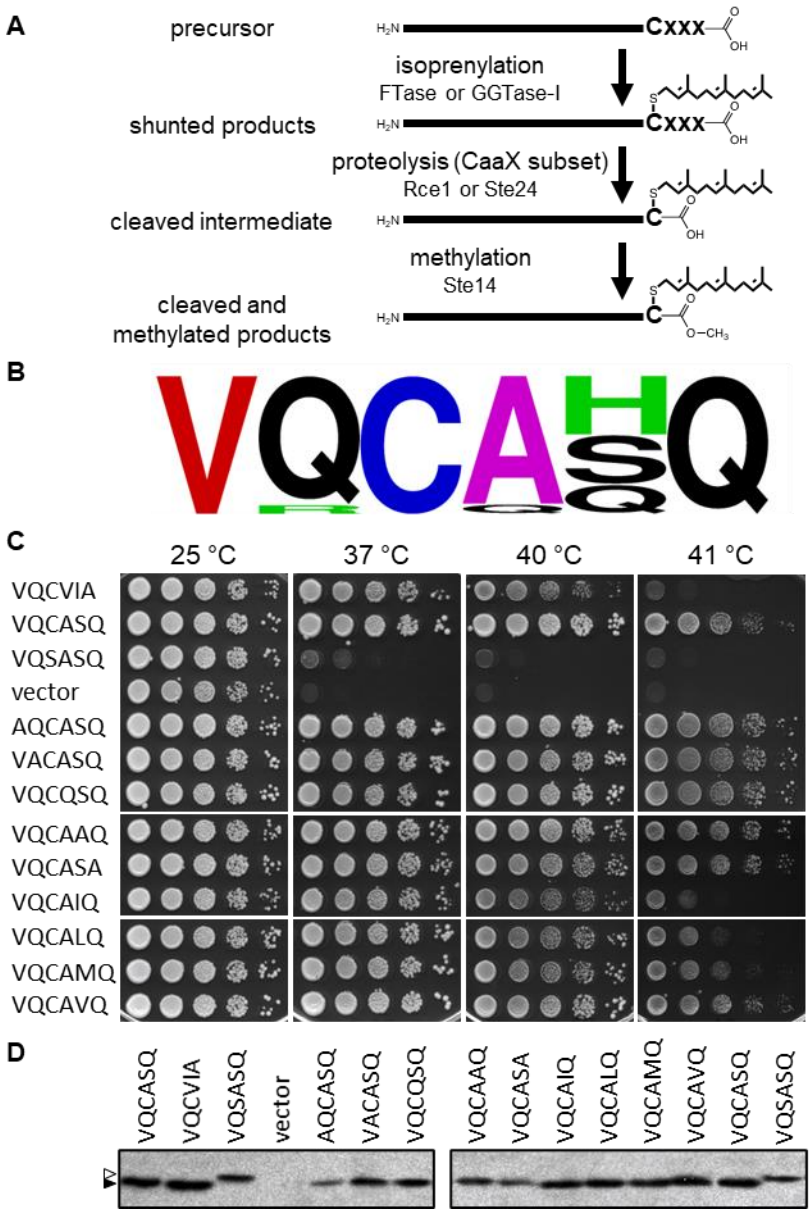


Figure 2.1. Site-directed mutation of the Ydj1 Cxxx motif reveals flexibility in sequence requirements for functional levels of isoprenylation. **A**) The Cxxx motif directs protein isoprenylation. Both farnesyl (C15) and geranylgeranyl (C20) can be added to Cxxx proteins; only C15 addition is shown for clarity. The isoprenylated species is either the endpoint modification (e.g., Ydj1; shunted proteins) or an intermediate that is additionally modified by proteolysis and carboxylmethylation (e.g., K-Ras4b; traditional CaaX proteins). Not diagrammed are more extensive modifications that can also occur, such as palmitoylation (e.g., H- and N-Ras) or distal proteolysis (e.g., lamin A; yeast **a**-factor). **B**) WebLogo frequency analysis of the last 7 amino acids associated with fourteen Ydj1 homologs retrieved from the Homologene database (http://www.ncbi.nlm.nih.gov/homologene). Color scheme is as described in *Methods*. See Figure S2.1 for specific sequence details. **C**) Ydj1 mutants were evaluated for their ability to support high temperature yeast growth. yWS304 yeast (*ydj1*Δ) expressing the

indicated plasmid-encoded Ydj1 mutant were cultured in selective SC-uracil media and pinned as 10-fold serial dilutions onto non-selective YPD; the leftmost spot in each panel is undiluted. Plates were incubated at the indicated temperature as described in *Methods*. **D**) The Ydj1 mutants indicated in Panel C were expressed in yWS304 (*ydj1*Δ) and cell lysates evaluated by anti-Ydj1 immunoblot. The specific plasmids used for Figure 2.1 are listed in Table 2.1. Unmodified Ydj1 (open triangle) migrates at a larger apparent Kda than prenylated Ydj1 (closed triangle).

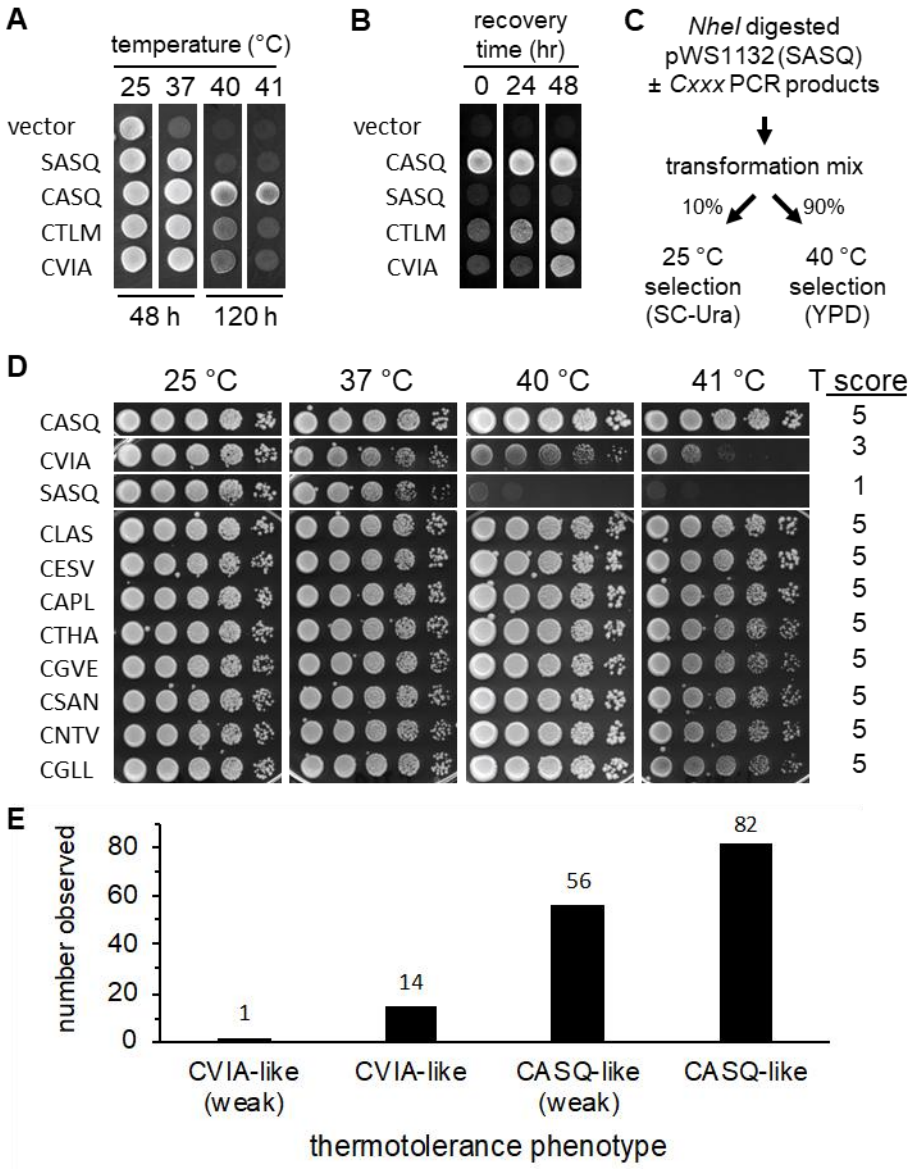


Figure 2.2. Impact of temperature and time on yeast thermotolerance. **A**) Yeast expressing the indicated Ydj1 variants were cultured to saturation in selective media, diluted into YPD, spotted using a multi-channel pipettor onto YPD solid media, and plates incubated at indicated temperatures and times. A 1:20 dilution was the source for spots incubated at 25 °C, and a 1:2 dilution was used for spots incubated at other temperatures. **B**) Yeast were cultured and processed as described for panel A using the 40 °C condition with the following alterations: the 1:2 dilution was into YPD; incubation at 40 °C was for 72 hours followed by recovery at 25 °C for the indicated times. **C**) Flow diagram of screen used to identify thermotolerant yeast expressing Ydj1 Cxxx variants in either the *ydj1*Δ or *ydj1*Δ *ste14*Δ background. **D**) Examples of thermotolerance profiles observed and thermotolerance (T) scores assigned when expressed in *ydj1*Δ background; the controls and panel of mutants are replicated in Figure S2.2. Yeast thermotolerance assays were performed as described for Figure 2.1. **E**) The

thermotolerance profiles of all recovered mutants were scored relative to yeast expressing Ydj1 with a SASQ, CVIA or CASQ motif, where the controls were assigned scores of 1, 3 and 5, respectively. The mutants were then binned according to their scores; see *Methods* for details on binning.

Table 2.1. Summary of thermotolerance phenotypes observed and associated prenylation prediction scores.

Category[a]	**number observed**	**average T score**[b]	**average FPB score**	**average RRS E score**	**average PrePS score**
all unique sequences	153	4.36 ± 0.57	-0.18 ± 1.41	0.62 ± 2.26	-3.31 ± 2.19
CASQ-like	82	4.76 ± 0.15	-0.26 ± 1.20	0.31 ± 1.62	-3.55 ± 1.99
CASQ-like (weak)	56	4.15 ± 0.22	0.20 ± 1.58	0.24 ± 0.79	-3.47 ± 1.97
CVIA-like	14	3.04 ± 0.24	-1.18 ± 1.35	3.24 ± 5.02	-1.61 ± 5.02
CVIA-like (weak)	1	2.47	-1.02	9.98	1.35
CASQ-like (x_2=ILV)	10	4.66 ± 0.12	-1.87 ± 0.50	1.63 ± 4.60	-1.93 ± 2.45
Ydj1/RRS overlap	8	3.30 ± 0.66	-1.83 ± 0.81	9.37 ± 4.23	0.94 ± 0.54

[a] Specific sequences and groupings are listed in Table S2.4.
[b] Thermotolerance (T) score averages were calculated using data from Figure S2.2. FlexPepBind (FPB) scores were derived from London *et al* (28); a score of -1.1 or less is predicted to have a high probability of prenylation. RRS Enrichment (E) scores were derived from Stein *et al* (25); a score of 3 or more is predicted to have a high probability of prenylation. PrePS scores were retrieved from the PrePS server; a score above 0 is predicted to have a high probability of prenylation.
[c] the scores reported for the single CVIA-like (weak) sequence are for that sequence alone. Scores for CASQ-like (x_2=ILV) are for the subset of CASQ-like sequences with BCAs at x_2. Scores for Ydj1/RRS overlap are for those sequences recovered in both the Ydj1 and RRS screens.
[d] NA – not applicable

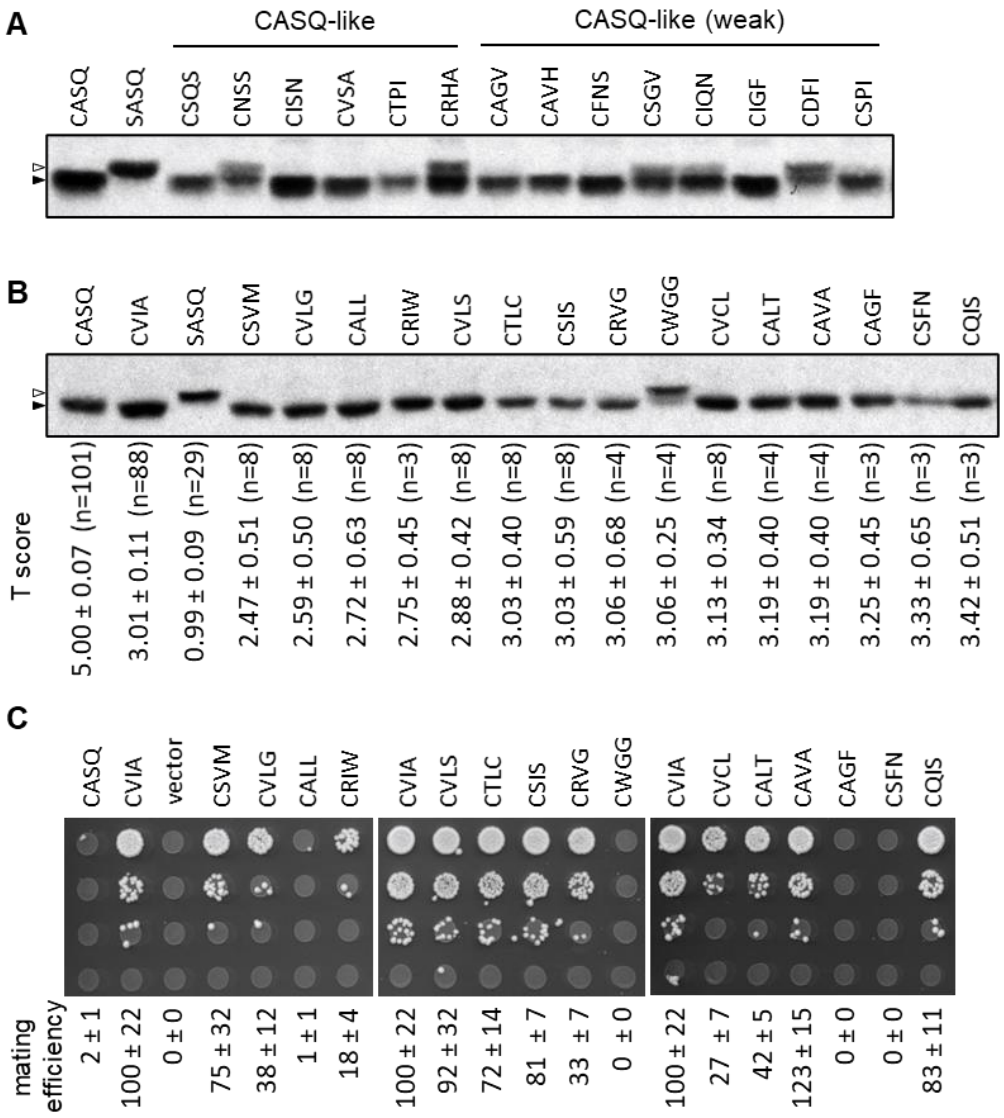


Figure 2.3. Isoprenylation and cleavage properties of Ydj1 Cxxx mutants identified through thermotolerance screening. **A-B**) Yeast expressing the indicated Ydj1 mutants were evaluated by a gel-shift assay as described in Figure 2.1C. The sequences evaluated were either randomly identified from the CASQ-like and CASQ-like (weak) groups (A) or reflect the combined set of sequences presenting with CVIA-like and CVIA-like (weak) thermotolerance phenotypes (B). Reference controls included on the blots are: farnesylated and uncleaved Ydj1 (CASQ); farnesylated, cleaved and carboxylmethylated Ydj1 (CVIA); unmodified Ydj1 (SASQ). The values in panel B reflect the thermotolerance scores observed for the corresponding mutant. **C**) Yeast expressing the indicated **a**-factor Cxxx variants in a *mfa1*Δ *mfa2*Δ background was assessed for mating competence using a serial dilution mating assay (panel) and quantitative mating assay (values). Values represent mating efficiency relative to wildtype CVIA, which was set to 100%; values were determined using 4 or more replicates from 2 or more individual experiments. The panel is representative of one of the replicates.

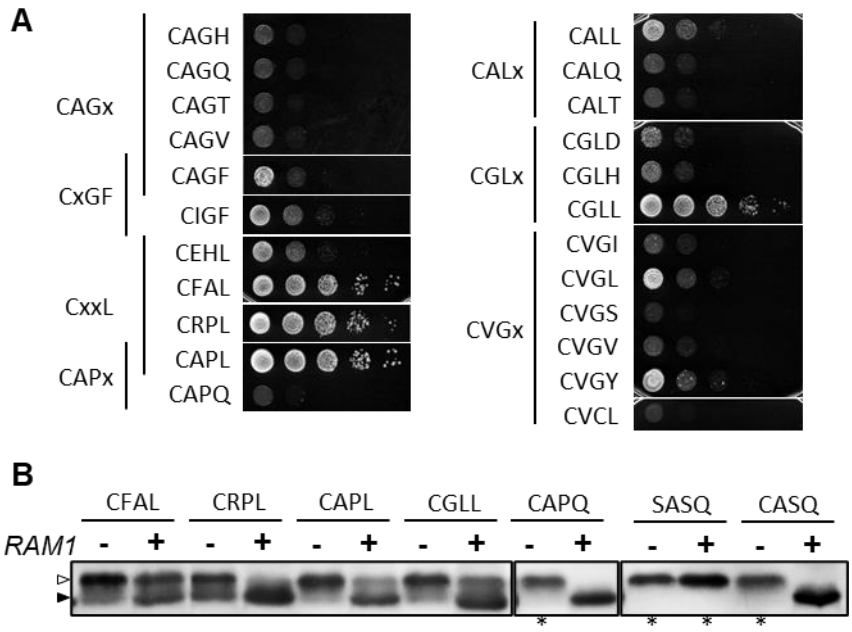


Figure 2.4. Investigations of the geranylgeranylation potential of Ydj1. **A**) The indicated Ydj1 Cxxx variants were assessed for their ability to restore thermotolerance to a *ydj1*Δ *ram1*Δ (yWS2542) background as described for Figure 2.1 except that the 40 °C incubation was ~80 hours without recovery. **B**) Lysates for gel-shift assays were prepared from mid-log cultures incubated at either 40 °C or 25 °C (the latter are marked with an asterisk). The strain backgrounds used were *ydj1*Δ *RAM1* (+; yWS2544) and *ydj1*Δ *ram1*Δ (-; yWS2542). Of note, the *ram1*Δ cultures were slow growing relative to *RAM1* cultures at 40 °C, taking 36 or more hours instead of 18-24 hours to achieve the same cell density. Unmodified Ydj1 (open triangle) migrates at a larger apparent KDa than prenylated Ydj1 (closed triangle).

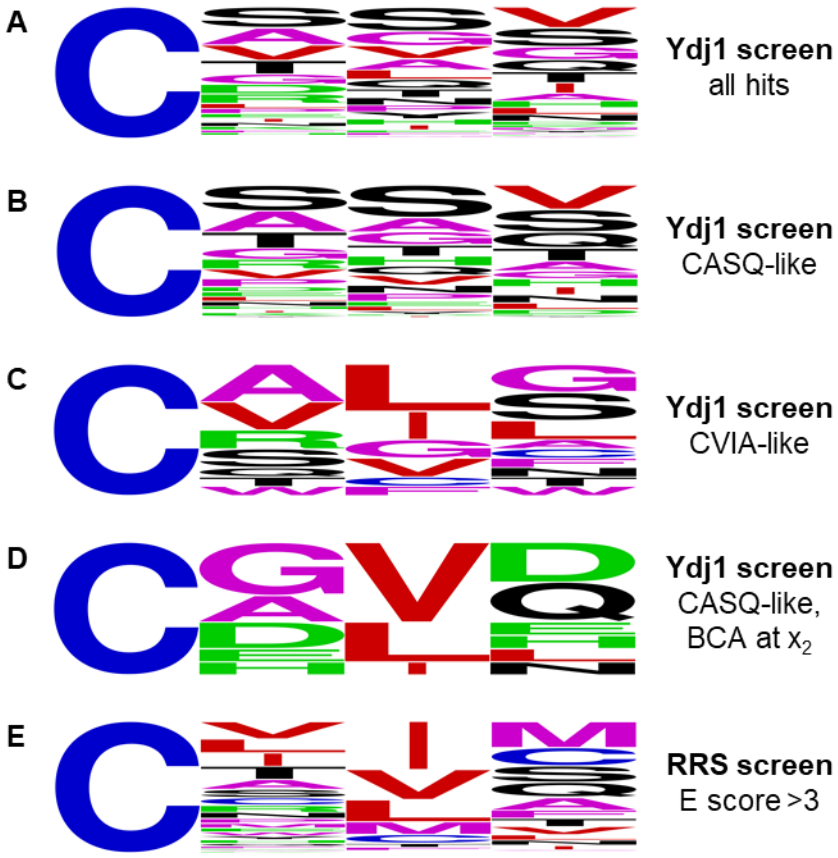


Figure 2.5. Frequency analysis of sequences identified by thermotolerance screening. Sequences were categorized by their associated thermotolerance scores (see Figure S2.2). For each grouping, a WebLogo analysis was performed (47). Groupings were **A**) all identified sequences (n=153), **B**) those most like CASQ (i.e. CASQ-like; T score range 4.5-5; n=82), **C**) those most like CVIA (i.e. CVIA-like; T score range 2.50–3.49; n=14), **D**) the subset of sequences from Panel B that had a branched chain amino acid at x_2 (n=10) and **E**) sequences identified by a Ras-based strategy as having high likelihood of prenylation (enrichment score >3; n=369) (25).

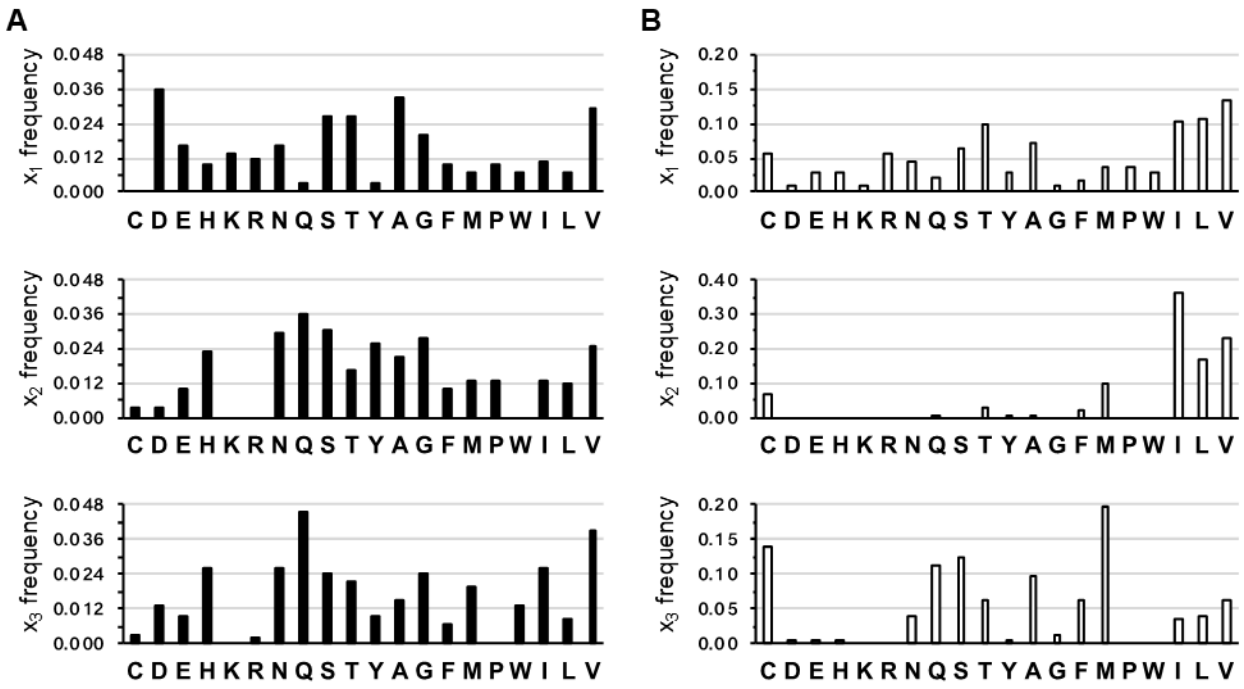


Figure 2.6. Amino acid frequency in hits recovered in Ydj1 and Ras-based screens. The frequency occurrences of each amino acid at the x_1, x_2 and x_3 positions were calculated for the set of Ydj1 (**A**) and Ras-based sequences (**B**). The complete set of Ydj1-based sequences (n=153) and the reduced-size set of Ras-based sequences (n=369) were used for the analysis. Frequency values for Ydj1-based sequences were normalized for codon bias (e.g., Leu codons are over-represented 6x relative to the Met codon); Ras-based sequences did not need normalization due to study design. Amino acids are clustered as reported for Weblogo analyses. Note that Y-axis scales differ for the three panels in B.

Table 2.2. High and low frequency amino acids in Ydj1 and Ras-based data sets.

Position	reporter	above CIL[a]	within CIL	below CIL
x_1	Ydj1	**A T V** D G S	**N** E I K R	**F H Q W Y** L M P; C (absent)
	Ras	**A T V** I L	**N** C R M S P	**F H Q W Y** D E G K
x_2	Ydj1	**V** A G H N Q S Y	**T** I L M P	**K R W** (all absent); **D E** C F
	Ras	**V** I L M	**T** C F	**K R W** G H N P S (all absent); **D E** A Q Y
x_3	Ydj1	**Q S** G H I N V	**T** A D M W	**K P** (all absent); **E R Y** C F L
	Ras	**Q S** A C M	**T** F I L N V	**K P R** W (all absent); **E Y** D G H

[a] The number of instances that each amino acid was observed at a particular position within the population of sequences was determined. For Ydj1-based sequences (n=153), the number of occurrences for each amino acid was normalized to adjust for codon bias (i.e., there are more Leu than Met codons). For Ras2-based sequences (n=369), normalization was not needed due to study design. A standard deviation and 95% confidence interval level (CIL) were determined for each population at each position that was used to determine amino acids above or below the interval. Amino acids that are in both the Ydj1 and Ras-based data sets for an indicated frequency group are bold.

Table 2.3. Results of isoprenylation prediction algorithms

		isoprenylation prediction (%)		
test set	**method**	**high**	**ambiguous**	**low**
Ydj1-based (n=153)	FlexPepBind	27[b]	20	54
	PrePS	7	25	68
	In-house	86	14	0
Ras-based (n=369)	FlexPepBind	60	15	25
	PrePS	56	32	12
	In-house	86	14	0
Cxxx (n=8000)	FlexPepBind	17	12	71
	PrePS	5	14	80
	In-house	42	28	30
SGD (n=89)[a]	FlexPepBind	28	27	46
	PrePS	26	8	66
	In-house	61	16	24

[a] This test set was identified using appropriate sequence patterns and the Pattern Match search function associated with the *Saccharomyces* Genome database (SGD). The set was culled of sequences annotated as dubious open reading frames.
[b] The summed value of the three categories for some sets exceeds 100% due to rounding errors.

Table S2.1. Yeast strains used in this study.

Strain	genotype	reference
BY4741	*MAT***a** *his3 leu2 met15 ura3*	(71)
IH1793; ATCC#204279	*MAT*α *lys1*	(72)
SM2331	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-*Δ*1 mfa2-*Δ*1*	(73)
yWS304	*MAT***a** *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 ydj1*Δ*::KAN^R*	(74)
yWS1632	*MAT***a** *his3 leu2 met15 ura3 ram1*Δ*::KAN^R*	(74, 75)
yWS1635	*MAT***a** *his3 leu2 met15 ura3 ydj1*Δ*::KAN^R ste14*Δ*::KAN^R*	(16)
yWS2542	*MAT***a** *his3 leu2 met15 ura3 ydj1*Δ*::NAT^R ram1*Δ*::KAN^R*	This study
yWS2544	*MAT***a** *his3 leu2 met15 ura3 ydj1*Δ*::NAT^R*	This study

Table S2.2. Plasmids used in this study.

gene	identifier	genotype	reference
vector	pRS315	*CEN LEU2*	(76)
	pRS316	*CEN URA3*	(76)
YDJ1	pWS942	*CEN URA3 YDJ1 (CASQ)*	(16)
	pWS1132	*CEN URA3 YDJ1 (SASQ)*	(16)
	pWS1246	*CEN URA3 YDJ1 (CTLM)*	(16)
	pWS1286	*CEN URA3 YDJ1 (CVIA)*	(16)
	pWS1339	*CEN URA3 YDJ1 (AQCASQ)*	This study
	pWS1341	*CEN URA3 YDJ1 (VACASQ*	This study
	pWS1373	*CEN URA3 YDJ1 (CQSQ)*	This study
	pWS1372	*CEN URA3 YDJ1 (CAAQ)*	This study
	pWS1343	*CEN URA3 YDJ1 (CASA)*	This study
	pWS1402	*CEN URA3 YDJ1 (CAIQ)*	This study
	pWS1403	*CEN URA3 YDJ1 (CALQ)*	This study
	pWS1404	*CEN URA3 YDJ1 (CAMQ)*	This study
	pWS1409	*CEN URA3 YDJ1 (CAVQ)*	This study
	pWS1623	*CEN URA3 ydj1::NAT^R*	This study
	various	*CEN URA3 YDJ1-Cxxx variants* (n=153)	This study
MFA1	pWS610	*CEN LEU2 MFA1 (CVIA)*	(77)
	pWS1587	*CEN LEU2 MFA1 (CRVG)*	This study
	pWS1588	*CEN LEU2 MFA1 (CVLS)*	This study
	pWS1589	*CEN LEU2 MFA1 (CVLG)*	This study
	pWS1590	*CEN LEU2 MFA1 (CTLC)*	This study
	pWS1591	*CEN LEU2 MFA1 (CSVM)*	This study
	pWS1616	*CEN LEU2 MFA1 (CRIW)*	This study
	pWS1617	*CEN LEU2 MFA1 (CALT)*	This study
	pWS1618	*CEN LEU2 MFA1 (CAVA)*	This study
	pWS1619	*CEN LEU2 MFA1 (CWGG)*	This study
	pWS1620	*CEN LEU2 MFA1 (CSIS)*	This study
	pWS1621	*CEN LEU2 MFA1 (CSFN)*	This study
	pWS1622	*CEN LEU2 MFA1 (CQIS)*	This study
	pWS1627	*CEN LEU2 MFA1 (CALL)*	This study
	pWS1628	*CEN LEU2 MFA1 (CVCL)*	This study
	pWS1629	*CEN LEU2 MFA1 (CAGF)*	This study

Table S2.3. PCR oligonucleotides used to mutate *YDJ1* and *MFA1* genes.

Gene	Oligo	Mutation	sequence (5´ → 3´) [c]
Plasmid UTR	oWS219[a]	NA[b]	TGACCATGATTACGCCAAGC
YDJ1	oWS983	AQCASQ	AACTATGATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGcTCAA **TGT**GCATCTCAATGATTTTCT
	oWS984	VQCASA	GAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCATCTgcATGATTTTCTTGATAAAAAAAGATCA
	oWS985	VACASQ	TATGATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTgcA **TGT**GCATCTCAATGATTTTCTTGA
	oWS986	Cxxx	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGc**nnnnnnnnnTGATTTTCTTGATAAAAAAAGA
	oWS990	VQCQSQ	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**caAagcCAATGATTTTCTTGATAAAAAAAGA
	oWS991	VQCAAQ	GATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCAgCTCAATGATTTTCTTGATAAAAAAAGA
	oWS993	VQCAIQ	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCtatcCAATGATTTTCTTGATAAAAAAAGATCA
	oWS994	VQCALQ	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCtctcCAATGATTTTCTTGATAAAAAAAGATCA
	oWS995	VQCAMQ	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCtatgCAATGATTTTCTTGATAAAAAAAGATCA
	oWS996	VQCAVQ	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA **TGT**GCtgtcCAATGATTTTCTTGATAAAAAAAGATCA
MFA1	oWS1145	CRIW	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**agaATTtggTAGTTTCTGCGTACAAAAACGCGT
	oWS1153	CRVG	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**agagTTGgTTAGTTTCTGCGTACAAAAACGCGT
	oWS1155	CALT	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GcTtTgaCTTAGTTTCTGCGTACAAAAACGCGT
	oWS1156	CAVA	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GcTgTTGCTTAGTTTCTGCGTACAAAAACGCGT
	oWS1157	CAGF	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GcTggTttTTAGTTTCTGCGTACAAAAACGCGT
	oWS1166	CVLS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GTTtTgtCTTAGTTTCTGCGTACAAAAACGCGT
	oWS1167	CVLG	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GTTtTgGgTTAGTTTCTGCGTACAAAAACGCGT
	oWS1168	CALL	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA **TGc**GcTtTgttgTAGTTTCTGCGTACAAAAACGCGT
	oWS1169	CTLC	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA

		TGcacTtTgtgTTAGTTTCTGCGTACAAAAACGCGT
oWS1170	CSVM	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGctcTgTTatgTAGTTTCTGCGTACAAAAACGCGT
oWS1211	CSIS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGctcTATTtCTTAGTTTCTGCGTACAAAAACGCGT
oWS1210	CWGG	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGctggggTGgTTAGTTTCTGCGTACAAAAACGCGT
oWS1212	CVCL	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGcGTTtgTttgTAGTTTCTGCGTACAAAAACGCGT
oWS1213	CSFN	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGctcTtTTaaTTAGTTTCTGCGTACAAAAACGCGT
oWS1214	CQIS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCA
		TGccaaATTtCTTAGTTTCTGCGTACAAAAACGCGT

[a] Reverse PCR oligonucleotide paired with *YDJ1* and *MFA1* mutagenic oligonucleotides.
[b] NA – not applicable
[c] n reflects random incorporation of A, C, G or T during synthesis; lowercase letters indicate positions different than the wildtype *YDJ1* or *MFA1* sequences in pWS942 and pWS610, respectively; bold TGT and TGC represent the position of the cysteine codon within the Cxxx encoding sequence.

Table S2.4. Categorization of 153 sequences recovered by Ydj1p-based thermoselection.

CASQ-like [a]			**CASQ-like (weak)**		**CVIA-like**
CAAG	CISN [b]	CSHV	CAHI	CLSI [b]	CAGF
CAGH	CISS [b]	CSQS	CAPQ	CLYT	CALL
CAGQ [b]	CKSQ	CSQT	CAVG	CMYS [b]	CALT
CAGT	CKYS [b]	CSQV	CDFI	CNQH	CAVA
CAGV (2)	CLAS	CSQY	CDLG	CNYQ	CQIS
CANI	CLNQ [b]	CSST (3) [c]	CDLT	CPVD [b]	CRIW
CAPL	CLST	CSSV (4) [c]	CDMS	CRSQ	CRVG
CASQ	CNAQ	CSTI	CDMV	CRSV	CSFN
CASS	CNSS [b]	CSTS [b]	CDQM	CRTA	CSIS (2)
CASV	CNTV	CSTT (2)	CDVG	CSAV (2) [c]	CTLC
CATG	CPNV	CTES [c]	CDVV	CSFE [b]	CVCL
CAVH	CPQI	CTGV	CEVG [b]	CSGV	CVLG
CAVQ [b]	CPSA	CTHA	CEYG	CSPI	CVLS (2)
CDID	CPST	CTPG [b]	CFAL	CSSM	CWGG [b]
CDLN [b]	CPSV [b]	CTPI	CFNG	CTAY	**CVIA-like (weak)**
CDTH	CRGS (2) [c]	CTPS [b]	CFNS	CTIR [b]	CSVM
CEHL	CRGV (2)	CTSH	CGGW	CTPV	
CESV (3)	CRHA [b]	CTSN	CGLH	CTSI (2)	
CEVQ	CRNV [b]	CTSQ [b]	CGNS	CTSV	
CGAV	CRPL	CTYS (2)	CGQN [b]	CVAT	
CGHA [b]	CRST [b]	CVAN (2) [c]	CGSH [b]	CVGI	
CGLD	CSAG	CVDH [b]	CHTQ	CVGL (2)	
CGLL	CSAI	CVES	CIGF [b]	CVGV	
CGQV (2)	CSAN	CVGS	CIQN [b]	CVGY [b]	
CGVE (2) [c]	CSAS	CVNT	CIYG	CVQT [b]	
CGVQ [b]	CSEQ	CVSA	CKIE	CVSI	
CHTG [b]	CSHA	CYSV [c]	CKYA	CVTS	
CHVD			CLNV	CVVR	

[a] Groups are based on thermotolerance score; see Figure S2.2 for average score of individual sequences; see Table 2.1 for distribution of scores. The CASQ-like group includes the CASQ and CAVQ sequences independently identified during screening. Values after certain sequences represent the number of instances the indicated sequence was recovered as a hit (if greater than one). Unless otherwise noted, the sequence was recovered using the *ydj1*Δ *ste14*Δ background.

[b] This sequence was identified using the *ydj1*Δ background.

[c] This sequence was identified independently using the *ydj1*Δ and *ydj1*Δ *ste14*Δ backgrounds.

Table S2.5. Categorization of yeast Cxxx proteins by predicted prenylation and cleavage status.

