STRUCTURAL AND FUNCTIONAL STUDIES OF THE DHHC9 SUBFAMILY OF PROTEIN ACYLTRANSFERASES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Phillip Lam Nguyen

May 2022

Structural and Functional Studies of the DHHC9 Subfamily of Protein Acyltransferases

Phillip Lam Nguyen, Ph. D.

Cornell University 2022

Protein S-acylation is a reversible lipid post-translational modification that allows dynamic regulation of processes such as protein stability, membrane association, and localization. Palmitoyltransferase ZDHHC9 (DHHC9) is one of the 23 human DHHC acyltransferases that catalyze protein S-acylation. Dysregulation of DHHC9 is associated with X-linked intellectual disability and increased epilepsy risk. Interestingly, activation of DHHC9 requires an accessory protein—GCP16. However, the exact role of GCP16 and the prevalence of a requirement for accessory proteins among other DHHC proteins remain unclear. Here, I report that one role of GCP16 is to stabilize DHHC9 by preventing its aggregation through formation of a protein complex. Using a combination of size-exclusion chromatography and palmitoyl acyltransferase assays, I demonstrate that only properly folded DHHC9-GCP16 complex is enzymatically active in vitro. Additionally, the ZDHHC9 mutations linked to X-linked intellectual disability result in reduced protein stability and DHHC9-GCP16 complex formation. Notably, I discovered that the C-terminal cysteine motif (CCM) that is conserved among the DHHC9 subfamily (DHHC14, -18, -5, and -8) is required for DHHC9 and GCP16 complex formation and activity in vitro. Co-expression of GCP16 with DHHCs containing the CCM improves DHHC protein stability. Like DHHC9, DHHC14 and DHHC18 require GCP16 for their enzymatic activity. Furthermore, GOLGA7B, an accessory protein with 75% sequence identity to GCP16, improves protein stability of DHHC5 and DHHC8, but not the other members of the DHHC9 subfamily, suggesting selectivity in accessory protein interactions. Additionally, I present my work towards a structural

basis for DHHC9, GCP16, and the DHHC9-GCP16 complex. My study supports a broader role for GCP16 and GOLGA7B in the function of human DHHCs.

BIOGRAPHICAL SKETCH

Phillip Lam Nguyen was born on May 21st, 1992, in Fountain Valley, California, son to Bao Quoc Nguyen and Nga Ngoc Lam. After completing high school in his hometown, he attended Orange Coast College, where worked as an organic chemistry stockroom technician and as a student instructor of mathematics for non-traditional students. His experiences with chemistry and teaching motivated him to continue his education at the University of California, Los Angeles (UCLA), where he earned his Bachelor of Science degree in Biochemistry. At UCLA, in the lab of Dr. David Dawson, he researched signaling pathways that contributed to pancreatic cancer adenocarcinoma. Additionally, he was a project coordinator and researcher for UCLA's team for the International Genetically-Engineered Machines (iGEM) competition, where he worked with Dr. Sri Kosuri and Dr. Todd Yeates. With iGEM, he worked to develop targeted-drug delivery systems using protein cages and biomaterials using synthetic spider silks. Phillip continued his graduate studies at Cornell University under the mentorship of Dr. Toshi Kawate and Dr. Maurine Linder. For his PhD, he studied the structure and function of several integral membrane proteins associated with disease.

ACKNOWLEDGMENTS

I thank Toshi Kawate and Maurine Linder for all their scientific support and guidance. I thank past and present members of the Kawate Lab and Linder Lab for being a pleasure to work with. I thank my family and friends who have offered support. Most importantly, I thank my wife, Shin-Ning, for providing a happy home environment and sharing my happiness.

TABLE OF CONTENTS

List of Figures ix

List of Tables xii

List of Abbreviations xiii

Preface 1

Chapter 1: General Introduction and Literature Review 2

Protein S-acylation 2

Discovery of DHHC Protein Acyltransferases 4

Structural Overview of DHHC Protein Acyltransferases 5

Mechanism of Catalysis 9

DHHC Proteins and Substrate Interactions 11

DHHC Protein Regulation by Accessory Proteins 13

S-acylation and DHHC Proteins in Disease 14

DHHC9 and GCP16 15

Remarks 17

Chapter 2: GCP16 Stabilizes the DHHC9 Subfamily of Protein Acyltransferases through a Conserved C-terminal Cysteine Motif 18

Introduction 18

Results 20

GCP16 Stabilizes DHHC9 through a Conserved C-terminal Cysteine Motif 20

DHHC9 CCM is Required for Complex Formation with GCP16 29

DHHC9 Enzymatic Activity for H-Ras Correlates with Protein Folding 31

DHHC9 Disease Mutations Result in Reduced Protein Stability 33

GCP16 and GOLGA7B Stabilize DHHCs in a Subtype-Specific manner 35

Discussion 42
Experimental Procedures 49
Chapter 3: Towards a Structural Basis for the DHHC9-GCP16 Complex 53
Introduction 53
Results 55
The Search for a DHHC9 Structural Candidate 55
Protein Engineering frDHHS9 for Improved Expression and Stability 64
frDHHS9 Protein Purification and Crystallization Challenges 68
Identification of a GCP16 Structure Candidate 77
zfGCP16 Purification and Crystallization 82
Exploring DHHC9-GCP16 Fusion Proteins 88
DHHC9-GCP16 Complex Protein Purification and Crystallization 97
Towards the Structure of DHHC9-GCP16 Using Cryogenic Electron Microscopy 105
Discussion 115
DHHC9 and GCP16 Structure Constructs and Purification 115
Crystallization Trials 118
DHHC9-GCP16 Cryo-EM 119
Experimental Procedures 120
Concluding Remarks 127
References 128

LIST OF FIGURES

Figure 1.1 Structural Overview of DHHC Proteins 8

Figure 2.1 DHHC9 is Stabilized by Co-expression with GCP16 Independent of the Catalytic Cysteine 21

Figure 2.2 Designing DHHC9 Deletion Constructs 24

Figure 2.3 A Conserved Region in the DHHC9 CTD is Required for GCP16-Mediated Stabilization 25

Figure 2.4 Defining the region of the DHHC9 CTD Required for GCP16-Mediated Stabilization 27

Figure 2.5 Cysteines in the DHHC9 CTD are Required for GCP16-Mediated Stabilization 28

Figure 2.6 DHHC9 CTD Cysteines are Required for Complex Formation with GCP16 30

Figure 2.7 DHHC9 Requires Proper Protein Folding for H-Ras Activity 32

Figure 2.8 DHHC9 Disease Mutations Exhibit Reduced Protein Stability 34

Figure 2.9 GCP16 Co-expression Stabilizes DHHC Proteins with the Conserved C-terminal Cysteine Motif 37

Figure 2.10 GOLGA7B Co-expression Stabilizes DHHC Proteins 38

Figure 2.11 DHHC14 and DHHC18 Each Forms a Complex with GCP16 to Confer PAT Activity 40

Figure 2.12 In Silico Model of DHHC9 Suggesting that the CCM is Located Where the α'2 Helix Exists in DHHC20 46

Figure 3.1 DHHC9 Ortholog Screen 59

Figure 3.2 Purification of DHHS9 First Pass Orthologs 61

Figure 3.3 frDHHS9 Detergent Screen 63

Figure 3.4 Designing DHHC9 Deletion Constructs 65

Figure 3.5 Screening frDHHS9 Deletion Constructs 67

Figure 3.6 frDHHS9 N2C2 is Non-Specifically Cleaved by Thrombin Protease 69

Figure 3.7 Purification of frog and human DHHC9 N2C2 Constructs 71

Figure 3.8 frDHHS9 N2C2 Forms Amorphous Quasi Crystals 72

Figure 3.9 CGFP-tagged DHHC9 Consistently Shows a Modest Improvement in Expression 75

Figure 3.10 frDHHS9 Extended C-terminus Constructs Aggregate When Applied to SEC 76

Figure 3.11 GCP16 Ortholog Screen 80

Figure 3.12 Purification of GCP16 Orthologs 81

Figure 3.13 Purification and Crystallization of zfGCP16 83

Figure 3.14 zfGCP16 Cysteine Mutation Analysis 85

Figure 3.15 Purification of zfGCP16 C24S 87

Figure 3.16 huDHHS9-GCP16 Fusion Protein Screen 89

Figure 3.17 Initial Screening of frDHHS9 N2-zfGCP16 Fusion Protein 91

Figure 3.18 Purification of frDHHS9 N2-zfGCP16 Fusion Protein 94

Figure 3.19 Purification of human DHHS9 N2C1-GCP16 Fusion Protein 96

Figure 3.20 Dual-Affinity Purification of the DHHC9-GCP16 Complex 98

Figure 3.21 Single-Step Strep-Affinity Purification Yields the Most DHHC9-GCP16 Complex 100