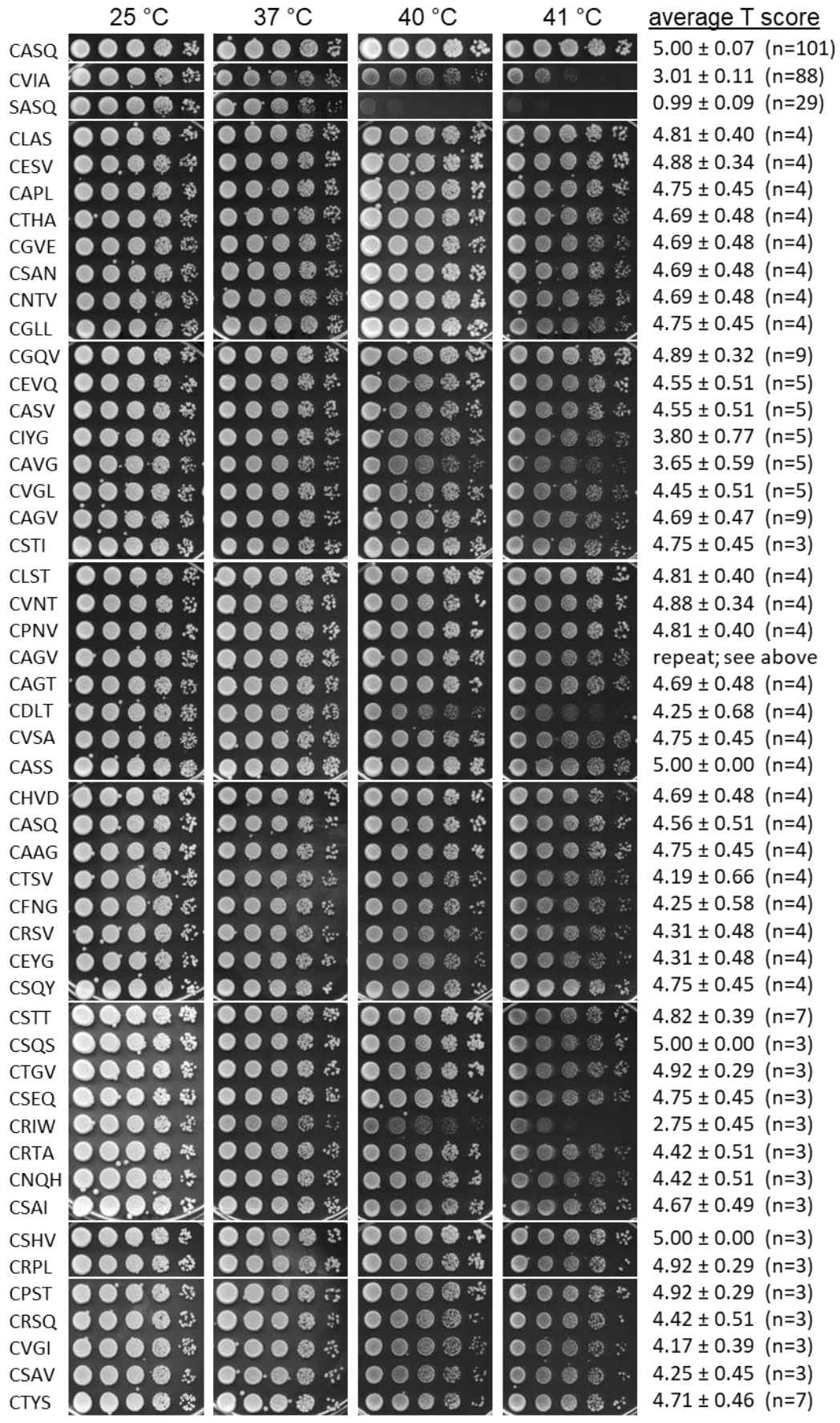
Locus ID	Gene	Motif	PS [b]	CS
Strongly Prenylated / Strongly Cleaved				
YML116W	*ATR1*	CTVA	0	3
YLR229C	*CDC42* [a]	CAIL	0	2
YIR032C	*DAL3*	CIII	0	2
YIR007W	*EGH1*	CVIS	0	3
YML006C	*GIS4*	CAIM	0	3
YJL062W	*LAS21*	CALD	0	1
YDR461W	*MFA1* [a]	CVIA	0	3
YNL145W	*MFA2* [a]	CVIA	0	3
YOR101W	*RAS1* [a]	CIIC	0	3
YNL098C	*RAS2* [a]	CIIS	0	3
YJL204C	*RCY1*	CCIM	0	2
YCR027C	*RHB1*	CSIM	0	2
YPR165W	*RHO1* [a]	CVLL	0	2
YNL090W	*RHO2* [a]	CIIL	0	2
YIL118W	*RHO3* [a]	CTIM	0	3
YKR055W	*RHO4* [a]	CIIM	0	3
YNL180C	*RHO5* [a]	CVIL	0	2
YGR152C	*RSR1* [a]	CTIL	0	2
YBL061C	*SKT5*	CVIM	0	3
YJR086W	*STE18* [a]	CTLM	0	3
YLR090W	*XDJ1*	CCIQ	0	2
YCR004C	*YCP4*	CTVM	0	3
YKL196C	*YKT6* [a]	CIIM	0	3
YDL009C		CAVS	0	3
YDL022C-A		CSII	0	1
YFL066C		CCVC	0	2
YGL082W		CVIM	0	3
YJL118W		CCCS	0	1
YJR128W		CMMI	0	1
YNL234W		CSIM	0	2
YOL014W		CIIL	0	2
YPL191C		CVVM	0	3
Strongly Prenylated / Weakly Cleaved				
YGR282C	*BGL2*	CDFS	0	0
YDR261C	*EXG2*	CASL	0	-1
YNL106C	*INP52*	CDPN	0	-3
YJR107W	*LIH1*	CSGL	0	-2
YKL176C	*LST4*	CNAG	0	-2
YBL049W	*MOH1*	CKCT	0	-1
YKR048C	*NAP1* [a]	CKQS	0	-1
YDL065C	*PEX19* [a]	CKQQ	0	-1
YBL018C	*POP8*	CKCI	0	-1
YBR087W	*RFC5*	CCLD	0	0
YMR060C	*SAM37*	CKYI	0	-2
YGL169W	*SUA5*	CIQF	0	0
YJR066W	*TOR1*	CPFW	0	-2
YKL203C	*TOR2*	CPFW	0	-2
YNL064C	*YDJ1* [a]	CASQ	0	0
YDR034W-B		CDVF	0	0
YFL065C		CCPS	0	-1
YHL049C		CCPS	0	-1
YMR265C		CSNA	0	-1
YMR272W-B		CMYV	0	-1
YOL164W-A		CIHH	0	-2
YPR203W		CCPS	0	-1

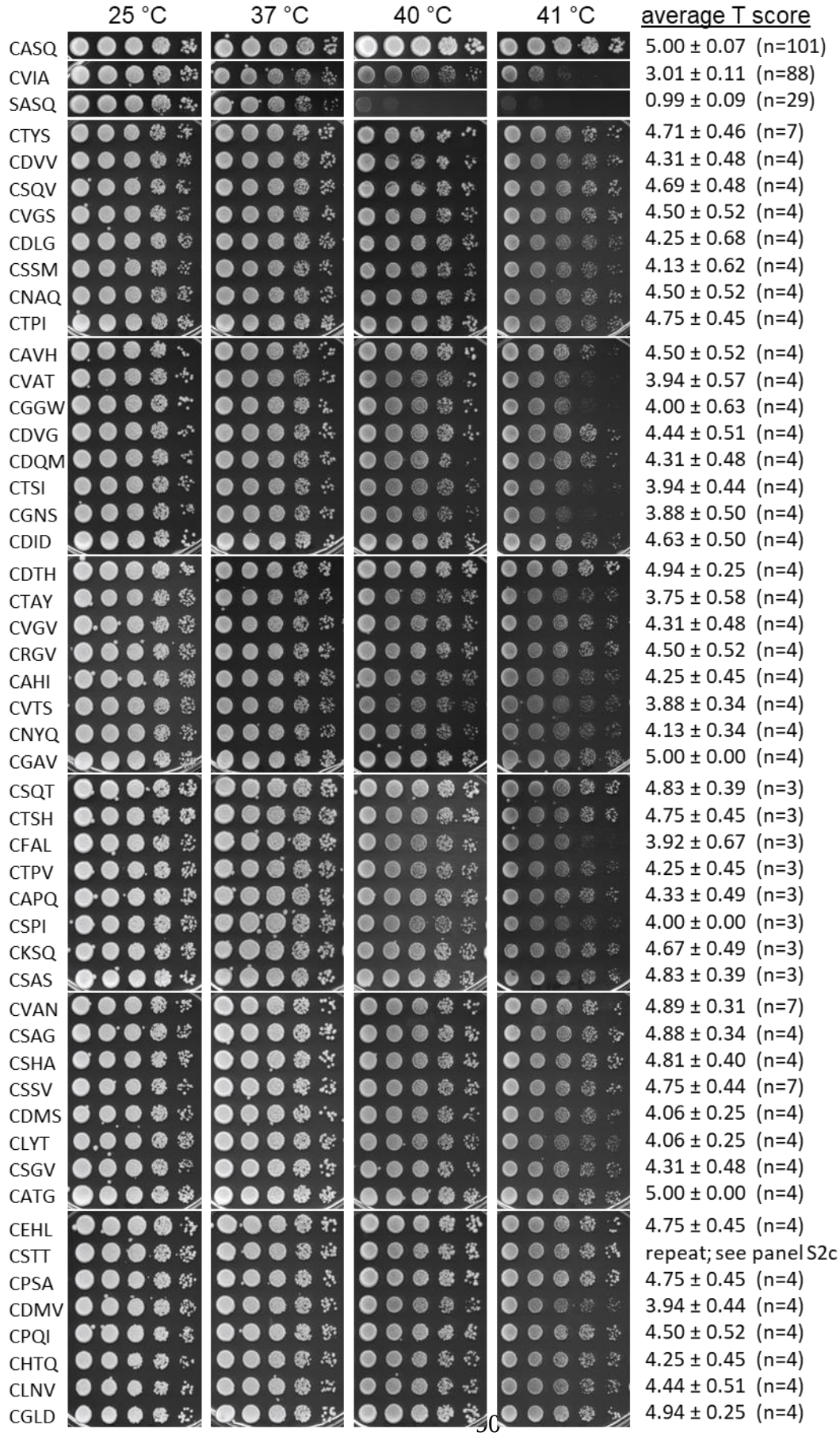
Locus ID	Gene	Motif	PS	CS
Weakly Prenylated / Strongly Cleaved				
YOR257W	*CDC31*	CTDS	-1	1
Weakly Prenylated / Weakly Cleaved				
YGR068C	*ART5*	CDDD	-1	-3
YGL263W	*COS12*	CNDV	-1	-1
YBR042C	*CST26*	CFIF	-1	0
YBR033W	*EDS1*	CFFN	-1	-1
YCR020C	*PET18*	CYNA	-1	-2
YOR242C	*SSP2*	CIDL	-1	0
YML041C	*VPS71*	CRNR	-1	-4
YJL059W	*YHC3*	CRME	-1	0
YBR096W		CSEI	-1	-1
YIL134C-A		CAPY	-1	-2
YPL109C		CPNY	-1	-3
YDL186W		CHHD	-1	-4
YMR187C		CKGE	-1	-4
Not Prenylated				
YMR300C	*ADE4*	CADY	-2	-1
YPR093C	*ASR1*	CHDE	-3	-3
YOR299W	*BUD7*	CYDA	-2	-1
YDR301W	*CFT1*	CQGK	-3	-5
YOR031W	*CRS5*	CEKC	-2	-2
YHR053C	*CUP1-1*	CSGK	-2	-4
YHR055C	*CUP1-2*	CSGK	-2	-4
YNL255C	*GIS2*	CPKA	-2	-1
YDR528W	*HLR1*	CTRK	-4	-3
YML075C	*HMG1*	CIKS	-2	0
YOR231W	*MKK1*	CWKD	-3	-4
YMR158W	*MRPS8*	CRVK	-2	-1
YMR023C	*MSS1*	CIGK	-2	-3
YDR307W	*PMT7*	CLAK	-2	-2
YGL045W	*RIM8*	CDDY	-2	-3
YDR257C	*RKM4*	CVKK	-4	-3
YBR150C	*TBS1*	CVKM	-2	0
YBR209W		CSKP	-4	-4
YKL069W		CVFK	-2	-1
YLR154C-G		CDGP	-2	-5
YOR034C-A		CQRK	-5	-5

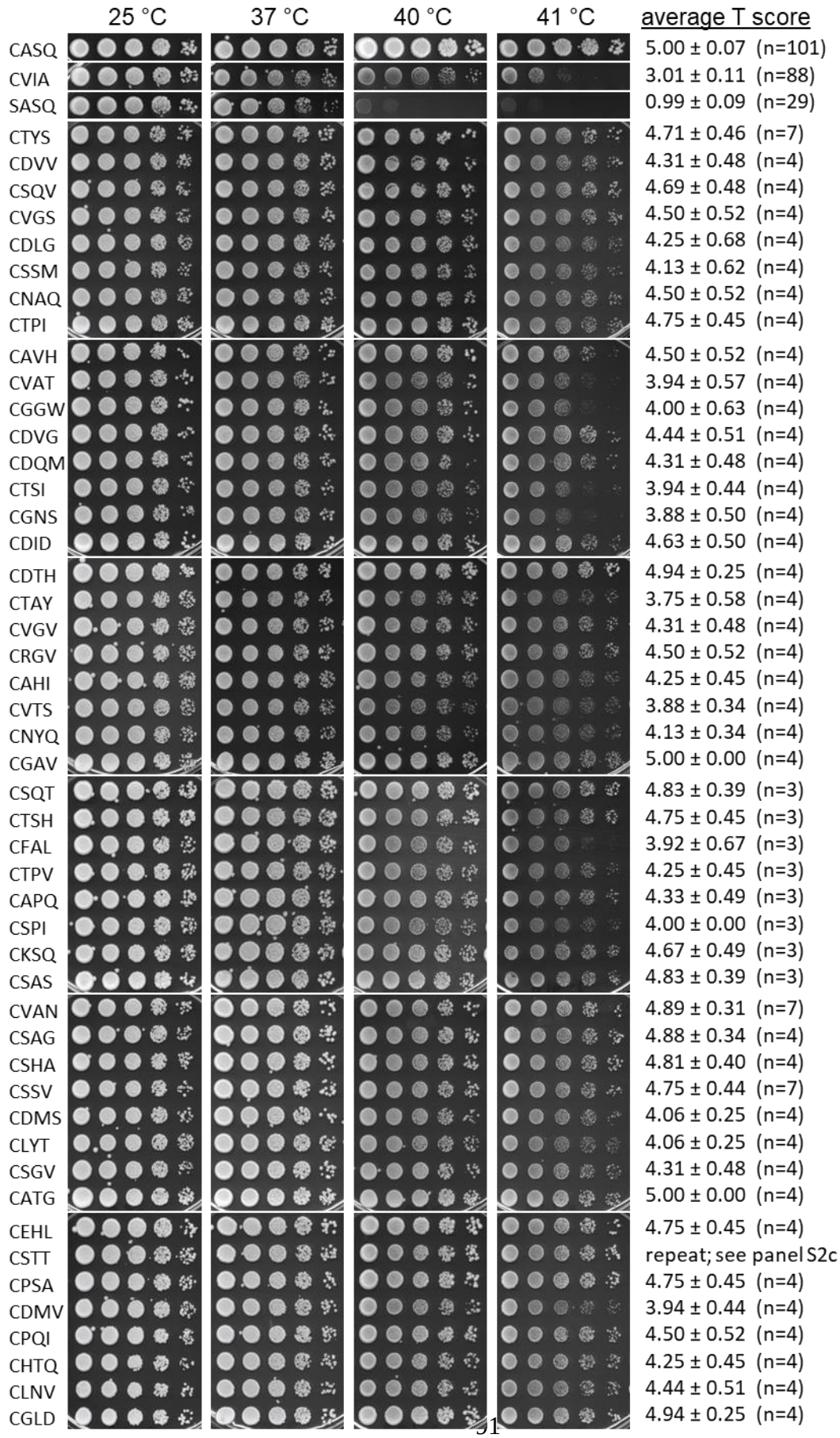
[a] Known to be farnesylated, geranylgeranylated or highly likely to be prenylated based on similarity to a homolog.
[b] PS – in-house algorithm prenylation score; CS – in-house algorithm cleavage score; see Materials and Methods for description of scoring systems.

Species	% Identity	COOH-terminus
S. cerevisiae (Ydj1p)	100.0	…VQCASQ
K. lactis	73.2	…VQCASQ
A. gossypii	71.4	…VQCASQ
S. pombe	55.6	…VQCAQQ
N. crassa	53.1	…VQCASQ
M. oryzae	52.0	…VQCASQ
G. gallus	51.5	…VQCAHQ
M. musculus	50.9	…VQCAHQ
H. sapiens	50.5	…VQCAHQ
D. rerio	50.1	…VQCAHQ
X. tropicalis	50.1	…VQCAHQ
O. sativa	48.1	…VQCAQQ
A. thaliana	47.3	…VQCAQQ
C. elegans	45.7	…VRCQHQ

Figure S2.1. COOH-terminal sequences of Ydj1p and related homologs.
"Ydj1" was used as a query to retrieve related sequences and overall percent identity scores from the Homologene database (http://www.ncbi.nlm.nih.gov/homologene). Only the last 6 amino acids of each homolog are shown. The retrieved list of homologs was culled to reduce over-representation of mammalian homologs and to limit each species to one representative when multiple entries were found. When multiple entries were observed, the homolog with the highest identity score relative to Ydj1p was retained.


25 °C
37 °C
40 °C
41 °C
average T score
CASQ 5.00 ± 0.07 (n=101)
CVIA 3.01 ± 0.11 (n=88)
SASQ 0.99 ± 0.09 (n=29)
CLAS 4.81 ± 0.40 (n=4)
CESV 4.88 ± 0.34 (n=4)
CAPL 4.75 ± 0.45 (n=4)
CTHA 4.69 ± 0.48 (n=4)
CGVE 4.69 ± 0.48 (n=4)
CSAN 4.69 ± 0.48 (n=4)
CNTV 4.69 ± 0.48 (n=4)
CGLL 4.75 ± 0.45 (n=4)
CGQV 4.89 ± 0.32 (n=9)
CEVQ 4.55 ± 0.51 (n=5)
CASV 4.55 ± 0.51 (n=5)
CIYG 3.80 ± 0.77 (n=5)
CAVG 3.65 ± 0.59 (n=5)
CVGL 4.45 ± 0.51 (n=5)
CAGV 4.69 ± 0.47 (n=9)
CSTI 4.75 ± 0.45 (n=3)
CLST 4.81 ± 0.40 (n=4)
CVNT 4.88 ± 0.34 (n=4)
CPNV 4.81 ± 0.40 (n=4)
CAGV repeat; see above
CAGT 4.69 ± 0.48 (n=4)
CDLT 4.25 ± 0.68 (n=4)
CVSA 4.75 ± 0.45 (n=4)
CASS 5.00 ± 0.00 (n=4)
CHVD 4.69 ± 0.48 (n=4)
CASQ 4.56 ± 0.51 (n=4)
CAAG 4.75 ± 0.45 (n=4)
CTSV 4.19 ± 0.66 (n=4)
CFNG 4.25 ± 0.58 (n=4)
CRSV 4.31 ± 0.48 (n=4)
CEYG 4.31 ± 0.48 (n=4)
CSQY 4.75 ± 0.45 (n=4)
CSTT 4.82 ± 0.39 (n=7)
CSQS 5.00 ± 0.00 (n=3)
CTGV 4.92 ± 0.29 (n=3)
CSEQ 4.75 ± 0.45 (n=3)
CRIW 2.75 ± 0.45 (n=3)
CRTA 4.42 ± 0.51 (n=3)
CNQH 4.42 ± 0.51 (n=3)
CSAI 4.67 ± 0.49 (n=3)
CSHV 5.00 ± 0.00 (n=3)
CRPL 4.92 ± 0.29 (n=3)
CPST 4.92 ± 0.29 (n=3)
CRSQ 4.42 ± 0.51 (n=3)
CVGI 4.17 ± 0.39 (n=3)
CSAV 4.25 ± 0.45 (n=3)
CTYS 4.71 ± 0.46 (n=7)


25 °C
37 °C
40 °C
41 °C
average T score
CASQ 5.00 ± 0.07 (n=101)
CVIA 3.01 ± 0.11 (n=88)
SASQ 0.99 ± 0.09 (n=29)
CTYS 4.71 ± 0.46 (n=7)
CDVV 4.31 ± 0.48 (n=4)
CSQV 4.69 ± 0.48 (n=4)
CVGS 4.50 ± 0.52 (n=4)
CDLG 4.25 ± 0.68 (n=4)
CSSM 4.13 ± 0.62 (n=4)
CNAQ 4.50 ± 0.52 (n=4)
CTPI 4.75 ± 0.45 (n=4)
CAVH 4.50 ± 0.52 (n=4)
CVAT 3.94 ± 0.57 (n=4)
CGGW 4.00 ± 0.63 (n=4)
CDVG 4.44 ± 0.51 (n=4)
CDQM 4.31 ± 0.48 (n=4)
CTSI 3.94 ± 0.44 (n=4)
CGNS 3.88 ± 0.50 (n=4)
CDID 4.63 ± 0.50 (n=4)
CDTH 4.94 ± 0.25 (n=4)
CTAY 3.75 ± 0.58 (n=4)
CVGV 4.31 ± 0.48 (n=4)
CRGV 4.50 ± 0.52 (n=4)
CAHI 4.25 ± 0.45 (n=4)
CVTS 3.88 ± 0.34 (n=4)
CNYQ 4.13 ± 0.34 (n=4)
CGAV 5.00 ± 0.00 (n=4)
CSQT 4.83 ± 0.39 (n=3)
CTSH 4.75 ± 0.45 (n=3)
CFAL 3.92 ± 0.67 (n=3)
CTPV 4.25 ± 0.45 (n=3)
CAPQ 4.33 ± 0.49 (n=3)
CSPI 4.00 ± 0.00 (n=3)
CKSQ 4.67 ± 0.49 (n=3)
CSAS 4.83 ± 0.39 (n=3)
CVAN 4.89 ± 0.31 (n=7)
CSAG 4.88 ± 0.34 (n=4)
CSHA 4.81 ± 0.40 (n=4)
CSSV 4.75 ± 0.44 (n=7)
CDMS 4.06 ± 0.25 (n=4)
CLYT 4.06 ± 0.25 (n=4)
CSGV 4.31 ± 0.48 (n=4)
CATG 5.00 ± 0.00 (n=4)
CEHL 4.75 ± 0.45 (n=4)
CSTT repeat; see panel S2c
CPSA 4.75 ± 0.45 (n=4)
CDMV 3.94 ± 0.44 (n=4)
CPQI 4.50 ± 0.52 (n=4)
CHTQ 4.25 ± 0.45 (n=4)
CLNV 4.44 ± 0.51 (n=4)
CGLD 4.94 ± 0.25 (n=4)


25 °C
37 °C
40 °C
41 °C
average T score
CASQ 5.00 ± 0.07 (n=101)
CVIA 3.01 ± 0.11 (n=88)
SASQ 0.99 ± 0.09 (n=29)
CTYS 4.71 ± 0.46 (n=7)
CDVV 4.31 ± 0.48 (n=4)
CSQV 4.69 ± 0.48 (n=4)
CVGS 4.50 ± 0.52 (n=4)
CDLG 4.25 ± 0.68 (n=4)
CSSM 4.13 ± 0.62 (n=4)
CNAQ 4.50 ± 0.52 (n=4)
CTPI 4.75 ± 0.45 (n=4)
CAVH 4.50 ± 0.52 (n=4)
CVAT 3.94 ± 0.57 (n=4)
CGGW 4.00 ± 0.63 (n=4)
CDVG 4.44 ± 0.51 (n=4)
CDQM 4.31 ± 0.48 (n=4)
CTSI 3.94 ± 0.44 (n=4)
CGNS 3.88 ± 0.50 (n=4)
CDID 4.63 ± 0.50 (n=4)
CDTH 4.94 ± 0.25 (n=4)
CTAY 3.75 ± 0.58 (n=4)
CVGV 4.31 ± 0.48 (n=4)
CRGV 4.50 ± 0.52 (n=4)
CAHI 4.25 ± 0.45 (n=4)
CVTS 3.88 ± 0.34 (n=4)
CNYQ 4.13 ± 0.34 (n=4)
CGAV 5.00 ± 0.00 (n=4)
CSQT 4.83 ± 0.39 (n=3)
CTSH 4.75 ± 0.45 (n=3)
CFAL 3.92 ± 0.67 (n=3)
CTPV 4.25 ± 0.45 (n=3)
CAPQ 4.33 ± 0.49 (n=3)
CSPI 4.00 ± 0.00 (n=3)
CKSQ 4.67 ± 0.49 (n=3)
CSAS 4.83 ± 0.39 (n=3)
CVAN 4.89 ± 0.31 (n=7)
CSAG 4.88 ± 0.34 (n=4)
CSHA 4.81 ± 0.40 (n=4)
CSSV 4.75 ± 0.44 (n=7)
CDMS 4.06 ± 0.25 (n=4)
CLYT 4.06 ± 0.25 (n=4)
CSGV 4.31 ± 0.48 (n=4)
CATG 5.00 ± 0.00 (n=4)
CEHL 4.75 ± 0.45 (n=4)
CSTT repeat; see panel S2c
CPSA 4.75 ± 0.45 (n=4)
CDMV 3.94 ± 0.44 (n=4)
CPQI 4.50 ± 0.52 (n=4)
CHTQ 4.25 ± 0.45 (n=4)
CLNV 4.44 ± 0.51 (n=4)
CGLD 4.94 ± 0.25 (n=4)

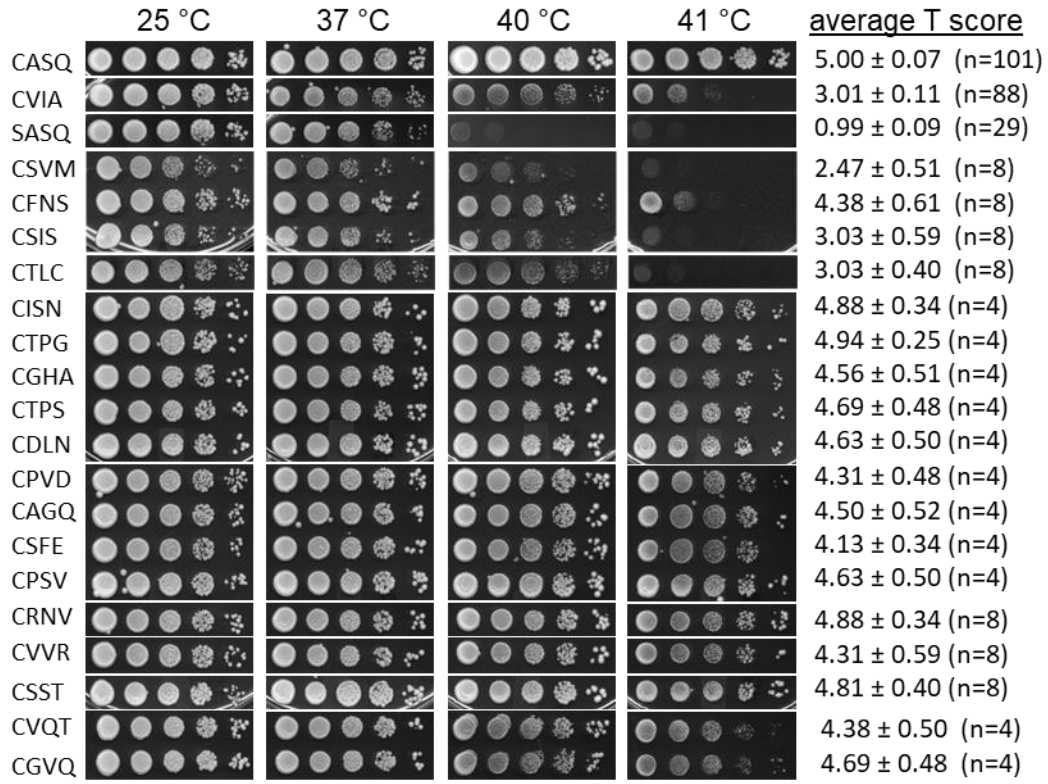


Figure S2.2. Phenotypes and isoprenylation status of Ydj1p Cxxx mutants identified by thermotolerance selection. Ydj1p Cxxx mutants were evaluated for their ability to support growth of a *ydj1*Δ strain at indicated temperatures as described for Figure 2.1. For clarity, only one representative dilution series is shown for each Cxxx mutant, including the reference controls that are replicated on different pages associated with this data set. The dilution series associated with the first 8 Ydj1p Cxxx mutants were used for Figure 2.2D. The thermotolerance profile of each dilution series was scored by 4 independent observers (range 1-5), where SASQ, CVIA and CASQ controls were set to values of 1, 3 and 5, respectively. Multiple replicates (not shown) were scored for each mutant, and scores were averaged to generate a thermotolerance (T) score and standard deviation. The total number of replicates evaluated follows the T score. Two independent yeast transformant colonies were scored for most analyses (n=136); only one transformant was scored in the remaining cases (n=17).

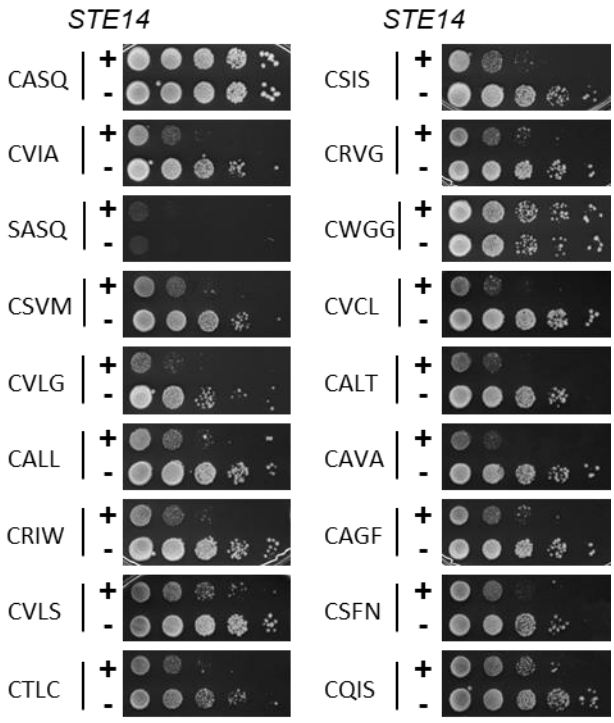


Figure S2.3. Impact of carboxylmethylation on thermotolerance properties of CVIA-like Ydj1p Cxxx mutants. The indicated Ydj1p Cxxx variants were expressed in *ydj1*Δ yeast with or without the Ste14p isoprenylcysteine methyltransferase (+ and -, respectively) and assessed for thermotolerance as described for Figure 2.1C, except that recovery at room temperature was for two days instead of one. Only the 41 °C condition is shown.

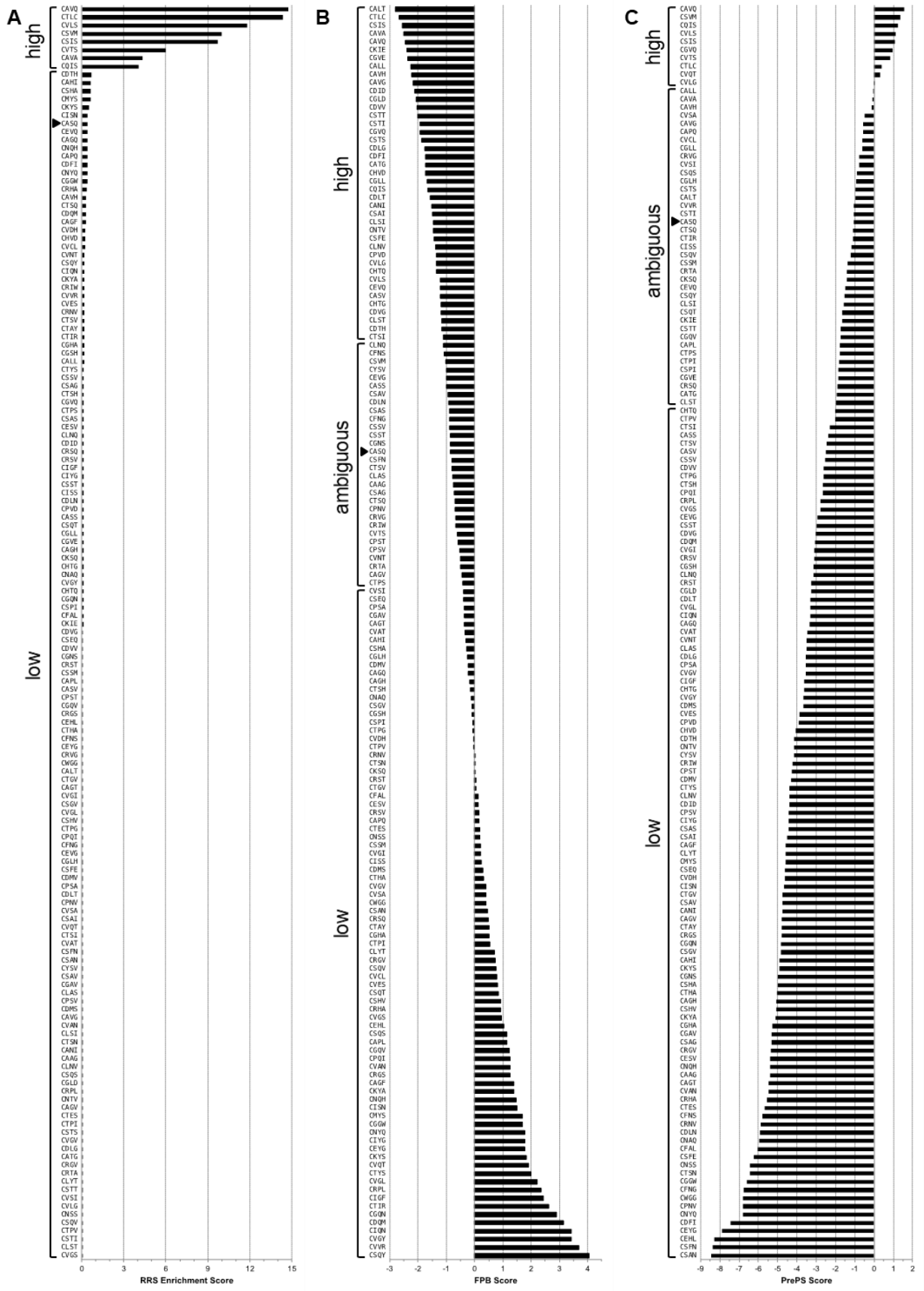

A
high
low
RRS Enrichment Score
B
high
ambiguous
low
FPB Score
C
high
ambiguous
low
PrePS Score

Figure S2.4. The predicted prenylation potential of individual Ydj1p-based hits. A) Hits were matched to a score derived from a Ras-based enrichment strategy for prenylatable sequences and graphed based on score (Stein et al. 2015). Hits are binned into 2 categories: high enrichment in screen (i.e. score greater than 3), and low enrichment (i.e. score less than 1); intermediate scores were not observed (i.e. score 1-3). B) Hits were matched to a score derived using a prenylation prediction algorithm (London et al. 2011). Hits are binned into 3 probability categories: high (i.e. scores less than -1.1); ambiguous (i.e. scores -1.1 to -0.4); low (i.e. scores greater than -0.4). C) Hits were matched to a score derived using the PrePS server (London et al. 2011). Hits are binned into 3 probability categories: high (i.e. scores more than 0); ambiguous (i.e. scores between 0 and -2); low (i.e. scores less than -2). For all graphs, scores were plotted such that sequences with the highest prenylation probability are at the top of each graph; the relative position of CASQ is noted (triangle).

References

1. Wang M, Casey PJ. Protein prenylation: Unique fats make their mark on biology. Nature reviews Molecular cell biology. 2016;17(2):110-22.

2. Hampton SE, Dore TM, Schmidt WK. Rce1: Mechanism and inhibition. Crit Rev Biochem Mol Biol. 2018;53(2):157-74.

3. Silvius JR. Mechanisms of ras protein targeting in mammalian cells. J Membrane Biol. 2002;190:83-92.

4. Winter-Vann AM, Casey PJ. Post-prenylation-processing enzymes as new targets in oncogenesis. Nature reviews Cancer. 2005;5(5):405-12.

5. Konstantinopoulos PA, Karamouzis MV, Papavassiliou AG. Post-translational modifications and regulation of the ras superfamily of gtpases as anticancer targets. Nat Rev Drug Discov. 2007;6(7):541-55.

6. Berndt N, Hamilton AD, Sebti SM. Targeting protein prenylation for cancer therapy. Nat Rev Cancer. 2011;11(11):775-91.

7. Cox AD, Der CJ, Philips MR. Targeting ras membrane association: Back to the future for anti-ras drug discovery? Clin Cancer Res. 2015;21(8):1819-27.

8. Finegold AA, Johnson DI, Farnsworth CC, Gelb MH, Judd SR, Glomset JA, et al. Protein geranylgeranyltransferase of *saccharomyces cerevisiae* is specific for cys-xaa-xaa-leu motif proteins and requires the cdc43 gene product but not the dpr1 gene product. Proc Natl Acad Sci USA. 1991;88(10):4448-52.

9. Moores SL, Schaber MD, Mosser SD, Rands E, O'Hara MB, Garsky VM, et al. Sequence dependence of protein isoprenylation. J Biol Chem. 1991;266(22):14603-10.

10. Yokoyama K, McGeady P, Gelb MH. Mammalian protein geranylgeranyltransferase-i: Substrate specificity, kinetic mechanism, metal requirements, and affinity labeling. Biochemistry. 1995;34(4):1344-54.

11. Hartman HL, Hicks KA, Fierke CA. Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity. Biochemistry. 2005;44(46):15314-24.

12. Maurer-Stroh S, and Eisenhaber, F. Refinement and prediction of protein prenylation motifs. Genome Biol. 2005;6(6):1-15.

13. Krzysiak AJ, Aditya AV, Hougland JL, Fierke CA, Gibbs RA. Synthesis and screening of a caal peptide library versus ftase reveals a surprising number of substrates. Bioorg Med Chem Lett. 2010;20(2):767-70.

14. Caplan AJ, Tsai J, Casey PJ, Douglas MG. Farnesylation of ydj1p is required for function at elevated growth temperatures in *saccharomyces cerevisiae*. J Biol Chem. 1992;267(26):18890-5.

15. Sapkota GP, Kieloch A, Lizcano JM, Lain S, Arthur JS, Williams MR, et al. Phosphorylation of the protein kinase mutated in peutz-jeghers cancer syndrome, lkb1/stk11, at ser431 by p90(rsk) and camp-dependent protein kinase, but not its farnesylation at cys(433), is essential for lkb1 to suppress cell vrowth. J Biol Chem. 2001;276(22):19469-82.

16. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

17. Caplin BE, Hettich LA, Marshall MS. Substrate characterization of the saccharomyces cerevisiae protein farnesyltransferase and type-i protein geranylgeranyltransferase. Biochim Biophys Acta. 1994;1205(1):39-48.

18. Reiss Y, Seabra MC, Armstrong SA, Slaughter CA, Goldstein JL, Brown MS. Nonidentical subunits of p21h-ras farnesyltransferase. Peptide binding and farnesyl pyrophosphate carrier functions. J Biol Chem. 1991;266(16):10672-7.

19. Yokoyama K, Goodwin GW, Ghomashchi F, Glomset JA, Gelb MH. A protein geranylgeranyltransferase from bovine brain: Implications for protein prenylation specificity. Proc Natl Acad Sci USA. 1991;88(12):5302-6.

20. Trueblood C, Ohya Y, Rine J. Genetic evidence for *in vivo* cross-specificity of the caax-box protein prenyltransferases farnesyltransferase and geranylgeranyltransferase-i in *saccharomyces cerevisiae*. Mol Cell Biol. 1993;13(7):4260-75.

21. Omer CA, Gibbs JB. Protein prenylation in eukaryotic microorganisms: Genetics, biology and biochemistry. Mol Microbiol. 1994;11(2):219-25.

22. Trueblood CE, Boyartchuk VL, Rine J. Substrate specificity determinants in the farnesyltransferase beta- subunit. Proc Natl Acad Sci U S A. 1997;94(20):10774-9.

23. Fu HW, Casey, P.J. Enzymology and biology of caax protein prenylation. Recent Prog Horm Res. 1999;54:315-42.

24. Roskoski R, Jr. Protein prenylation: A pivotal posttranslational process. Biochemical and biophysical research communications. 2003;303(1):1-7.

25. Stein V, Kubala MH, Steen J, Grimmond SM, Alexandrov K. Towards the systematic mapping and engineering of the protein prenylation machinery in saccharomyces cerevisiae. PloS one. 2015;10(3):e0120716.

26. Hougland JL, Gangopadhyay SA, Fierke CA. Expansion of protein farnesyltransferase specificity using "tunable" active site interactions: Development of bioengineered prenylation pathways. J Biol Chem. 2012;287(45):38090-100.

27. Reid TS, Terry KL, Casey PJ, Beese LS. Crystallographic analysis of caax prenyltransferases complexed with substrates defines rules of protein substrate selectivity. Journal of molecular biology. 2004;343(2):417-33.

28. London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O. Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS computational biology. 2011;7(10):e1002170.

29. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA. Context-dependent substrate recognition by protein farnesyltransferase. Biochemistry. 2009;48(8):1691-701.

30. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA. Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities. Journal of molecular biology. 2010;395(1):176-90.

31. Wang YC, Dozier JK, Beese LS, Distefano MD. Rapid analysis of protein farnesyltransferase substrate specificity using peptide libraries and isoprenoid diphosphate analogues. ACS Chem Biol. 2014;9(8):1726-35.

32. Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, et al. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. Proc Natl Acad Sci U S A. 2004;101(34):12479-84.

33. Onono FO, Morgan MA, Spielmann HP, Andres DA, Subramanian T, Ganser A, et al. A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with western blotting. Molecular & cellular proteomics : MCP. 2010;9(4):742-51.

34. Suazo KF, Schaber C, Palsuledesai CC, Odom John AR, Distefano MD. Global proteomic analysis of prenylated proteins in plasmodium falciparum using an alkyne-modified isoprenoid analogue. Sci Rep. 2016;6:38615.

35. Wang YC, Distefano MD. Synthetic isoprenoid analogues for the study of prenylated proteins: Fluorescent imaging and proteomic applications. Bioorg Chem. 2016;64:59-65.

36. Jennings BC, Danowitz AM, Wang YC, Gibbs RA, Distefano MD, Fierke CA. Analogs of farnesyl diphosphate alter caax substrate specificity and reactions rates of protein farnesyltransferase. Bioorg Med Chem Lett. 2016;26(4):1333-6.

37. Flom GA, Lemieszek M, Fortunato EA, Johnson JL. Farnesylation of ydj1 is required for in vivo interaction with hsp90 client proteins. Molecular biology of the cell. 2008;19(12):5249-58.

38. Michaelis S, Herskowitz I. The **a**-factor pheromone of *saccharomyces cerevisiae* is essential for mating. Mol Cell Biol. 1988;8(3):1309-18.

39. Chen P, Sapperstein SK, Choi JD, Michaelis S. Biogenesis of the *saccharomyces cerevisiae* mating pheromone **a**-factor. J Cell Biol. 1997;136(2):251-69.

40. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, et al. Designer deletion strains derived from saccharomyces cerevisiae s288c: A useful set of strains and plasmids for pcr-mediated gene disruption and other applications. Yeast (Chichester, England). 1998;14(2):115-32.

41. Ashok S, Hildebrandt ER, Samuelson-Ruiz MC, Hardgrove DS, Schmidt WK, Hougland JL. Expanding the potential prenylome through unanticipated prenylation of non-canonical c-terminal peptide sequences. in preparation.

42. Elble R. A simple and efficient procedure for transformation of yeasts. BioTechniques. 1992;13:18-20.

43. Oldenburg KR, Vo KT, Michaelis S, Paddon C. Recombination-mediated pcr-directed plasmid construction *in vivo* in yeast. Nucleic Acids Res. 1997;25(2):451-2.

44. Blanden MJ, Suazo KF, Hildebrandt ER, Hardgrove DS, Patel M, Saunders WP, et al. Efficient farnesylation of an extended c-terminal c(x)3x sequence motif expands the scope of the prenylated proteome. The Journal of biological chemistry. 2018;293(8):2770-85.

45. Kim S, Lapham A, Freedman C, Reed T, Schmidt W. Yeast as a tractable genetic system for functional studies of the insulin-degrading enzyme. J Biol Chem. 2005;280(30):27481-90.

46. Alper BJ, Nienow TE, Schmidt WK. A common genetic system for functional studies of pitrilysin and related m16a proteases. Biochem J. 2006;398(1):145-52.

47. Crooks GE, Hon G, Chandonia JM, Brenner SE. Weblogo: A sequence logo generator. Genome research. 2004;14(6):1188-90.

48. . !!! INVALID CITATION !!! .

49. Caplan AJ, Tsai J, Casey PJ, Douglas MG. Farnesylation of ydj1p is required for function at elevated growth temperatures in saccharomyces cerevisiae. The Journal of biological chemistry. 1992;267(26):18890-5.

50. Caplan AJ, Cyr DM, Douglas MG. Ydj1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism. Cell. 1992;71:1143-55.

51. Atencio D, Yaffe M. *Mas5*, a yeast homolog of dnaj involved in mitochondrial protein import. Mol Cell Biol. 1992;12:283-91.

52. Lu Z, Cyr DM. The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone ydj1 assist hsp70 in protein folding. J Biol Chem. 1998;273(10):5970-8.

53. Lu Z, Cyr DM. Protein folding activity of hsp70 is modified differentially by the hsp40 co-chaperones sis1 and ydj1. J Biol Chem. 1998;273(43):27824-30.

54. Summers DW, Douglas PM, Ren HY, Cyr DM. The type i hsp40 ydj1 utilizes a farnesyl moiety and zinc finger-like region to suppress prion toxicity. J Biol Chem. 2009;284(6):3628-39.

55. Firth AE, Patrick WM. Glue-it and pedel-aa: New programmes for analyzing protein diversity in randomized libraries. Nucleic Acids Res. 2008;36(Web Server issue):W281-5.

56. Urbaniak G, Plous S. Research randomizer. https://wwwrandomizerorg/. 2013.

57. He B, Chen P, Chen SY, Vancura KL, Michaelis S, Powers S. *Ram2*, an essential gene of yeast, and *ram1* encode the two polypeptide components of the farnesyltransferase that prenylates **a**-factor and ras proteins. Proc Natl Acad Sci USA. 1991;88(24):11373-7.

58. Hrycyna CA, Sapperstein SK, Clarke S, Michaelis S. The *saccharomyces cerevisiae ste14* gene encodes a methyltransferase that mediates c-terminal methylation of **a**-factor and ras proteins. EMBO J. 1991;10:1699-709.

59. Boyartchuk VL, Ashby MN, Rine J. Modulation of ras and **a**-factor function by carboxyl-terminal proteolysis. Science. 1997;275(5307):1796-800.

60. Caldwell GA, Wang S-H, Naider F, Becker J. Consequences of altered isoprenylation targets on **a**-factor export and bioactivity. Proc Natl Acad Sci USA. 1994;91:1275-9.

61. Gangopadhyay SA, Losito EL, Hougland JL. Targeted reengineering of protein geranylgeranyltransferase type i selectivity functionally implicates active-site residues in protein-substrate recognition. Biochemistry. 2014;53(2):434-46.

62. Gotte K, Girzalsky W, Linkert M, Baumgart E, Kammerer S, Kunau WH, et al. Pex19p, a farnesylated protein essential for peroxisome biogenesis. Mol Cell Biol. 1998;18(1):616-28.

63. Emmanouilidis L, Schutz U, Tripsianes K, Madl T, Radke J, Rucktaschel R, et al. Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor pex19. Nature communications. 2017;8:14635.

64. Stephenson RC, Clarke S. Identification of a c-terminal protein carboxyl methyltransferase in rat liver membranes utilizing a synthetic farnesyl cysteine-containing peptide substrate. J Biol Chem. 1990;265:16248-54.

65. Kinsella BT, Erdman RA, Maltese WA. Carboxyl-terminal isoprenylation of ras-related gtp-binding proteins encoded by rac1, rac2, and rala. Journal of Biological Chemistry. 1991;266(15):9786-94.

66. Hrycyna CA, Clarke S. Maturation of isoprenylated proteins in *saccharomyces cerevisiae*. J Biol Chem. 1992;267(15):10457-64.

67. James GL, Goldstein JL, Pathak RK, Anderson RG, Brown MS. Pxf, a prenylated protein of peroxisomes. J Biol Chem. 1994;269(19):14182-90.

68. Galichet A, Gruissem W. Developmentally controlled farnesylation modulates atnap1;1 function in cell proliferation and cell expansion during arabidopsis leaf development. Plant Physiol. 2006;142(4):1412-26.

69. Gomez R, Goodman LE, Tripathy SK, O'Rourke E, Manne V, Tamanoi F. Purified yeast protein farnesyltransferase is structurally and functionally similar to its mammalian counterpart. Biochem J. 1993;289:25-31.

70. Omer CA, Kral AM, Diehl RE, Prendergast GC, Powers S, Allen CM, et al. Characterization of recombinant human farnesyl-protein transferase: Cloning, expression, farnesyl diphosphate binding, and functional homology with yeast prenyl-protein transferases. Biochemistry. 1993;32(19):5167-76.

71. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nature genetics. 1996;14(4):450-6.

72. Michaelis S, Herskowitz I. The a-factor pheromone of saccharomyces cerevisiae is essential for mating. Molecular and cellular biology. 1988;8(3):1309-18.

73. Chen P, Sapperstein SK, Choi JD, Michaelis S. Biogenesis of the saccharomyces cerevisiae mating pheromone a-factor. The Journal of cell biology. 1997;136(2):251-69.

74. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the saccharomyces cerevisiae genome. Nature. 2002;418(6896):387-91.

75. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the saccharomyces cerevisiae genome. Nature. 2002;418(6896):387-91.

76. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in saccharomyces cerevisiae. Genetics. 1989;122(1):19-27.

77. Krishnankutty RK, Kukday SS, Castleberry AJ, Breevoort SR, Schmidt WK. Proteolytic processing of certain caax motifs can occur in the absence of the rce1p and ste24p caax proteases. Yeast (Chichester, England). 2009;26(8):451-63.

CHAPTER 3

Functional classification and validation of yeast prenylation motifs using machine learning and genetic reporters

Berger, B.M., Yeung W., Goyal A., Zhou Z., Hildebrandt E.R., Kannan N., Schmidt W.K.

Submitted to PLOS One, October 20, 2021.

Abstract

Protein prenylation by farnesyltransferase (Ftase) is often described as the targeting of a cysteine-containing motif (CaaX) that is enriched for aliphatic amino acids at the a_1 and a_2 positions, while quite flexible at the X position. Prenylation prediction methods often rely on these features despite emerging evidence that Ftase has broader target specificity than previously considered. Using a machine learning approach and training sets based on canonical (prenylated, proteolyzed, and carboxymethylated) and recently identified shunted motifs (prenylation only), this study aims to improve prenylation predictions with the goal of determining the full scope of prenylation potential among the 8000 possible Cxxx sequence combinations. Further, this study aims to subdivide the prenylated sequences as either shunted (i.e., uncleaved) or cleaved (i.e., canonical). Predictions were determined for *Saccharomyces cerevisiae* Ftase and compared to results derived using currently available prenylation prediction methods. *In silico* predictions were further evaluated using *in vivo* methods coupled to two yeast reporters, the yeast mating pheromone **a**-factor and Hsp40 Ydj1, that represent proteins with canonical and shunted CaaX motifs, respectively. Our machine learning based approach expands the repertoire of predicted Ftase targets and provides a framework for functional classification.

Introduction

CaaX-type protein prenylation refers to the covalent linkage of a farnesyl or geranylgeranyl isoprenoid group (C15 and C20, respectively) to proteins containing a COOH-terminal CaaX motif, where C is an invariant cysteine, a_1 and a_2 are typically

aliphatic residues, and X is one of many amino acids (1). Farnesyltransferase (Ftase) and geranylgeranyltransferase-I (GGTase-I) facilitate the isoprenoid addition to the CaaX cysteine thiol, with GGTase-I targeting the subset of CaaX sequences having Leu, Phe or Met at the X position (2-4). For many CaaX proteins, initial isoprenylation is followed by proteolysis that removes the aaX tripeptide, mediated by Rce1 or Ste24, and carboxymethylation of the isoprenylated cysteine, mediated by isoprenylcysteine carboxyl methyltransferase (ICMT; Ste14 in yeast) (5). These modifications increase the overall COOH-terminal hydrophobicity of modified proteins and often occur to CaaX proteins well-known to be membrane associated (e.g., Ras GTPases).

Despite Ftase arguably being the most well characterized enzyme in the CaaX modification pathway, its specificity still remains unclear. Early primary sequence comparisons of known Ftase targets often outlined the standard, aliphatic-enriched consensus motif termed CaaX. One of the first methods to predict Ftase substrates was developed into the Prenylation Prediction Suite (PrePS) (6). This method evaluated the last 15 amino acids of known prenylated targets, including many Ras and Ras-related GTPases and a few non-canonical sequences for which evidence of prenylation was previously established, to determine a consensus of physio-biochemical properties important for prenylation, which was then used to predict prenylation. PrePS was then applied to create a database of all prenylation predictions across all known proteins, regardless of species (7). The prenylation potential of nearly all 8000 possible CaaX sequences has also been investigated using genetics and high throughput NextGen Sequencing (NGS) in the context of a mutated form of H-Ras (Ras61) that was

heterologously expressed in yeast (8). The identified target sequences were consistent with the initially described consensus CaaX motif. Parallel *in vitro* and *in silico* studies have suggested, however, that Ftase may be able to accommodate substantially broader substrates than initially proposed (9-13). A broader consensus for human Ftase was also proposed using FlexPepBind (FPB), an approach involving structure-based molecular docking and energy minimization constraints (12). This approach identified several sequences that were not initially expected to be prenylated but subsequently biochemically validated as Ftase targets. Despite these new experimental observations and advancements in prenylation prediction methods, many prenylated sequences still fail to be accurately predicted as Ftase substrates. Past approaches involving *in vitro* peptide libraries and metabolic labeling with farnesyl analogs suitable for click-chemistry have been able to identify additional non-canonical sequences as Ftase targets, however, peptide libraries are often costly and can be labor intensive and metabolic labeling is limited to cell specific sequences (9, 10, 14-17). Thus, limitations still prevent exploration of the full scope of prenylation for all 8000 Cxxx sequences.

While the specificity of Ftase is emerging to be more flexible than anticipated, the CaaX proteases that mediate subsequent cleavage of the aaX tripeptide appear more stringent, requiring aliphatic residues at a_1 and/or a_2 positions (18). This observation identifies an inherent bias in many Ftase assays due to the use of canonical reporters such as Ras and **a**-factor where the specificity of the downstream proteases may limit the prenylatable sequences that can be identified. To overcome this bias, we recently developed *S. cerevisiae* Hsp40 Ydj1 into a novel *in vivo* reporter for yeast Ftase activity

(19). Unlike canonical reporters previously used *in vivo*, the non-canonical CaaX sequence of Ydj1 (CASQ) is farnesylated, then “shunted” out of the canonical CaaX pathway without being further proteolyzed and carboxymethylated. Previous studies have shown that yeast require Ydj1 prenylation for growth at high temperatures (i.e., thermotolerance), as evident by a reduced thermotolerant phenotype observed when canonical modification occurs (i.e., prenylation, proteolysis and carboxymethylation), and a further reduction in thermotolerance with lack of prenylation (19, 20). This thermotolerant phenotype was used to identify 153 sequences that supported Ydj1 prenylation-dependent yeast growth at high temperatures (21). The recovered sequences were vastly different than standard canonical CaaX sequences, lacking characteristic aliphatic amino acids but consistent with specificities observed through *in vitro* and *in silico* studies. For clarity, all 8000 sequences are referred to as Cxxx sequences in this study, while predicted prenylated sequences are referred to as CaaX motifs with qualifiers added to specify those that are canonically modified (i.e., cleaved) or shunted (i.e., uncleaved).

In this study, we used machine learning and yeast genetic data derived from both Ras61 and Ydj1 *in vivo* reporters to develop methods for predicting the prenylation potential of all 8000 Cxxx sequences within the yeast system. Predictions were then compared to those derived using PrePS, FPB, and Freq. The latter is a frequency-based, in-house method developed in our previous study of Cxxx sequences that support Ydj1-dependent thermotolerance. Our findings suggest that the use of machine learning with data derived from both canonical and non-canonical reporters results in

improved prediction of yeast Ftase targets. This approach was also used to develop a first-ever prediction for CaaX proteolysis, leading to effective predictions for establishing whether a prenylated sequence follows the canonical or shunted pathway (i.e., cleaved vs. uncleaved).

Materials and Methods

Training set curation

Prenylation: Training sets can be found in Supplemental File S3.1 and were derived from previously published datasets. The positive set initially included 369 sequences identified through a Ras61 prenylation screen (enrichment score >3 at 37 °C; ≥5 occurrences) and 153 sequences identified through a Ydj1 prenylation screen (8, 21). The positive training set was curated to form a reduced set of 489 unique sequences by removing duplicate sequences that overlapped between the sets (n=8), sequences found naturally in the *Saccharomyces cerevisiae* proteome (n=21), and sequences that had previously been incorporated into reporters (n=4). The negative set initially consisted of 514 sequences that were lowest scoring in the Ras61 prenylation screen (enrichment score ≤0.036 at 37 °C; ≥5 occurrences at 25 °C). The negative set was curated to form a reduced set of 508 unique sequences by removing 6 sequences found naturally in the *Saccharomyces cerevisiae* proteome.

Cleavage: Training sets can be found in Supplemental File S3.1 and were derived from previously published datasets (8, 21). The positive set initially included 153 top scoring

Ras61 sequences (enrichment score >3 at 37 °C; ≥5 occurrences). From this, the positive training set was reduced to a unique set of 140 by removing duplicate sequences that overlapped with the Ydj1 set (n=2), sequences found naturally in the *Saccharomyces cerevisiae* proteome (n=8), and sequences that had previously been incorporated into reporters (n=3). The negative set initially included 153 sequences recovered in theYdj1 screen. The negative set was reduced to 136 sequences by removing sequences that were genetically confirmed to be canonically modified (n=15), sequences found naturally in the *Saccharomyces cerevisiae* proteome (n=1), and sequences that had previously been incorporated into reporters (n=1).

Feature generation & pre-processing

Feature generation: In order to generate features for machine learning, we explored three different ways of representing Cxxx sequences: 1) the specific amino acid sequence represented by one-hot encoding, 2) the physico-biochemical features retrieved from the Aaindex database (ftp://ftp.genome.jp/pub/db/community/aaindex/; downloaded 1/17/2021) (22), and 3) sequence embedding generated by ESM-1b (https://github.com/facebookresearch/esm; downloaded 2/9/2021), a state-of-the-art Transformer model that was pre-trained on roughly 250M protein sequences (23). Sequence features were represented by an array of size 60, which accounts for one-hot encoding of 20 amino acid residues at the 3 variable “x” positions of the Cxxx sequence. Aaindex features were represented by an array of size 1659, which accounts for all 553 physico-biochemical features defined by the database for each of the 3 positions. These

features were normalized to a range of 0 to 1 in order to equalize their scales. ESM-1b features were generated by taking advantage of the model's ability to account for contextual information, capturing the potential effects of neighboring residues. We represented the COOH-terminal localization of the Cxxx sequence by front-padding with 100 unspecified "x" residues. In addition, the model added two special characters to represent the beginning and end of the amino acid sequence. This sequence was used to generate an embedding of size (1280, 106), which represents a 1280-dimensional abstract description of 104 residue positions plus two special symbols. ESM-1b features were extracted from this embedding by retrieving the positions corresponding with the Cxxx sequence and end-of-sequence character, which resulted in an array of size (1280, 5), flattened to size 6400. We retained the positional encoding corresponding to the invariant cysteine due to the model's unique ability to capture contextual information.

Dimensionality reduction: Redundant features were removed through principal component analysis, a standard dimensionality reduction technique (24). This resulted in the reduction of sequence features from 60 to 53 dimensions, Aaindex features from 1659 to 50 dimensions, and ESM-1b features from 6400 to 276 dimensions. These reduced features captured 99% of total variance in each feature set.

Prediction of Cxxx prenylation & cleavage

Scoring: We quantified the performance of all prediction models based on accuracy, precision, recall, and F1-score. Reported values indicate the mean across 10-fold cross validation while confidence intervals indicate the standard deviation.

Position-specific scoring matrix (PSSM): We constructed a PSSM based on prenylated or cleaved motifs. The amino acid distribution was normalized against a background amino acid distribution defined by the BLOSUM62 substitution matrix (25) with a pseudo-count of 0.05. The resulting model was used to calculate the log probability of a given sequence being prenylated or cleaved. In order to obtain binary predictions, we defined a cutoff log probability that best separated the positive from the negative examples.

Machine learning algorithms: We tested the performance of various machine learning algorithms as implemented by Scikit-learn (26). The parameters of individual predictors were optimized by grid search. Specific algorithms tested were support vector machine (SVM), Naïve Bayes, k-nearest neighbor (kNN), and Gradient Boosted Decision Tree (GBDT). In subsequent analyses, we estimated the probabilities of each prediction for SVM through Platt scaling (27).

Software: All computational analyses, unless otherwise mentioned, were implemented in Python 3 using NumPy (28) and PyTorch (29). Figure plots were created using

Matplotlib (30), seaborn (31), WebLogo3(32), and Adobe Illustrator. For WebLogo3, a custom color scheme was used where cysteine I was blue, polar charged amino acids (H, K, R, E, D) were green, polar uncharged amino acids (N, Q, S, T, Y) were black, branched-chain amino acids (L, I, V) were red, and all other amino acids (F, A, P, G, M, W) were purple. This scheme matches that used in a previously published study of FTase specificity by our group (21).

Cut-offs used for predictions by prenylation methods: For analysis with the Prenylation Prediction Suite, (PrePS; https://mendel.imp.ac.at/PrePS), all 8000 Cxxx sequences were evaluated in the context of human H-Ras (RQHKLRKLNPPDESGPGCMSCKCxxx). While PrePS only requires 15 amino acids for scoring, 26 were used to remain consistent with previous studies (19, 21). For PrePS, sequences scoring greater than -2 were deemed positive predictions. For FlexPepBind, sequences scoring greater than -1.1 were deemed positive predictions, consistent with the stringent threshold defined by the original study (12). For Freq, prenylation sequences scoring greater than -1 were deemed positive predictions, while sequences scoring greater than 0 were deemed positive predictions for cleavage (21).

Experimental validation

Yeast strains: Strains used in this study are listed in Supplementary Materials **Table S3.3**. Lithium acetate-based transformation methods were used to introduce plasmids into yeast strains (21, 33). All strains were propagated at 25 °C unless otherwise stated,

in YPD or appropriate selection media. For yWS2393, deletion of *STE24* was carried out in strain yWS44 (*mfa1*Δ *mfa2*Δ) using: a DNA fragment from pWS405 (*CEN URA3 ste24::KanMX4)* that was transformed into yWS44 (34). G418 resistant colonies were checked by PCR for integration of *ste24::KANMX4* at the *STE24* locus. For yWS2462, deletion of *RCE1* was carried out in strain yWS44 using a *rce1::KAN* fragment recovered by PCR from the haploid yeast gene deletion collection (35), and integration at the *RCE1* locus was confirmed by PCR.

Plasmids: Plasmids used in this study are listed in Supplementary Material **Table S3.4**. All plasmids newly created for this study were constructed using methods previously reported (19, 21, 36). Briefly, new plasmids encoding Ydj1 or **a**-factor reporters were constructed using PCR-directed recombination. Mutagenic oligonucleotides (**Table S3.5**) encoding desired Cxxx sequences were co-transformed with linearized or gapped parent plasmids, transformation mixes plated onto appropriate selection media, and plasmids recovered from surviving colonies. Plasmids were sequenced through the entire open reading frame of the reporter using an appropriate DNA sequencing primer and a sequencing service (Genewiz, Southfield NJ; Eurofins Genomics, Louisville, Kentucky). pWS130 (*2µ URA3* P_{PGK}*-HsRce1*Δ*22*) was constructed by subcloning a PCR-derived fragment from a baculovirus expression vector encoding *HsRce1*Δ*22* (courtesy of P. Casey, Duke University). The PCR fragment was designed to contain 5´ BamHI and 3´ PstI sites that were used for subcloning, where the latter was blunted with T4 Polymerase prior to cloning into the BamHI and SacII sites of pWS28 (2µ URA3 PPGK) (37). pWS1609 was created from pWS1275 (*2µ URA3* P_{PGK}*-HA-HsSTE24*) by

PCR-directed, plasmid-based recombination to eliminate the HA-tag, followed by subcloning P_{PGK}-*HsSTE24* into pRS316 (*CEN URA3*) (36, 38).

Ydj1 gel shift assay: The prenylation status of Ydj1 was examined as described previously. Briefly, yeast strains expressing Ydj1 were cultured to A_{600} 0.9-1.1 at 30 °C in synthetic complete media lacking uracil (SC-U). Cell pellets of the same mass were collected by centrifugation, washed with water, and cell extracts prepared by alkaline hydrolysis followed by TCA precipitation (39). Cell extracts were resuspended in Sample Buffer (250 mM Tris, 6 M Urea, 5% β-mercaptoethanol, 4% SDS, 0.01% bromophenol blue, pH 8) and analyzed by SDS-PAGE and immunoblotting with rabbit anti-Ydj1 antibody (courtesy of Dr. Avrom Caplan) and HRP conjugate antibody in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween-20; pH 7.5) with 1% milk/TBST. Blots were developed with WesternBright ECL Spray (Advansta Inc, San Jose, California), and images captured using X-ray film or a digital imager (Kwikquant, Kindle Biosciences, Greenwich, Connecticut).

Yeast mating assay: Mating assays were performed as previously described (21). Briefly, *MAT***a** and *MAT*α strains were cultured to saturation at 30 °C in synthetic complete media lacking leucine (SC-L) and YPD, respectively, then normalized to an A_{600} value of 1 by dilution with appropriate sterile media. *MAT***a** cultures were mixed individually 1:10 with the *MAT*α cultures, each mixture was serially diluted 10-fold using 115dentify115lized *MAT*α culture as the diluent, and serial dilutions were pinned onto minimal (SD) and synthetic complete media lacking lysine plates (SC-K). Plates were

incubated for 72 hours and imaged against a black background using flat-bed scanner. Images were adjusted using Photoshop to optimize the dynamic range of signal by adjusting input levels to a fixed range of 25-150.

Data Availability

Strains and plasmids are available upon request. All relevant datasets for this study are included in the supplemental files of manuscript. The coding used is publicly available at the GitHub repository: https://github.com/waylandy/prenylation_cleavage_prediction.

Results

Prenylated and cleaved Cxxx sequences can be distinguished based on primary amino acid sequence feature

To evaluate whether the information encoded in primary sequences can be used to distinguish prenylated and cleaved sequences, we first curated a training dataset from two previously published genetic screens that used Ras61 and Ydj1 as reporters (8, 21). As prenylation is necessary for the optimal function of both Ras61 and Ydj1 reporter activities, we curated 489 prenylated sequences by combining the top performing sequences from both screens. Another 508 low performing sequences from the Ras61 study served as the non-prenylated set; the Ydj1-based study did not yield information for low-performing sequences. Notably, prenylation and proteolysis have historically been considered coupled events, and as such, previous methodologies do not report on proteolysis. However, the Ydj1 reporter is uniquely able to differentiate

between shunted (i.e., only prenylated) and cleaved sequences (i.e., canonically modified; prenylated, cleaved and carboxymethylated). Thus, we curated 136 sequences from the Ydj1 screen and 140 sequences from the Ras61 screen to serve as shunted and cleaved sets, respectively (21).

We next evaluated the contribution of three sequence representation methods: one hot encoding of primary sequence (sequence-only), AAindex, and ESM-1b. These methods capture different aspects of Cxxx sequences (see Materials & Methods for additional details) in classifying prenylated and non-prenylated sequences. Two-dimensional projections of each set of features revealed that sequence-only and AAindex features readily distinguish prenylated and non-prenylated sequences, while ESM-1b exhibited poor separation (**Fig 3.1A**). As AAindex appeared to best separate the prenylated and non-prenylated sequences, we used Weblogo to analyze the sequences clustered with the right and left sides of the projection (**Fig 3.1B**). The right-side cluster was mostly composed of prenylated sequences that closely resembled the canonical definition of CaaX, with a clear enrichment of aliphatic amino acids at the a_2 position, and to some extent the a_1 position. By comparison, the left-side cluster was a mixed population of prenylated and non-prenylated sequences lacking these canonical aliphatic residues. Although ESM-1b encodes more information (276 dimensions to capture 99% variance in data compared to 50 dimensions for sequence and AAindex (see Materials & Methods)), the poor separation observed with ESM-1b is likely a consequence of the additional contextual information which could not be sufficiently compressed into two-

dimensional space. All three sequence representation methods, meanwhile, are suitable for separating cleaved and uncleaved sequences (**Fig 3.1C**).

SVM-ESM-1b outperforms several machine learning based models for prenylation and cleavage predictions

A position-specific scoring matrix model (PSSM) is a common bioinformatics method employed for motif detection (40). A variation of this method is used by the PrePS model (6). We thus constructed a PSSM model based on the Cxxx sequences from our curated datasets to establish a baseline for comparisons of other prenylation and cleavage prediction models. The PSSM model applied to a curated dataset of both canonical and non-canonical sequences achieved 83.8 ± 3.3% accuracy for prenylation predictions, and a second PSSM model to predict cleavage achieved 93.8 ± 4.6% accuracy, based on 10-fold cross validation (**Table 3.1**). We next evaluated whether the baseline PSSM classification accuracy could be improved through different representations of Cxxx sequences using machine learning (see Materials & Methods for details on methods used).

For prenylation, most of the 12 machine learning methods evaluated scored above 80% in all categories. We selected the best model based on F1-score, defined as the harmonic mean of precision and recall. Based on this criterion, support vector machine (SVM) paired with ESM-1b features was the best overall performer. We next evaluated how well each model predicted prenylation of a validation set of 31 Cxxx sequences that

were not part of training sets (**Table S3.1**). Within this validation set, 19 of the 31 sequences naturally occur in the yeast proteome. The reasons for choosing these 19 sequences varied: 12 formed 6 pairs that differ by only one amino acid, for example Ras2 (CIIS) and Hmg1 (CIKS); 7 exhibited varying predictions for prenylation with multiple prediction methods (e.g., PrePS, Freq, etc.). The remaining 12 sequences were chosen due to differing predictions by SVM-ESM-1b, PrePS, and the frequency-based scoring system (Freq). The sequences representing the validation set were incorporated onto Ydj1 and prenylation evaluated by a gel shift assay (**Fig 3.4A, Table 3.3**), with the exception of one sequence (CQSQ) that had been previously evaluated (21). Relative to PSSM, most machine learning methods improved at predicting actual prenylation (**Table 3.1**; Validation score). SVM was repeatedly the best overall performer when paired with ESM-1b features. Considering the results of performance testing with training and naïve test sets, SVM paired with ESM-1b features was chosen as the preferred machine learning method for additional prenylation prediction studies.

We also explored sequence cleavage using similar methods (**Table 3.2**). All models performed comparably well based on 10-fold cross validation, with most scoring above 90% in all categories. As observed for prenylation prediction, many of the models surpassed the PSSM model for accuracy and recall, and only 1 bettered PSSM for precision (**Table 3.2**). Overall, SVM paired with either sequence or ESM-1b features achieved the best F1-score for predicting cleavage. As SVM-ESM-1b had the smaller standard deviation, it was chosen as the preferred method for cleavage prediction. We next evaluated how well each model predicted cleavage of the validation set of 19

naturally occurring Cxxx sequences. We incorporated these 19 sequences onto the **a**-factor reporter that conditionally requires both prenylation and cleavage for bioactivity (**Fig 3.4B**). Because 5 of the sequences were not observed to be prenylated by gel-shift assay (CIKS, CIDL, CSEI, CSGL, CSGK), these sequences were not expected to exhibit any **a**-factor activity, which was indeed the case. For this reason, these 5 sequences were not included statistically in the **a**-factor validation set. The remaining 14 sequences either possessed **a**-factor activity, indicative of cleavage, or lacked bioactivity, indicative of only being prenylated. Surprisingly, we found that several models outperformed SVM-ESM-1b on the validation set when considering the 14 prenylated sequences (**Table 3.3, Table S3.1**). We caution, however, that the small size of the validation set may lack sufficient statistical power to make proper comparisons and conclusions.

Global predictions for prenylation and cleavage of Cxxx sequence space

After evaluating different models for prenylation and cleavage with our curated training and validation sets, we chose SVM paired with ESM-1b to predict both prenylation and cleavage for the full scope of Cxxx sequences (Supplemental File S3.2). In the case of prenylation, our model was trained to make binary predictions, but these sequence predictions are better represented on a continuum as partial prenylation could occur, resulting in sequences with fractions of the protein population being prenylated. In order to model this continuum, we obtained probabilistic outputs for the SVM model by Platt scaling (**Fig 3.2**) (27). We note that this method only provides an estimated probability, which does not perfectly translate to a strict cutoff value for the actual binary

classification. Altogether, our analysis of all 8000 Cxxx sequences predicts that 67% (n=5373) are unmodified, 18% (n=1420) are shunted (i.e., prenylation only), and 15% (n=1217) cleaved (i.e., canonically modified; prenylated, cleaved, and carboxylmethylated). (**Fig 3.3A, D**). We also made global predictions using the SVM-ESM-1b prenylation model paired with our previously published Freq method that outperformed all machine learning models on cleavage validation score (**Fig 3.3B,E**), as well as using Freq for both prenylation predictions and cleavage (**Fig 3.3C,F**) (21). All predictions were qualitatively similar, with the majority of the 8000 sequences being unmodified, and more shunted sequences predicted relative to canonical sequences.