Figure 3.22 Protein Concentration Prior to SEC Affects DHHC9-GCP16 Complex Elution and Protein Molar Ratios 102

Figure 3.23 DHHC9-GCP16 Complex Forms Amorphous Crystals 104

Figure 3.24 Cryo-EM of DHHC9-GCP16 in Nanodiscs Reconstruction #1 107

Figure 3.25 Purification of DHHC9 and DHHC9-GCP16 Complex from HEK GnTI- Cells 109

Figure 3.26 Reconstitution of DHHC9-GCP16 into Lipid Nanodiscs 111

Figure 3.27 Cryo-EM of DHHC9-GCP16 in Nanodiscs Reconstruction #2 114

LIST OF TABLES

Table 2.1 DHHC9 Constructs 23

Table 3.1 DHHC9 Orthologs Percent Identity Matrix 57

Table 3.2 GCP16 Orthologs Percent Identity Matrix 78

Table 3.3 Table of DHHC9 and GCP16 Genes and Codenames 121

LIST OF ABBREVIATIONS

C-terminal Cysteine Motif (CCM)

Effect on Ras Function 2 (Erf2)

cysteine-rich domain (CRD)

Asp-Pro-Gly (DPG)

Thre-Thre-x-Glu (TTxE)

Palmitoyltransferase conserved C-terminus (PaCCT)

ankyrin repeat (AR)

AR domain of DHHC17 (ANK17)

Membrane-associated Guanylate Kinase (MaGUK)

Src-homology 3 (SH3)

selenoprotein K (SelK)

huntington (HTT)

polyglutamine (polyQ)

Interferon-induced transmembrane protein 3 (IFITM3)

X-linked intellectual disability (XLID)

critical micelle concentration (CMC)

cryogenic electron microscopy (cryo-EM)

single-particle analysis (SPA)

Basic Local Alignment Search Tool (BLAST)

fluorescence-detection size-exclusion chromatography (FSEC)

dodecyl-maltoside (DDM)

frog DHHS9 (frDHHS9)

coral snake DHHS9 (csDHHS9)

chameleon DHHS9 (chDHHS9)

zebrafish DHHS9 (zfDHHS9)

decyl-maltoside (C10M)

nonyl-glucoside (C9G)

lauryldimethylamine oxide (LDAO)

full-length (FL)

polyethylene glycol (PEG)

1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoethanolamine (POPE)

1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoglycerol (POPG)

1,2-Distearoyl-sn-glycero-3-phosphoglycerol (DSPG)

two-dimensional (2D)

three-dimensional (3D)

styrene maleic acid (SMA)

protein acyltransferase (PAT)

2-bromopalmitate-DHHC20 (2-BP-DHHC20)

PREFACE

This dissertation consists of three chapters. Chapter 1 is a general introduction and literature review of S-acylation and DHHC protein acyltransferases. Chapter 2 presents my findings that GCP16 stabilizes the DHHC9 subfamily of protein acyltransferases through a conserved C-terminal motif. Chapter 2 is written as a standalone chapter for publication. Chapter 3 describes my work and progress towards solving the structures of DHHC9, GCP16, and the DHHC9-GCP16 complex. The thesis finishes with a brief summary and concluding remarks.

Chapter 1: General Introduction and Literature Review

Cells utilize a wide array of protein modifications to extend the chemical properties of the 20 standard amino acids, dramatically expanding the ways in which proteins regulate cellular processes. Protein lipidation is a unique type of co- or post-translational modification where lipid moieties are covalently attached to nucleophilic side chains of specific amino acid residues, which drastically affects how a protein ultimately functions. The fundamental physicochemical effect of protein lipidation is a local increase in hydrophobicity. In turn, this modulates the function of target proteins by increasing their binding affinity to biological membranes, allowing changes in subcellular localization, affecting protein folding and stability, and modulating protein-protein interactions. While there are at least six types of lipid modifications including fatty acids, isoprenoids, sterols, phospholipids, glycosylphosphatidylinositol (GPI) anchors, and lipid-derived electrophiles (1) the focus of this work is on protein S-acylation and the protein acyltransferases the catalyze this reaction. For reviews on other lipid modifications, please refer to references (1–4).

Protein S-acylation

Over the last half-century, researchers have made significant advances in our understanding of protein S-acylation and its occurrence in biology. In 1971, Stoffyn and Folch-Pi first observed that fatty acids constituted 2.5% by weight of brain white matter proteolipid apoprotein (5). While the authors acknowledged they could not localize these fatty acids in the proteolipids, they showed the lipid mixture was composed mainly of palmitic, stearic, and oleic acids attached via ester linkage (5). In 1979, Schmidt et al. discovered that viral glycoproteins were covalently modified with fatty acids that included palmitate, stearate, and oleate during the viral maturation cycle (6). By 1980, Schlesinger et al. showed evidence that protein S-acylation is a broad cellular activity that is present in membrane-bound and membrane-associating

proteins in mouse and human cultured cell lines (7). For decades, there was a steady growth in the number of S-acylation publications and an increase in the appreciation for the functional importance of this modification, with explosive progress since the development of global S-acylation profiling techniques (8, 9). Today, it is understood that protein S-acylation is a highly conserved post-translational modification found in all known eukaryotic organisms (10). In humans, it is one of the most common posttranslational modifications, with over 4,500 proteins annotated on the SwissPalm database (11).

Chemically, protein S-acylation involves the attachment of long-chain fatty acids at cysteine residues via a thioester linkage. Early lines of evidence cleverly uncovered that the fatty acid attachment occurred on cysteine residues. Magee et al. showed protease digestion and removal of a small, cytoplasmic-facing portion of the Sindbis virus PE2/E2 carboxy terminus containing cysteine residues removed a significant amount of fatty acid ultimately detected (12). Additionally, Kaufman et al. showed HLA-B and -DR heavy chains of human B lymphoblastoid cells could be labeled with exogeneous radioactive palmitic acid (13). The authors localized the signal to a transmembrane region that contained cysteine residues, and showed the release of fatty acids exhibited pH sensitivity and kinetics similar to those of palmitoyl-CoA (13). In many of these experiments, radioactive palmitic acid was used to study S-acylaltion, leading to S-acylation commonly referred to as palmitoylation for many years. While the most common fatty acid modification is palmitate (C16:0), other fatty acids such as stearate (C18:0), oleate (C18:1), myristate (C14:0), palmitoleate (C16:1), arachidonate (C20:4), and eicosapentaenoate (C20:5) have been observed (14–18).

S-acylation can have broad regulatory effects on modified proteins. It can confer membrane binding for peripheral membrane proteins, target proteins to membrane microdomains and subdomains such as lipid rafts, affect protein trafficking and intracellular localization, improve protein stability, facilitate protein interactions, and more (10). One key property that distinguishes S-acylation from other lipid modifications is its reversibility and

dynamic nature. The labile nature of thioester bonds allows S-acylated proteins to be regulated by cycles of S-acylation and deacylation (19). For example, a cycle of palmitoylation and depalmitoylation regulates the localization and activity of palmitoylated Ras isoforms (20–22). Farnesylated H- and N-Ras show weak membrane affinity with Golgi membranes. Ras palmitoylation promotes interaction with Golgi membranes and facilitates stable association and trafficking from the Golgi to the plasma membrane. Depalmitoylation releases Ras from the plasma membrane, and the process of palmitoylation and trafficking from the Golgi to the plasma membrane is repeated. This cycle of palmitoylation and deplamitoylation is thought to regulate the time course and sites of Ras signaling by allowing rapid redistribution of Ras on intracellular membranes (20). While the reversibility of S-acylation allows dynamic regulation of proteins such as Ras isoforms, rapid cycles of acylation and de-acylation are not generalizable for all S-acylated proteins. Global profiling of dynamic protein palmitoylation suggested some S-acylation events turn over at very low rates or not at all within the contexts studied (9). It is an ongoing process to better understand the functional consequences of S-acylation in the context of individual proteins and their effects on cellular physiology.

Discovery of DHHC Protein Acyltransferases

Although S-acylation was first described in the 1970s (5), it was debated whether this process occurred spontaneously or through enzymatic catalysis. Non-enzymatic S-acylation, often called autoacylation, can occur on a number of proteins including G protein α subunits, myelin proteolipid protein, myelin P_0 glycoprotein, the G-protein-coupled receptor (GPCR) rhodopsin, and the transport protein particle component Bet3 (23–27). Confusingly, one study with $G_{i\alpha1}$ showed the *in vitro* reaction occurred on the same residue that is palmitoylated *in vivo* and the reaction proceeded to completion at a rate close to the rate of palmitate loss observed in pulse-chase experiments *in vivo* (23). These observations showed that S-acylation could occur, at least in these proteins, without the need for an enzyme.

Thirty years after S-acylation was first described, the enzymes that catalyze S-acylation were discovered. By 1999, it was known that *S. cerevisiae* contained two Ras proteins (Ras1 and Ras2) that affected diverse processes related to cell growth (28). Ras-dependent growth requires plasma membrane localization, which depends on a series of posttranslational modifications on a carboxyl-terminal CaaX box where C is cysteine, a is an aliphatic residue, and X is the terminal residue (29–31). First, protein farnesyl transferase attaches a farnesyl group to the cysteine residue of the CaaX box. Second, farnesylation targets Ras to the endoplasmic reticulum, where an endopeptidase removes the -aaX peptide and the farnesylated carboxy-terminus is methylated by a methyltransferase. However, prenylation alone is insufficient for efficient plasma membrane targeting of Ras. Yeast that fail to palmitoylate Ras have reduced Ras at the plasma membrane (32). Through genetic studies in yeast, Deschenes and colleagues identified the first protein acyltransferase (PAT), a complex of Erf2 (effect on Ras function 2), and a second protein affecting Ras palmitoylation, Erf4 (33). A follow up study showed Erf2 and Erf4 form an endoplasmic reticulum-associate complex required for Ras plasma membrane localization in yeast (34). Most importantly, purified Erf2-Erf4 complex was shown to catalyze the transfer of radiolabeled palmitate from palmitoyl CoA to the substrate Ras2 *in vitro*, demonstrating that Erf2-Er4 complex is a *bona fide* PAT (35). Concurrently, Roth et al. identified Akr1p as a PAT for its substrate Yck2p (36). Together, Erf2 and Akr1p represent the founding members of the DHHC PAT family whose family members range in number from 5-7 in yeast to 23 in humans (37).

Structural Overview of DHHC Protein Acyltransferases

DHHC proteins are multipass transmembrane proteins aptly named for the conserved Asp-His-His-Cys motif embedded in a cysteine-rich domain (CRD) involved in zinc ion binding (38–40). In addition to the DHHC-CRD, other conserved motifs have been identified such as the Asp-Pro-Gly (DPG) motif, Thr-Thr-x-Glu (TTxE), and palmitoyltransferase conserved C-terminus

(PaCCT) (37, 41). Outside the DHHC-CRD, DHHC proteins show considerable sequence variation, especially in the N- and C-terminal domains. DHHC proteins range in size from ~263 amino acid residues (DHHC22) to over ~750 amino acids (DHHC8). In humans, most family members possess a core membrane topology with at least four transmembrane domains (42, 43). Some members with variations are DHHC13, DHHC17, and DHHC23, which are predicted to have six transmembrane domains, and DHHC4 and DHHC24, which are predicted to have five transmembrane domains. Both DHHC4 and DHHC24 are interesting cases, as they imply one of the termini faces the luminal or extracellular space. Lastly, DHHC22 is predicted to have only two transmembrane domains.

In 2018, Rana et al. solved the first atomic structures of two DHHC PATs, human DHHC20 (hDHHC20) and a catalytically inactive mutant of zebrafish DHHC15 (zfDHHS15) (44). These structures represent DHHC proteins with a core four transmembrane topology. The structures revealed that the transmembrane helices form a tepee-like arrangement, with the DHHC-CRD and the C-terminus oriented towards the cytoplasm (Fig. 1.1). The structures showed that the C-terminal domain interacts extensively with both the DHHC-CRD and the transmembrane domains. Residues in the C-terminal TTxE and PaCCT motifs make crucial contacts that provide local stability. Interestingly, the second threonine of the TTxE motif directly interacts with the aspartic acid of the DHHC motif, though the exact chemical role of this contact in catalysis remains unclear, as DHHC3 and DHHC7 have TGIE and TEIE, respectively, but retain enzymatic activity (44–46). The glutamate of TTxE forms a salt-bridge with a conserved residue arginine in the DHHC-CRD. The conserved Asp of the PaCCT motif forms extensive hydrogen bonds with its neighboring residues and is important for structural integrity of the enzyme (44). The structure also showed two zinc ions coordinated by CCHC zinc-finger domains in the DHHC-CRD to likely serve a structural role, supporting prior studies on the role of zinc binding (38, 39, 44). The overall DHHC-CRD precisely positions of the DHHC catalytic cysteine close to the membrane to cytoplasm interface; this explains why many S-acylated cysteines are located close to the membrane (11,

47, 48). The authors also solved the structure hDHHC20 irreversibly inhibited by 2-bromopalmitate, which allowed the identification of the lipid binding cavity (44). Above the active site, the four transmembrane helices form a cavity where the fatty acyl chain binds (44). Even though these three structures represent two DHHC members from two different organisms and were obtained using different crystallization conditions (lipidic cubic phase vs. hanging drop vapor diffusion), they are almost identical in overall architecture (44). This suggests that these structural features are likely shared by other members of the DHHC family, at least with the core four transmembrane topology.