Comparisons to previous prenylation methods and evaluation of yeast proteome predictions

Several prenylation predictors have been developed previously. These include PrePS, a PSSM-based model; FlexPepBind (FPB), a molecular docking-based model encompassing energy scores; and Freq, an in-house method developed by scoring the frequency of residues at each position in the positive and negative testing sets used for machine learning in this study. Relative to all 8000 Cxxx sequence space, our SVM-ESM-1b based model predicts prenylation for more sequences (33%) in comparison to PrePS (20%) and FlexPepBind (17%), but less by comparison to Freq (42%). While Freq predicts more prenylated sequences, it is important to note that this method overpredicts prenylation in the negative training set relative to the SVM-ESM-1b model (~40% vs. 3%, respectively). A potential explanation for the higher false positive rate of

Freq may be that this method does not explicitly encode contextual information when generating features. Overall, we conclude that the SVM-ESM-1b based machine learning model predicts more prenylatable space as compared to PrePS and FlexPepBind and may more accurately predict prenylation than our previously reported Freq method. Regarding CaaX cleavage prediction, Freq has been the only available method for binning prenylated sequences as either shunted or cleaved. Freq predicts more shunted sequences relative to PSSM-based predictions (30% vs. 21%, respectively), while the prediction for cleaved sequences is the same in both cases (12%).

Altogether, the yeast genome contains 89 proteins having Cxxx at the COOH-terminus. Prenylation and cleavage predictions were determined for the Cxxx sequences associated with these proteins using our SVM-ESM-1b and PSSM models, respectively. SVM predicted 41 yeast Cxxx proteins to be prenylated, where 32 were canonically modified and 9 were shunted (**Fig 3.3B**). While many of the canonically modified CaaX proteins have been previously characterized (**a**-factor, Ras, etc.), some have non-canonical Cxxx sequences and have not been previously evaluated for their prenylation status, including Cst26 (CFIF; an acyltransferase) and Sua5 (CIQF; involved in threonylcarbamoyladenosine synthesis). Of the 89 Cxxx sequences associated with the yeast proteome, 19 were directly evaluated in this study in the context of the Ydj1 reporter (**Table 3.3, Fig 3.4A, S3.1**). The SVM-ESM-1b model correctly predicted the prenylation (both positive and negative) for 84% of the sequences. By comparison, PrePS was next best, correctly predicting 79%, followed by Freq correctly predicting

74%, and FPB correctly predicting 58%. Because SVM-ESM-1b, PrePS, and Freq performed similarly in predicting prenylation of naturally occurring Cxxx sequences, we evaluated additional sequences to better differentiate the prediction methods. Our lab possesses a large collection of plasmids encoding Ydj1-Cxxx variants (n > 200). Excluding those with Cxxx sequences that were part of machine learning training sets and others for which SVM-ESM-1b and PrePS had the same prediction led us to 12 plasmids with varying differential predictions by SVM, PrePS, and Freq. (**Table 3.4**). For these 12 Cxxx sequences, Freq correctly predicted 10, SVM-ESM-1b correctly predicted 9, and PrePS correctly predicted 4 (**Table 3.4, Fig 3.4B**). All 12 sequences were prenylated to so°ree, resulting in a high percentage of false negatives for PrePS and SVM. Thus, for the combined set of 31 sequences evaluated, SVM correctly predicted 81% (25/31), Freq correctly identified 77% (24/31), and PrePS correctly predicted 61% (19/31) (**Table 3.5**).

For assessing cleavage, we used the yeast **a**-factor mating pheromone as a reporter (**Fig 3.4B**). Canonical modification of **a**-factor (i.e., prenylation, cleavage, and carboxylmethylation) is required for mating of haploid yeast, which can be quantified as an indirect measure of **a**-factor production. As noted previously, for this assessment, we only evaluated the 14 sequences that were confirmed as being prenylated by Ydj1 gel-shift, regardless of whether they were predicted to be prenylated by any computational method. In this case, Freq outperformed SVM-ESM-1b, correctly predicting cleavage for 93% of sequences compared to 71%, respectively; FBP and PrePS are not able to predict cleavage, so they were not evaluated (**Table 3.3**). For sequences where mating

is observed, the mating levels are comparable to that of the wild type **a**-factor sequence (CVIA) (**Fig 3.4B**), indicative of complete rather than partial cleavage.

Limitations of machine learning for predicting CaaX protein PTMs

While SVM-ESM-1b can predict prenylation and cleavage, one limitation is that it does not provide any information about enzyme specificity due to the lack of enzyme-specific training information. For both prenylation and proteolysis, there are two possible enzymes for each reaction. For prenylation, FTase and GGTase-I 124dentify124dprenylate a wide array of CaaX proteins with C15 farnesyl and C20 geranylgeranyl, respectively, while for proteolysis, Rce1 and Ste24 are both able to cleave the farnesylated CVIA motif of **a**-factor, but selectivity is observed for other motifs. The determinants of substrate specificity have not been fully ascertained for the aforementioned enzymes. A case in point is proteolysis of the CaaX motif CSIM, a sequence found on human prelamin A that has long thought to be a substrate of both CaaX proteases. SVM-ESM-1b and PSSM both predict that CSIM is cleaved, which we confirmed by using the **a**-factor reporter. When both proteases were present, comparable mating levels were observed between strains expressing **a**-factor in the context of the native CVIA motif that is cleaved by both Rce1 and Ste24, the CTLM motif that is Rce1-specific, and the CSIM motif (**Fig 3.5A**). When evaluated in the context of just one CaaX protease, we observed that all three motifs could be cleaved by Rce1, but only CVIA was cleaved by Ste24 (**Fig 3.5B**). A similar result was observed when evaluating the human CaaX proteases in our yeast system (**Fig 3.5C**). Our observations are consistent with multiple reports challenging the role of Ste24 as an authentic CaaX protease, including a recent *in vitro* study demonstrating the inability of

the human Ste24 ortholog, ZMPSTE24 to cleave at the Cys(farnesyl)-Ser bond of the CSIM motif, as would be expected for a CaaX protease (41).

Discussion

A collection of *in vivo, in silico* and *in vitro* observations support a wider array of prenylation substrates than those previously defined by the COOH-terminal CaaX motif (8-12, 21). Among the new substrates are those that lack aliphatic amino acids at the a_1 and a_2 position, leading to a broader definition for the prenylation motif. Using the machine learning platform SVM paired with ESM-1b training on CaaX motifs identified using both shunted and canonical reporters, we have developed a robust prediction algorithm for protein prenylation. SVM-based predictions suggest that approximately 33% of all 8000 Cxxx motifs are prenylatable. This estimate is approximately 50% higher than the number of potential targets predicted by PrePS and is approximately double the number of sequences predicted by FlexPepBind (FPB). These findings are not meant to be indicative of the number of prenylated proteins in a cell since far fewer than all 8000 possible Cxxx motifs are encoded in genomes. For example, *S. cerevisiae* encodes only 89 proteins that end in Cxxx. Of these, SVM-ESM-1b predicted 46% (n=41) to be prenylated. By comparison, FPB and PrePS predicted 27% (n=24) and 32% (n=29) of yeast proteins to be prenylated, respectively. Confirmation of SVM-predicted prenylation will need to be evaluated on a case-by-case basis or by application of emerging methods for *in vivo* labeling of prenyl proteins to firmly establish whether SVM is an improvement over previous methods. We fully expect that the predicted SVM-based prenylation in some cases may not be possible in natural proteins

due to inaccessibility of the COOH terminus to cytosolic prenyltransferases, either due to steric constraints for cytosolic proteins or the incompatible topology of membrane or secreted proteins. Despite the potential limitations of our prediction method, it is clear that SVM-ESM-1b predicted prenylation of known, non-canonical Cxxx sequences in instances where other methods did not (e.g., Ydj1 CASQ and Pex19 CKQQ), suggesting that SVM is an improvement for identifying prenylated proteins as a whole. Moreover, the non-canonical CKQS sequence associated with the histone chaperone Nap1 is also predicted to be prenylated by our SVM-ESM-1b model. To date, there exists no direct evidence for yeast Nap1 prenylation, but such evidence does exist for human and plant Nap1 homologs, which both possess a similar CKQQ motif (42, 43). Notably, the CKQQ sequence is also present on the human tumor suppressor Lkb1, another well documented prenylprotein (44).

As part of this study, we were also able to develop SVM-ESM-1b into a first-ever method for distinguishing between shunted (i.e., prenylation only) and cleaved sequences (i.e., canonical). Of the approximately 2600 sequences predicted to be prenylated by SVM, approximately 63% are predicted to be shunted and the remaining 37% cleaved. Again, these findings are not meant to reflect the actual ratio of shunted and cleaved prenylated proteins in cells. In fact, we observe that the predictions are somewhat inversed within the yeast proteome. Of the 41 sequences predicted to be prenylated, 27% are predicted to be shunted and the remaining 73% cleaved. This observation suggests that the cleavage and carboxymethylation of the prenylated COOH terminus may serve an important role *in vivo*, potentially increasing membrane

association, as historically expected for canonical CaaX modifications. While the role of the isoprenyl group on shunted proteins remains unclear, we posit that this PTM may help mediate protein-protein interactions and/or provide a structural role rather than contribute to membrane association. This is supported by observations made on the human protein Spindly, whose Cxxx motif of CPQQ was predicted to be shunted by our SVM model, and for which a farnesyl-dependent protein complex interaction has been proposed (45, 46).

An unexpected result from this study was the observation that Freq and SVM-ESM-1b had a similar level of accuracy for prenylation prediction of the validation set (77% and 80%, respectively). As noted previously, Freq globally predicted more prenylated sequences than SVM-ESM-1b (42% and 33%, respectively), which is consistent with Freq having a higher false positive rate compared to SVM for our negative training set (40% and 3%, respectively). This suggests to us that Freq overpredicts prenylation. It's also worth noting that while Freq and SVM-ESM1b rely on the same data set for predictions, their predictions are not coincident, indicating that predictions are fundamentally different for the two methods. Long term, we expect that future advancements in machine learning will lead to better prediction performance relative to the Freq-based method.

To further improve our prediction methods, one aspect that we wish to especially improve upon is the high false negative rate for prenylation predictions that was determined empirically by evaluating a small subset of test sequences (n=31; **Table**

3.5). While a larger test set may yield a more accurate false negative rate, it remains possible that the high negative false rate is simply due to the training datasets themselves being too small or somehow compromised. We have high confidence that our positive prenylation training set is composed of prenylated sequences that, importantly, were derived from studies involving both canonical and shunted reporters. Our negative training test set, however, was derived from a single study that relied on a canonical reporter, and it is suspected that shunted sequences may be among the negative hits in that study, thus poisoning the quality of our negative test set. Our future studies are aimed at identifying a set of sequences that better reflect non-prenylatable sequences for use as an improved negative training set that we expect to lead to improved prenylation predictions and a lower false negative rate.

Interestingly, we observed that several models out-performed SVM-ESM-1b for cleavage prediction (e.g., PSSM, Freq). As previously noted, a larger set of test sequences may be needed to better assess performance. Alternatively, it may be that a better genetic test for cleavage is required. Previous studies have reported that geranylgeranylated **a**-factor has less mating activity *in vivo* (18, 21, 47), suggesting that the genetic mating assay may only work well in the context of farnesylated **a**-factor. This potentially impacts results associated with the CFIF and CIQF sequences in our test set; the terminal Phe is a preferred GGTase-I feature. SVM-ESM-1b predicted prenylation of both sequences while SVM-ESM-1b, PSSM and Freq methods all predicted cleavage. Prenylation was confirmed in the context of Ydj1, but neither sequence supported **a**-factor mating activity that would be indicative of cleavage. It

remains unclear whether lack of mating activity is due to shunting or geranylgeranylation. Because of this issue, it is difficult to fully assess the accuracy of any of the cleavage predictors described in this study. In terms of the CaaX proteases, while CSIM was identified as a canonical motif, additional genetic studies utilizing **a**-factor were needed to resolve whether cleavage was mediated by Rce1 or Ste24. As the yeast **a**-factor mating pheromone is the only known substrate of Ste24 to date, it is tempting to speculate that Rce1 is the main and possibly only relevant CaaX protease. If that eventually bears out to be the case, then our cleavage predictors could be used to infer Rce1 specificity.

Altogether, we have demonstrated that machine learning can be developed into a useful tool to prediction prenylation and cleavage events associated with CaaX proteins. The utility of this tool is reflected by its ability to better identify possible shunted sequences relative to other publicly available prediction methods, in addition to identifying canonically modified sequences. These findings represent an important step in expanding the full scope of prenylatable motifs in yeast. Given °gh degree of target specificity exhibited by both prenyltransferases and CaaX proteases across species, it is likely that the pren129dentify space identified by this study also represents the full scope of prenylated motifs in humans. Among these are sequences associated with proteins that represent potential new additions to the prenylome, which has implications for the impact of prenyltransferase and protease inhibitors being developed as therapeutics.

Acknowledgements

We thank Avrom Caplan (City College of New York) for anti-Ydj1 antibody, Patrick Casey (Duke University) for the baculovirus vector encoding human Rce1, and Ora Furman-Schueler (Hebrew University of Jerusalem) for sharing FlexPepBind scores. We also thank Jacob Greenway (Schmidt Lab, UGA) and members of the Schmidt Lab for their assistance with methods, reagent preparation, and critical discussions.

Funding: This work was supported by NIH funds to WKS and NK (NIH NIGMS GM132606, https://www.nih.gov/) and funds to WKS (NIH NIGMS R01GM117148, https://www.nih.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared no competing interests exist.

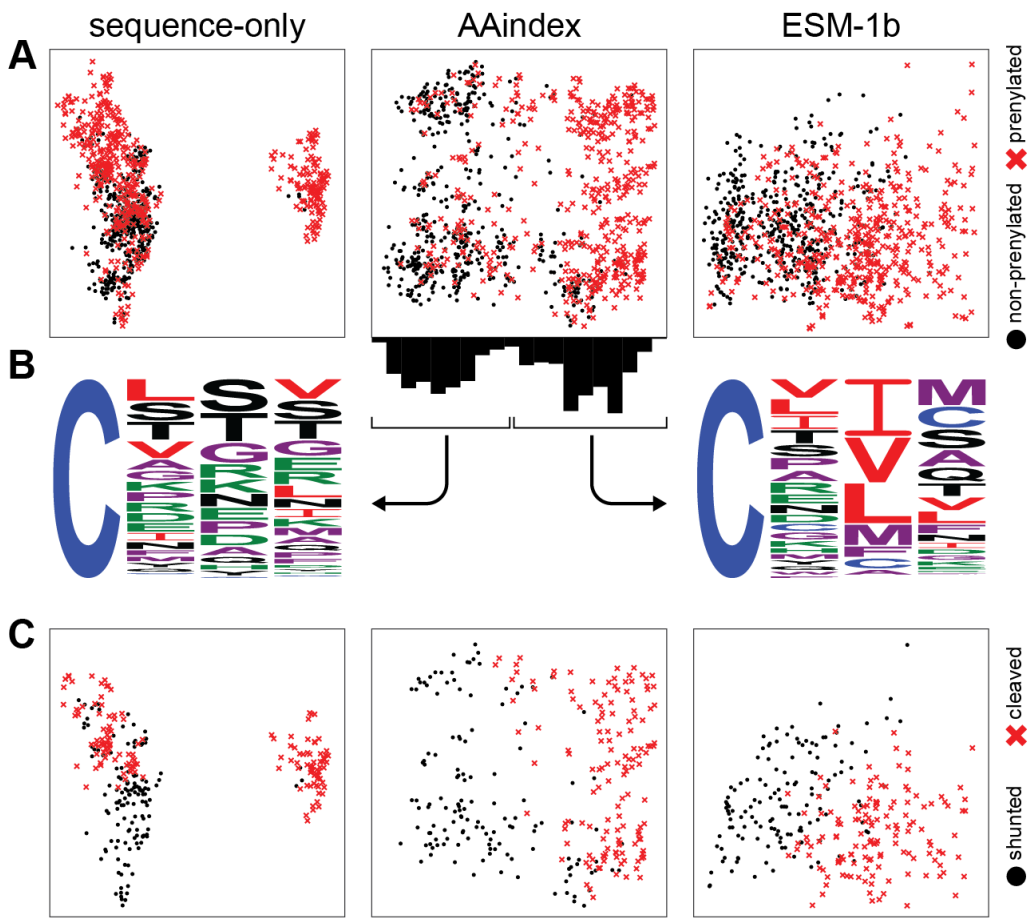


Figure 3.1. Separation of sequences by machine learning-based methods. A) Data points from all three features sets: sequence only, AAindex and ESM-1b, are represented as a two-dimensional projection of prenylated (red x) and non-prenylated sequences (black dot). The axes are not shown as they represent a linear combination of all features that maximizes variance. B) Bimodal distribution of sequences across the X-axis from the AAindex manifold were graphed as sequence logos. The distribution shown on the left contains a mix of non-prenylated Cxxx sequences and prenylated, non-canonical sequences, while the one on the right mostly consists of prenylated, canonical CaaX sequences. C) A similar two-dimensional projection was used to represent cleaved (red x) and shunted (i.e., uncleaved) sequences (black dot).

Table 3.1. Performance of various models for prenylation prediction.

Model[a]	Features[b]	Accuracy[c]	Precision	Recall	F1	Validation[d]
PSSM	sequence	83.8 ± 3.3	87.7 ± 3.5	77.9 ± 5.9	82.4 ± 3.8	68.4 (13/19)
SVM	sequence	86.0 ± 2.7	86.5 ± 4.0	84.9 ± 3.8	85.6 ± 2.8	84.2 (16/19)
SVM	AAindex	85.1 ± 3.5	86.6 ± 4.1	82.4 ± 3.6	84.4 ± 3.6	73.7 (14/19)
SVM	ESM-1b	86.4 ± 3.0	86.6 ± 3.3	85.5 ± 4.1	86.0 ± 3.1	84.2 (16/19)
GBDT	sequence	86.2 ± 2.4	87.9 ± 3.5	83.4 ± 3.5	85.5 ± 2.6	68.4 (13/19)
GBDT	AAindex	86.2 ± 2.8	87.2 ± 3.5	84.3 ± 4.3	85.6 ± 3.0	73.7 (14/19)
GBDT	ESM-1b	85.0 ± 2.9	85.8 ± 3.8	83.2 ± 3.6	84.4 ± 3.0	78.9 (15/19)
Näive Bayes	sequence	82.9 ± 1.8	85.5 ± 3.1	78.7 ± 3.3	81.9 ± 1.9	63.2 (12/19)
Näive Bayes	AAindex	82.1 ± 3.0	82.2 ± 4.0	81.4 ± 3.6	81.7 ± 3.0	73.7 (14/19)
Näive Bayes	ESM-1b	73.2 ± 2.3	70.4 ± 1.9	78.3 ± 3.9	74.1 ± 2.5	57.9 (11/19)
kNN	sequence	84.1 ± 3.7	82.7 ± 4.3	85.5 ± 4.3	84.0 ± 3.7	78.9 (15/19)
kNN	AAindex	82.7 ± 2.3	83.5 ± 3.1	81.0 ± 3.0	82.2 ± 2.3	78.9 (15/19)
kNN	ESM-1b	83.0 ± 2.3	82.4 ± 3.5	83.4 ± 2.4	82.9 ± 2.1	78.9(15/19)

[a]PSSM – Position-specific Scoring Matrix; SVM – support vector machine; GBDT – Gradient Boost Decision Tree; kNN – k-Nearest Neighbors.
[b]Features for predicting sequence prenylation were based on one-hot encoding (sequence), physico-biochemical properties of amino acids (AAindex), and the ESM-1b Transformer model (ESM-1b).
[c]Reported percentages indicate the mean across 10-fold cross validation, while confidence intervals indicate the standard deviation.
[d]Reported percentages based off validation set tested *in vivo*

Table 3.2. Performance of various models for cleavage prediction.

Model[a]	Features	Accuracy	Precision	Recall	F1	Validation
PSSM	sequence	93.8 ± 4.6	97.1 ± 4.5	90.7 ± 7.9	93.6 ± 4.9	89.4 (12 / 14)
SVM	sequence	97.5 ± 2.3	96.7 ± 4.3	98.6 ± 2.9	97.5 ± 2.2	78.9 (10 / 14)
SVM	AAindex	96.4 ± 2.8	95.3 ± 4.1	97.9 ± 3.3	96.5 ± 2.7	78.9 (10 / 14)
SVM	ESM-1b	97.5 ± 1.6	97.3 ± 3.3	97.9 ± 3.3	97.5 ± 1.6	78.9 (10 / 14)
GBDT	sequence	94.9 ± 3.4	94.6 ± 3.9	95.8 ± 5.7	95.0 ± 3.4	52.6 (8 / 14)
GBDT	AAindex	86.9 ± 3.2	87.7 ± 4.3	85.3 ± 3.1	86.4 ± 3.2	73.7 (11 / 14)
GBDT	ESM-1b	86.2 ± 1.9	87.0 ± 2.8	84.5 ± 2.3	85.7 ± 1.9	78.9 (10 / 14)
Näive Bayes	sequence	89.9 ± 6.0	89.1 ± 6.6	91.5 ± 7.0	90.1 ± 5.9	68.4 (9 / 14)
Näive Bayes	AAindex	94.2 ± 2.4	94.5 ± 4.0	94.3 ± 4.3	94.3 ± 2.3	89.4 (12 / 14)
Näive Bayes	ESM-1b	85.5 ± 7.6	85.3 ± 7.3	86.4 ± 10.3	85.7 ± 7.9	68.4 (9 / 14)
kNN	sequence	94.9 ± 4.3	96.5 ± 4.7	93.6 ± 5.9	94.9 ± 4.4	84.2 (12 / 14)
kNN	AAindex	94.6 ± 3.3	92.4 ± 5.9	97.9 ± 3.3	94.9 ± 3.0	78.9 (10 / 14)
kNN	ESM-1b	95.3 ± 4.5	94.7 ± 5.6	96.4 ± 5.8	95.4 ± 4.5	78.9 (10 / 14)

[a]Terms, definitions, and calculations are as described for Table 3.1.

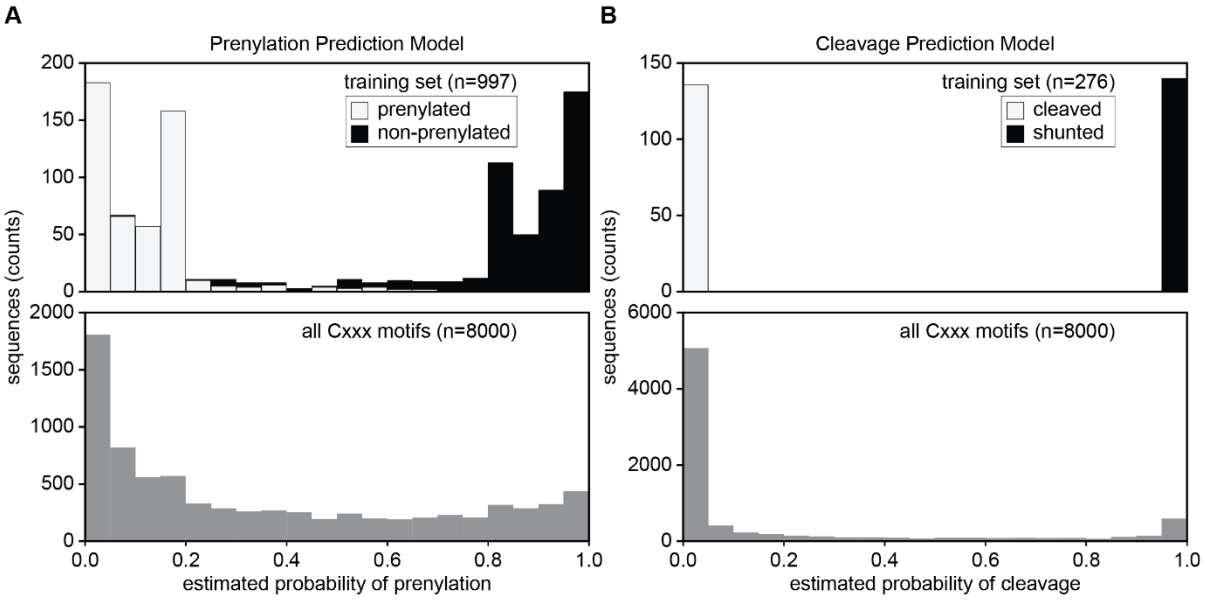


Figure 3.2. Probability distributions for prenylation and cleavage predictions made by SVM-ESM-1b. Probability distributions for both prenylation (A) and cleavage (B) determined for the training sets (top) and for all 8000 Cxxx motifs (bottom). A) For prenylation, the training set distribution is represented as a stacked bar plot where prenylated sequences are white, while non-prenylated sequences are black. B) For cleavage, the training set distribution is represented as a stacked bar plot where shunted sequences (prenylation only) are black and cleaved sequences for proteolysis. The probability distributions were determined for the training sets (top) and for all 8000 Cxxx

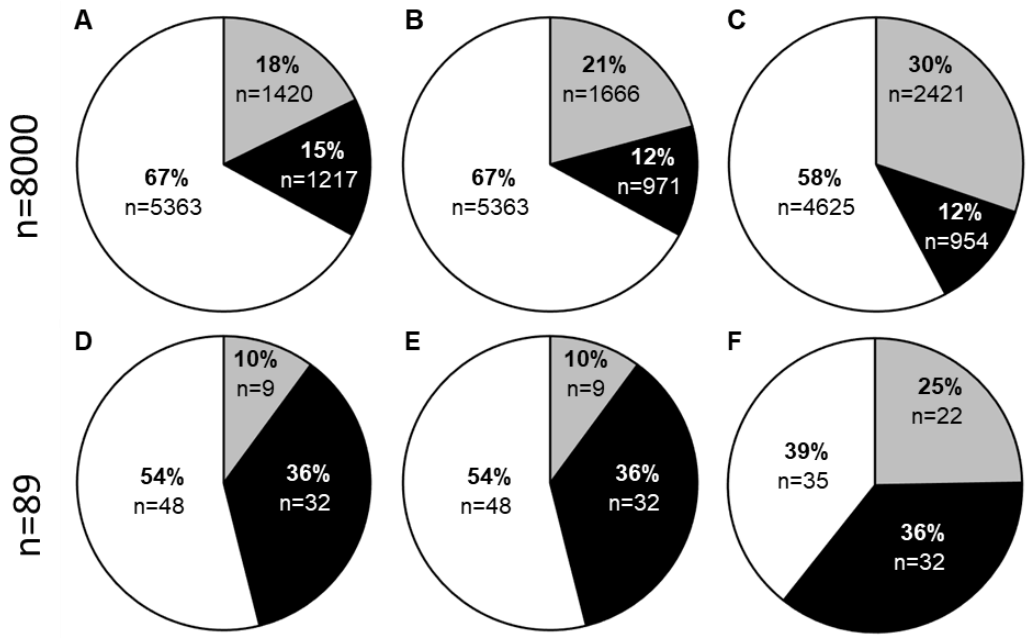


Figure 3.3. Predictions for modification of Cxxx sequences based on various methods. Predictions for prenylation and cleavage for all 8000 Cxxx sequences (A-C) and 89 naturally occurring yeast Cxxx sequences (D-F). Models used were SVM-ESM-1b for both predictions (A, D), SVM-ESM-1b for prenylation and Freq for cleavage (B, E), and Freq for both (C, F). Predictions are binned as non-prenylated (white), shunted (gray), and cleaved sequences (black).

Table 3.3. Comparison of prenylation and cleavage prediction models with empirical observation.

			Prenylation					Cleavage		
	yeast protein	CaaX	SVM[a,b]	PrePS	Freq	FPB	Observed[c]	SVM[a]	Freq	Observed[d]
similar sequences	Ras2	CIIS	+	+	+	+	+	+	+	+
	Hmg1	CIKS	-	-	-	-	-	NA	NA	NA
	Rho2	CIIL	+	+	+	-	+	+	+	+
	Ssp2	CIDL	-	-	-	-	-	NA	NA	NA
	Skt5, MiY1	CVIM	+	+	+	-	+	+	+	+
	Tbs1	CVKM	-	-	-	-	+	+	-	-
	YDL022C-A	CSII	+	+	+	+	+	+	+	+
	YBR096W	CSEI	-	-	-	-	-	NA	NA	NA
	YMR265C	CSNA	-	-	+	-	+	-	-	-
	Pet18	CYNA	-	-	-	+	+	-	-	-
	Lih1	CSGL	-	-	+	-	-	NA	NA	NA
	Cup1	CSGK	-	-	-	-	-	NA	NA	NA
other sequences	Nap1	CKQS	+	+	+	-	+	-	-	-
	Cst26	CFIF	+	+	-	-	+	+	-	-
	YIL134C-A	CAPY	+	+	-	-	+	-	-	-
	Atr1	CTVA	+	+	+	+	+	+	+	+
	Las21	CALD	+	-	+	+	+	-	+	-
	YDL009C	CAVS	+	+	+	+	+	-	+	+
	Sua5	CIQF	+	+	+	-	+	+	-	-
number observed/predicted			16/19	15/19	14/19	11/19		10/14	13/14	

[a]Signs represent predictions of prenylation and cleavage that were reported as positive (+) or negative (-) by the indicated model. NA – not applicable.
[b]SVM – SVM-ESM-1b; PrePS – Prenylation Prediction Suite; Freq – in-house, frequency-based; FPB – FlexPepBind.
[c]Observed by Ydj1 prenylation gel shift – see Figure 3.4A, Figure S3.1.
[d]Observed by **a**-factor mating – see Figure 3.4C.

Table 3.4. Comparison of SVM-ESM-1b and PrePS prenylation predictions with empirical observations.

Reporter	**Prenylation**				
Ydj1-Cxxx	SVM[a,b]	PrePS	Freq	FPB	Observed[c]
CAAQ	+	-	+	-	+
CAHQ	+	-	+	-	+
CASA	+	-	+	-	+
CKQH	+	-	+	-	+
CNLI	+	-	+	-	+
CSFL	+	-	+	-	+
CVAA	+	-	+	-	+
CVFM	+	-	+	-	+
CKQG	-	+	+	-	+
CKQL	-	+	+	-	+
CQTS	-	+	-	-	+
CQSQ[d]	+	+	-	-	+
number observed/predicted	9/12	4/12	10/12	0/12	

[a]Signs represent predictions of prenylation and cleavage that were reported as positive (+) or negative (-) by the indicated model. NA – not applicable.
[b]SVM – SVM-ESM-1b; PrePS – Prenylation Prediction Suite; Freq – in-house, frequency-based; FPB – FlexPepBind.
[c]Observed by Ydj1 prenylation gel shift – see Figure 3.4B.
[d]Observation previously reported (21).

Table 3.5. Summary of prenylation and cleavage predictions.

	Prenylation				Cleavage	
	SVM[a,]	PrePS	Freq	FPB	SVM[a]	Freq
number observed/predicted[b]	**25/31**	**19/31**	**24/31**	**11/31**	**10/14**	**13/14**
%observed/predicted	81%	61%	77%	28%	71.4%	92.9%
number false positive	0/20	0/14	1/22	0/6	3/7	1/7
% false positive	0	0	4.5%	0	42.8%	14.3%
number false negative	6/11	12/17	5/9	20/25	1/7	0/7
% false negative	54.5%	70.6%	55.5%	80%	14.3%	0

[a] SVM – SVM-ESM-1b; PrePS – Prenylation Prediction Suite; Freq – in-house, frequency-based; FPB – FlexPepBind.
[b]Values determined by empirical data via Ydj1 prenylation gel shift (prenylation, Figure 3.4A,B, Figure S3.1) or **a**-factor mating (cleavage, Figure 3.4C).

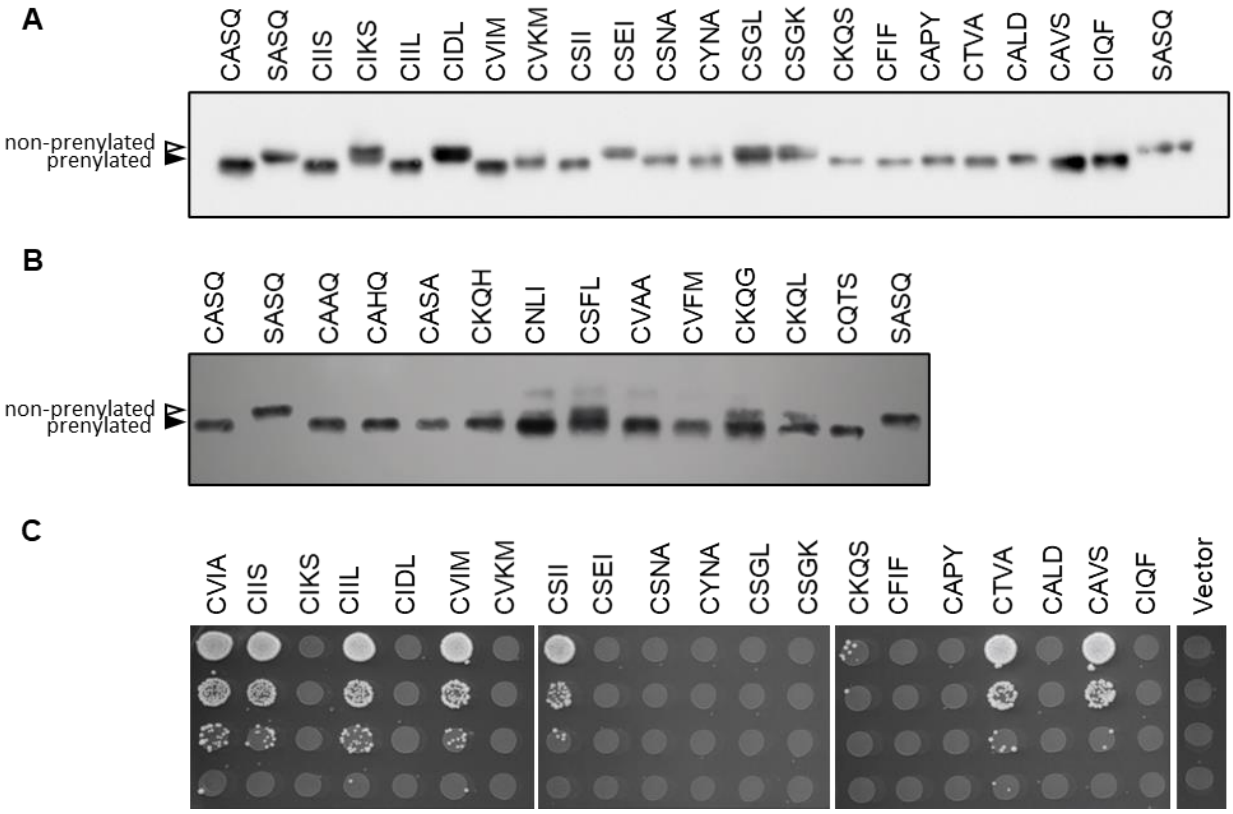


Figure 3.4. Empirically determined prenylation and cleavage of various Cxxx sequences. Yeast strains lacking chromosomally encoded *YDJ1* (yWS304 or yWS2544, *ydj1*Δ) or *MFA1* and *MFA2* (SM2331, *mfa1Δ mfa2Δ*) were engineered to individually express the indicated Ydj1-Cxxx or a-factor-Cxxx variant, respectively, using a plasmid-based expression system (Table S3.2). A, B) Prenylation of the indicated naturally occurring Cxxx sequences in yeast (A) or global Cxxx sequences (B) were determined by Ydj1-gel shift assay. Yeast extracts were evaluated by SDS-PAGE and anti-Ydj1 immunoblot to reveal prenylated (closed triangle) and non-prenylated sequences (open triangle). Partial prenylation (i.e., doublet bands) were counted as a positive result. C) Cleavage of the indicated Cxxx sequences was determined by the a-factor mating assay. *MAT***a** yeast cultures were serial diluted 10-fold in the presence of excess *MATα* yeast (IH1793) and plated on SD media. Mating is indicated by diploid growth and is reported relative to mating exhibited by wildtype a-factor (CVIA).

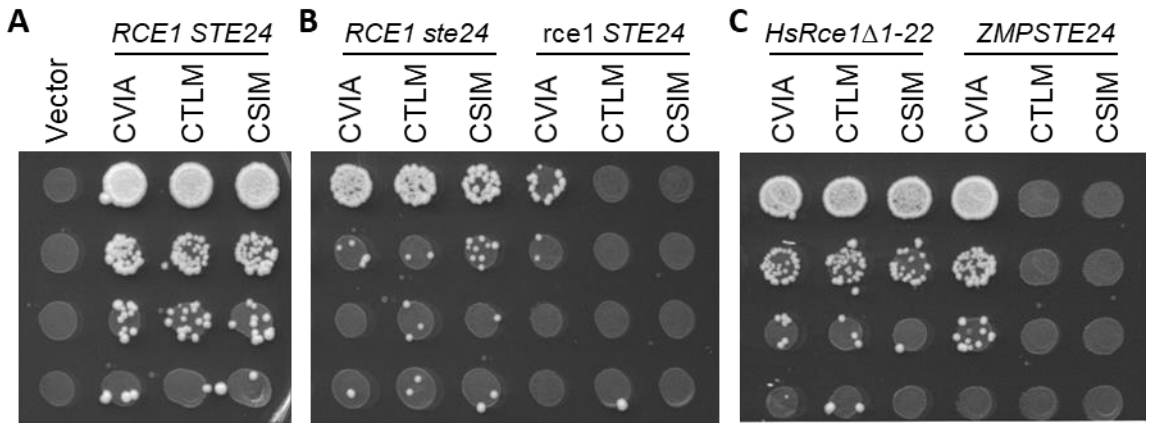


Figure 3.5. Rce1 is responsible for cleavage of yeast a-factor-CSIM. Yeast strains expressing the indicated a-factor Cxxx variant as the sole source of a-factor were evaluated as described for Figure 3.4 in the context of yeast and human CaaX proteases. Yeast strains expressing A) both yeast CaaX proteases (SM2331, *mfa1Δ mfa2Δ*), B) one or the other yeast CaaX protease (yWS2393, *mfa1*Δ *mfa2*Δ *ste24*; yWS2462, *mfa1*Δ *mfa2*Δ *rce1*), or C) plasmid-based human CaaX proteases (pWS130, *HsRce1*Δ*1-22*; pWS1609, *ZMPSTE24*) in a strain lacking both yeast CaaX proteases (yWS164, *mfa1*Δ *mfa2*Δ *rce1 ste24*).

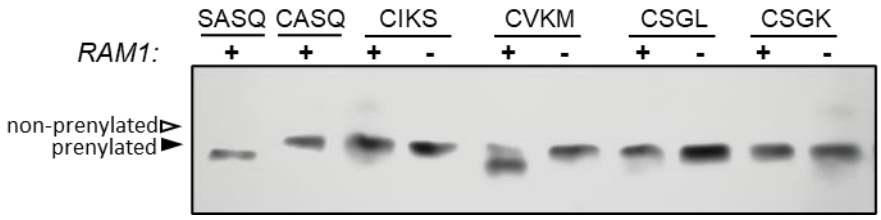


Figure S3.1. Confirmation of prenylation status on ambiguous Cxxx sequences. Yeast strains lacking chromosomally encoded *YDJ1* +/- *RAM1* (yWS304, *ydj1*Δ or yWS2542, *ydj1*Δ*ram1*Δ) expressing Ydj1-Cxxx plasmids of sequences were evaluated in the presence/absence of FTase (*RAM1* gene) as described in Fig 3.4. Sequences were selected from Fig 3.4A for further evaluation due to unclear gel shift or prenylation status.

Supplementary Table S3.1. Probability estimates and prediction calls for prenylation and cleavage of naturally occurring yeast Cxxx sequences as reported by the SVM-ESM-1b model.

		prenylation		cleavage	
yeast protein	**motif**	**score[a]**	**prediction**	**score**	**prediction**
Ras2	CIIS	0.9969	+	6.1594	+
Hmg1	CIKS	0.2587	-	NA	NA
Rho2	CIIL	0.9880	+	5.0875	+
Ssp2	CIDL	0.0534	-	NA	NA
Skt5, MiY1	CVIM	1.0000	+	8.9377	+
Tbs1	CVKM	0.1811	-[b]	1.8026	-
YDL022C-A	CSII	0.9762	+	5.9957	+
YBR096W	CSEI	0.1155	-	NA	NA
YMR265C	CSNA	0.1339	-[b]	-0.6976	-
Pet18	CYNA	0.0738	-[b]	-0.2218	-
Lih1	CSGL	0.1847	-	NA	NA
Cup1	CSGK	0.0034	-	NA	NA
Nap1	CKQS	0.5680	+	1.8671	-
Cst26	CFIF	0.9800	+	5.1732	+[b]
YIL134C-A	CAPY	0.6833	+	-1.2698	-
Atr1	CTVA	0.9860	+	4.8219	+
Las21	CALD	0.6017	+	2.4882	-
YDL009C	CAVS	0.9831	+	4.3306	+
Sua5	CIQF	0.9018	+	4.0671	-

[a]Probability estimates were determined for the SVM-ESM-1b model using Platt-scaling. Signs represent predictions of prenylation and that were reported as positive (+) or negative (-) by the SVM-ESM-1b model. NA – not applicable.
[b]Prediction differs from empirical observation.

Supplementary Table S3.2. Prediction calls for cleavage of naturally occurring yeast Cxxx sequences by indicated model.

yeast protein	motif	PSSM[a,b]	Naïve Bayes	kNN
Ras2	CIIS	+	+	+
Hmg1	CIKS	NA	NA	NA
Rho2	CIIL	+	+	+
Ssp2	CIDL	NA	NA	NA
Skt5, MiY1	CVIM	+	+	+
Tbs1	CVKM	-	-	-
YDL022C-A	CSII	+	+	+
YBR096W	CSEI	NA	NA	NA
YMR265C	CSNA	-	-	-
Pet18	CYNA	-	-	-
Lih1	CSGL	NA	NA	NA
Cup1	CSGK	NA	NA	NA
Nap1	CKQS	-	-	-
Cst26	CFIF	+[c]	+[c]	+[c]
YIL134C-A	CAPY	-	-	-
Atr1	CTVA	+	+	+
Las21	CALD	-	-	-
YDL009C	CAVS	-[c]	-[c]	-[c]
Sua5	CIQF	-	-	-
number observed/predicted		12/14	12/14	12/14
% observed/predicted		85.7	85.7	85.7
number false positive		1/7	1/7	1/7
% false positive		14.3	14.3	14.3
number false negative		1/7	1/7	1/7
% false negative		14.3	14.3	14.3

[a]PSSM – PSSM sequence; Naïve Bayes – Naïve Bayes AAindex; kNN – kNN sequence.
[b]Signs represent predictions of cleavage that were reported as positive (+) or negative (-) by the indicated model. NA – not applicable.
[c]Prediction differs from empirical observation.

Supplementary Table S3.3. Yeast strains used in this study.

Strain	Genotype	Reference
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(48)
IH1793; ATCC#204279	*MAT*α *lys1*	(49)
SM2331	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-Δ1 mfa2-Δ1*	(50)
yWS164	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-Δ1 mfa2-Δ1 rce1::TRP1 ste24::KAN^R*	(37)
yWS304	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ydj1Δ::KAN^R*	(35)
yWS2393	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-Δ1 mfa2-Δ1 ste24::KAN^R*	This study
yWS2462	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-Δ1 mfa2-Δ1 rce1::KAN^R*	This study
yWS2542	*MAT***a** *his3 leu2 met15 ura3 ydj1Δ::NAT^R ram1Δ::KAN^R*	(21)
yWS2544	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ydj1Δ::NAT^R*	(21)

Supplementary Table S3.4 Yeast expression plasmids used in this study.

Gene	Identifier	Genotype	Reference
vectors	pRS315	*CEN LEU2*	(51)
	pRS316	*CEN URA3*	(51)
	pRS415	*CEN LEU2*	(51)
HsRce1	pWS130	*2μ URA3 P_{PGK} HsRce1Δ22*	This study
HsSte24	pWS1609	*CEN URA3 P_{PGK} HsSTE24*	This study
YDJ1	pWS942	*CEN URA3 YDJ1*	(19)
	pWS1132	*CEN URA3 YDJ1-SASQ*	(19)
	pWS1343	*CEN URA3 YDJ1-CASA*	(21)
	pWS1372	*CEN URA3 YDJ1-CAAQ*	(21)
	pWS1410	*CEN URA3 YDJ1-CVAA*	This study
	pWS1411	*CEN URA3 YDJ1-CKQS*	This study
	pWS1437	*CEN URA3 YDJ1-CAHQ*	This study
	pWS1456	*CEN URA3 YDJ1-CAKS*	This study
	pWS1460	*CEN URA3 YDJ1-CQTS*	This study
	pWS1461	*CEN URA3 YDJ1-CSFL*	This study
	pWS1463	*CEN URA3 YDJ1-CVIM*	This study
	pWS1729	*CEN URA3 YDJ1-CTDS*	This study
	pWS1745	*CEN URA3 YDJ1-CALD*	This study
	pWS1746	*CEN URA3 YDJ1-CAPY*	This study
	pWS1747	*CEN URA3 YDJ1-CAVS*	This study
	pWS1748	*CEN URA3 YDJ1-CFIF*	This study
	pWS1749	*CEN URA3 YDJ1-CIDL*	This study
	pWS1751	*CEN URA3 YDJ1-CIIL*	This study
	pWS1752	*CEN URA3 YDJ1-CIKS*	This study
	pWS1753	*CEN URA3 YDJ1-CIQF*	This study
	pWS1757	*CEN URA3 YDJ1-CSEI*	This study
	pWS1758	*CEN URA3 YDJ1-CSGK*	This study
	pWS1759	*CEN URA3 YDJ1-CSGL*	This study
	pWS1760	*CEN URA3 YDJ1-CSII*	This study
	pWS1761	*CEN URA3 YDJ1-CSNA*	This study
	pWS1762	*CEN URA3 YDJ1-CTVA*	This study
	pWS1763	*CEN URA3 YDJ1-CVKM*	This study
	pWS1764	*CEN URA3 YDJ1-CYNA*	This study
	pWS1830	*CEN URA3 YDJ1-CNLI*	This study
	pWS1834	*CEN URA3 YDJ1-CVFM*	This study
	pWS2021	*CEN URA3 YDJ1-CKQG*	This study
	pWS2022	*CEN URA3 YDJ1-CKQH*	This study
	pWS2025	*CEN URA3 YDJ1-CKQL*	This study
MFA1	pWS610	*CEN LEU2 MFA1*	(52)
	pWS613	*CEN LEU2 MFA1-CTLM*	(19)
	pWS846	*CEN LEU2 MFA1-CKQS*	(52)
	pWS1561	*CEN LEU2 MFA1-CSIM*	This study
	pWS1562	*CEN LEU2 MFA1-CIIS*	This study
	pWS1671	*CEN LEU2 MFA1-CTVA*	This study
	pWS1730	*CEN LEU2 MFA1-CALD*	This study
	pWS1733	*CEN LEU2 MFA1-CIQF*	This study
	pWS1734	*CEN LEU2 MFA1-CYNA*	This study
	pWS1738	*CEN LEU2 MFA1-CVIM*	This study
	pWS1739	*CEN LEU2 MFA1-CSGL*	This study
	pWS1777	*CEN LEU2 MFA1-CAVS*	This study
	pWS1778	*CEN LEU2 MFA1-CFIF*	This study
	pWS1779	*CEN LEU2 MFA1-CIIL*	This study

	pWS1780	*CEN LEU2 MFA1-CIKS*	This study
	pWS1781	*CEN LEU2 MFA1-CIDL*	This study
	pWS1782	*CEN LEU2 MFA1-CSII*	This study
	pWS1783	*CEN LEU2 MFA1-CSEI*	This study
	pWS1784	*CEN LEU2 MFA1-CSNA*	This study
	pWS1785	*CEN LEU2 MFA1-CAPY*	This study
	pWS1788	*CEN LEU2 MFA1-CSGK*	This study
	pWS1809	*CEN LEU2 MFA1-CVKM*	This study

Supplementary Table S3.5. PCR Oligonucleotides used in this study.

Gene	Oligo	Mutation	Sequence (5' to 3')
Plasmid UTR	oWS219[a]	NA[b]	tgaCCATGATTACGCCAAGC
YDJ1	oWS999	CVAA	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGTTGCCGCATGAttttcttgataaaaaaagatca
	oWS1000	CKQS	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTAAGCAGAGCTGAttttcttgataaaaaaagatca
	oWS1008	CAHQ	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTT CAATGTGCTCAtCAATGAttttcttgataaaaaaagatcaac
	oWS1028	CIIS	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTT CAATGTatcatttctTGAttttcttgataaaaaaagatca
	oWS1032	CQTS	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTT CAATGTcaaacatctTGAttttcttgataaaaaaagatca
	oWS1033	CSFL	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTT CAATGTtctttttgTGAttttcttgataaaaaaagatca
	oWS1035	CVIM	GATTCCGATGAAGAAGAACAAGGTGGCGAAGGTGTT CAATGTgttatcatgTGAttttcttgataaaaaaagatca
	oWS1051	NA[b]	GGTATGAAGTGGAGGGAGGAT
	oWS1334	CALD	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGCTTTGGATTGAttttcttgataaaaaaagatc
	oWS1335	CAPY	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGCTCCATATTGAttttcttgataaaaaaagatc
	oWS1336	CAVS	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGCTGTTTCTTGAttttcttgataaaaaaagatc
	oWS1337	CFIF	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTTTATTTTTTGAttttcttgataaaaaaagatc
	oWS1338	CIDL	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTATTGATTTGTGAttttcttgataaaaaaagatc
	oWS1340	CIIL	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTATTATTTTGTGAttttcttgataaaaaaagatc

	oWS1341	CIKS	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTATTAAATCTTGAttttcttgataaaaaaagatc
	oWS1342	CIQF	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTATTCAATTTTGAttttcttgataaaaaaagatc
	oWS1346	CSEI	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTCTGAAATTTGAttttcttgataaaaaaagatc
	oWS1347	CSGK	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTCTGGTAAATGAttttcttgataaaaaaagatc
	oWS1348	CSGL	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTCTGGTTTGTGAttttcttgataaaaaaagatc
	oWS1349	CSII	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTCTATTATTTGAttttcttgataaaaaaagatc
	oWS1350	CSNA	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTCTAATGCTTGAttttcttgataaaaaaagatc
	oWS1351	CTVA	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTACTGTTGCTTGAttttcttgataaaaaaagatc
	oWS1352	CVKM	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGTTAAAATGTGAttttcttgataaaaaaagatc
	oWS1353	CYNA	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTTATAATGCTTGAttttcttgataaaaaaagatc
	oWS1423	CNLI	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTAATTTGATTTGAttttcttgataaaaaaagatc
	oWS1444	CVFM	TCCGATGAAGAAGAACAAGGTGGCGAAGGTGTTCAA TGTGTTTTTATGTGAttttcttgataaaaaaagatc
	oWS1583	CKQG	tccgatgaagaagaacaaggtggcgaaggtgttcaatgtAAACAAGG Ttgattttcttgataaaaaaagatc
	oWS1584	CKQH	tccgatgaagaagaacaaggtggcgaaggtgttcaatgtAAACAACAT tgattttcttgataaaaaaagatc
	oWS1587	CKQL	tccgatgaagaagaacaaggtggcgaaggtgttcaatgtAAACAATTG tgattttcttgataaaaaaagatc
MFA1	oWS356	CVKM	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GcgtaaaaatgTAGTTTCTGCGTACAAAAACGCGT

	oWS1178	CSIM	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GctctatcatgTAGTTTCTGCGTACAAAAACGCGT
	oWS1179	CIIS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GCatcatttctTAGTTTCTGCGTACAAAAACGCGT
	oWS1224	CALD	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTgctttggatTAGtttctgcgtacaaaaacgCGT
	oWS1232	CIIL	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTattattttgTAGtttctgcgtacaaaaacgCGT
	oWS1234	CIQF	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTattcaatttTAGtttctgcgtacaaaaacgCGT
	oWS1236	CSGL	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTtctggtttgTAGtttctgcgtacaaaaacgCGT
	oWS1239	CTVA	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTactgttgctTAGtttctgcgtacaaaaacgCGT
	oWS1241	CVIM	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTgttattatgTAGtttctgcgtacaaaaacgCGT
	oWS1244	CYNA	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTtataatgctTAGtttctgcgtacaaaaacgCGT
	oWS1357	CAVS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTgctgtttctTAGtttctgcgtacaaaaacgCGT
	oWS1358	CFIF	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GTtttatttttTAGtttctgcgtacaaaaacgCGT
	oWS1362	CIKS	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GcattaaatctTAGtttctgcgtacaaaaacgCGT
	oWS1363	CIDL	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GcattgatttgTAGtttctgcgtacaaaaacgCGT
	oWS1364	CSII	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GctctattattTAGtttctgcgtacaaaaacgCGT
	oWS1365	CSEI	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GctctgaaattTAGtttctgcgtacaaaaacgCGT
	oWS1366	CSNA	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GctctaatgctTAGtttctgcgtacaaaaacgCGT

	oWS1367	CSGK	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GctctggtaaaTAGtttctgcgtacaaaaacgCGT
	oWS1369	CAPY	AACTATATTATCAAAGGTGTCTTCTGGGACCCAGCAT GcgctccttatTAGtttctgcgtacaaaaacgCGT

[a]Reverse oligonucleotide paired with listed *MFA1* or *YDJ1* mutagenic oligonucleotides
[b]Not applicable.
[c]Lowercase letters indicate differences from wildtype sequences.

.

References

1. Hartman HL, Hicks KA, Fierke CA. Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity. Biochemistry. 2005;44(46):15314-24.

2. Lane KT, Beese LS. Thematic review series: Lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type i. Journal of lipid research. 2006;47(4):681-99.

3. Wang M, Casey PJ. Protein prenylation: Unique fats make their mark on biology. Nature reviews Molecular cell biology. 2016;17(2):110-22.

4. Wright LP, Philips MR. Thematic review series: Lipid posttranslational modifications. Caax modification and membrane targeting of ras. Journal of lipid research. 2006;47(5):883-91.

5. Hampton SE, Dore TM, Schmidt WK. Rce1: Mechanism and inhibition. Critical reviews in biochemistry and molecular biology. 2018:1-18.

6. Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome biology. 62005. p. R55.

7. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS computational biology. 2007;3(4):e66.

8. Stein V, Kubala MH, Steen J, Grimmond SM, Alexandrov K. Towards the systematic mapping and engineering of the protein prenylation machinery in saccharomyces cerevisiae. PloS one. 2015;10(3):e0120716.

9. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA. Context-dependent substrate recognition by protein farnesyltransferase. Biochemistry. 2009;48(8):1691-701.

10. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA. Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities. Journal of molecular biology. 2010;395(1):176-90.

11. Reid TS, Terry KL, Casey PJ, Beese LS. Crystallographic analysis of caax prenyltransferases complexed with substrates defines rules of protein substrate selectivity. Journal of molecular biology. 2004;343(2):417-33.

12. London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O. Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS computational biology. 2011;7(10):e1002170.

13. Sousa SF, Coimbra JT, Paramos D, Pinto R, Guimarães RS, Teixeira V, et al. Molecular dynamics analysis of a series of 22 potential farnesyltransferase substrates containing a caax-motif. Journal of molecular modeling. 2013;19(2):673-88.

14. DeGraw AJ, Palsuledesai C, Ochocki JD, Dozier JK, Lenevich S, Rashidian M, et al. Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. Chemical biology & drug design. 2010;76(6):460-71.

15. Onono FO, Morgan MA, Spielmann HP, Andres DA, Subramanian T, Ganser A, et al. A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with western blotting. Molecular & cellular proteomics : MCP. 2010;9(4):742-51.

16. Palsuledesai CC, Ochocki JD, Kuhns MM, Wang YC, Warmka JK, Chernick DS, et al. Metabolic labeling with an alkyne-modified isoprenoid analog facilitates imaging and quantification of the prenylome in cells. ACS chemical biology. 2016;11(10):2820-8.

17. Storck EM, Morales-Sanfrutos J, Serwa RA, Panyain N, Lanyon-Hogg T, Tolmachova T, et al. Dual chemical probes enable quantitative system-wide analysis of protein prenylation and prenylation dynamics. Nature Chemistry. 2019.

18. Trueblood CE, Boyartchuk VL, Picologlou EA, Rozema D, Poulter CD, Rine J. The caax proteases, afc1p and rce1p, have overlapping but distinct substrate specificities. Molecular and cellular biology. 2000;20(12):4381-92.

19. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

20. Caplan AJ, Tsai J, Casey PJ, Douglas MG. Farnesylation of ydj1p is required for function at elevated growth temperatures in saccharomyces cerevisiae. The Journal of biological chemistry. 1992;267(26):18890-5.

21. Berger BM, Kim JH, Hildebrandt ER, Davis IC, Morgan MC, Hougland JL, et al. Protein isoprenylation in yeast targets cooh-terminal sequences not adhering to the caax consensus. Genetics. 2018;210(4):1301-16.

22. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. Aaindex: Amino acid index database, progress report 2008. Nucleic Acids Res. 2008;36(Database issue):D202-5.

23. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences. 2021;118(15).

24. Tipping ME, Bishop CM. Mixtures of probabilistic principal component analyzers. Neural Computation. 1999;11(2):443-82.

25. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(22):10915-9.

26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. J Mach Learn Res. 2011;12(null):2825–30.

27. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers. 1999;10(3):61-74.

28. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with numpy. Nature. 2020;585(7825):357-62.

29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library2019 December 01, 2019:[arXiv:1912.01703 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv191201703P.

30. Hunter JD. Matplotlib: A 2d graphics environment. Computing in Science and Engg. 2007;9(3):90–5.

31. Waskom ML. Seaborn: Statistical data visualization. Journal of Open Source Software. 2021;6(60):3021.

32. Crooks GE, Hon G, Chandonia JM, Brenner SE. Weblogo: A sequence logo generator. Genome research. 2004;14(6):1188-90.

33. Elble R. A simple and efficient procedure for transformation of yeasts. BioTechniques. 1992;13(1):18-20.

34. Plummer LJ, Hildebrandt ER, Porter SB, Rogers VA, McCracken J, Schmidt WK. Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues. The Journal of biological chemistry. 2006;281(8):4596-605.

35. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the saccharomyces cerevisiae genome. Nature. 2002;418(6896):387-91.

36. Oldenburg KR, Vo KT, Michaelis S, Paddon C. Recombination-mediated pcr-directed plasmid construction in vivo in yeast. Nucleic Acids Res. 1997;25(2):451-2.

37. Cadiñanos J, Schmidt WK, Fueyo A, Varela I, López-Otín C, Freije JMP. Identification, functional expression and enzymic analysis of two distinct caax proteases from caenorhabditis elegans. The biochemical journal. 3702003. p. 1047-54.

38. Hildebrandt ER, Arachea BT, Wiener MC, Schmidt WK. Ste24p mediates proteolysis of both isoprenylated and non-prenylated oligopeptides. The Journal of biological chemistry. 2016;291(27):14185-98.

39. Kim S, Lapham AN, Freedman CG, Reed TL, Schmidt WK. Yeast as a tractable genetic system for functional studies of the insulin-degrading enzyme. The Journal of biological chemistry. 2005;280(30):27481-90.

40. Stormo GD, Schneider TD, Gold L, Ehrenfeucht A. Use of the 'perceptron' algorithm to distinguish translational initiation sites in e. Coli. Nucleic Acids Res. 1982;10(9):2997-3011.

41. Nie L, Spear E, Babatz TD, Quigley A, Dong YY, Chu A, et al. A new paradigm for prelamin a proteolytic processing by zmpste24: The upstream sy^ll cleavage occurs first and there is no caax processing by zmpste24. bioRxiv. 2020:2020.05.13.093849.

42. Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, et al. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. Proceedings

of the National Academy of Sciences of the United States of America. 2004;101(34):12479-84.

43. Galichet A, Gruissem W. Developmentally controlled farnesylation modulates atnap1;1 function in cell proliferation and cell expansion during arabidopsis leaf development. Plant Physiol. 2006;142(4):1412-26.

44. Collins SP, Reoma JL, Gamm DM, Uhler MD. Lkb1, a novel serine/threonine protein kinase and potential tumour suppressor, is phosphorylated by camp-dependent protein kinase (pka) and prenylated in vivo. The Biochemical journal. 2000;345 Pt 3(Pt 3):673-80.

45. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, et al. A novel role of farnesylation in targeting a mitotic checkpoint protein, human spindly, to kinetochores. The Journal of cell biology. 2015;208(7):881-96.

46. Mosalaganti S, Keller J, Altenfeld A, Winzker M, Rombaut P, Saur M, et al. Structure of the rzz complex and molecular basis of its interaction with spindly. The Journal of cell biology. 2017;216(4):961-81.

47. Caldwell GA, Wang SH, Naider F, Becker JM. Consequences of altered isoprenylation targets on a-factor export and bioactivity. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(4):1275-9.

48. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nature genetics. 1996;14(4):450-6.

49. Michaelis S, Herskowitz I. The a-factor pheromone of saccharomyces cerevisiae is essential for mating. Molecular and cellular biology. 1988;8(3):1309-18.

50. Chen P, Sapperstein SK, Choi JD, Michaelis S. Biogenesis of the saccharomyces cerevisiae mating pheromone a-factor. The Journal of cell biology. 1997;136(2):251-69.

51. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in saccharomyces cerevisiae. Genetics. 1989;122(1):19-27.

52. Krishnankutty RK, Kukday SS, Castleberry AJ, Breevoort SR, Schmidt WK. Proteolytic processing of certain caax motifs can occur in the absence of the rce1p and ste24p caax proteases. Yeast (Chichester, England). 2009;26(8):451-63.

CHAPTER 4

Investigations of the Histone Chaperone Nap1 as a Shunt Protein

The CaaX pathway is a three step post-translational modification pathway that occurs to proteins containing a CaaX motif at their COOH terminus, where “C” is a Cysteine, “a” is an aliphatic amino acid, and “X” can be one of several amino acids. This pathway, originally described as isoprenylation of the Cysteine, followed by proteolysis of the aaX, and carboxyl methylation of the lipidated Cysteine, has long been associated with increasing hydrophobicity and membrane association of proteins (1-3). However, in recent years, an alternative outcome, termed the shunt pathway, was investigated using the yeast Hsp40 chaperone, Ydj1 (4). In this shunt pathway, proteins go through the first step of the CaaX pathway, isoprenylation, but then are “shunted” out of the pathway, omitting the proteolysis and methylation. While proteins have been suggested previously to not undergo full CaaX processing, Ydj1 was the first case where this isoprenylation only modification was shown to be essential for proper function (5-7). Since this finding, there has been an interest in identifying additional sequences and proteins that may undergo this shunt pathway (8).

One protein that has been suspected of being shunted is Nucleosome Assembly Protein 1 (Nap1). Nap1 is a member of the NAP/SET family of histone chaperones, which are universally conserved across eukaryotes (9). Nap1 was first identified for its activity *in vitro* assembling nucleosomes, although it has since been implicated in

several other functions *in vivo*, including cell cycle regulation, transport of histones H2A/H2B, transcription, and several others (9-13). Importantly, Nap1 contains a conserved CKQ[Q/S] motif (CKQQ: human, *A. thaliana*; CKQS: yeast) at its COOH terminus that has been predicted to be prenylated (Chapter 3) (14). Additionally, studies in *Arabidopsis thaliana* and humans imply that these orthologs can be prenylated (15, 16). While the function of this prenyl group is unclear, previous studies have shown that protein lipidation is important for localization and protein-protein interactions for Pex19 and Spindly found in humans (17-20).

In addition to prenylation, Nap1 orthologs are known to undergo several other post-translational modifications, including phosphorylation, acetylation, and polyglutamylation (21-24). The PTMs of Nap1 are under investigated, with the effects of glutamylation and prenylation remaining undefined. The phosphorylation and acetylation modifications have been proposed to be cell cycle dependent, implicating a possible role for Nap1 in cell cycle regulation. Supporting this, Nap1 has been shown to interact with several cyclins, as well as Gin4, a protein known to be involved in septin formation (12, 25). It is suspected that Nap1's localization may also be dependent on the cell cycle and phosphorylation. Nap1 is primarily cytosolic but contains both a nuclear localization signal and nuclear export signal (NLS/NES) (26). In *Drosophila*, the Nap1 homolog was shown to localize to the nucleus during S phase but is cytosolic for the remainder of the cell cycle (27). In yeast, Nap1's nucleocytoplasmic shuttling remains unclear. In fact, it has been challenging to observe yeast Nap1 in the nucleus at all, with the most success coming from mutating the NES, effectively trapping Nap1 within the

nucleus (28). From this method, it was observed that Nap1's phosphorylation plays a role in its nucleocytoplasmic cycling.

While the mechanism leading to Nap1's nuclear localization remains elusive, previous yeast studies have shown that for transport into the nucleus, Nap1 binds to histones H2A/H2B in the cytosol and the complex is transported into the nucleus by the karyopherin Kap114 (13, 29, 30). Nap1 appears to increase the affinity of histones H2A/H2B for Kap114; but when Nap1 is not present, the histones are able to be transported by alternative nuclear transporters, including Kap121, Kap123, and Kap95. Interestingly, Nap1 phosphorylation was not shown to have any effect *in vivo* or *in vitro* on Kap114 affinity. As previous studies have shown that this phosphorylation impacts nuclear transport, this suggests that Nap1 may have an additional Kap114-independent mechanism for nuclear localization that has yet to be determined.

In this study, we aimed to establish that *Saccharomyces cerevisiae* Nap1 (*Sc*Nap1) as a shunted (prenylated and non-proteolyzed) protein. While biochemical evidence by mass spectrometry is pending, we demonstrate by alternate methods such as gel shift and indirect genetic reporters that Nap1 is indeed prenylated and shunted. We then investigated the impact of the prenyl group on Nap1 function and localization. It was found that lipidation has an effect on Nap1 nuclear localization. Furthermore, by looking at a strain lacking Nap1's known nuclear transporter (*kap114Δ)*, we observed that Nap1-CKQS remains at similar levels within the nucleus, unlike our non-modified Nap1 mutant (Nap1-SKQS), indicating that Nap1 may have an additional nuclear transport mechanism.

Materials and Methods

Strains: Strains used in this study are listed in **Supplemental Table S4.1**. Lithium acetate-based transformations were used to introduce plasmids, as described previously (8, 31). All strains were propagated in SC-Uracil (SC-U) media at 30 °C unless otherwise stated. The *nap1Δ kap114Δ* was constructed by inducing a mating type switch using Gal-HO in the *MAT***a** *nap1Δ* strain (yWS1666) to *MAT*α *nap1Δ* (yWS1772) (32). A genetic cross of yWS1772 (*MAT*α *nap1Δ*) and yWS2201 *(MAT***a** *kap114Δ*) was then performed, followed by sporulation and evaluation of spores by PCR.

Plasmids: Plasmids used in this study are listed in **Supplemental Table S4.2**. All plasmids constructed for this study were created using previously reported methods (8, 33). Briefly, mutagenic oligonucleotides encoding desired CaaX motifs or NH_2-terminal tags were co-transformed with linearized/gapped parent plasmids. Transformation mixes were plated onto appropriate media, and plasmids recovered from emerging colonies. Restriction digest and sequencing were used to verify proper sequence throughout the entire open reading frame.

Assays: Ydj1 thermotolerance and yeast mating assays were done as reported previously (4, 8). For phenotypic screens using various chemicals, strains of interests were grown to saturation overnight at 25 °C and were serially diluted 1:10 (unless otherwise stated) and pinned on appropriate media/conditions. Plates were then incubated at 25 °C, 30 °C, and/or 37 °C for 48-72 hours.

Microscopy: Strains were grown at 30 °C to mid to late log. When appropriate, Hoechst 33342 (Invitrogen) was added at a concentration at 15 µg/mL, cultures were protected from light and incubated an additional 30 minutes. Cells were harvested, washed with PBS, and the pellets resuspended in residual SC-U media. 2 µL of culture were spotted onto charged slides (Diamond White Glass Charged Slides, Globe Scientific Inc.) and imaged using Axio Observer Z1 microscope equipped with 63x oil immersion objective. Images were captured using an AxioCam MRm controlled with AxioVision 4.6 software. Images were adjusted using identical ImageJ settings.

Spheroplast preparation: Strains of interest were grown to log phase in SC-U media. Cells were then harvested by centrifugation, washed with water, and resuspended in 100 mM Tris pH 9.4, 10 mM DTT. Cells were incubated on ice for 10 minutes, recovered by centrifugation, and washed with 10 mM NaN_3. Recovered cells were then incubated with Zymolyase in OB buffer (50 mM KPi pH 7.5, 1.4 M sorbitol, 10 mM NaN_3) at 30 °C for 30 minutes, followed by a 10-minute recovery on ice. Spheroplasts were then harvested by centrifugation and used for various purposes..

Nuclei Isolation: A nuclei isolation protocol using Ficoll buffers was adapted from previously reported methods (34). Briefly, 100mL cultures were grown to OD_{600} of ~1. Cells were harvested and spheroplasts prepared using Zymolyase treatment as previously described using slight buffer modifications (spheroplast buffer containing 50 mM KPi pH 7.4, 0.6 M Sorbitol, 10 mM DTT and Zymolyase). Spheroplasts were then overlaid onto 7.5% Ficoll-Sorbitol solution (7.5% Ficoll, 0.6 M Sorbitol, 20 mM KPi pH

6.5, 1 mM $MgCl_2$, 1 mM DTT, and 1X protease inhibitors (1X protease inhibitors- 1 mM phenylmethanesulfonylfluoride fluoride (PMSF) and aprotinin, 1 µg/mL each of chymostatin, leupeptin, and pepstatin) and centrifuged. The supernatant was aspirated and spheroplasts were resuspended in wash buffer (20 mM KPi pH 6.5, 0.6 M Sorbitol, 1 mM $MgCl_2$, 1 mM DTT and 1X protease inhibitors, then dounced in 18% Ficoll solution (18% Ficoll, 20 mM KPi, pH 6.5, 1 mM $MgCl_2$, 1 mM DTT, 0.01% NP-40, 1X protease inhibitors and harvested by centrifugation. The supernatant containing the cytosol was then removed, and the pellet was resuspended in wash buffer. Protein loading volumes were normalized by Ponceau S staining of nitrocellulose membrane and Nap1 protein levels checked by immunoblot.

Preparation of cell extracts for SDS-PAGE: Unless otherwise stated, cell lysates for immunoblotting were prepared by TCA precipitation (35). Briefly, yeast strains expressing Nap1 were grown to OD_{600} ~1 and equal masses harvested by centrifugation, washed with water, and cell extracts were prepared by alkaline hydrolysis followed by TCA precipitation. Cell pellets were then resuspended in urea sample buffer (250 mM Tris, 6 M Urea, 5% β-mercaptoethanol, 4% SDS, 0.01% bromophenol blue, pH 8).

Immunoblotting: Samples were analyzed by SDS-PAGE followed by immunoblotting on nitrocellulose with anti-His or anti-GFP and HRP conjugate in 1% TBST (100 mM Tris, 400 mM NaCl, 0.1% Tween 20, pH 7.5). Protein levels were detected by WesternBright

ECL Spray (Advansta Inc, San Jose, California), and images captured using X-ray film or a digital imager (Kwikquant, Kindle Biosciences, Greenwich, Connecticut).

MNase digests: The Nap1 MNase digest protocol was adapted from previously reported (36, 37). Briefly, 200mL cultures of Nap1 strains of interest were grown to mid-log and harvested by centrifugation. Spheroplasts were prepared as described with slight buffer modifications (50 mM Tris-HCl, pH 7.5, 1 M sorbitol, 10 mM β-mercaptoethanol) and were resuspended in MNase Digestion Buffer (10 mM Tris-HCl, 1 M Sorbitol, 50 mM NaCl, 5 mM $MgCl_2$, 0.075% NP-40). 30 U of MNase (Takara Bio) were added and samples were incubated at 37 °C. Reactions were stopped by adding 5% SDS, 50 mM EDTA to tubes after 10 minutes, and followed by proteinase K treatment and Phenol: Chloroform extraction of DNA (38). DNA pellets were resuspended in TE/RNase solution and analyzed by 1.5-2% agarose gels.

Ty1 retromobility assay: The Nap1 Ty1 mobility assay protocol was adapted from previously reported methods (39, 40). Briefly, the Ty1-H3HIS3-AI plasmid under the endogenous Ty1 promoter (pOY1 courtesy of Dr. David Garfinkel) was transformed into Nap1 strains as described above and plated on appropriate media. Transformations were incubated at 30 °C. Single colonies from transformation plates were patched onto SC-Uracil and Leucine (SC-UL) plates and grown at room temperature for 72 hours to allow for Ty1 mobility events. After 72 hours, patches were replica plated onto SC-Histidine, Uracil, and Leucine (SC-HUL) plates and grown at 30 °C for 72 hours. For a

semi-quantitative analysis, colonies from each patch were counted and statistical analysis done using Student's t-test.

Protein purification: 2 L SC-U liquid culture of strain IH 1783 (*MAT***a** *trp1 leu2 ura3 his4 can1*) expressing pWS1323 (*2μ* P_{PGK} *URA3 His-Nap1*) was grown to mid-log phase at 30 °C. Spheroplasts were prepared as described, resuspended in lysis buffer (50 mM KPi pH 7.5, 0.5 M sorbitol, 0.02% NaN_3, and 1X protease inhibitors) and subjected to bead beating. The cell lysate was clarified by centrifugation and purified using poly-histidine tag and Talon resin. Resin was washed with lysis buffer containing 5 mM imidazole and eluted using lysis buffer with 300 mM imidazole. After purification, the protein was concentrated using Amicon Ultra-15 10k centrifugal filter unit (Millipore).

E. coli protein expression: BL21 DE3 strain expressing a Ydj1 plasmid (courtesy of Dr. Avrom Caplan) or Nap1 plasmid (NorClone Biotech) were co-expressed with pETDuet plasmid +/- FTase α/β subunits (Novagen) or a CDF-Duet plasmid encoding FTase α/β subunits (Courtesy of Dr. Mark Distefano). *E. coli* growth conditions and induction was adapted from previously published protocols (41, 42). Briefly, Ydj1 overnight cultures were diluted into fresh media and grown at 37 °C. After 1 hour, protein expression was induced by the addition of 1mM IPTG and cultures were grown for 3-6 hours. Then, cells were harvested by centrifugation, the pellet resuspended in Laemmli Sample buffer and evaluated by immunoblotting. For Nap1, further optimization was attempted, including induction times up to 19 hours, growth at different temperatures, and TB media instead of LB.