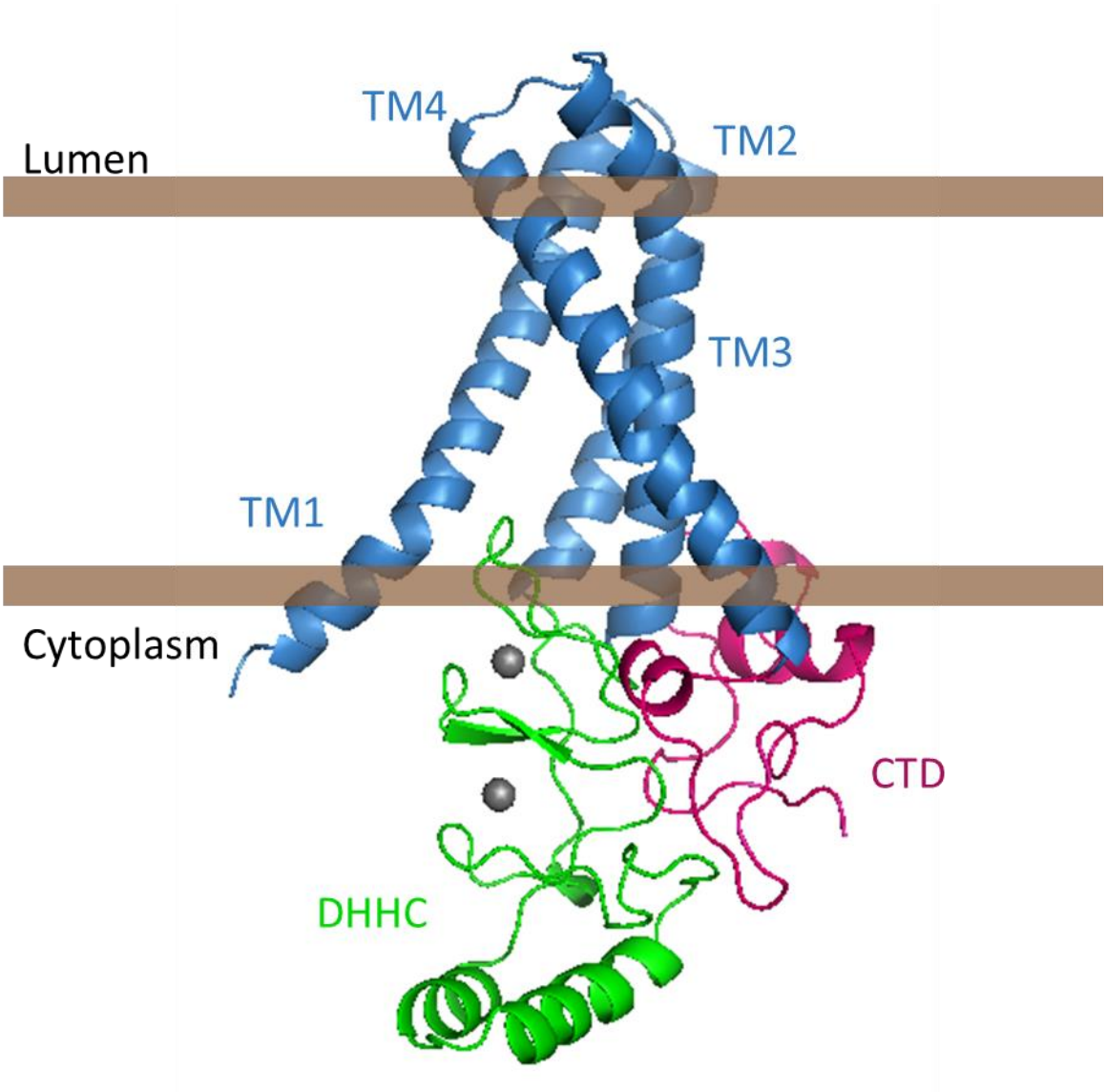

Figure 1.1. Structural Overview of DHHC Proteins. Cartoon representation of human DHHC20 (PDB ID: 6BML). The four transmembrane helices (TM) are shown in blue, the DHHC cystine-rich domain (DHHC) is shown in green, and the C-terminal domain (CTD) is shown in magenta. Zn^{2+} ions are shown as grey spheres.

Mechanism of Catalysis

Researchers using yeast Erf2-Er4 and mammalian DHHC2 and DHHC3 revealed that DHHC proteins utilize a two-step catalytic mechanism (45, 49). In the first step, the enzyme uses acyl-CoA as an acyl donor to form an acyl-enzyme intermediate. In the second step, palmitate is transferred to the substrate protein. Early experiments with Erf2-Er4 and Akr1p showed mutation of the DHHC cysteine prevents both autoacylation and substrate transfer, supporting that the DHHC cysteine is required for catalysis (35, 36). With the structure of hDHHC20, the catalytic triad was finally visualized, and a catalytic mechanism was proposed. DHHC enzymes use a catalytic triad consisting of residues in the DHHC domain (Asp153, His154, and Cys156 in hDHHC20) (47). The triad residues are arranged sequentially and linearly, compared to the three dimensional arrangements seen in the catalytic triad in serine proteases (50) and cysteine proteases (51). Nonetheless, the geometry is overall conserved. The distances between Cys-His and His-Asp pair are ~4 and 3 angstroms, respectively (47). In the catalytic mechanism, His154 is polarized by Asp153 to act as a base in removing a proton from Cys156, generating a thiolate nucleophile. The Cys156 thiolate then attacks the carbonyl carbon of the fatty acyl-CoA thioester, forming the autoacylated DHHC enzyme intermediate. Then, the fatty acyl chain is transferred to a protein substrate, regenerating the DHHC enzyme. This proposed mechanism requires the catalytic cysteine to act as a thiolate nucleophile. However, in the hDHHC20 structure, this cysteine is within the hydrophobic transmembrane region of the protein (44). To explain this observation, Stix et al. used molecular dynamics simulations to propose that hDHHC20 induces a local deformation of the membrane where the catalytic cysteine resides, allowing it to access water and form a thiolate nucleophile (52).

While aspects of above mechanism are likely used by all DHHC proteins, previous biochemical studies showed variations in the requirement for the DHHC motif. Gonzales et al. showed yeast Swf1 and Pfa4 with the catalytic cysteine mutated to arginine retained activity (53). One proposed explanation is that the reaction proceeds through a ternary complex with

the DHHC protein, acyl-CoA, and protein substrate, rather than a ping-pong mechanism (47). This result suggests S-acylation may not necessarily require an autoacylated intermediate, and it was proposed to explain why several groups have been unable to detect autoacylated DHHC proteins including DHHC13, 19, 22, and 17 (54–56). Additionally, DHHC13 contains a DQHC motif, rather than a DHHC motif. Though DHHC13 can tightly bind S-acylated proteins SNAP25 and CSP, mutation of DHHC13 DQHC to DHHC did not enable DHHC13 to S-acylate the two proteins; however, mutation of DHHC17 DHHC to DHQC abolished its activity for SNAP25 and CSP (55). These studies show that DHHC and DHQC are not interchangeable. Along those lines, Mitchell et al. showed yeast Erf2-Erf4 H201A with a DAHC mutation could autopalmitoylate but not transfer palmitate (49). Together, these lines of evidence suggest there may be differences among more diversified DHHC proteins.

DHHC proteins display distinct fatty acyl chain selectivity. In 2012, Jennings and Linder used a competition assay where nonradiolabeled acyl-CoAs of different chain lengths and saturations were used to compete with radiolabeled palmitate for acylation of enzyme and substrate (45). They observed a difference in acyl-CoA specificity between DHHC3 and DHHC2 where DHHC2 could utilize acyl chains 14 carbons and longer, whereas DHHC3 selectively used 14-16 carbon acyl chains (45). This study further showed that fatty acyl CoA selection was determined at the autoacylation stage, as the reduction of radiolabeled palmitate incorporation into substrate was determined by whether it was incorporated into the acyl-enzyme intermediate (45). In a follow up study, Greaves et al. used a cell-based S-acylation assay with fatty acid-azide/alkyne labeling to demonstrate marked differences in the fatty acid selectivity of the DHHC enzymes including DHHC2, 3, 5, 7, 11, 15, and 17 (46). Strikingly, this difference in selectivity was present in two highly related enzymes, DHHC3 and DHHC7. DHHC7 had a greater ability to incorporate C18:0 chains compared to DHHC3 (46). Through domain swapping analysis, the authors identified a single amino acid, Ile182, on transmembrane domain 3 of DHHC3; when this isoleucine was mutated to serine, DHHC3 was

more able to incorporate C18:0 chains (46). The structural basis of acyl chain selectivity was revealed in the structure of hDHHC20 irreversibly inhibited with 2-bromopalmitate (44). The fatty acid chain is inserted into the cavity formed by the four transmembrane helices, with transmembrane domain 3 contributing the most residues in contact with the fatty acid chain. The homologous residue to Ile182 in DHHC3 is Tyr181 in hDHHC20, which forms a hydrogen bond with Ser29 on transmembrane domain 1. Together, the Tyr-Ser hydrogen bond pair closes off the end of the fatty acid chain binding cavity. Mutation of Ser29 to a bulky side-chain results in hDHHC20 preferring shorter fatty acid chains whereas mutation of Tyr181 to a small amino acid results in DHHC20 preferring longer fatty acid chains (44). One consideration is that different DHHCs contain different residues in transmembrane domain 3, and different membranes can be enriched for different lipids. Therefore, high-resolution structures of additional DHHC enzyme family members and thorough functional analyses will be required for a more complete understanding of DHHC fatty acyl selectivity.

DHHC Proteins and Substrate Interactions

In 2006, Roth et al. performed proteomic analysis to determine how individual DHHC proteins affected cellular S-acylation in yeast (57). In yeast strains deficient of the palmitoyltransferase Ark1, palmitoylation of Yck1, Yck2, Meh1, Yp199c, and Yk1047w were decreased or abolished (57). Depletion of Erf2 resulted in reduced, but not depleted, levels of Ras palmitoylation (57). While the palmitoylation of some S-acylated proteins were dependent on specific DHHC proteins, knockout of individual DHHCs showed only a modest effect on global cellular palmitoylation. The authors also noted that Akr1p targets were soluble proteins that are exclusively palmitoylated, while Erf2 targets were modified by other lipid groups (57). Overall, this study showed clear evidence of DHHC protein substrate specificity, as well as varying degrees of substrate overlap.

In humans, the number of S-acylated proteins currently exceeds 4500, while there are

only 23 DHHC proteins (11). This suggests DHHC proteins are very likely to act on multiple substrates. In 2006, Fukata et al used a systematic screening approach to identify specific DHHC protein-substrate pairs in mammalian cells (58). By overexpressing DHHC proteins with specific substrate pairs and detecting the levels of palmitoylated substrate, the authors identified specific DHHC proteins that increased the levels of palmitoylated substrate (58). This approach was complemented with knockdown studies, where DHHC protein knockdown was used to observe the effect on substrate palmitoylation (58). Many mammalian DHHC protein-substrate pairs were largely identified based on results from co-overexpression studies. Overall, these studies have shown that DHHC proteins S-acylate a broad range of substrates, with varying degrees of selectivity and substrate overlap (59). However, it is worth noting that conclusions based on overexpression studies should be validated orthogonally, as overexpression of membrane proteins can cause non-physiological interactions and membrane localizations. Currently, there is no clear indication that a universal mechanism drives how DHHC proteins recognize specific substrates. Nonetheless, I will highlight a few cases of DHHC protein-substrate interactions and S-acylation.

While the DHHC-CRD is responsible for catalyzing S-acylation, the N- and C-terminal domains of DHHC proteins enable substrate-specific and other protein-protein interactions. Several DHHC proteins contain modular protein binding domains that mediate interactions with substrates and other proteins. DHHC17 and DHHC13 contain ankyrin repeat (AR) domains in their N-termini. Lemonidis et al. identified an evolutionarily conserved sequence [VIAP][VIT]xxQP in SNAP23, cysteine string protein, Huntingtin, cytoplasmic linker protein 3, and microtuble-associated protein 6 that is recognized by the DHHC17 and DHHC13 AR domain (60). In a follow up study, Verardi et al. solved the structure of the AR domain of DHHC17 (ANK17) bound to a Snap25b peptide fragment, revealing the ANK17 domain forms a concave shape with a conserved Trp residue forming crucial contacts dipeptide “QP” proline residue in the substrate (56). Another substrate interaction motif is the PDZ binding motif found

in the C-terminus of human DHHC3, 5, 8, 7, 14, 16, 17, 20, and 21. The PDZ binding domain is required for DHHC8 to bind to and palmitoylate PICK1 (61) and for DHHC5 and DHHC8 to bind and palmitoylate GRIP1b (62). DHHC14 contains a Type-I PDZ domain, which is required for its direct interaction with the PDZ domaining-containing Membrane-associated Guanylate Kinase (MaGUK) PSD93 and its palmitoylation of PSD93 and Kv1 potassium channels (63). Finally, DHHC6 contains a Src-homology 3 (SH3) domain, which is required for its interaction with selenoprotein K (SelK) in the ER membrane; together, DHHC6/SelK palmitoylate the Inositol 1,4,5-triphosphate receptor, IP3R (64).

DHHC Protein Regulation by Accessory Proteins

Most DHHC proteins are thought to be a functional enzyme on their own, and, in some cases, DHHCs have been observed to form a complex of homomers that regulate their activity (65, 66). However, an increasing amount of evidence suggests that several DHHC proteins are regulated by accessory proteins that control their activity, stability, and localization (67).

The yeast palmitoyltransferase, Erf2 and its accessory protein, Erf4, form a protein complex that is responsible for Ras palmitoylation (33–35). Mitchell et al. showed, in the absence of Erf4, Erf2 steady-state levels and half-life are decreased, suggesting a role for Erf4 in Erf2 stability (68). However, maintaining Erf2 stability is not the only role for Erf4, as yeast with a stable Erf2 mutant still show growth defects (68). The authors used a coupled-fluorescence-based autoacylation assay to show that Erf2, in the absence of Erf4, exhibited rapid hydrolysis of the acyl-Erf2 intermediate; this result supports that Erf4 plays a role in catalysis by shielding the active site from water in order to prevent a futile cycle of Erf2 autoacylation and hydrolysis (68). Lastly, Erf4 was also shown to be critical for acyl transfer to Erf2 substrate Ras2, although the exact mechanism is to be determined (68). Based on homology to Erf2-Erf4, the mammalian homologues DHHC9-GCP16 were discovered (69). DHHC9 and GCP16 are major focuses of this thesis work. They will be revisited at the end of

this literature review, and their functional relationship will be explored in later chapters.

Recently, DHHC5 was also shown to be regulated by GCP16 and its related paralog, GOLGA7B. Ko et al. reported a functional interaction between GOLGA7 (GCP16) and DHHC5 in the context of nonapoptotic cell death by the small-molecule CIL56 (70). They showed that DHHC5-GCP16 complex was required for CIL56-mediated toxicity, possibly by inhibiting anterograde Golgi transport and disruption of trafficking pathways (70). Additionally, a human interactome study reported DHHC5 and GOLGA7B interaction (71). Woodley and Collins studied the role of DHHC5 and GOLGA7B in regulating cell adhesion (72). The authors found that DHHC5 binds to and palmitoylates GOLGA7B, which stabilized DHHC5 at the plasma membrane. GOLGA7B was also found to modulate the DHHC5 interactome, by promoting its association with components of desmosomes and proteins in cell adhesion (72).

DHHC17 protein palmitoylation and activity was shown to be regulated by Huntingtin protein (HTT), which is a key protein that causes Huntington's disease (73). Huntington's disease is an autosomal-dominant neurodegenerative disorder caused by polyglutamine (polyQ) expansion in the HTT protein (74). HTT has been shown to interact with DHHC17 and DHHC13 through their ankyrin-repeat domains, and C214 of HTT is S-acylated. Mutation of the HTT S-acylation site or knockdown of DHHC17 increased inclusion bodies formed by mutant HTT, a hallmark of Huntington's disease (75). In the brains of mice lacking one *HTT* allele or when HTT protein expression is knocked down in cortical neurons, DHHC17 palmitoylation is decreased and palmitoylation of its substrates SNAP25 and GluR1 are decreased. While this study showed HTT can positively modulate DHHC17 palmitoylation and activity, it remains to be determined whether this is a direct interaction.

S-acylation and DHHC Proteins in Disease

DHHC proteins catalyze S-acylation on an extensive number of proteins that are key players in physiological processes and disease progression. S-acylated proteins include

membrane receptors involved in cell signaling such as G-protein-coupled receptors (GPCRs), proteins in membrane and vesicle trafficking, ion channels and transporters, adapters and chaperones, transcriptional regulators, enzymes and kinases, and cell adhesion proteins (10). It is no surprise that dysregulation of S-acylation and DHHC proteins is associated with diseases such as cancers and neurodegenerative disorders. For examples, DHHC17 and DHHC13, also known as HIP14 and HIP14L, respectively, are associated with Huntington disease, DHHC9 and DHHC15 are associated with X-linked intellectual disability, DHHC8 is linked to Schizophrenia, and DHHC12 is linked to Alzheimer disease (76). Furthermore, many DHHCs including DHHC 2, 3, 7, 9, and 21 are linked to cancers such as colorectal, breast, lung, prostate, ovarian, and leukemia among others (77). Most recently,DHHC20 and DHHC9 S-acylation activity were shown to affect the membrane lipid organization and infectivity of SARS-CoV-2, the viral agent of the COVID-19 global pandemic (78). For a more comprehensive review of DHHC protein substrates and disease associations, the reader is referred to references (10, 76, 77, 79).

DHHC9 and GCP16

In humans, DHHC9 was the first DHHC protein discovered to require an accessory protein—GCP16 (also known as GOLGA7)—for its enzymatic function, based on sequence homology with the yeast Ras PAT Erf2/Er4 (69). DHHC9 shares 31% sequence identity with Erf2, while GCP16 shares 17% sequence identity with Erf4. DHHC9 colocalizes with and requires GCP16 for its enzymatic activity (69). While the exact sequence identity between the pair is limited, they possess functional similarities. For example, both form heteromeric DHHC protein complexes, localize in early secretory pathway membranes, and are putative Ras PATs in their respective organisms (35, 69). GCP16 is a small peripheral membrane protein that is itself palmitoylated, and it has associations with proteins involved in vesicular transport at the Golgi (80). However, it is unclear what the exact role of GCP16 is in DHHC9 protein regulation,

how GCP16 supports DHHC PAT activity, and to what extent GCP16 or other accessory proteins function with DHHC proteins in general.