Results

Nap1 is a proposed shunted protein

The Nap1 family of proteins share a conserved, COOH terminal Cysteine across 23 orthologs in 19 eukaryotic species (**Fig 4.1A**). While not fitting the traditional definition of a “CaaX” motif, existing evidence from plants and human show that these motifs can be prenylated, suggesting that *Sc*Nap1 is also a prenylprotein (15, 16). This hypothesis is supported by prediction methods from two previously published methods predicting the yeast Nap1 motif, CKQS, to be prenylated (14) (Chapter 3). In order to confirm this, an immunoblot gel shift assay comparing wildtype Nap1 (Nap1-CKQS) to a Nap1 mutant containing a single point mutation (C414S) abolishing the prenylation site (Nap1-SKQS) was evaluated. From this, a shift was evident between Nap1-CKQS and Nap1-SKQS, confirming that *Sc*Nap1 is indeed prenylated (**Fig 4.1B**). After establishing Nap1’s prenylation status, the next question was whether Nap1 undergoes the canonical, 3-step CaaX pathway, or the prenylation-only shunted outcome. As the motif CKQS lacks the characteristic aliphatic amino acids at the a_1 and a_2 position, it was hypothesized that Nap1 would be a shunted protein (prenylation only). Supporting this hypothesis that Nap1 is shunted, emerging data from a recent study using metabolic labeling followed by mass spectrometry indicated that the human Nap1 orthologs, Nap1L1 and Nap1L4 (CKQQ), were farnesylated but unprocessed (not cleaved or carboxyl methylated) (16). While we aim to directly show this prenylation only modification on *Sc*Nap1, the hydrophobic nature of the farnesyl, together with the fact that these modifications are located at the COOH terminus of the protein create a challenge for detection by mass spectrometry. Nonetheless, we have purified *Sc*Nap1

from *Saccharomyces cerevisiae* and provided this to collaborators for MS determination of the COOH-terminus (**Fig S4.1**).While these results are pending, we provide indirect evidence that *Sc*Nap1 undergoes prenylation only using the well-established genetic reporters, Ydj1 and **a**-factor, which allowed for the evaluation of shunted and cleaved sequences, respectively. In the Ydj1 thermotolerance assay, the Nap1 sequence, CKQS, appears to grow better than the known cleaved sequence, CVIA, at 40 °C, suggesting that CKQS is shunted (**Fig. 4.1C**). Additionally, when comparing MFA1-CKQS to the canonical, wild type **a**-factor sequence, CVIA, we can see that there is no growth, indicating a lack of cleavage and mating (**Fig. 4.1D**). Taking all this information together provides evidence that Nap1 is a prenylated, shunted protein.

nap1Δ has no obvious phenotype

As the Nap1 prenylation site is so well conserved, there was significant interest in determining the function of the modified COOH terminus for Nap1 function. We first sought a phenotype visible between wildtype Nap1 and a *nap1Δ* strain, hypothesizing that it may be possible to observe an intermediate phenotype of modified Nap1. For this reason, we searched previous studies for *nap1Δ* phenotypes and repeated several previously reported observations based on cell shape and various growth assays. The differences in cell shape were described as elongated buds in *nap1Δ* strains, however, we did not observe any noticeable abnormalities in cell shape in *nap1Δ* compared to wild type in our strains (**Fig 4.2A**) (23, 43). While there may be a slight difference in growth between wildtype Nap1 and *nap1Δ*, there was no difference between wildtype Nap1-CKQS and the non-prenylatable Nap1-SKQS (**Fig 4.2B**). Contrary to previous

publications, our genetic screens showed no visible growth difference in the presence of acetic acid or 6-azauracil (**Fig 4.2C**) (44-46). As Nap1 has been shown by various protein interaction studies to interact with proteins involved in lysine and cell wall biosynthesis, we tested media lacking lysine or containing Calcofluor White, a dye used for staining cell wall chitin but observed no difference (47-49). Lastly, we tested Nap1 strains under a variety of stress conditions, including heat shock, thermotolerance, glucose deprivation, NaN_3 or DMSO stress, and long-term viability and saw no observable difference between Nap1-CKQS, Nap1-SKQS, or *nap1Δ* growth (**Fig 4.2C**, data not shown). From this, we concluded that there may not be enough dynamic range between Nap1-CKQS and *nap1Δ* to utilize a genetic phenotype to see the impact of shunting for Nap1's function.

Prenylation near the COOH-terminus impacts Nap1 nuclear localization

While Nap1 is known to be a nucleocytoplasmic shuttle, interestingly, localization studies show it to be primarily localized to the cytosol. However, as the prenylation status of Nap1 is not commonly acknowledged, majority of localization studies continue to tag the Nap1 COOH-terminus, effectively blocking Nap1 prenylation. For this reason, we hypothesized that prenylation may affect Nap1 localization, which was vastly unobservable due to the COOH-terminal tags. To test this, we used NH_2-terminal GFP-Nap1 and GFP-Nap1-SKQS constructs to determine if there were any localization differences between prenylated and non-modified Nap1, respectively. While Nap1 was mainly diffused throughout the cell for both constructs, GFP-Nap1-CKQS could on occasion be observed colocalizing with Hoechst DNA stain **(Fig 4.3A**). This nuclear

localization for Nap1 appears to be more prevalent in stationary cultures (up to 1 week after inoculation), and it was estimated that after one week nearly 20% of cells displayed this enriched Nap1 nuclear localization (n=97, cells counted if determined viable by Hoechst staining). GFP-Nap1-CKQS nuclear localization could also be seen in rapidly diving cells (log phase), although less often. Notably, colocalization was evidently lacking in GFP-Nap1-SKQS. Indeed, while Nap1 is generally visibly diffused throughout all the cytosol, in many cases for GFP-Nap1-SKQS there is a distinct lack of Nap1 present where the nucleus is located, with an accumulation of GFP-Nap1-SKQS sometimes visible surrounding the Hoechst nuclear staining. As we were unable to determine the optimal conditions or timing for this nuclear localization by microscopy cell cycle staging, nutrient conditions, etc.), we investigated if prenylation impacted nuclear localization by isolating nuclei and evaluating by western blot. From this, we observed that Nap1-SKQS nuclear levels were substantially reduced compared to wild type (**Fig 4.3B**). Notably, this nuclear localization likely represents a small fraction of the Nap1 pool present in the cell. To observe this localization, approximately 20% of the total nuclear fraction was loaded, compared to approximately 3% and 1.5% of samples loaded for total cell lysate and cytosolic fractions, respectively.

After determining that prenylation affects Nap1 nuclear levels, the next question was whether the prenyl group impacts nuclear import. Previous studies have shown that Kap114 is involved in nuclear transport of Nap1 together with histone H2A/H2B, and is the only known nuclear transporter (karyopherin) of Nap1 to date (13). To determine if prenylation of Nap1 affects Kap114 transport, a *nap1Δ kap114Δ* strain was constructed and evaluated. By immunoblot, while wildtype Nap1-CKQS nuclear levels remain

similar, Nap1-SKQS has a noticeable decrease in nuclear levels (**Fig 4.3C**). These findings suggest two points: the first, that Kap114 trafficking of Nap1 is a prenyl-independent mechanism, as it appears to be the primary means of nuclear import for Nap1-SKQS. Secondly, these findings suggest that wild type Nap1 has another method of nuclear import that has not yet been determined.

Nap1 nuclear levels do not cause any discernible differences in nuclear assays

After determining that prenylation impacts nuclear levels, we then wanted to investigate if this would affect Nap1's functions within the nucleus. Nap1 is known to be involved in chromatin packing and rearrangement, with differences by micrococcal nuclease (MNase) digests being reported previously between Nap1 and *nap1Δ* strains (37). As MNase is a nonspecific endonuclease able to cleave open chromatin regions, we hypothesized that if prenylation or lack thereof, affected Nap1's chromatin associated functions, this may result in a shifted band pattern by electrophoresis after MNase treatment. However, after evaluating Nap1-CKQS, Nap1-SKQS, and *nap1Δ* by MNase digests, we did not see any discernible difference between the three strains (**Fig 4.4A**). We then turned to the use of the retrotransposon Ty1 to determine if Nap1 prenylation has any effect on chromatin packing based on Ty1 mobility. In this case, a plasmid encoding a Ty1 element with the yeast HIS3 gene interrupted by an antisense artificial intron (Ty1*his*3*AI*) was transformed into Nap1 strains of interest (40). Following Ty1 RNA transposition, this antisense intron is spliced out, allowing for growth on plates lacking histidine (39). While Ty1 is known to have preferred integration sites, such as tRNA encoding regions, we hypothesized that if Nap1 influences chromatin packing and

rearrangement, it may result in more accessible insertion sites depending on the Nap1 strains (50, 51). While there was a difference in Ty1 mobility in *nap1Δ* strains, the difference between Nap1-CKQS and Nap1-SKQS was not significant (**Fig 4.4B, C**). This implies that Nap1 could be important for retrotransposition, but for the purposes of this work, we were unable to observe any differences in function between wildtype and non-prenylated Nap1 in either assay.

Nap1 is not efficiently farnesylated in *E. coli* system

As all *in vivo* observations between Nap1-CKQS and Nap1-SKQS resulted in subtle to no observable change, we then hypothesized that this difference may be more substantial by *in vitro* assays. While several Nap1 *in vitro* assays have been established, namely histone binding, general methods have involved recombinant Nap1 orthologs purified from *E. coli*, abolishing any effects of PTMs. Previously, hGBP1, a large GTPase that is known to be farnesylated, was co-expressed with plasmids expressing FTase in *E. coli*, allowing for purification of recombinant, farnesylated hGBP1 (41). We sought to use a similar system with yeast Nap1 to obtain enough purified farnesylated protein to use for *in vitro* assays. As a proof of concept, we initially optimized this system using the shunt protein Ydj1 and two different plasmids expressing yeast FTase. With Ydj1, an evident double band was visible with FTase 1 (pWS1693), indicating partial prenylation, while a single lower band was seen with FTase 2 (pWS1790), suggesting full prenylation (**Fig 4.5A**). After these results, we then moved on to farnesylation of Nap1 co-expressed with FTase 2. Under similar conditions as Ydj1, a small upper band was visible relative to the non-prenylated control, indicating

a small population of farnesylated Nap1 (**Fig 4.5B**). To increase farnesylation, we also tested FTase 1, as well as co-expression of Nap1 with both FTase (**Fig 4.5C**). While no band was visible with FTase 1, for FTase 2 again a subtle, small upper band was visible. When both FTase plasmids were expressed, there seems to be less Nap1 produced, however, the amount farnesylated Nap1 appears to be similar to that of FTase 2. Further optimization attempts to improve Nap1 farnesylation included increased time after induction, use of TB media, and lowering incubation temperature. However, from all these attempts, there was little improvement, with the majority of protein remaining unmodified by immunoblot. As it was determined Nap1 would require further optimization for purification, this project has not been completed to date.

Discussion

While previous studies suggest that *Sc*Nap1 may be a prenylprotein, here we provide evidence that Nap1 is indeed farnesylated by gel shift (14-16, 52). Additionally, while biochemical evidence is still pending, our Ydj1 thermotolerance data of the CKQS motif, together with the data gleaned from metabolic labeling showing Nap1L1 is farnesylated and unprocessed (i.e., no additional cleavage or methylation) provide supporting evidence that Nap1 is shunted (16). While the purpose of shunting remains unclear, this data raises the possibility that Nap1 is the second confirmed shunted protein in yeast. Although our other confirmed shunt protein, Ydj1, demonstrates a clear shunt-dependent phenotype, our studies to date have yet to reveal a similar situation for Nap1. Interestingly, while isoprenylation has generally been associated with membrane localization, in recent years examples of different functions have emerged. These

include a structural function in the case of hGBP1 and farnesyl-dependent protein interactions for both Pex19 and Spindly in humans (18, 20, 53, 54). We hypothesize from these observations that the function of farnesylation for Nap1 may provide a structural or affinity role. Regardless, the conservation of this farnesylation site in Nap1 orthologs emphasize that this lipid likely provides some function that has yet to be determined.

Nap1 has been shown to be involved in many different functions, from histone binding, chromatin packing and cell cycle regulation, and yet *nap1Δ* strains show very little discernible differences. This suggests that Nap1 may function redundantly with other yeast proteins, which minimizes the effects of *nap1Δ*. Indeed, this has been previously shown in the case of Nap1's role interacting with cyclin Bs, requiring a triple knockout in order for *nap1Δ* phenotype to be visible (11). While several previous studies report various *nap1Δ* phenotypes, we found these to be mostly irreproducible in our hands (11, 37, 44-46). While in some cases, like cell shape, the number of abnormal cells has already been noted to vary depending on publication, differences in strain backgrounds, haploid versus diploid strains, or plasmid based Nap1 versus chromosomal levels may provide explanations for some of these differences. More comprehensive screening may require securing additional strains from previous studies to resolve some of these factors.

While Nap1's localization has long been described as cytosolic by microscopy, which our findings largely agree with, this appears to be the first case of Nap1 microscopy where *Sc*Nap1 nuclear colocalization is visible. As noted above, previous studies using fluorophore tags place these at the COOH terminus, blocking

farnesylation of Nap1. Using NH_2-terminal GFP-tags, we were able to see this colocalization in wild type cells compared to a non-prenylated C414S mutant, Nap1-SKQS, which was later confirmed by immunoblot. While this phenotype was first identified in stationary cultures (1-8 days old), studies investigating nutrient stress, long-term viability, and additional stress conditions yielded no differences (**Fig 4.1C**, data not shown). It's also worth noting that there are additional subtle microscopy observations, including stress granule incorporation and septin localization, that could also be associated with farnesylation (data not shown). However, these would require additional studies for confirmation. It's possible that the farnesyl may be involved in proper shuttling or protein-protein interactions as the cell cycle progresses, which may be visible by optimized microscopy attempts. Several attempts were made to observe Nap1 localization in a cell-cycle dependent manner to see if there was a farnesyl-dependent effect, including cell cycle arrests, live cell microscopy, and synchronization of cultures. However, limited progress was made during all these attempts. In some cases, such as live cell microscopy, technical issues with microscopes resulted in cells shifting out of focus or bleaching. In others, additional strains and/or further optimization may be necessary for cell cycle arrests or synchronization. Due to the investment of time and resources necessary for proper optimization for these microscopy studies, it was decided not to move forward with these directions.

The differences observed in Nap1 nuclear levels between Nap1-CKQS and Nap1-SKQS raise several questions as to why the farnesyl matters for nuclear trafficking. It's possible that farnesylation is necessary for proper nuclear import or export of Nap1. Additionally, as Nap1 is a phosphoprotein with phosphorylation sites

highly regulated by the cell cycle, it's possible that this farnesyl group may be allosterically affecting the NLS or a phosphorylation site. It is interesting that Nap1's main known nuclear importer, Kap114, appears to be farnesyl-independent, with the absence of Kap114 mostly affecting unfarnesylated Nap1 trafficking. A previous study utilizing a COOH-terminally tagged Nap1 (effectively blocking farnesylation) to observe Nap1 transport provides support that this Nap1/Kap114 interaction does not need a farnesyl group for Nap1 trafficking with histones H2A/H2B (28). It is possible that Nap1's other method of import may be impacted by this lipid modification. While no other karyopherin has been identified for Nap1, one wide-scale study using yeast 2-hybrid assays to identify interactions found that Nap1 interacts with Kap95, the importin β which dimerizes with the importin α, Kap60. This Kap60α /Kap95β complex is perhaps the best characterized of the karyopherins and responsible for nuclear import of classical nuclear localization signals. Indeed, Nap1's NLS was predicted to be imported by this Kap60α/Kap95β complex using an importin α/β prediction server (55). It may be interesting in the future to determine if there is an interaction between Nap1 and Kap60/Kap95 and if so, if prenylation plays a role. Alternatively, it's possible that prenylation could affect nuclear export, with Nap1-SKQS having a higher affinity for the export chaperone. Regardless of the mechanism causing this Nap1 nuclear level imbalance, we did not see any differences in Nap1 nuclear functions by MNase digests or Ty1 mobility (**Fig 4.4**). However, our Ty1 mobility findings showing a difference in Nap1-CKQS and *nap1Δ* are consistent with previous findings using a Ty1-lacZ fusion reporter. This study, utilizing a β-galactosidase output, showed that deletion of the NAP1 gene resulted in an approximately 29% decrease in β-galactosidase levels

compared to wildtype Nap1 (13). Interestingly, Ty1 has been shown to prefer integration into nucleosome bound DNA near the H2A/H2B interface, suggesting that Ty1 integration is sensitive to nucleosomal changes and packing (51, 56, 57). Although our Ty1 mobility assay did show these clear differences between Nap1-CKQS and *nap1Δ*, the differences between prenylated and non-prenylated were deemed insignificant. Whether this might suggest Nap1-SKQS is fully functional in chromatin packing, there is redundant cellular machinery that prevents a phenotype, or that the dynamic range for these assays is not wide enough remains unclear.

Lastly, it's unclear why our attempts at farnesylating Nap1 *in vivo* were unsuccessful in *E. coli*. While this method has been successful for both hGBP1 and Ydj1, Nap1 appears to present more of a challenge (**Fig 4.5**) (41). While partial prenylation and full prenylation are visible for Ydj1 expressed with FTase 1 and 2, respectively, only a small upper band is visible for Nap1 when expressed with FTase 2. It was thought that possible co-expression with both FTase plasmids would increase Nap1 farnesylation, but it appears that this reduced the Nap1 expression. A gel of Nap1 expressed in yeast was ran parallel to the *E. coli* extracts to confirm a gel shift was visible under the same gel conditions (data not shown). Additionally, the single higher band present in our yeast gel shift assay suggests that Nap1 is fully prenylated in *Saccharomyces cerevisiae* (**Fig 4.1B**). To date, there has been no improvement of Nap1 farnesylation using the *E. coli* expression system.

Altogether, this study aimed at identifying two concepts of Nap1: first, *Sc*Nap1 is a farnesylated, shunted protein; second, that the COOH terminal modifications serve a function. While we were able to show evidence by gel shift and genetic reporters that

Nap1 is a shunted prenylprotein, we have yet to determine the impact of this modification beyond nuclear localization. While our data suggests that the lipid group has an effect on Nap1 nuclear levels, hinting at an importance for proper shuttling, further studies would be necessary to define further impacts of Nap1 farnesylation.

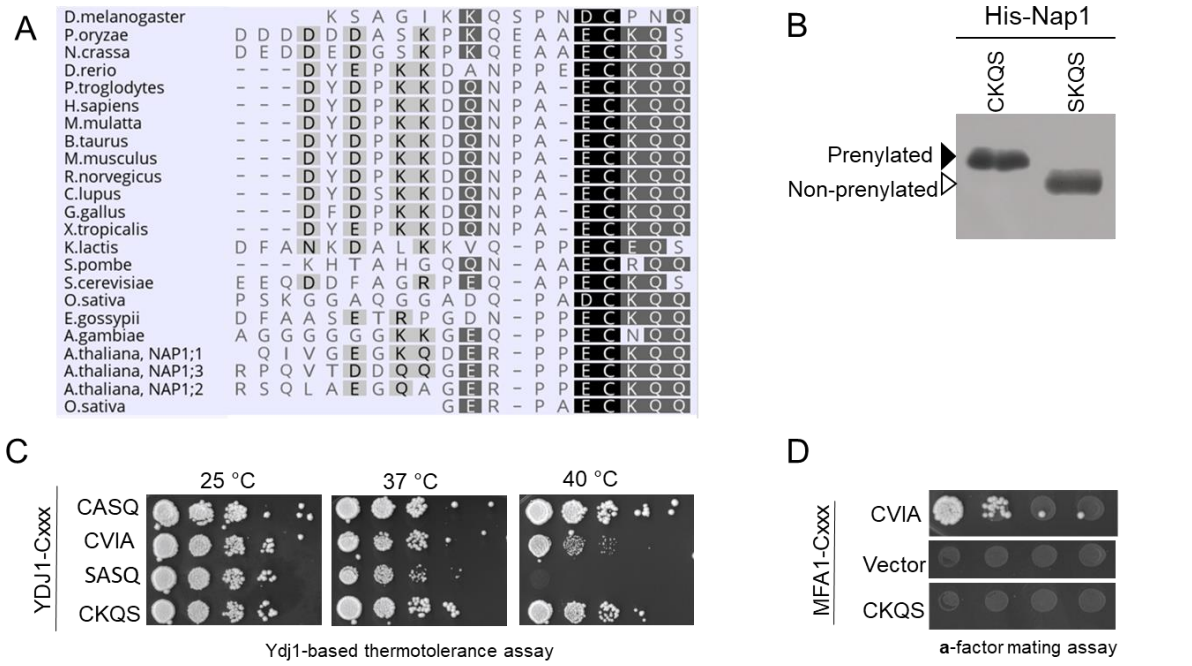


Figure 4.1: *Sc*Nap1 is a prenylprotein. A) Nap1 orthologs contain a conserved CKQ[Q/S] motif at their COOH terminus, indicating that they may be a target for prenylation. While this has been implied for *Hs* and *At*Nap1 orthologs, no data has been shown yet for *Sc*Nap1. B) Yeast strains lacking chromosomally encoded *NAP1* (yWS1666, *nap1Δ*) containing a His-tagged wildtype Nap1 (His-Nap1-CKQS) or a C414S mutant abolishing the potential prenylation site (His-Nap1-SKQS) were evaluated by gel shift. Yeast extracts were run on SDS-PAGE gel and evaluated by anti-His immunoblot to show prenylation (closed triangle) for wildtype Nap1, compared to the unprenylated His-Nap1-SKQS mutant (open triangle). C, D) To determine if Nap1 is shunted (prenylation only) versus canonically processed (cleaved and carboxyl methylated), the Nap1 Cxxx motif was evaluated on the shunted and canonical genetic reporters, Ydj1 and **a**-factor, respectively. Yeast strains lacking chromosomally encoded *YDJ1* (yWS304, *ydj1*Δ) or *MFA1* and *MFA2* (SM2331, *mfa1Δ mfa2Δ*) but expressing various Ydj1-Cxxx or **a**-factor-Cxxx variants. C) Thermotolerance of Ydj1-CKQS compared to the known shunted sequence (CASQ) and canonical sequence (CVIA). Saturated cultures were pinned in 10-fold serial dilutions and incubated at the temperatures indicated. D) Mating of MFA1-CKQS compared to the canonical sequence CVIA. Saturated *MAT***a** cultures diluted in the presence of *MATα* cells were pinned on SD media in 10-fold serial dilutions. Mating is indicated by diploid cell growth.

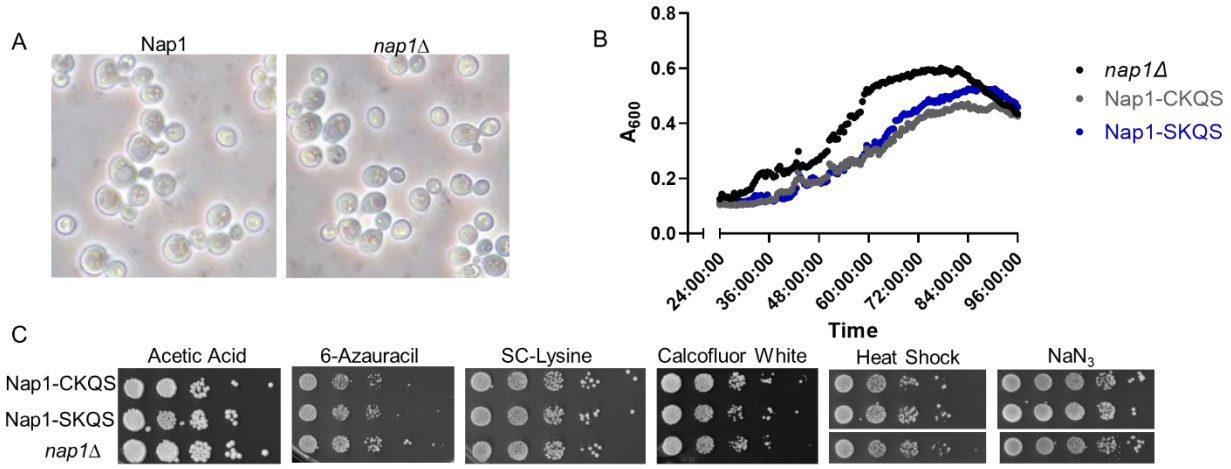


Figure 4.2: *Sc*Nap1 shows no obvious phenotypes under various growth conditions. Yeast strains lacking chromosomally encoded *NAP1* (yWS1666, *nap1Δ*) containing plasmids expressing wildtype Nap1 (Nap1-CKQS), a C414S mutant abolishing the potential prenylation site (Nap1-SKQS), or Uracil vector (*nap1Δ*) were evaluated under various conditions to identify any apparent Nap1 phenotypes. A) Brightfield microscopy using 63X lens show similar cell shapes between wildtype Nap1 and *nap1Δ.* B) Growth curve of indicated Nap1 strains at 38 °C as measured by A_{600} reading every 30 minutes. C) Nap1 cultures were pinned in 10-fold serial dilutions onto media with various chemicals/conditions, including YPD plates containing 60mM Acetic Acid, SC-Ura plates containing 100µg/mL 6-Azauracil, SC-Lysine plates, and YPD media with 15 µg/mL Calcofluor white stain. For NaN_3 and heat shock, cultures were incubated with 0.5% NaN_3 for approximately one cell cycle (90 minutes) or heated at 55 °C for 5 minutes prior to pinning on YPD media, respectively.

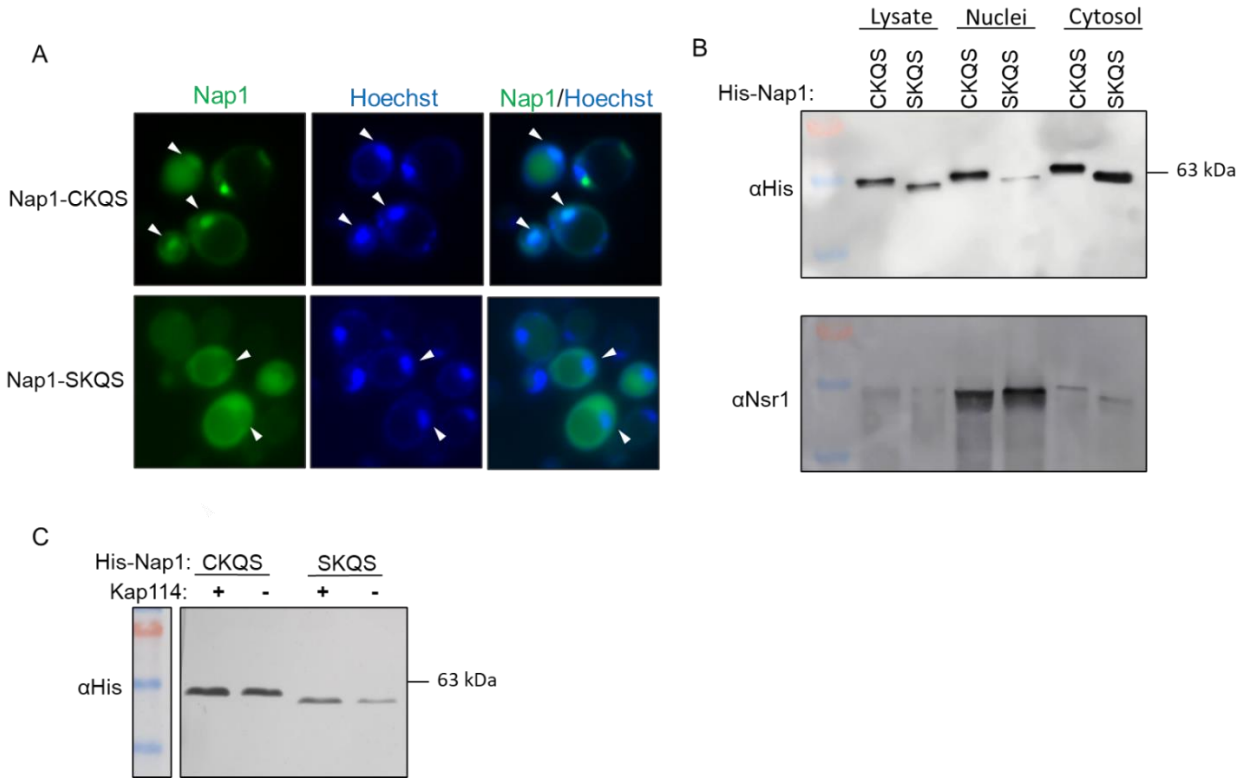


Figure 4.3: Nap1 prenylation impacts nuclear localization. A) Microscopy localization studies of yeast strains lacking chromosomally encoded *NAP1* (yWS1666, *nap1Δ*) containing plasmids expressing GFP-tagged wildtype Nap1 (GFP-Nap1-CKQS), or non-prenylated Nap1 (GFP-Nap1-SKQS) show that GFP-Nap1-CKQS is able to colocalize with Hoechst 33342 stain in a small fraction of cells, whereas colocalization is never visible in GFP-Nap1-SKQS strains. B, C) Cell fractionation of His-tagged Nap1 strains (as described in Fig 4.1) were evaluated by SDS-PAGE and immunoblot showing Nap1 levels in total cell lysate, nuclear fraction, and cytosol (B). Cell fractionation immunoblot of Nap1. Cell fractions are not proportional, with sample loading normalized by Ponceau S staining for each CKQS/SKQS fraction pairing. Nuclear fraction is indicated by nuclear protein, Nsr1. C) Evaluation of Nap1 nuclear levels in the presence or absence of the Nap1 nuclear transporter, Kap114.

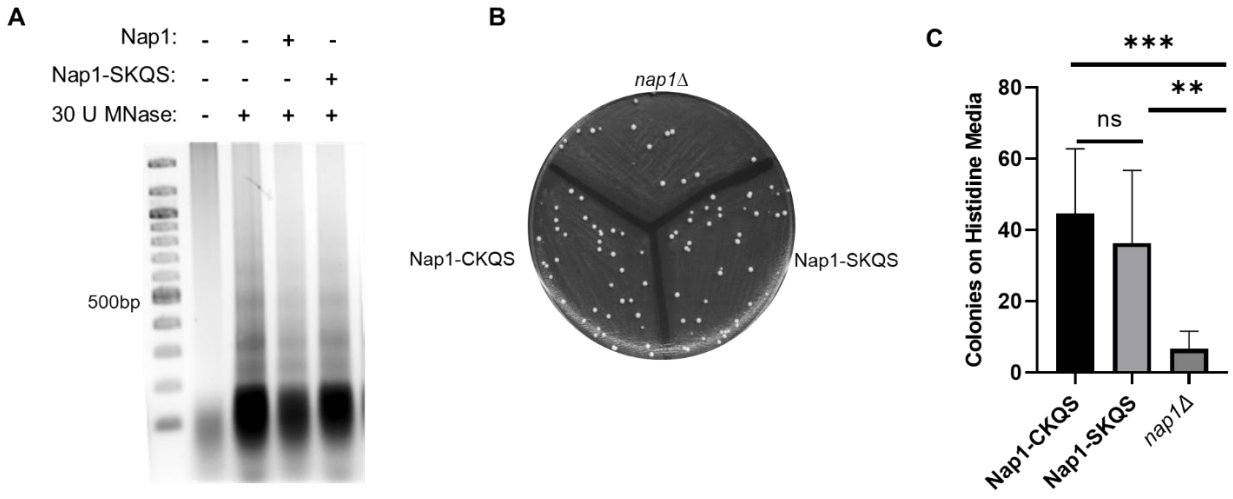


Figure 4.4: Prenylation shows no obvious impact on Nap1 nuclear functions. Nap1 strains described in Figure 4.2 were evaluated for any functional differences. A) Yeast spheroplasts were treated with 30 U MNase and incubated at 37 °C for 10 minutes. DNA was recovered using phenol: chloroform extraction and banding patterns evaluated by electrophoresis for any shifts. B,C) Transposition of Ty1 in varying Nap1 strains. Single colonies were patched onto SC media lacking Uracil and Leucine (SC-UL), grown for 72 hours at 25 °C and then replica plated onto SC-Histidine, Uracil, and Leucine (SC-HUL). After an additional 72 hours, plates were scanned (B) and colonies quantified (n=8) (C).

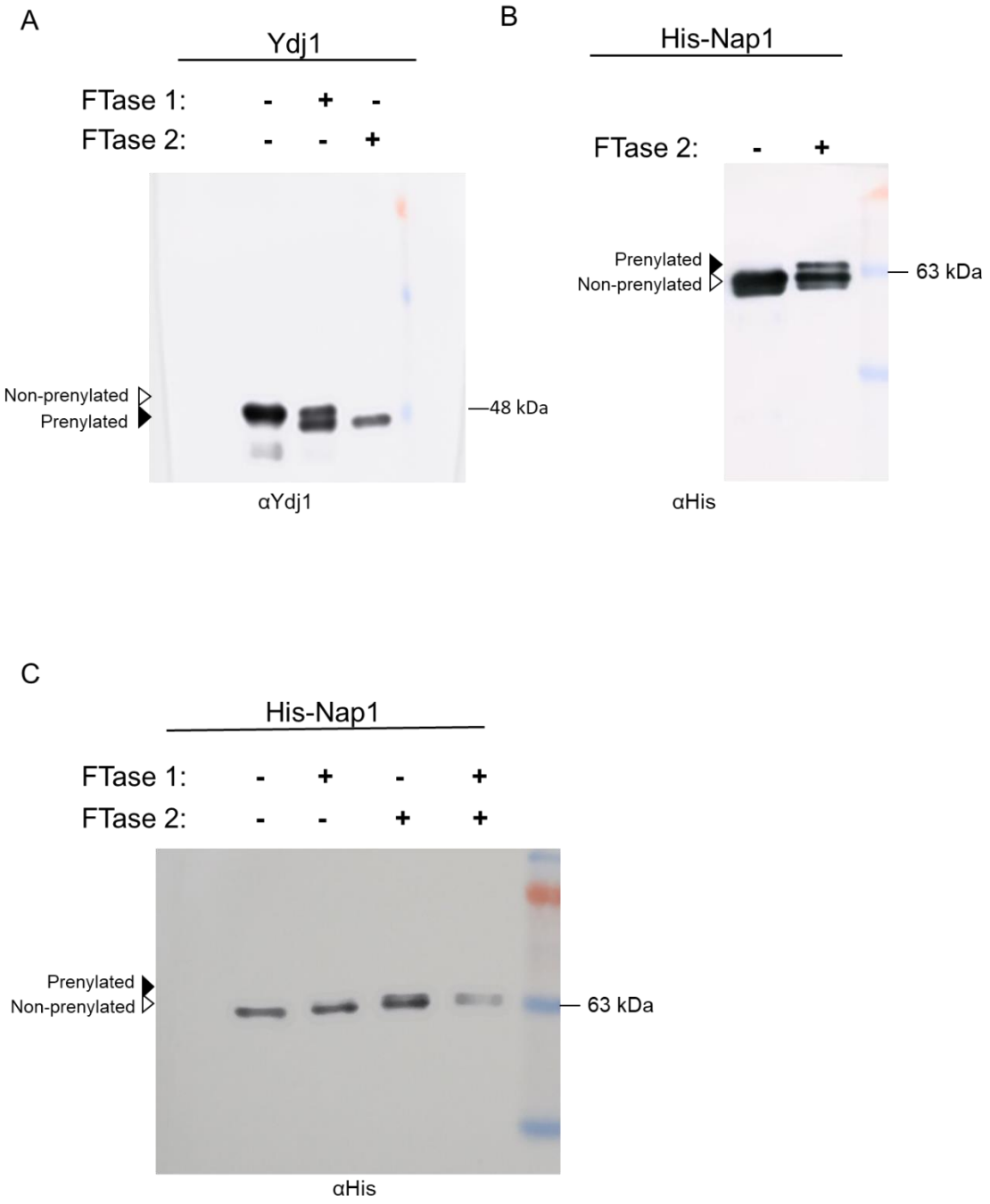


Figure 4.5: *In vivo* farnesylation of shunt proteins. Plasmids encoding recombinant *Sc*Ydj1 (A) and *Sc*Nap1 (B, C) were co-expressed in BL21 DE3 with pETDuet (FTase 1, pWS1693) and/or CDF-Duet (FTase 2, pWS1790) plasmids expressing both *Sc*FTase subunits. Protein expression was induced with 1mM IPTG. After induction, cells were harvested by centrifugation and resuspended in Laemmli buffer for immunoblot evaluation. Prenylation (closed triangle) is indicated by downward shift for Ydj1 and upward shift for Nap1.