To date, seven DHHC9 substrates have been identified. The most well-characterized DHHC9 substrate is Ras. DHHC9 was discovered based on its relation to the yeast Ras PAT, and DHHC9 has been shown to act on H-Ras and N-Ras *in vitro* (69). Several follow up studies support DHHC9's role as a human Ras palmitoyltransferase. Shimell et al. showed DHHC9 and N-Ras likely function through the same pathway to promote dendritic cell growth in neuronal cell culture (81). In human fibrosarcoma HT1080 cells, overexpressing *ZDHHC9* promotes, while repressing it diminishes, Ras PM localization (82). *ZDHHC9* silencing reduces oncogenic H-Ras-mediated cell transformation and colony formation in murine NIH 3T3 cells (82). Chai et al. reported microRNA-134 regulates DHHC9-dependent H-Ras localization in SST^+ neurons that is dependent upon the palmitoylated cysteines (83). Using *ZDHHC9* knockout mice, Liu et al. showed Ras palmitoylation is decreased in bone marrow cells (84). Furthermore, ZDHHC9 inactivation inhibits oncogenic N-Ras cellular transformation in hematopoietic cells and slows down the development of myeloid leukemia in N-Ras-induced chronic myelomonocytic leukemia mice models (84). Other DHHC9 substrates are involved in a variety of cellular pathways. DHHC9 S-acylates and contributes to membrane association of stress-regulated exon (STREX) variant of the intracellular C-terminal domain of the large conductance calcium- and voltage-activated potassium (BK) channel (85). DHHC9 S-acylates β2-adrenergic receptor as part of its trafficking to and from the plasma membrane (86). DHHC9 can act on Interferon-induced transmembrane protein 3 (IFITM3), an endosome- and lysosome-localized protein involved in restricting viral infections (87). DHHC9 S-acylates Programmed cell death 1 ligand 1 (PD-L1) to regulate PD-L1 stability in breast cancer cells (88). DHHC9 S-acylates glucose transporter GLUT1 in the promotion of glioblastoma tumorigenesis (89). Lastly, DHHC9 S-acylates SARC-Cov-2 spike protein to allow the formation of viruses with enhanced fusion capacity (78).

Dysregulation of DHHC9 is associated with several diseases. Loss of function mutations

in ZDHHC9 result in X-linked intellectual disability (XLID), with affected individuals displaying neurodevelopmental delay, seizures, and facial dysmorphism (90–93). Furthermore, *ZDHHC9* gene expression is upregulated in many cancers including , myeloma, lung, prostate, ovarian, and gastric cancers (77, 79). Accordingly, DHHC9 has clear biomedical importance and merits further investigation.

Remarks

In my thesis work, I focused on obtaining a better understanding of DHHC9 and GCP16. Additionally, I sought to better understand the role of GCP16 in the DHHC9-GCP16 complex and to determine the extent accessory protein requirements in the context of other DHHC proteins. Chapter 2 discusses my findings that GCP16 stabilizes the DHHC9 subfamily of protein acyltransferases through a conserved C-terminal motif. Chapter 2 is written as a standalone chapter for publication. As of the date of writing this thesis, the manuscript is in revision. Chapter 3 is a report of my efforts towards solving the structure of DHHC9, GCP16, and the DHHC9-GCP16 complex by X-ray crystallography and cryogenic electron microscopy.

Chapter 2: GCP16 Stabilizes the DHHC9 Subfamily of Protein Acyltransferases through a Conserved C-terminal Cysteine Motif

Introduction

Cells utilize a wide array of protein post-translational modifications to extend the chemical properties of the 20 standard amino acids, dramatically expanding the ways in which proteins regulate cellular processes. In eukaryotes, protein S-acylation is the process in which long-chain fatty acids (with palmitate being the most prevalent) are added to proteins at cysteine residues via a labile thioester linkage. While the fundamental chemical effect of S-acylation is a local increase in protein hydrophobicity, the reversible nature of S-acylation can result in a range of consequences such as dynamic changes protein stability, membrane affinity and binding, and protein trafficking and localization (10).

DHHC protein acyltransferases (PATs) are the enzymes responsible for catalyzing the addition of long-chain fatty acids to substrate proteins (42). Members of the DHHC family are conserved throughout eukaryotic evolution, with 5-7 members in yeast and as many as 23 members in humans (37, 94). Furthermore, dysregulation of DHHC proteins are associated with a myriad of diseases that include cancers and neurodegenerative disorders (76, 95). DHHC proteins are aptly named for the conserved Asp-His-His-Cys motif required for their PAT activity, which is embedded in a cysteine-rich domain (CRD) involved in zinc ion binding (38–40). Additionally, all DHHC proteins are multipass transmembrane proteins, with subfamily diversity stemming from varied membrane topologies, sequence divergence in the amino- and carboxy-terminal regions, and protein partner requirements (35, 67, 69). It is generally accepted that DHHC proteins utilize a two-step ping-pong mechanism (45, 49). First, the DHHC protein uses acyl-CoA as an acyl group donor to form an acyl-enzyme intermediate; then, the acyl group is transferred from the DHHC cysteine to the target cysteine on the protein substrate. While this mechanism is thought to be shared by all DHHCs, a majority of mechanistic and structural insight

has been determined from studies on one subfamily of DHHCs that include DHHC2, 3, and 20 (39, 40, 45). Thus, this leaves many other DHHC proteins underrepresented, and potentially subfamily-specific insights are yet to be discovered.

In humans, DHHC9 was the first DHHC protein discovered to require an accessory protein—GCP16 (also known as GOLGA7)—for its enzymatic function, based on sequence homology with the yeast Ras PAT Erf2/Er4 (69). DHHC9 colocalizes with and requires GCP16 for its enzymatic activity (69). GCP16 is a small peripheral membrane protein that is itself palmitoylated, and it has associations with proteins involved in vesicular transport at the Golgi (80). However, it is unclear what the exact role of GCP16 is in DHHC9 protein regulation, how GCP16 supports DHHC PAT activity, and to what extent GCP16 or other accessory proteins function with DHHC proteins in general. Furthermore, loss of function mutations in ZDHHC9 result in X-linked intellectual disability (XLID), with affected individuals displaying neurodevelopmental delay, seizures, and facial dysmorphism (90–93). Despite the strong clinical relevance, the molecular basis for how *ZDHHC9* mutations affect DHHC9 protein and its regulation is incompletely understood. In this study, I sought to better understand the role of GCP16 in the DHHC9-GCP16 PAT complex, and I investigated the stability of the DHHC9 disease mutants. I also assessed whether the accessory proteins are involved with other DHHC proteins.

Results

GCP16 Stabilizes DHHC9 through a Conserved C-terminal Cysteine Motif

My initial attempts to express and purify DHHC9 from Sf9 cells were met with low protein yield, reduced purity, and extensive protein aggregation. This was consistent with a previous report that DHHC9 purified without GCP16 resulted in an increased fraction of DHHC9 protein being proteolyzed (69). Thus, I reasoned that one function of GCP16 is to stabilize DHHC9 and prevent misfolding or aggregation. I used fluorescence-detection size-exclusion chromatography (FSEC) to examine the effect of GCP16 on DHHC9 protein stability in cell lysates (96). DHHC9 fused with N-terminal mNeonGreen was expressed in HEK293 cells with or without GCP16. I chose to fuse the fluorescent protein to the DHHC9 N-terminus to minimize perturbing its C-terminus, based on previous reports that N-terminal modifications to Erf2 had no effect on its stability and that the C-terminus of DHHC proteins are generally thought to mediate protein-protein interactions (68, 97). Cells were solubilized in a dodecylmaltoside (DDM)-containing buffer, and cleared lysates were applied to a gel-filtration column coupled to a fluorometer to monitor fluorescence (Fig. 2.1). DHHC9 expressed alone showed a prominent void peak characteristic of high-molecular weight protein aggregate and only a minor peak at the expected retention time for a well-folded, monomeric protein (Fig. 2.1 A blue trace). By contrast, DHHC9 co-expressed with GCP16 showed improved monodispersity and a 4-fold increase in magnitude (Fig. 2.1 A orange trace, C, D). To determine if the GCP16-mediated stabilizing effect is dependent on the catalytic cysteine, I performed the same assay using DHHS9, in which the catalytic cysteine was mutated to serine. Compared to DHHS9 alone, DHHS9 co-expressed with GCP16 showed an almost identical improvement in monodispersity and expression as DHHC9 (Fig. 2.1 B, C, D). These data suggest that DHHC9 is stabilized by GCP16 co-expression, and this effect is not dependent on DHHC9 catalytic activity.

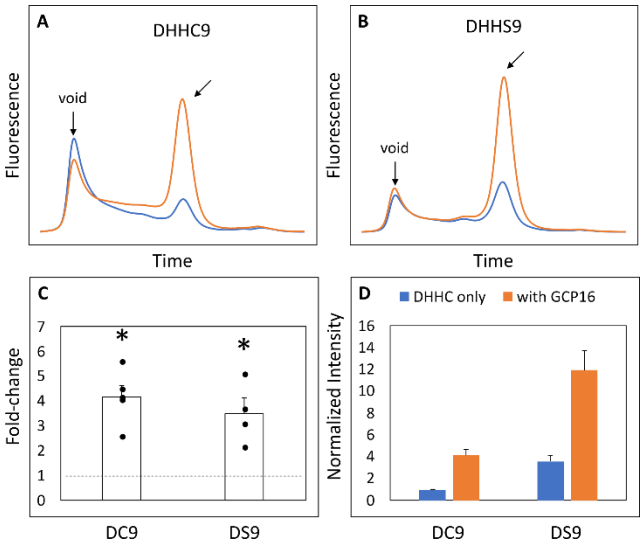

Figure 2.1. DHHC9 is Stabilized by Co-expression with GCP16 Independent of the Catalytic Cysteine. FSEC profiles of (A) DHHC9 and (B) DHHS9. HEK cells were transfected with vector to express the indicated constructs without or with GCP16 co-expression shown with blue and orange traces, respectively. Cells were solubilized in a DDM-containing buffer, and the cleared lysate was analyzed via SEC by detecting fluorescence at excitation and emission wavelengths of 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the approximate retention time corresponding to the size of DHHC9. Bar charts showing (C) average fold-change and (D) normalized max intensity upon GCP16 co-expression for the indicated constructs for n≥4 experiments. Normalization was done as described in the experimental procedures. The dashed line indicates a fold-change of 1. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

Next, I asked which regions and amino acid residues of DHHC9 are required for the GCP16-mediated stabilizing effect. I used evolutionary conservation to generate combinatorial N- and C-terminal deletions of DHHC9 (Table 2.1 and Fig. 2.2) and assessed the effect of GCP16 co-expression for each construct by FSEC (Fig. 2.3). For DHHC9 with N-terminal truncation N2, C-terminal truncation C1, or both, GCP16 co-expression improved protein behavior, similar to the full-length construct (Fig. 2.3 B, C, E, G, H). However, for DHHC9 constructs containing truncation C2, GCP16 co-expression showed no significant difference when compared to DHHC9 alone (Fig. 2.3 D, F). I observed that some individual truncations alter total levels of DHHC9 detected (Fig. 2.3 H). For example, constructs N2 and N2C1 improved DHHC9 protein stability in the absence of GCP16, possibly by removing disordered/aggregation-prone regions of DHHC9 (Fig. 2.3 B, E, G, H). Nonetheless, the addition of GCP16 further increased the stabilizing-effect in constructs that included truncation C1. Thus, DHHC9 residues between truncation C1 and C2 are required for GCP16-mediated stabilization.

Construct	Description
DHHC9 (full-length)	
DHHC9 N2	Δ2-19
DHHC9 C1	Δ301-364
DHHC9 C2	Δ278-364
DHHC9 N2C1	Δ2-19, Δ301-364
DHHC9 N2C1.1	Δ2-19, Δ295-364
DHHC9 N2C1.2	Δ2-19, Δ289-364
DHHC9 N2C1.3	Δ2-19, Δ282-364
DHHC9 N2C2	Δ2-19, Δ278-364

Table 2.1. DHHC9 Constructs. Table denoting the generated constructs (left) and amino acid changes (right).

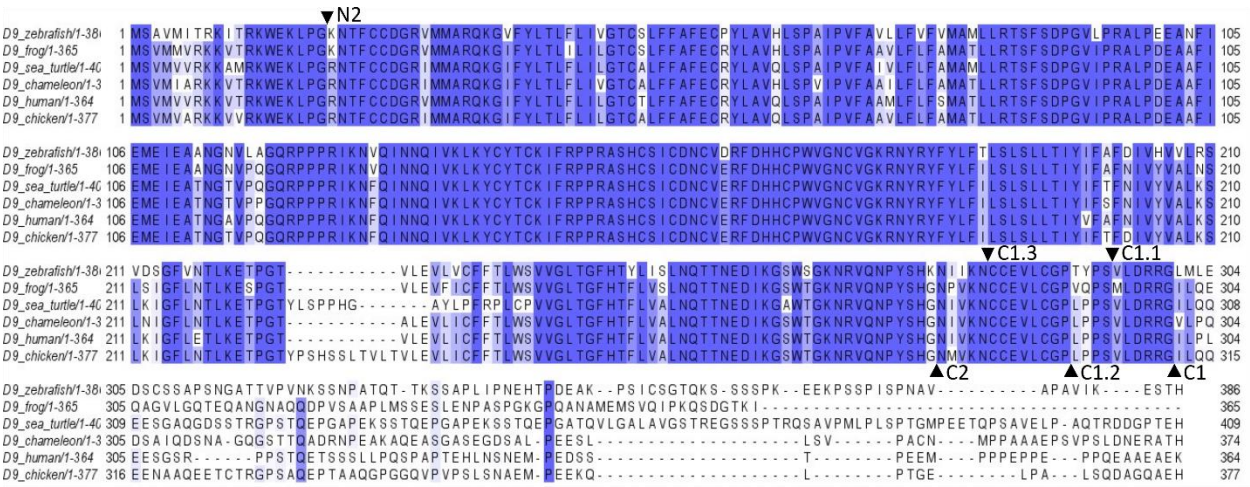

Figure 2.2. Designing DHHC9 Deletion Constructs. Amino acid sequence of human DHHC9 compared to various orthologs. Amino acids highlighted in dark blue are completely conserved, in light blue are somewhat conserved, and in white are not conserved. Arrows indicate where the designated deletions were made.

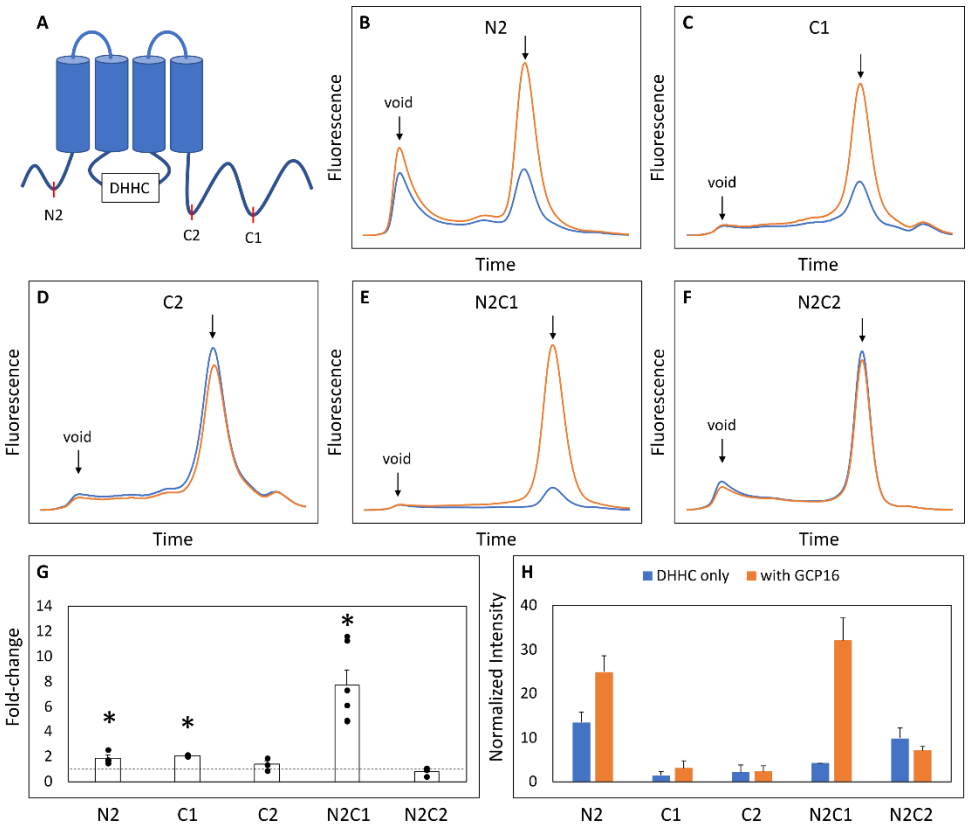

Figure 2.3. A Conserved Region in the DHHC9 CTD is Required for GCP16-Mediated Stabilization. (A) Cartoon depicting the tested DHHC9 truncations. (B-F) FSEC profiles for the indicated constructs. HEK cells were transfected with DHHC9 without (blue) or with GCP16 (orange). Cells were solubilized in a DDM-containing buffer, and the cleared lysate was analyzed by FSEC. Bar charts showing (G) average fold-change and (H) normalized max intensity upon GCP16 co-expression for the indicated constructs for n≥3 experiments. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