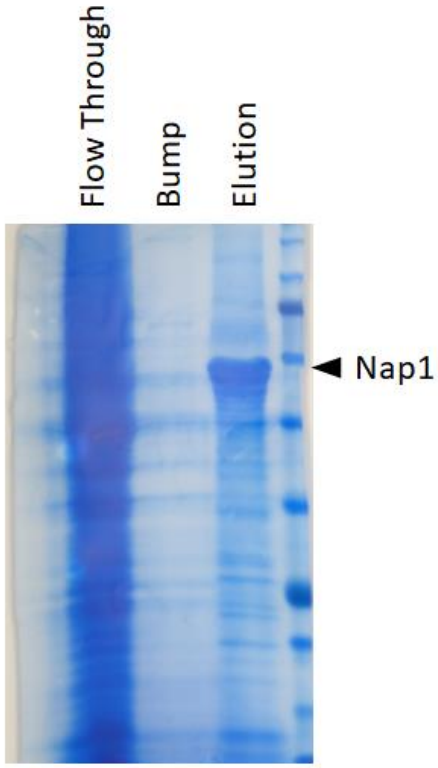


Figure S4.1: Purification of *Sc*Nap1. His-Nap1 (pWS1323) cloned under constitutively active, high expression PGK promoter was expressed in BY4174 yeast strain and purified using Talon resin and concentrated using Amicon Ultra 10k centrifugal filter and evaluated by SDS-PAGE/Coomassie.

Supplementary Table S4.1. Yeast strains used in this study

Strain	Genotype	Reference
IH1783; ATCC#204278	*MAT***a** *trp1 leu2 ura3 his4 can1*	(58)
SM2331	*MAT***a** *trp1 leu2 ura3 his4 can1 mfa1-∆1 mfa2-∆*	(58)
yWS304	*MAT***a** *his leu2 met15 ura3 ydj1::KAN*	(59)
yWS1666	*MAT***a** *his3 leu2 met15 ura3 nap1::KAN*	(60)
yWS1772	*MAT*α *his3 leu2 met15 ura3 nap1::KAN*	This study
yWS2173	*MAT***a** *his3 leu2 met15 ura3 kap114::KAN*	(60)
yWS2201	*his3 leu2 met15 ura3 nap1::KAN kap114::KAN*	This study

Supplementary Table S4.2. Plasmids used in this study

Gene	Identifier	Genotype	Reference
vectors	pRS315	*CEN LEU2*	(61)
	pRS316	*CEN URA3*	(61)
	pWS1648	*pETDuet-1 DNA*	This study
NAP1	pWS1318	*CEN URA3 NAP1*	This study
	pWS1323	*2μ URA3 His-NAP1*	This study
	pWS1474	*CEN URA3 His-NAP1*	This study
	pWS1475	*CEN URA3 NAP1-SKQS*	This study
	pWS1479	*CEN URA3 GFP-NAP1*	This study
	pWS1497	*CEN LEU2 NAP1*	This study
	pWS1498	*CEN LEU2 NAP1-SKQS*	This study
	pWS1769	*CEN URA3 GFP-NAP1-SKQS*	This study
	pWS1770	*CEN URA3 His-NAP1-SKQS*	This study
	pWS1938	*pET28-His-TEV-Nap1*	This study
YDJ1	pWS1411	*YDJ1-CKQS*	Chapter 3
	pWS 1132	*YDJ1-SASQ*	(4)
	pWS942	*YDJ1*	(4)
	pWS1286	*YDJ1-CVIA*	(4)
	pET9.YDJ1	*YDJ1*	(42)
MFA1	pWS610	*MFA*	(62)
	pWS846	*MFA1-CKQS*	Chapter 3
RAM1/RAM2	pWS1790	*CDF-Duet1*	(63)
	pWS1693	*pET-Duet RAM2 - RAM1*	This study
TY1	pOY1	*Ty1-H3his3-AI*	(40)

References

1. Zhang FL, Casey PJ. Protein prenylation: Molecular mechanisms and functional consequences. Annual review of biochemistry. 1996;65:241-69.

2. Wright LP, Philips MR. Thematic review series: Lipid posttranslational modifications. Caax modification and membrane targeting of ras. Journal of lipid research. 2006;47(5):883-91.

3. Wang M, Casey PJ. Protein prenylation: Unique fats make their mark on biology. Nature reviews Molecular cell biology. 2016;17(2):110-22.

4. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

5. Heilmeyer LM, Jr., Serwe M, Weber C, Metzger J, Hoffmann-Posorske E, Meyer HE. Farnesylcysteine, a constituent of the alpha and beta subunits of rabbit skeletal muscle phosphorylase kinase: Localization by conversion to s-ethylcysteine and by tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(20):9554-8.

6. Kilpatrick EL, Hildebrandt JD. Sequence dependence and differential expression of ggamma5 subunit isoforms of the heterotrimeric g proteins variably processed after prenylation in mammalian cells. The Journal of biological chemistry. 2007;282(19):14038-47.

7. Leung KF, Baron R, Ali BR, Magee AI, Seabra MC. Rab gtpases containing a caax motif are processed post-geranylgeranylation by proteolysis and methylation. The Journal of biological chemistry. 2007;282(2):1487-97.

8. Berger BM, Kim JH, Hildebrandt ER, Davis IC, Morgan MC, Hougland JL, et al. Protein isoprenylation in yeast targets cooh-terminal sequences not adhering to the caax consensus. Genetics. 2018;210(4):1301-16.

9. Kellogg DR, Kikuchi A, Fujii-Nakata T, Turck CW, Murray AW. Members of the nap/set family of proteins interact specifically with b-type cyclins. The Journal of cell biology. 1995;130(3):661-73.

10. Ishimi Y, Yasuda H, Hirosumi J, Hanaoka F, Yamada M. A protein which facilitates assembly of nucleosome-like structures in vitro in mammalian cells. Journal of biochemistry. 1983;94(3):735-44.

11. Kellogg DR, Murray AW. Nap1 acts with clb1 to perform mitotic functions and to suppress polar bud growth in budding yeast. The Journal of cell biology. 1995;130(3):675.

12. Altman R, Kellogg D. Control of mitotic events by nap1 and the gin4 kinase. The Journal of cell biology. 1997;138(1):119-30.

13. Mosammaparast N, Ewart CS, Pemberton LF. A role for nucleosome assembly protein 1 in the nuclear transport of histones h2a and h2b. The EMBO Journal. 2002;21(23):6527-38.

14. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS computational biology. 2007;3(4):e66.

15. Galichet A, Gruissem W. Developmentally controlled farnesylation modulates atnap1;1 function in cell proliferation and cell expansion during arabidopsis leaf development. Plant Physiol. 2006;142(4):1412-26.

16. Storck EM, Morales-Sanfrutos J, Serwa RA, Panyain N, Lanyon-Hogg T, Tolmachova T, et al. Dual chemical probes enable quantitative system-wide analysis of protein prenylation and prenylation dynamics. Nature Chemistry. 2019.

17. Jones JM, Morrell JC, Gould SJ. Pex19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. The Journal of cell biology. 2004;164(1):57-67.

18. Emmanouilidis L, Schutz U, Tripsianes K, Madl T, Radke J, Rucktaschel R, et al. Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor pex19. Nature communications. 2017;8:14635.

19. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, et al. A novel role of farnesylation in targeting a mitotic checkpoint protein, human spindly, to kinetochores. The Journal of cell biology. 2015;208(7):881-96.

20. Mosalaganti S, Keller J, Altenfeld A, Winzker M, Rombaut P, Saur M, et al. Structure of the rzz complex and molecular basis of its interaction with spindly. The Journal of cell biology. 2017;216(4):961-81.

21. Regnard C, Desbruyères E, Huet JC, Beauvallet C, Pernollet JC, Eddé B. Polyglutamylation of nucleosome assembly proteins. The Journal of biological chemistry. 2000;275(21):15969-76.

22. Asahara H, Tartare-Deckert S, Nakagawa T, Ikehara T, Hirose F, Hunter T, et al. Dual roles of p300 in chromatin assembly and transcriptional activation in cooperation with nucleosome assembly protein 1 in vitro. Molecular and cellular biology. 2002;22(9):2974-83.

23. Calvert MEK, Keck KM, Ptak C, Shabanowitz J, Hunt DF, Pemberton LF. Phosphorylation by casein kinase 2 regulates nap1 localization and function. Molecular and cellular biology. 2008;28(4):1313-25.

24. Huang Z-X, Zhao P, Zeng G-S, Wang Y-M, Sudbery I, Wang Y. Phosphoregulation of nap1 plays a role in septin ring dynamics and morphogenesis in candida albicans. mBio. 2014;5(1):e00915-13.

25. Mortensen EM, McDonald H, Yates J, 3rd, Kellogg DR. Cell cycle-dependent assembly of a gin4-septin complex. Molecular biology of the cell. 2002;13(6):2091-105.

26. Fujii-Nakata T, Ishimi Y, Okuda A, Kikuchi A. Functional analysis of nucleosome assembly protein, nap-1. The negatively charged cooh-terminal region is not necessary for the intrinsic assembly activity. The Journal of biological chemistry. 1992;267(29):20980-6.

27. Ito T, Bulger M, Kobayashi R, Kadonaga JT. Drosophila nap-1 is a core histone chaperone that functions in atp-facilitated assembly of regularly spaced nucleosomal arrays. Molecular and cellular biology. 1996;16(6):3112-24.

28. Miyaji-Yamaguchi M, Kato K, Nakano R, Akashi T, Kikuchi A, Nagata K. Involvement of nucleocytoplasmic shuttling of yeast nap1 in mitotic progression. Molecular and cellular biology. 2003;23(18):6672-84.

29. Mosammaparast N, Jackson KR, Guo Y, Brame CJ, Shabanowitz J, Hunt DF, et al. Nuclear import of histone h2a and h2b is mediated by a network of karyopherins. The Journal of cell biology. 2001;153(2):251-62.

30. Mosammaparast N, Del Rosario BC, Pemberton LF. Modulation of histone deposition by the karyopherin kap114. Molecular and cellular biology. 2005;25(5):1764-78.

31. Elble R. A simple and efficient procedure for transformation of yeasts. BioTechniques. 1992;13(1):18-20.

32. Herskowitz I, Jensen RE. [8] putting the ho gene to work: Practical uses for mating-type switching. Methods in enzymology. 194: Academic Press; 1991. p. 132-46.

33. Oldenburg KR, Vo KT, Michaelis S, Paddon C. Recombination-mediated pcr-directed plasmid construction in vivo in yeast. Nucleic Acids Res. 1997;25(2):451-2.

34. Kubota T, Stead DA, Hiraga S-i, ten Have S, Donaldson AD. Quantitative proteomic analysis of yeast DNA replication proteins. Methods (San Diego, Calif). 2012;57(2):196-202.

35. Kim S, Lapham AN, Freedman CG, Reed TL, Schmidt WK. Yeast as a tractable genetic system for functional studies of the insulin-degrading enzyme. The Journal of biological chemistry. 2005;280(30):27481-90.

36. Zaret K. Micrococcal nuclease analysis of chromatin structure. Current protocols in molecular biology. 1999;45(1):21.1. 1-.1. 17.

37. Aguilar-Gurrieri C, Larabi A, Vinayachandran V, Patel NA, Yen K, Reja R, et al. Structural evidence for nap1-dependent h2a-h2b deposition and nucleosome assembly. Embo j. 2016;35(13):1465-82.

38. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:Chloroform. CSH protocols. 2006;2006(1).

39. Curcio MJ, Garfinkel DJ. Single-step selection for ty1 element retrotransposition. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(3):936-40.

40. Lee BS, Lichtenstein CP, Faiola B, Rinckel LA, Wysock W, Curcio MJ, et al. Posttranslational inhibition of ty1 retrotransposition by nucleotide excision repair/transcription factor tfiih subunits ssl2p and rad3p. Genetics. 1998;148(4):1743-61.

41. Fres JM, Müller S, Praefcke GJK. Purification of the caax-modified, dynamin-related large gtpase hgbp1 by coexpression with farnesyltransferase. Journal of lipid research. 2010;51(8):2454-9.

42. Caplan AJ, Tsai J, Casey PJ, Douglas MG. Farnesylation of ydj1p is required for function at elevated growth temperatures in saccharomyces cerevisiae. The Journal of biological chemistry. 1992;267(26):18890-5.

43. Longtine MS, Theesfeld CL, McMillan JN, Weaver E, Pringle JR, Lew DJ. Septin-dependent assembly of a cell cycle-regulatory module in saccharomyces cerevisiae. Molecular and cellular biology. 2000;20(11):4049-61.

44. Keck KM, Pemberton LF. Interaction with the histone chaperone vps75 promotes nuclear localization and hat activity of rtt109 in vivo. Traffic (Copenhagen, Denmark). 2011;12(7):826-39.

45. D'Arcy S, Martin KW, Panchenko T, Chen X, Bergeron S, Stargell LA, et al. Chaperone nap1 shields histone surfaces used in a nucleosome and can put h2a-h2b in an unconventional tetrameric form. Mol Cell. 2013;51(5):662-77.

46. Del Rosario BC, Pemberton LF. Nap1 links transcription elongation, chromatin assembly, and messenger rnp complex biogenesis. Molecular and cellular biology. 2008;28(7):2113-24.

47. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. Nature. 2006;440(7084):637-43.

48. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences. 2001;98(8):4569-74.

49. Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002;415(6868):141-7.

50. Brodeur GM, Sandmeyer SB, Olson MV. Consistent association between sigma elements and trna genes in yeast. Proceedings of the National Academy of Sciences. 1983;80(11):3292-6.

51. Mularoni L, Zhou Y, Bowen T, Gangadharan S, Wheelan SJ, Boeke JD. Retrotransposon ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in trna hotspots. Genome research. 2012;22(4):693-703.

52. Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome biology. 62005. p. R55.

53. Lorenz C, Ince S, Zhang T, Cousin A, Batra-Safferling R, Nagel-Steger L, et al. Farnesylation of human guanylate-binding protein 1 as safety mechanism preventing structural rearrangements and uninduced dimerization. The FEBS Journal. 2019;n/a(n/a).

54. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, et al. A novel role of farnesylation in targeting a mitotic checkpoint protein, human spindly, to kinetochores. Journal of Cell Biology. 2015;208(7):881-96.

55. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(25):10171-6.

56. Baller JA, Gao J, Stamenova R, Curcio MJ, Voytas DF. A nucleosomal surface defines an integration hotspot for the saccharomyces cerevisiae ty1 retrotransposon. Genome research. 2012;22(4):704-13.

57. Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, et al. Structural basis for retroviral integration into nucleosomes. Nature. 2015;523(7560):366-9.

58. Michaelis S, Herskowitz I. The a-factor pheromone of saccharomyces cerevisiae is essential for mating. Molecular and cellular biology. 1988;8(3):1309-18.

59. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the saccharomyces cerevisiae genome. Nature. 2002;418(6896):387-91.

60. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, et al. Designer deletion strains derived from saccharomyces cerevisiae s288c: A useful set of strains and plasmids for pcr-mediated gene disruption and other applications. Yeast (Chichester, England). 1998;14(2):115-32.

61. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in saccharomyces cerevisiae. Genetics. 1989;122(1):19-27.

62. Krishnankutty RK, Kukday SS, Castleberry AJ, Breevoort SR, Schmidt WK. Proteolytic processing of certain caax motifs can occur in the absence of the rce1p and ste24p caax proteases. Yeast (Chichester, England). 2009;26(8):451-63.

63. DeGraw AJ, Hast MA, Xu J, Mullen D, Beese LS, Barany G, et al. Caged protein prenyltransferase substrates: Tools for understanding protein prenylation. Chemical biology & drug design. 2008;72(3):171-81.

CHAPTER 5

CONCLUSIONS

Since the establishment of the Shunt pathway as an alternative outcome of the CaaX pathway, it has opened the possibility of additional, non-canonical sequences being targets of prenylation, with some potentially being shunted *in vivo* (1). In Chapter 2, we were able to use Ydj1 to identify additional shunted sequences and expand upon the number of known prenylatable sequences. By identifying approximately 140 shunted sequences with no clear motif consensus, this study illustrates the broad specificity of these CaaX prenyltransferases (**Fig 2.5**). However, by abolishing the initial rules of the CaaX pathway of aliphatic enrichment at the a_1 and a_2 positions, this expands the number of possible prenylatable Cxxx sequences to 8000 (20 x 20 x 20 possible residues at each x position). This number presents a challenge to probe all these sequences experimentally. While peptides tested *in vitro* have been successful at identifying sequences outside the traditional CaaX motif, testing all 8000 would be cost prohibitive (2). In recent years, click chemistry utilizing metabolic labeling have also been useful in identifying prenylation *in vivo*¸ but they are limited to naturally occurring proteins and motifs (3-5). An *in vivo* study containing a library containing nearly all 8000 Cxxx sequences followed by Next-Gen Sequencing was done in yeast, however this study used Ras as a reporter, limiting the output of positive hits to sequences that also undergo proteolysis and methylation (6). While our new Ydj1 reporter may prove to be

useful to overcome this limitation, the screen done in Chapter 2 only resulted in 67,000 colonies screened. It was determined that using this method would require approximately 3.2 million colonies to ensure full coverage of all 8000 sequences, creating a labor extensive experiment. While it may be possible to find a way to utilize Ydj1 using NGS as used in the Ras screen this is beyond the scope of this study, although is being pursued by others in the lab.

In order to circumvent the challenge of testing all 8000 sequences, in Chapter 3 we instead aimed to develop a method to predict prenylation for all possible Cxxx sequences using machine learning. Although previous methods to predict prenylation have been established, by using the sequences we identified in Chapter 2 together with previously published data, we designed a prenylation prediction method that would avoid an aliphatic enriched bias and allow for better predictions for non-canonical sequences using the machine learning platform, Support Vector Machine (SVM) (**Fig 3.4, Tables 3.3-3.5**) (6-9). Indeed, by evaluating 31 sequences for prenylation by gel shift, were able to see that our prediction method using SVM outperformed previously established predictions. It is interesting that the Freq, a scoring system developed in Chapter 2, performed nearly as well as (SVM). A possible explanation for this is that both Freq and SVM rely on the sequences identified by Ydj1. However, as the machine learning field continues to rapidly evolve, it's likely that these machine learning based predictions can be improved, while the Freq predictions will remain unchanged. This implies that machine learning may eventually outperform Freq.

After identifying more prenylatable sequences, either *in vivo* or as predicted prenylation targets, we next wanted to investigate Nap1, a protein predicted by our

machine learning to be prenylated and shunted. In Chapter 4, we were able to show by gel shift that Nap1 was indeed prenylated in yeast and provide indirect evidence that this motif (CKQS) is shunted. Unfortunately, due to the subtle nature of Nap1 phenotypes, we were unable to identify a definitive role for this lipid PTM. However, microscopy hinted at a possible nuclear localization defect in our non-farnesylated mutant, which we were then able to show more clearly by western blot. This suggests that the farnesyl plays a role in nuclear localization of Nap1, and yet we were able to show that farnesylation does not seem to be necessary for nuclear import by Kap114, the only identified karyopherin for Nap1 (**Fig 4.3**). This implies that there is an additional mechanism for Nap1 nuclear trafficking that has yet to be determined where farnesylation may potentially play a role. Interestingly, a study done in *Arabidopsis thaliana* showed that during leaf development, farnesylated *At*Nap1 was found in the nucleus, supporting our finding's that farnesylation may impact nuclear trafficking in yeast (10).

Traditionally, prenylation has always been associated with membrane association, although in recent years, farnesyl groups have been shown to play a role in localization away from the cell membrane, as well as protein-protein interactions. In humans, the dynein regulator, Spindly, has been shown to be farnesylated and is predicted to be shunted with the motif CPQQ (Chapter 3). Providing evidence for farnesylation playing a role of farnesyl groups in protein-protein interactions, Spindly's involvement in the RZZ complex, a conserved group of mitotic checkpoint proteins, was shown to occur in a farnesyl-dependent manner. Additionally, Spindly's farnesylation was essential for proper localization to the kinetochores (11, 12) Although farnesylated

proteins are often shown with the lipid group inserted into the membrane, seemingly not interacting with the rest of the protein structure, an NMR study on the peroxisomal protein, Pex19, showed that the addition of the farnesyl induces substantial conformational changes throughout the protein. While the unfarnesylated COOH terminus of Pex19 was noted to be flexible and structurally autonomous, addition of the farnesyl group provided more rigidity to this region and was proposed to form a high affinity binding site for interactors of Pex19 (13). Notably, as Ydj1 and Nap1 are both noted to have disordered COOH termini, it's possible that farnesylation would have a similar affect, reducing the flexibility of this domain and perhaps inducing additional conformational changes. Due to the disordered nature of Nap1's COOH terminus, it's often truncated for *in vitro* and structural studies. Because of this, it's unclear what this region's function is, with conflicting data on the role of the COOH terminus in histone binding emerging from previous studies (14-16). To determine if the farnesyl would have any impact *in vitro* on Nap1's histone binding or overall structure, we initially hoped to be able to obtain enough purified, farnesylated protein for Nap1 using the *E. coli* expression system. However, while it may be possible to utilize this method for farnesylated Ydj1, it seems to be less efficient for Nap1 (**Fig 4.5**)(17). If this method or *in vitro* farnesylation of Nap1 can be used to obtain modified Nap1, it may be worthwhile repeating histone binding from previous studies with the lipid group present to see if there are any differences. In the future, structural studies may provide insight on the role of lipid groups on shunted proteins, as well as support the emerging evidence of farnesylation having a functional impact outside of membrane association.

Throughout this study we were able to expand upon the recently characterized shunted branch of the CaaX pathway by identifying shunted sequences *in vivo.* We were then able to use these sequences to help predict prenylation by machine learning for sequences both found naturally in yeast, as well as all 8000 possible combinations. All in all, we found that the total number of prenylatable sequences is much greater than previously expected, with nearly 1/3 of all Cxxx sequences predicted to be prenylated. Moreover, we investigated another shunt protein, the histone chaperone Nap1, and found that farnesylation may play a role in nuclear trafficking of Nap1. Altogether, this thesis demonstrates the broad specificity of CaaX-type prenylation and identifies examples of shunting *in vivo* through Ydj1 and Nap1.

References

1. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

2. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA. Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities. Journal of molecular biology. 2010;395(1):176-90.

3. DeGraw AJ, Palsuledesai C, Ochocki JD, Dozier JK, Lenevich S, Rashidian M, et al. Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. Chemical biology & drug design. 2010;76(6):460-71.

4. Storck EM, Morales-Sanfrutos J, Serwa RA, Panyain N, Lanyon-Hogg T, Tolmachova T, et al. Dual chemical probes enable quantitative system-wide analysis of protein prenylation and prenylation dynamics. Nature Chemistry. 2019.

5. Onono FO, Morgan MA, Spielmann HP, Andres DA, Subramanian T, Ganser A, et al. A tagging-via-substrate approach to detect the farnesylated proteome using two-dimensional electrophoresis coupled with western blotting. Molecular & cellular proteomics : MCP. 2010;9(4):742-51.

6. Stein V, Kubala MH, Steen J, Grimmond SM, Alexandrov K. Towards the systematic mapping and engineering of the protein prenylation machinery in saccharomyces cerevisiae. PloS one. 2015;10(3):e0120716.

7. Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome biology. 62005. p. R55.

8. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS computational biology. 2007;3(4):e66.

9. London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O. Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS computational biology. 2011;7(10):e1002170.

10. Galichet A, Gruissem W. Developmentally controlled farnesylation modulates atnap1;1 function in cell proliferation and cell expansion during arabidopsis leaf development. Plant Physiol. 2006;142(4):1412-26.

11. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, et al. A novel role of farnesylation in targeting a mitotic checkpoint protein, human spindly, to kinetochores. The Journal of cell biology. 2015;208(7):881-96.

12. Mosalaganti S, Keller J, Altenfeld A, Winzker M, Rombaut P, Saur M, et al. Structure of the rzz complex and molecular basis of its interaction with spindly. The Journal of cell biology. 2017;216(4):961-81.

13. Emmanouilidis L, Schutz U, Tripsianes K, Madl T, Radke J, Rucktaschel R, et al. Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor pex19. Nature communications. 2017;8:14635.

14. Park YJ, Luger K. The structure of nucleosome assembly protein 1. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(5):1248-53.

15. Zlatanova J, Seebart C, Tomschik M. Nap1: Taking a closer look at a juggler protein of extraordinary skills. The FASEB Journal. 2007;21(7):1294-310.

16. Fujii-Nakata T, Ishimi Y, Okuda A, Kikuchi A. Functional analysis of nucleosome assembly protein, nap-1. The negatively charged cooh-terminal region is not necessary for the intrinsic assembly activity. The Journal of biological chemistry. 1992;267(29):20980-6.

17. Fres JM, Müller S, Praefcke GJK. Purification of the caax-modified, dynamin-related large gtpase hgbp1 by coexpression with farnesyltransferase. Journal of lipid research. 2010;51(8):2454-9.

APPENDIX

Structure/Function Investigations of the CaaX Protease Rce1

Ras Converting Enzyme 1 (Rce1) is an integral membrane protease and key contributor to the CaaX post translational modification pathway. In the canonical pathway, CaaX proteins (where C is cysteine, a is an aliphatic residue, and X is any amino acid) undergo a three-step modification beginning with addition of either a C15 farnesyl or C20 geranylgeranyl isoprenyl lipid to the Cysteine residue catalyzed by Farnesyltransferase (FTase) or Geranylgeranyltransferase-I (GGTase-I) respectively. This is followed by proteolysis of the -aaX tripeptide by Rce1 or Ste24 and carboxymethylation of the isoprenylated cysteine by Isoprenylcysteine Carboxy Methyltransferase (ICMT), or Ste14 in yeast. These modifications increase hydrophobicity of proteins and promote membrane association (1).

CaaX proteins have a wide variety of functions within the cell, but Ras and other small GTPases are the most notable. Ras has been estimated to be involved in 30% of all cancer, with higher percentages in different tumor types such as pancreatic cancer (>95%)(2). Previous studies have determined that Ras undergoes all three processing steps to ensure proper localization and protein function (3-5). Therefore, understanding this pathway is of interest for anti-cancer therapies. Prenylation inhibitors were developed throughout the 1990’s and early 2000s; however, after reaching clinical trials, FTase/GGTase inhibitors were mostly unsuccessful because GGTase-I adds an alternative isoprenyl group to Ras in the absence of FTase (6). Recently, the primary

targets have moved on to inhibiting proteolysis or carboxymethylation by Rce1 and ICMT, respectively (7).

Several studies have been done examining the effect inhibition of Rce1 has on Ras. After Rce1 was initially identified in yeast, it was determined that Rce1 deficient yeast strains significantly reduced the effect of hyperactive Ras (8). A full genetic knockout of Rce1 resulted in an embryonic lethality in mice; however, conditional knockouts and cell lines have been developed. Using Cre-Lox to create a *rce1Δ* cell line in both normal cells and in a melanoma cell line, Rce1 deletion resulted in slower cell growth and inhibited progression in melanoma cells (4, 9). Inhibition by Rce1 Protease Inhibitors (RPI) showed that the Ras isomers are mislocalized to the cytosol, further supporting the idea of a promising anti-cancer therapy (10). Meanwhile, tissue specific studies show that the importance of Rce1 on other proteins is still unclear. As noted above, a Rce1 mouse knockout was embryonic lethal, but major organ systems showed no obvious cause of death (4). In oncogenic hematopoietic cells, Rce1 knockout accelerated the progression of myeloproliferative disease (11). In a conditional knockout of Rce1 in the heart, where expression levels are high, cardiomyopathies resulted in premature death of mice. Notably, in that same study, Rce1 knockout in the liver had no negative effects on mice (12). In neural cells, Rce1 deficiency caused a rapid degradation of photoreceptor cells (13).. It is thought that most of these effects are due to disrupting processing of all CaaX proteins, but in most cases it is unclear what specific proteins are involved (14).

While there has been evidence supporting the idea that Rce1 inhibitors may be a viable Ras therapeutic, the lack of structural information for Rce1 has slowed drug

development. Rce1 is localized to the ER membrane, with purification attempts dating back to the early 1990s remaining largely unsuccessful (15, 16). To date, only one structure of Rce1 has been solved by X-ray crystallography using an antibody from *Methanococcus maripaludis*, a prokaryotic methanogen in 2013 (17). The 2.5 Å structure showed Rce1 has eight transmembrane helices. *Sc*Rce1, meanwhile, has been predicted to have seven transmembrane domains by topology studies.(18). The *Mm*Rce1 crystal structure also provided insight on the mechanism of action of Rce1, which was unclear until that point. Rce1 was initially proposed to be a metalloenzyme or Cysteine protease, due to the conversed, essential Histidine and Glutamate residues and sensitivity to Cysteine protease inhibitors, respectively. However, no metal was detected within the enzyme, as well as metal chelators having no effect on activity. Additionally, in yeast, Rce1 was still active after mutagenesis of all its Cysteine residues. This led to the thought that Rce1 may have a novel mechanism. From the solved crystal structure, Manolaridis *et. al*. proposed that Rce1 utilizes a glutamate activated water molecule by E140 *Mm*Rce1, E156 in *Sc*Rce1;(17-22).

While *Mm*Rce1 has no known substrates *in vivo*, the substrate specificity is generally conserved among species, with 3 highly conserved residues proposed to be essential for catalytic activity: E140, H173, and H227 of *Mm*Rce1 (E156, H194 and H248 in *S. cerevisiae*) (17, 21). The functional relevance of this structure remains questionable, as this structure comes from a prokaryotic organism and there are no known native substrates. When examined *in vitro*, *Mm*Rce1 was able to cleave a farnesylated peptide based off the human CaaX motif from, RhoA, but was unable to cleave geranylgeranylated peptides. Additionally, *Mm*Rce1 cleaved at both the a_1 (P1)

and a_2 (P1´) positions of the CaaX motif rather than just a_1, suggesting that this ortholog is slightly more promiscuous than human or yeast Rce1 (17).

In general, membrane proteins have been notoriously difficult to characterize because they must be maintained in a hydrophobic environment throughout purification. Detergents have been the main technology used for solubilization of membrane proteins for decades; however, the range of critical micelle concentrations for each detergent, the number of detergents available, price, and optimization of the detergents, salts, and many other factors for each protein, solubilization can be a costly step. Additionally, even after solubilization with detergents, the stability of the protein is often affected due to of the loss of its surrounding lipid environment and stability provided while embedded within the membrane. This instability often causes problems in downstream steps such as purification and structural studies (23). Recently, there has been a surge in solved membrane protein structures using nanodiscs, which are able to act as a membrane mimetic during purification and maintain stability of membrane proteins (24-27). Nanodiscs, which consist of synthetic or natural phospholipids enclosed in a discoidal shape by membrane scaffold proteins, allow for the isolation and further characterization of membrane proteins by a variety of different techniques in addition to X-ray crystallography, such as cryo-EM and NMR. The downside to these synthetic nanodiscs is that they do not necessarily assist with the initial solubilization, which still necessitates the use of detergents. However, a variation of nanodiscs, SMALPs (Styrene-Maleic Acid Lipid Polymers), also referred to as lipodiscs or native nanodiscs, are able to solubilize membrane proteins without detergents (28). SMALPs are made of a repeating ratio of styrene to maleic acid and bind and disrupt the membrane

environment before partitioning the membrane components into soluble discs approximately 10 nm in diameter without the need for detergents. Additionally, the proteins themselves are reconstituted in their native lipid and accessory protein environment, preserving protein function(29, 30). Because of these characteristics, SMALPs have become especially useful in studying particularly insoluble or unstable proteins, as well as membrane protein complexes (27, 30, 31). While Rce1 has proved itself to be a challenging protein to study over the past twenty years, SMALPs may be a viable option to help stabilize it during purification and future structural studies.

Materials and Methods

Strains: All plasmids were expressed in SM3614 (32). Lithium acetate-based transformations were used to introduce plasmids, as described previously (33, 34). All strains were propagated in SC-Uracil (SC-U) media at 30 °C unless otherwise stated.

Plasmids: Plasmids used in this study are listed in **Supplemental Table SA.1**. All plasmids constructed for this study were created using previously reported methods (34, 35). Briefly, mutagenic oligonucleotides encoding desired NH_2 or COOH-terminal tags were co-transformed with linearized/gapped parent plasmids. Transformation mixes were plated onto appropriate media, and plasmids recovered from emerging colonies. Restriction digest and sequencing were used to verify proper sequence throughout the entire open reading frame.

Membrane isolation: Strains of interest were grown to log phase in appropriate media. Cells were then harvested by centrifugation, washed with water, and resuspended in 100 mM Tris, pH 9.4, 10 mM DTT. Cells were incubated on ice 10 minutes, recovered by centrifugation, and washed with 10 mM NaN_3. Recovered cells were incubated with Zymolyase in OB buffer (50 mM KPi, pH 7.5, 1.4 M sorbitol, 10 mM NaN_3) at 30 °C for 30 minutes, followed by a 10-minute recovery on ice. Spheroplasts were then harvested by centrifugation and resuspended in Lysis Buffer (50 mM Tris, pH 7.5, 0.2 M Sorbitol, 0.02% NaN_3, and protease inhibitors [1 mM phenylmethanesulfonylfluoride fluoride and aprotinin, 1 µg/mL each of chymostatin, leupeptin, and pepstatin]). Samples were then subjected to bead beating, 4 x 4 minutes. Cell membranes were isolated using differential centrifugation spins, starting with a 1000 x g spin. The supernatant was then collected and spun at 16,000 x g. The pellet containing membranes was then resuspended in lysis buffer, and protein concentration determined by Bradford assay if necessary.

SMALPs solubilization: SMALPs solubilization performed similar to as described (28). Briefly, isolated membranes expressing Rce1 were treated with SMALPs (XIRAN SL400005 S40, XIRAN SL25010 S25, and XIRAN SL300010 S35; Polyscope) at 1:1 volume and incubated at room temperature for one hour with gentle agitation. Membranes were then centrifuged at 100,000 or 16,000 x g for 1 hour, and the supernatant extracted. To determine solubility, insoluble pellet was resuspended to same volume and supernatant, and samples run on SDS-PAGE. DIBMA solubilization was performed similarly to SMALPs. Briefly, DIBMA sodium salts (Anatrace) were

resuspended in membrane lysis buffer. DIBMA suspension was then added to isolated membranes expressing Rce1 at 1:1 volume and incubated overnight. Samples were then centrifuged and the supernatant extracted.

SDS-PAGE and immunoblotting: Immunoblotting was performed as described previously (34, 36). Samples were analyzed by SDS-PAGE followed by immunoblotting on nitrocellulose with anti-HA or anti-GST and HRP conjugate in 1% TBST (100 mM Tris, 400 mM NaCl, 0.1% Tween 20, pH 7.5). Protein levels were detected by WesternBright ECL Spray (Advansta Inc, San Jose, California), and images captured using X-ray film or a digital imager (Kwikquant, Kindle Biosciences, Greenwich, Connecticut).

Purification: After solubilization with SMALPs, small scale purifications was performed using Ni-NTA or GST SpinTrap columns (GE healthcare) following manufacturer's protocol under native conditions, with the exception that SMA-Rce1 was incubated with resin overnight.

PEG-Maleimide labeling assay: The PEG-Maleimide (Peg-Mal) assay was performed as previously described (18). Briefly, isolated membranes expressing Rce1 or SMA-Rce1 samples with/without 0.5% Triton X were treated with 5 mM Peg-Mal and incubated on ice for 1 hour. Reactions were then quenched with 300 mM β-Mercaptoethanol for 10 minutes. Samples were then evaluated by immunoblotting.

Fluorescence based proteolysis assay: This assay was performed as previously described (10). Briefly, isolated yeast membranes expressing Rce1 were mixed with a fluorogenic peptide substrate based on the K-Ras CaaX motif (ABZ-KSKTKC(farnesyl)QLIM) (Anaspec, San Jose, CA) diluted in 4% DMSO. Membranes diluted in HM buffer (100 mM Hepes, 5 mM $MgCl_2$) were dispensed in clear bottom black 96 well plate, and the assay initiated by adding peptide substrate to membranes for a total concentration of 0.25mg/mL membranes and 20uM peptide. Fluorescence was measured every 30-60 seconds by Biotek Synergy HT or Biotek H1 with 320/420-nm excitation/emission filter. The fluorescence data was then graphed in Prism.

in vitro ***a****-factor Assay*: Proteolysis/methylation assays using SMA-Rce1 or membrane extracts were performed as described previously (37). Briefly, a farnesylated peptide resembling **a**-factor (YIIKGVFWDPA(farnesyl)CVIA) solubilized in MeOH was incubated with either purified SMA-Rce1 or membranes expressing Rce1 diluted in reaction buffer (200 mM Hepes, pH 7.5, 200 mM NaCl) to initiate cleavage. Reaction was incubated at 30 °C for 10 minutes and terminated by heating samples to 95 °C. For methylation, membranes containing Ste14 and SAM were added to sample containing heat killed proteases and peptide. The reaction was once again incubated at 30 °C for 60 minutes before terminating by heating mixture to 95 °C. The reaction mixes were then spotted onto a lawn of *MAT*α *sst2* cells (SM1086) spread on a YPD and incubated overnight. Proper **a**-factor processing results in a halo or lack of growth on *MAT*α lawn.

Results

Rce1 is soluble using SMALPS

Solubilization of membrane proteins is often one of the main limitations in advancements of structure characterization. Unfortunately, solubilization of each protein requires its own optimization of detergents and conditions, and even under optimal conditions, protein stability once detergent solubilized does not ensure activity. Styrene Maleic Acid Lipid Polymers (SMALPs) are an emerging new technique in membrane protein solubilization. SMALPs are able to directly bind to and break membranes, with membrane components, into discoidal segments approximately 10 nm in diameter (30). By this method, proteins and membrane complexes stay in their native environment and remain functional. Detergent solubilization of Rce1 results in a loss of enzymatic activity, as Rce1 purification attempts dating back over twenty years have shown. Instead, attempts often settled for a low yield, partial purification (20, 38). Yet, the stability provided by SMALPs may be what is required to purify and stabilize this challenging protein.

The size and solubility of SMALPs is dependent on the styrene: maleic acid styrene, with the most commonly used somewhere within the 2:1 to 3:1 range. Three readily available, hydrolyzed SMALPs are currently being tested: 1.3:1, 2.3:1, and 3:1. For Rce1 solubilization, membranes enriched with Rce1 were incubated with 2.5% (w/v) SMALPs and then centrifuged to remove insoluble proteins. A western blot of His- and GST- tagged Rce1 after centrifugation showed that both 2.3:1 and 3:1 SMALPs were able to solubilize Rce1 approximately 60% and 50%, respectively (**Fig A.1**). Due to this slight advantage of 2.3:1 SMALPs during solubilization, this ratio was used moving

forward unless stated otherwise. While it was hypothesized that GST tag would assist with solubility, the opposite seemed to occur. SMALP-treated GST-Rce1 appeared be consistently soluble but was more susceptible to pelleting during high-speed centrifugation (100,000g) compared to His-Rce1 (**Fig A.1B**). Nonetheless, from this set of experiments it is apparent that Rce1 can be solubilized using SMALPs.

His-tagged SMA-Rce1 can be purified successfully

After sufficient solubilization, SMALP-treated Rce1 (SMA-Rce1) pilot purifications were attempted using 10xHis-HA-*Sc*Rce1, and GST-*Sc*Rce1 using His and GST single use SpinTraps. Additionally, a His-tagged human ortholog of Rce1 was attempted (10xHis-HA-*Hs*Rce1Δ1-22). Previous studies have shown this human Rce1 ortholog to be active in yeast (10, 21). Using His SpinTraps, both His-*Sc*Rce1 and *Hs*Rce1 were able to be purified and visualized by immunoblot and silver stain (*Hs*Rce1 not shown) (**Fig A.2**). The GST-ScRce1 construct was less successful, with no detectable Rce1 by immunoblot. Due to inconsistent and inefficient solubility together with the unsuccessful purification attempt, the use of GST-Rce1 was discontinued.

SMA-Rce1 does not show activity by fluorescence

As Rce1 was able to be purified, the next question is if this SMA-Rce1 is active. Fluorescence-based assays are a commonly used Rce1 activity assays, where a quencher placed at the a_1 position in the CaaX motif (10). When the peptide is uncleaved, the fluorophore and quencher are close enough to remained quenched, whereas when Rce1 cleaves the -aaX of the peptide, fluorescence is emitted. Isolated

membranes containing Rce1 show an increase in fluorescence indicating Rce1 is active, however, SMALPs were added to the membranes, there was no fluorescence (**Fig A.3A**). Previous research showed excess SMALPs decreasing or inhibiting enzyme activity, and in many cases, removing excess SMALPs or decreasing the amount used to solubilize the protein resolved it (27, 39). Unfortunately, neither of these steps restored Rce1 activity. We hypothesized that there were three possible explanations for this lack of activity: (1) the SMALPs were inhibiting Rce1, (2) SMALPs were incompatible with the fluorescence assay, or (3) Rce1 was unfolded and therefore inactive. We first attempted to decrease the amount of SMALPs through a variety of conditions, including reducing the amount of SMALPs used for solubilization, using purified enzyme, using a concentrator to remove excess SMALPS, and adding salts, detergents, and lipids. As none of these methods indicated Rce1 was catalytically active, we then investigated whether the SMALPs were incompatible with the fluorescence assay. To test if SMALPS compatibility, SMALPs were added to two other proteases: membranes containing the other membrane bound CaaX protease, Ste24, and Trypsin.. Interestingly, the addition of SMALPs to both also resulted in a lack of activity. As Ste24 had previously been shown to cleave both farnesylated and unfarnesylated proteins, unlike Rce1 (40). Hypothesizing that the SMALPs may be preventing access to the hydrophobic isoprenyl group, both farnesylated and unmodified peptides were used with Ste24, and in both cases the results remained the same with no fluorescence. The lack of fluorescence in three different protease assays may indicate that for some reason, SMALPs may be incompatible with this fluorescence-based assay.

SMA-Rce1 demonstrates no enzymatic activity

After determining that SMALPs were incompatible with our fluorescence assay, we moved to an alternative approach, an *in vitro* **a**-factor proteolysis/methylation assay, to determine if Rce1 is active within the SMALPs. This assay uses a farnesylated peptide based on the yeast mating pheromone, **a**-factor, to show enzymatic activity. The farnesylated peptide is first incubated with membranes enriched for Rce1 or Ste24, which cleave the -aaX from the **a-**factor CaaX motif, before the addition of membranes expressing the carboxyl methyltransferase, Ste14. The processed **a**-factor is then plated on a lawn of sensitive *MAT*α cells containing a *sst2* mutation, leading to the surrounding cells going into growth arrest, forming a halo in the presence of properly processed **a**-factor (37). This method is significantly more sensitive than the fluorescence assay, able to detect 12 pg/µL of **a**-factor, suggesting that we should be able to detect even low levels of Rce1 activity. The farnesylated **a**-factor peptide was incubated with both membranes expressing Rce1 and purified SMA-Rce1 for both human and yeast Rce1 orthologs. While a halo was visible for the Rce1 membranes, again no activity was visible for SMA-Rce1 (**Fig A.3B**). We once again hypothesized that the SMALPs may be inhibiting Rce1, and attempted to add several factors such as salts, lipids, detergents, and several other conditions but saw no halo for the purified Rce1 under any conditions.

Use of alternative lipid nanodiscs, DIBMA, showed no improvement

Due to the fluorescence incompatibility and lack of activity in SMA-Rce1, we decided to also use a newer, alternative lipid nanodisc, DIBMA. For DIBMA copolymers, the

aromatic styrene polymer is replaced by diisobutylene (41). This allows for use in fluorescence-based assays, as well as absorbance-based quantification of proteins, a challenge we observed when using SMALPs. Additionally, DIBMA copolymers have a less disruptive impact on lipids during solubilization, which we felt may be helpful for Rce1's stability. DIBMA follows a similar method of solubilization as SMALPs and were shown to have comparable solubilization levels for Rce1 (**Fig A.4A**). However, purification using DIBMA resulted in less purified Rce1, as shown by silver stain (**Fig A.4B**). Nonetheless, we tested Rce1 activity for DIBMA-treated samples and once again saw no activity for our fluorescence-based assay or *in vitro* **a**-factor processing (not shown, **Fig A.3B**).

SMA-Rce1 is susceptible to additional Peg-Mal labeling

After not being able to detect any Rce1 activity, we sought to determine if Rce1 is properly folded within the SMALPs. A previously study using Peg-Maleimide (Peg-Mal) labelling with Rce1 membranes showed that Rce1's Cysteines are buried within the cell membrane and are inaccessible to Peg-Mal labeling (18). Properly folded Rce1 showed two bands by immunoblot, one at its usual size (approximately 35 kDa) and one 5 kDa higher from the accessible Cysteine located in the HA tag. Meanwhile, as Rce1 is known slightly denature in the presence of detergents, such as Triton-X, this was used as a negative control for Peg-Mal labeling. Indeed, when Triton-X is present, additional bands are visible, forming a ladder pattern. We then used Peg-Mal labeling in our SMA-Rce1 and compared the labeling pattern by immunoblot to Rce1 membranes +/- Triton X. From this, we were able to see several bands within the SMA-Rce1, suggesting that

SMA-Rce1 may not maintain its proper fold after SMALP treatment and removal from the membrane (**Fig A.5**). We again added lipids and other factors in attempts to stabilize Rce1, as well as testing DIBMA-treated Rce1 by Peg-Mal labeling, but all resulted in a similar ladder banding pattern.

Although Peg-Mal labeling was also attempted on *Hs*Rce1, while *Sc*Rce1 has 7 total cysteines, *Hs*Rce1 has 16 Cysteines. For this reason, it was challenging to determine the proper folding pattern for *Hs*Rce1 to compare purified SMA-*Hs*Rce1. However, the data from *Sc*Rce1, together with the lack of enzymatic activity for *Hs*Rce1 by the *in vitro* **a**-factor assay suggest that *Hs*Rce1 may not be functional.

Discussion

Structural characterization of the CaaX protease, Rce1, has remained a challenge over the past three decades. Here, we attempted to use a new method in membrane protein biology, Styrene-Maleic Acid Lipid Polymers (SMALPs), as a detergent free method to solubilize Rce1 while keeping it within its native lipid environment. While this method has been successful for many membrane proteins and complexes, it showed limited success for Rce1.

As opposed to other membrane proteins in the CaaX pathway, Rce1 has been shown to become inactive in the presence of detergents, suggesting that it loses its structural integrity. This makes any structural information difficult to obtain, with one published structure from a distantly related prokaryote. Using SMALPS, we were able to show that approximately 50% of Rce1 can be solubilized from the membrane and purified using a His-tag. However, the question became if this SMALP-treated (SMA-

Rce1) maintained its proper fold after solubilization and purification. Comparing SMA-Rce1 Peg-Mal labeling to membrane-bound Rce1 Peg-Mal labeling, it's evident that the cysteines are more accessible within the SMALPs. This data, together with the lack of visible activity, suggests that Rce1 may not maintain its structural integrity throughout the solubilization process. Although many attempts were made to help stabilize Rce1, including additions of salts, detergents below CMC levels, various lipids, there was no indication that Rce1 was properly folded or active. We also attempted to use an Rce1 ortholog from the yeast *Kluyveromyces marxianus*, which is able to grow from 5-40 °C, with the thought that this temperature adaptability may provide more stability (42). Unfortunately, *Km*Rce1 exhibited the same lack of activity. As a last resort, since Rce1 activity was already not shown, Rce1 membranes were saturated with a peptidomimetic inhibitor in attempt to stabilize Rce1's active site during solubilization/purification, however this Peg-Mal labeling still revealed this to have extra labelling. Through all these attempts, we were unable to see any activity of solubilized Rce1 or any way to show that the enzyme was properly folded, leading us to discontinue this project.

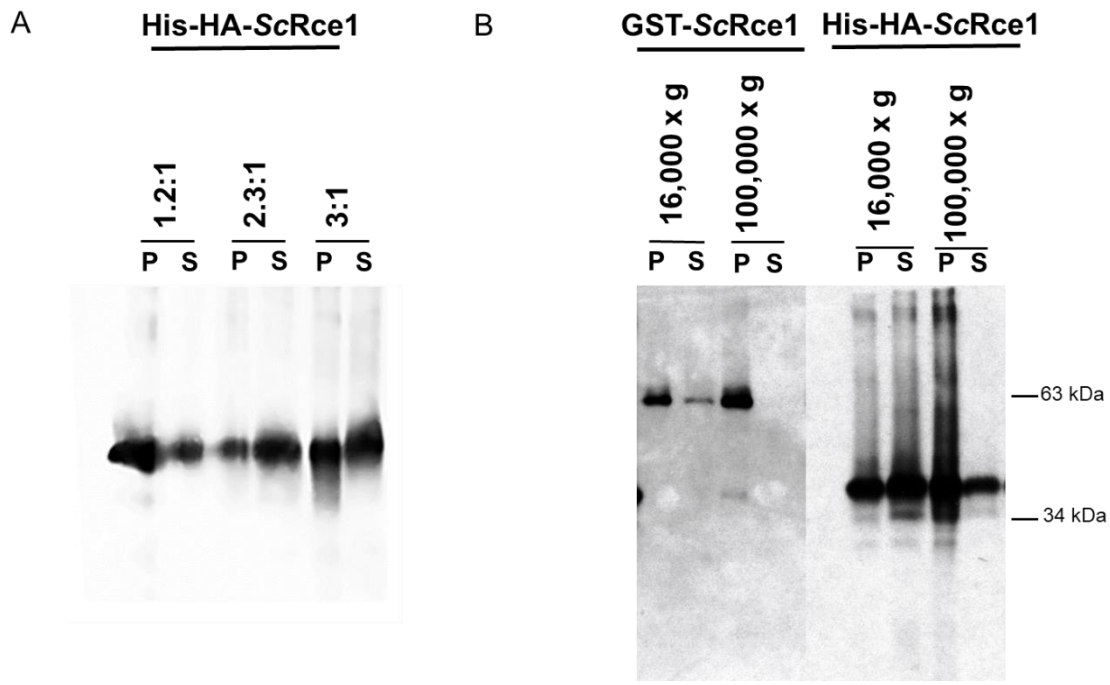


Fig A.1: Rce1 can be solubilized using Styrene: Maleic Acid Lipid Polymers (SMALPs). SMALPs are a new technology in membrane protein biology with the ability to solubilize membrane proteins straight from the membrane without detergents. Rce1 was treated with various sized SMALPs and the amount of Rce1 remaining in the insoluble pellet (P) versus the supernatant (S) was evaluated by immunoblot. A) Solubilization of Rce1 using SMALPs of differing Styrene: Maleic Acid ratios. B) Solubility of GST and His-tagged Rce1 following SMALPs treatment and clarifying spins at different speeds.

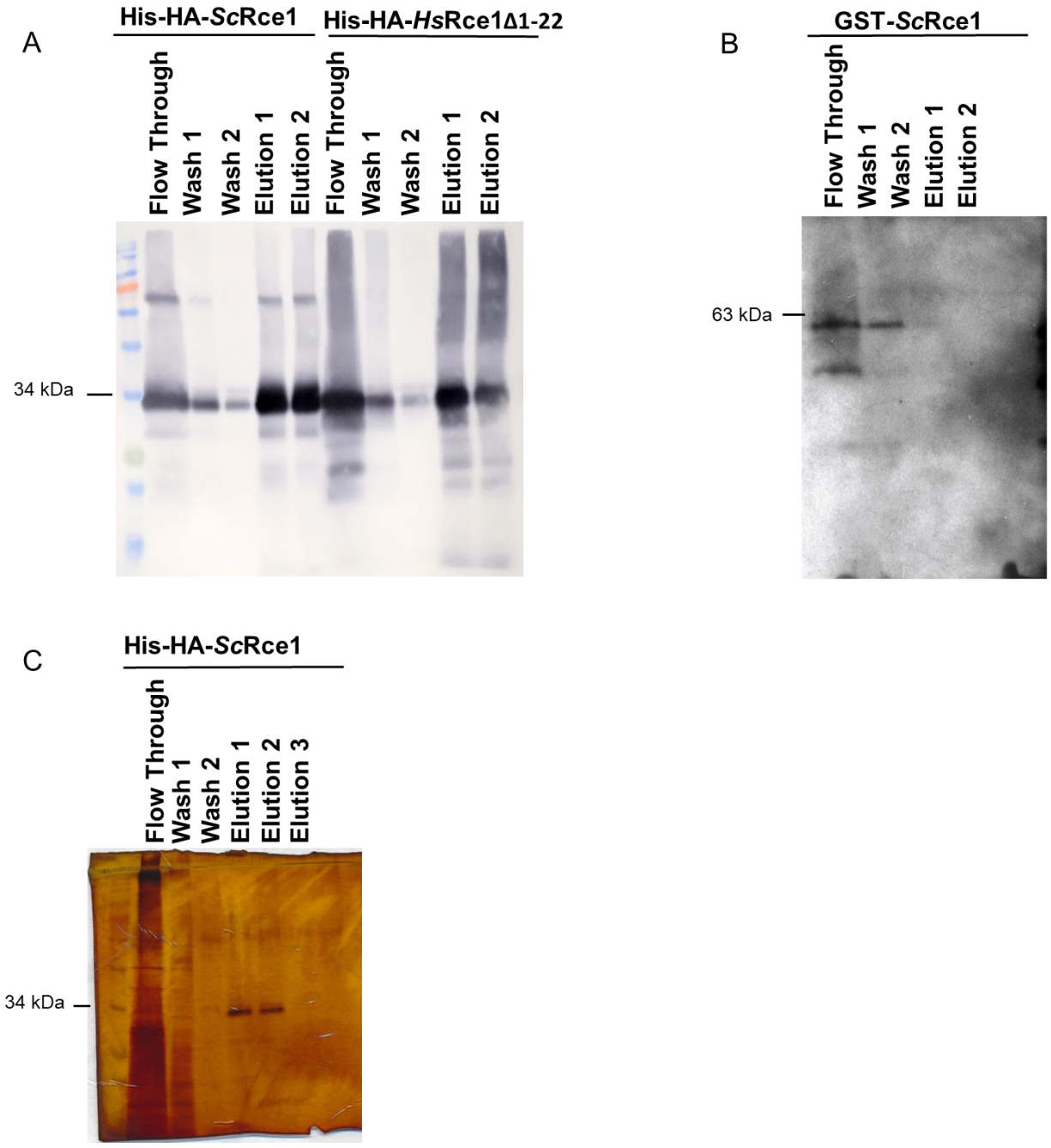


Fig A.2: Purification of Rce1 using SMALPs. His and GST-tagged ScRce1 and His-tagged HsRce1 were solubilized using 2.3:1 SMALPs and purified. A) Purification of His-tagged Rce1 orthologs using His-SpinTrap. Samples were eluted using 300 mM imidazole. B) Purification of GST-tagged Rce1 using GST-SpinTraps using 10 mM Glutathione. C) Silver staining of *Sc*Rce1 purification.

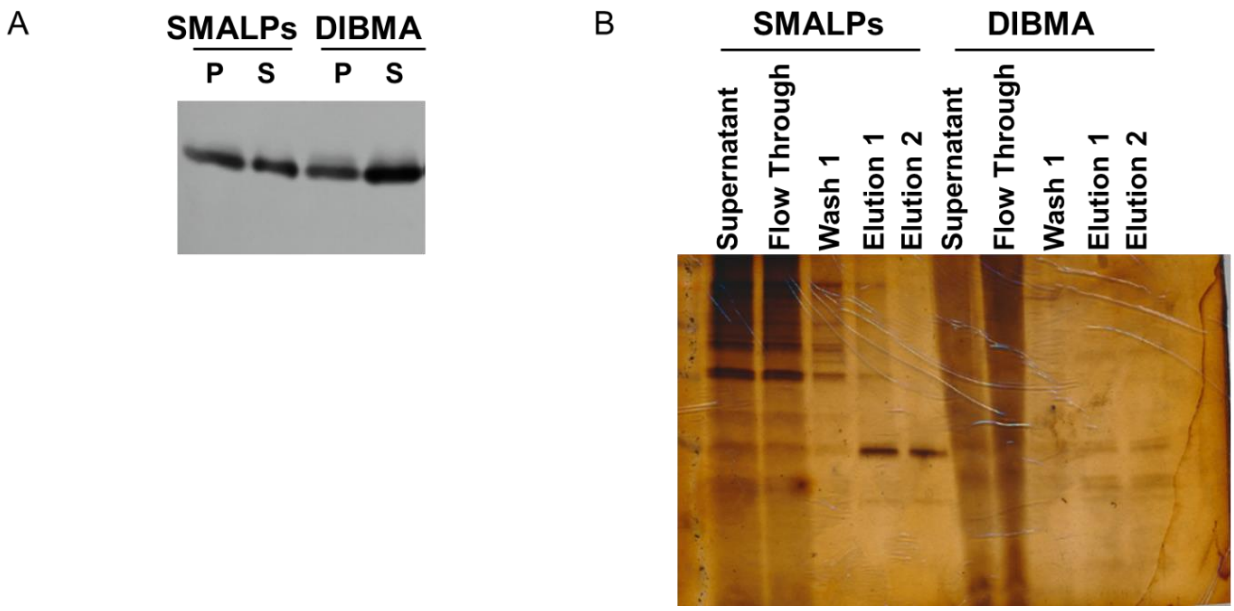
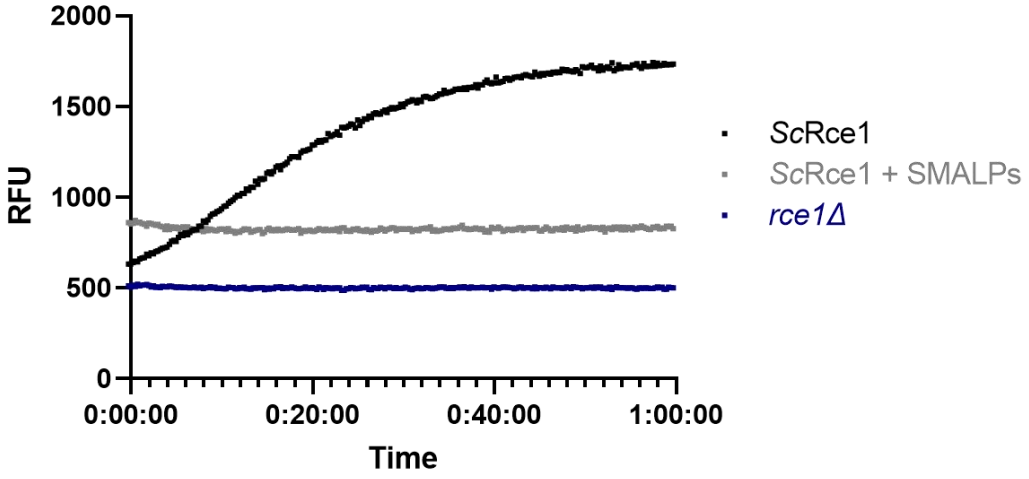


Fig A.3: DIBMA Solubilization and purification of His-HA-*Sc*Rce1. A) *Sc*Rce1 solubilization using alternative lipid nanodiscs, DIBMA, was compared to 2.3:1 SMALPs solubility by immunoblot. B) Purification of SMALPs and DIBMA solubilized Rce1.

A



B

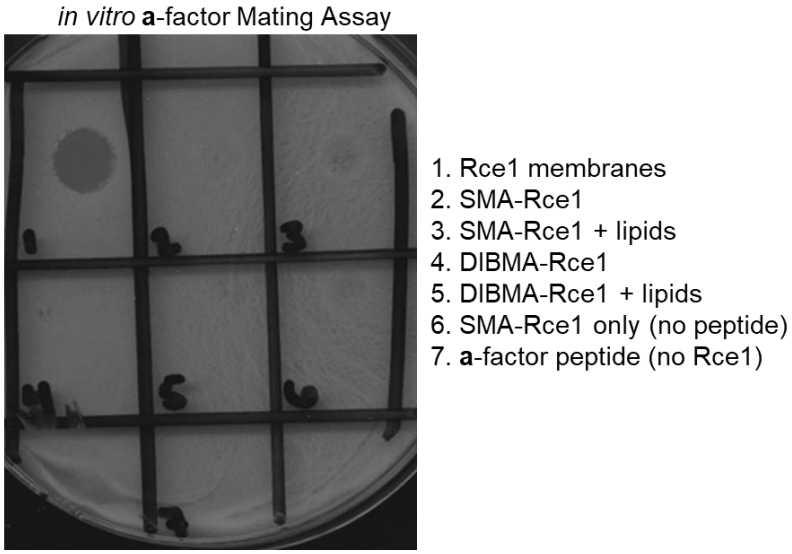


Fig A.4: Solubilized Rce1 shows no enzymatic activity. A) SMALP solubilized Rce1 activity was investigated using a fluorescence-based assay. Proteolytic cleavage of the peptide results in increased fluorescence. B) An *in vitro* **a**-factor processing assay shows no activity for SMALP and DIBMA purified Rce1, with and without additional lipids. Activity is indicated by a lack of growth or “halo” on the lawn of *MATα ss2* cells, as shown in box 1.

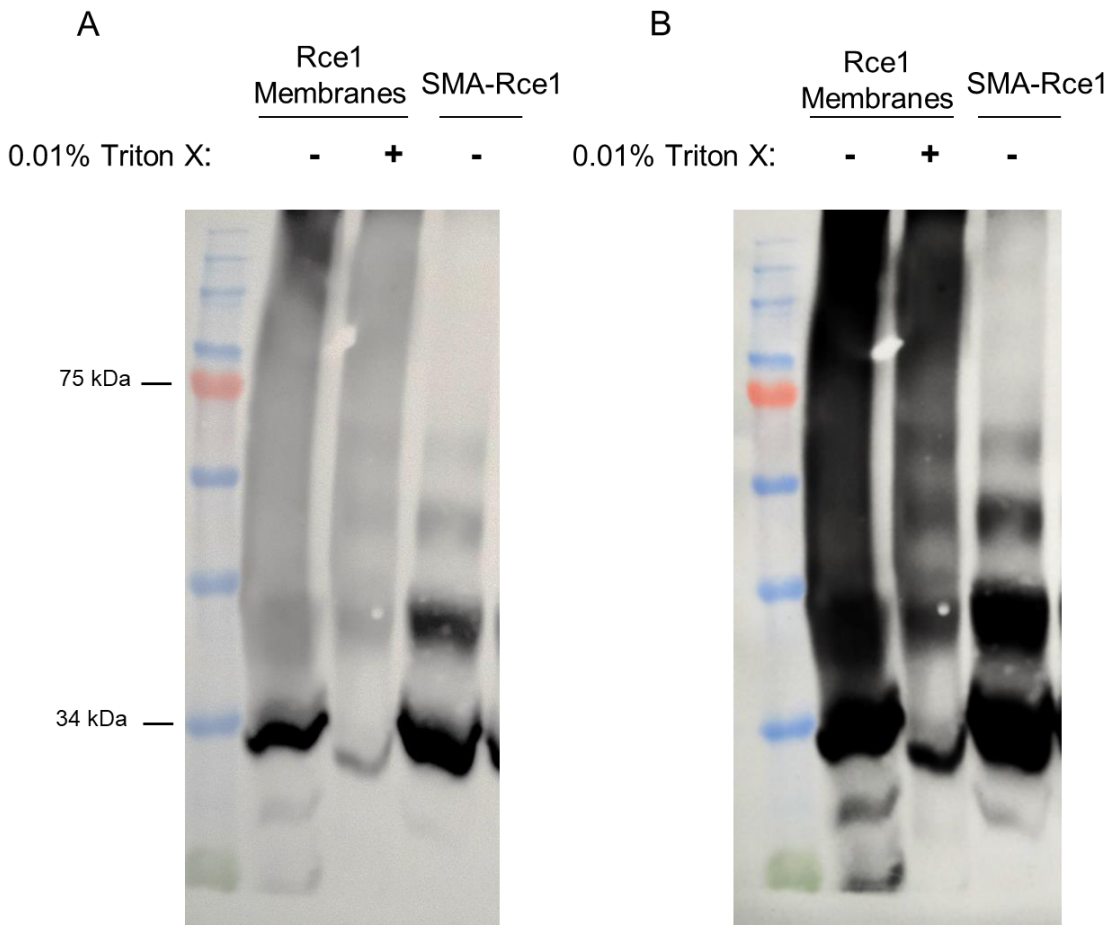


Fig A.5: Peg-Maleimide treatment of Rce1 shows SMA-Rce1 is susceptible to additional labeling. Membranes expressing Rce1 and purified SMA-Rce1 were incubated with 5mM Peg-Maleimide and the samples evaluated by immunoblot. Cysteines found in membrane bound Rce1 are buried, with only 1-2 bands present: unlabeled Rce1 (~35 kDa, HA tag ~40 kDa). Addition of detergents results in loss of Rce1 structural integrity, leading to additional Cysteine labeling and banding pattern with multiple bands visible. Two exposures of the same immunoblot are shown. A) Short exposure shows only one band present for membrane-bound Rce1 in the absence of detergent, as well as faint laddering pattern in membrane Rce1 + detergent and SMA-Rce1 B) Longer exposure showing cysteine labeling for both membrane Rce1 + detergent and SMA-Rce1.

Table SA.1 Rce1 Plasmids used in this study

Identifier	Genotype	Reference
pWS127	*2μ P_{PGK} HIS::HA::RCE1 (L177I)*	This study
pWS374	*2μ P_{PGK} HIS::HA::FXa::Hs Rce1Δ22*	This study
pWS636	*2μ P_{PGK} RCE1::GST*	This study

References

1. Gao J, Liao J, Yang GY. Caax-box protein, prenylation process and carcinogenesis. Am J Transl Res. 2009;1(3):312-25.

2. The ras initiative: National Cancer Institute; [Available from: https://www.cancer.gov/research/key-initiatives/ras.

3. Der JHJ, Cochrane CG, Bourne JR, Solski PA, Buss JE, C J. Farnesol modification of kirsten-ras exon 4b protein is essential for transformation. 1990.

4. Kim E, Ambroziak P, Otto JC, Taylor B, Ashby M, Shannon K, et al. Disruption of the mouse rce1 gene results in defective ras processing and mislocalization of ras within cells. The Journal of biological chemistry. 1999;274(13):8383-90.

5. Michaelson D, Ali W, Chiu VK, Bergo M, Silletti J, Wright L, et al. Postprenylation caax processing is required for proper localization of ras but not rho gtpases. Molecular biology of the cell. 2005;16(4):1606-16.

6. Pai DBW, Paul K, Tish NH, Irma N-O, Linda J, Joseph JC, et al. K- and n-ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors. 1997.

7. Winter-Vann AM, Casey PJ. Post-prenylation-processing enzymes as new targets in oncogenesis. Nature reviews Cancer. 2005;5(5):405-12.

8. Boyartchuk VL, Ashby M, N. , Jasper R. Modulation of ras and a-factor function by carboxyl-terminal proteolysis. 1997.

9. Bergo MO, Ambroziak P, Gregory C, George A, Otto JC, Kim E, et al. Absence of the caax endoprotease rce1: Effects on cell growth and transformation. Molecular and cellular biology. 2002;22(1):171-81.

10. Mohammed I, Hampton SE, Ashall L, Hildebrandt ER, Kutlik RA, Manandhar SP, et al. 8-hydroxyquinoline-based inhibitors of the rce1 protease disrupt ras membrane localization in human cells. Bioorganic & medicinal chemistry. 2016;24(2):160-78.

11. Wahlstrom AM, Cutts BA, Karlsson C, Andersson KM, Liu M, Sjogren AK, et al. Rce1 deficiency accelerates the development of k-ras-induced myeloproliferative disease. Blood. 2007;109(2):763-8.

12. Bergo MO, Lieu HD, Gavino BJ, Ambroziak P, Otto JC, Casey PJ, et al. On the physiological importance of endoproteolysis of caax proteins: Heart-specific rce1 knockout mice develop a lethal cardiomyopathy. The Journal of biological chemistry. 2004;279(6):4729-36.

13. Christiansen JR, Kolandaivelu S, Bergo MO, Ramamurthy V. Ras-converting enzyme 1-mediated endoproteolysis is required for trafficking of rod phosphodiesterase 6 to photoreceptor outer segments. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(21):8862-6.

14. Karlsson C, Akula MK, Staffas A, Cisowski J, Sayin VI, Ibrahim MX, et al. Knockout of the ras endoprotease rce1 accelerates myeloid leukemia by downregulating gadd45b. Leukemia. 2021;35(2):606-9.

15. Schmidt WK, Tam A, Fujimura-Kamada K, Michaelis S. Endoplasmic reticulum membrane localization of rce1p and ste24p, yeast proteases involved in carboxyl-terminal caax protein processing and amino-terminal a-factor cleavage. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(19):11175-80.

16. Jang GF, Gelb MH. Substrate specificity of mammalian prenyl protein-specific endoprotease activity. Biochemistry. 1998;37(13):4473-81.

17. Manolaridis I, Kulkarni K, Dodd RB, Ogasawara S, Zhang Z, Bineva G, et al. Mechanism of farnesylated caax protein processing by the intramembrane protease rce1. Nature. 2013;504(7479):301-5.

18. Hildebrandt ER, Davis DM, Deaton J, Krishnankutty RK, Lilla E, Schmidt WK. Topology of the yeast ras converting enzyme as inferred from cysteine accessibility studies. Biochemistry. 2013;52(38):6601-14.

19. Ma YT, Gilbert BA, Rando RR. Inhibitors of the isoprenylated protein endoprotease. Biochemistry. 1993;32(9):2386-93.

20. Chen Y, Ma YT, Rando RR. Solubilization, partial purification, and affinity labeling of the membrane-bound isoprenylated protein endoprotease. Biochemistry. 1996;35(10):3227-37.

21. Plummer LJ, Hildebrandt ER, Porter SB, Rogers VA, McCracken J, Schmidt WK. Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues. The Journal of biological chemistry. 2006;281(8):4596-605.

22. Pei J, Grishin NV. Type ii caax prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. Trends in Biochemical Sciences. 2001;26(5):275-7.

23. Hardy D, Desuzinges Mandon E, Rothnie A, Jawhari A. The yin and yang of solubilization and stabilization for wild-type and full-length membrane protein. Methods (San Diego, Calif). 2018.

24. Nikolaev M, Round E, Gushchin I, Polovinkin V, Balandin T, Kuzmichev P, et al. Integral membrane proteins can be crystallized directly from nanodiscs. Crystal Growth & Design. 2017;17(3):945-8.

25. Parmar M, Rawson S, Scarff CA, Goldman A, Dafforn TR, Muench SP, et al. Using a smalp platform to determine a sub-nm single particle cryo-em membrane protein structure. Biochimica et biophysica acta. 2018;1860(2):378-83.

26. Denisov IG, Sligar SG. Nanodiscs in membrane biochemistry and biophysics. Chemical reviews. 2017;117(6):4669-713.

27. Smirnova IA, Sjostrand D, Li F, Bjorck M, Schafer J, Ostbye H, et al. Isolation of yeast complex iv in native lipid nanodiscs. Biochimica et biophysica acta. 2016;1858(12):2984-92.

28. Lee SC, Knowles TJ, Postis VL, Jamshad M, Parslow RA, Lin YP, et al. A method for detergent-free isolation of membrane proteins in their local lipid environment. Nature protocols. 2016;11(7):1149-62.

29. Swainsbury DJK, Scheidelaar S, Foster N, van Grondelle R, Killian JA, Jones MR. The effectiveness of styrene-maleic acid (sma) copolymers for solubilisation of integral membrane proteins from sma-accessible and sma-resistant membranes. Biochimica et biophysica acta. 2017;1859(10):2133-43.

30. Knowles TJ, Finka R, Smith C, Lin YP, Dafforn T, Overduin M. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. Journal of the American Chemical Society. 2009;131(22):7484-5.

31. Dorr JM, Koorengevel MC, Schafer M, Prokofyev AV, Scheidelaar S, van der Cruijsen EA, et al. Detergent-free isolation, characterization, and functional reconstitution of a tetrameric k+ channel: The power of native nanodiscs. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(52):18607-12.

32. Tam A, Nouvet FJ, Fujimura-Kamada K, Slunt H, Sisodia SS, Michaelis S. Dual roles for ste24p in yeast a-factor maturation: Nh2-terminal proteolysis and cooh-terminal caax processing. The Journal of cell biology. 1998;142(3):635-49.

33. Elble R. A simple and efficient procedure for transformation of yeasts. BioTechniques. 1992;13(1):18-20.

34. Berger BM, Kim JH, Hildebrandt ER, Davis IC, Morgan MC, Hougland JL, et al. Protein isoprenylation in yeast targets cooh-terminal sequences not adhering to the caax consensus. Genetics. 2018;210(4):1301-16.

35. Oldenburg KR, Vo KT, Michaelis S, Paddon C. Recombination-mediated pcr-directed plasmid construction in vivo in yeast. Nucleic Acids Res. 1997;25(2):451-2.

36. Hildebrandt ER, Cheng M, Zhao P, Kim JH, Wells L, Schmidt WK. A shunt pathway limits the caax processing of hsp40 ydj1p and regulates ydj1p-dependent phenotypes. eLife. 2016;5.

37. Tam A, Schmidt WK, Michaelis S. The multispanning membrane protein ste24p catalyzes caax proteolysis and nh2-terminal processing of the yeast a-factor precursor. The Journal of biological chemistry. 2001;276(50):46798-806.

38. Nishii W, Muramatsu T, Kuchino Y, Yokoyama S, Takahashi K. Partial purification and characterization of a caax-motif-specific protease from bovine brain using a novel fluorometric assay. Journal of biochemistry. 1997;122(2):402-8.

39. Rehan S, Paavilainen VO, Jaakola VP. Functional reconstitution of human equilibrative nucleoside transporter-1 into styrene maleic acid co-polymer lipid particles. Biochimica et biophysica acta. 2017;1859(5):1059-65.

40. Hildebrandt ER, Arachea BT, Wiener MC, Schmidt WK. Ste24p mediates proteolysis of both isoprenylated and non-prenylated oligopeptides. The Journal of biological chemistry. 2016;291(27):14185-98.

41. Oluwole AO, Danielczak B, Meister A, Babalola JO, Vargas C, Keller S. Solubilization of membrane proteins into functional lipid-bilayer nanodiscs using a diisobutylene/maleic acid copolymer. Angewandte Chemie (International ed in English). 2017;56(7):1919-24.

42. Raimondi S, Zanni E, Amaretti A, Palleschi C, Uccelletti D, Rossi M. Thermal adaptability of kluyveromyces marxianus in recombinant protein production. Microbial Cell Factories. 2013;12(1):34.