I next sought to determine the exact residues between DHHC9 C1 and C2 that are required for GCP16-mediated stabilization. I made finer truncations between regions C1 and C2 in the N2C1 background and assayed for the effect of GCP16 co-expression using FSEC (Fig. 2.4). Much like the parent construct, truncations C1.1 and C1.2 showed an increase in expression and monodispersity with GCP16 co-expression (Fig. 2.4 B, C, E, F). However, truncation C1.3 displayed no significant difference whether GCP16 was co-expressed or not (Fig. 2.4D, E, F). Between truncations C1.2 and C1.3, I identified a conserved sequence “CCXXXC” at residues 283-288, which I refer to as the C-terminal cysteine motif (CCM). Given that cysteines may play a role in protein-protein interactions, I tested whether mutation of these CCM cysteines would abolish the stabilizing effect. I generated serine mutants at each individual cysteine in the CCM using the DHHS9 N2C1 background and assessed the effect of GCP16 co-expression on DHHC9 protein (Fig. 2.5). Mutations C283S and C284S showed a diminished but statistically significant improvement in protein expression when co-expressed with GCP16 (Fig. 2.5 A, B, E, F). Interestingly, C288S or the triple mutations at all three cysteines (TM) showed no effect of GCP16 co-expression on DHHC9 protein quality (Fig. 2.5 C, D, E, F). Taken together, these data show that DHHC9 protein expression and monodispersity is significantly improved by GCP16 co-expression, and this stabilization effect requires a conserved DHHC9 C-terminal cysteine motif, with cysteine 288 being the most critical.

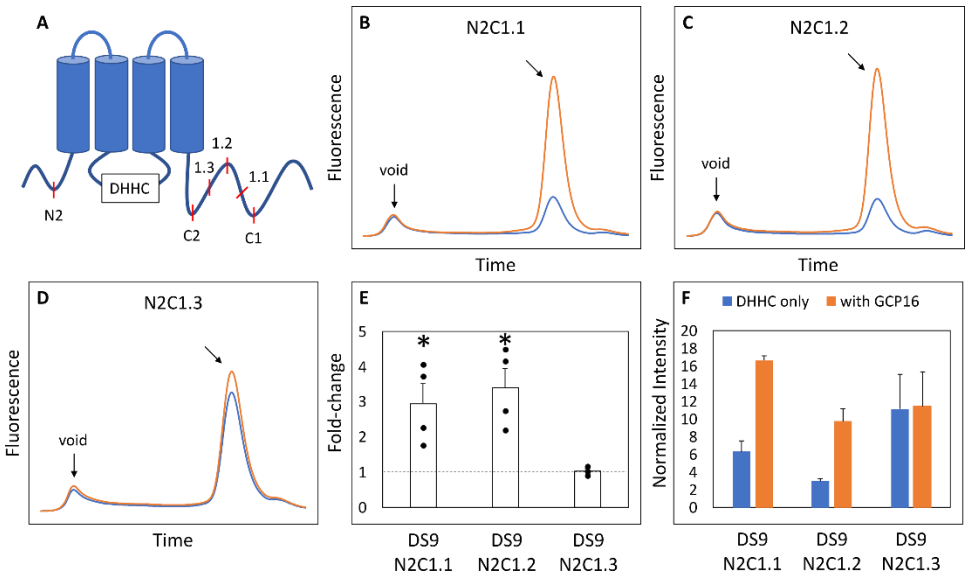

Figure 2.4. Defining the region of the DHHC9 CTD Required for GCP16-Mediated Stabilization. (A) Cartoon depicting the tested DHHC9 truncations. (B-D) FSEC profiles for the indicated constructs. HEK cells were transfected with DHHC9 without (blue) or with GCP16 (orange). Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed by FSEC. Bar charts showing (E) average fold-change and (F) normalized max intensity upon GCP16 co-expression for the indicated constructs for n≥3 experiments. The dashed line indicates a fold-change of 1. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

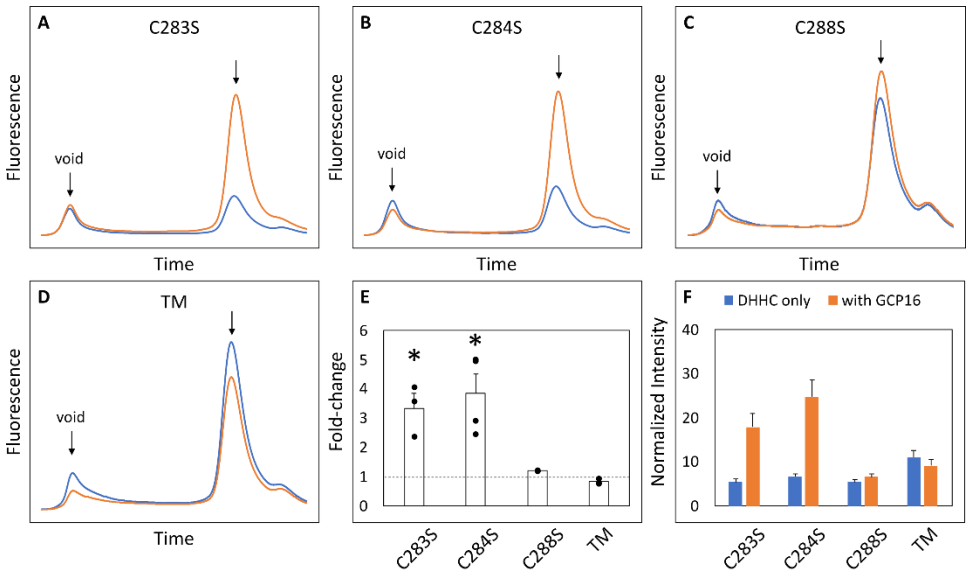

Figure 2.5. Cysteines in the DHHC9 CTD are Required for GCP16-mediated Stabilization. (A-C) FSEC profiles for the indicated cysteine to serine mutation and (D) the triple mutation (TM) in the DHHS9 N2C1 parent construct. HEK cells were transfected with DHHC9 without (blue) or with GCP16 (orange). Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via FSEC. Bar charts showing (E) average fold-change and (F) normalized max intensity upon GCP16 co-expression for the indicated constructs for n≥3 experiments. The dashed line indicates a fold-change of 1. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

The DHHC9 CCM is Required for Complex Formation with GCP16

While FSEC is an efficient way to probe whether GCP16 expression influences DHHC9 stability, it cannot address whether the two proteins are interacting. To determine if DHHC9 and GCP16 form a complex, I co-expressed these two proteins in Sf9 insect cells, affinity purified, and assessed whether they co-elute in SEC (Fig. 2.6). Across multiple experiments, affinity-purified DHHC9 in the absence of GCP16 consistently eluted at the void volume (Fig. 2.6 A). This and the observation that the protein remains impure after tandem affinity chromatography and SEC (Fig. 2.6 B), suggested that purified DHHC9 primarily exists as a high molecular weight protein aggregate. GCP16 alone eluted around ~13 mL (Fig. 2.6 C, D). When DHHC9 and GCP16 were co-expressed, both proteins co-eluted at a volume of ~11 mL (Fig. 2.6 E, F) with a monodispersed peak, supporting that DHHC9-GCP16 exists as a folded complex.

Next, I tested whether the CCM of DHHC9 is required for complex formation with GCP16. I co-purified DHHC9 TM with GCP16 and analyzed the SEC profile. I confirmed by SDS-PAGE that both DHHC9 TM and GCP16 were present in the elution from affinity purification (Fig. 2.6 H lane 1). However, the SEC profile exhibited two distinct peaks at the void volume and at ~13 mL (Fig. 2.6 G), paralleling what was seen for DHHC9 and GCP16 when purified separately. Furthermore, SDS-PAGE analysis showed enrichment for DHHC9-strep in the fraction corresponding to the void peak fraction and GCP16-strep in the fraction at ~13 mL (Fig. 2.6 H lane 2-3). Together, these experiments support that DHHC9 and GCP16 form a complex when co-expressed and co-purified. Furthermore, mutation of the conserved DHHC9 C-terminal cysteines prevents complex formation under these conditions of detergent solubilization.

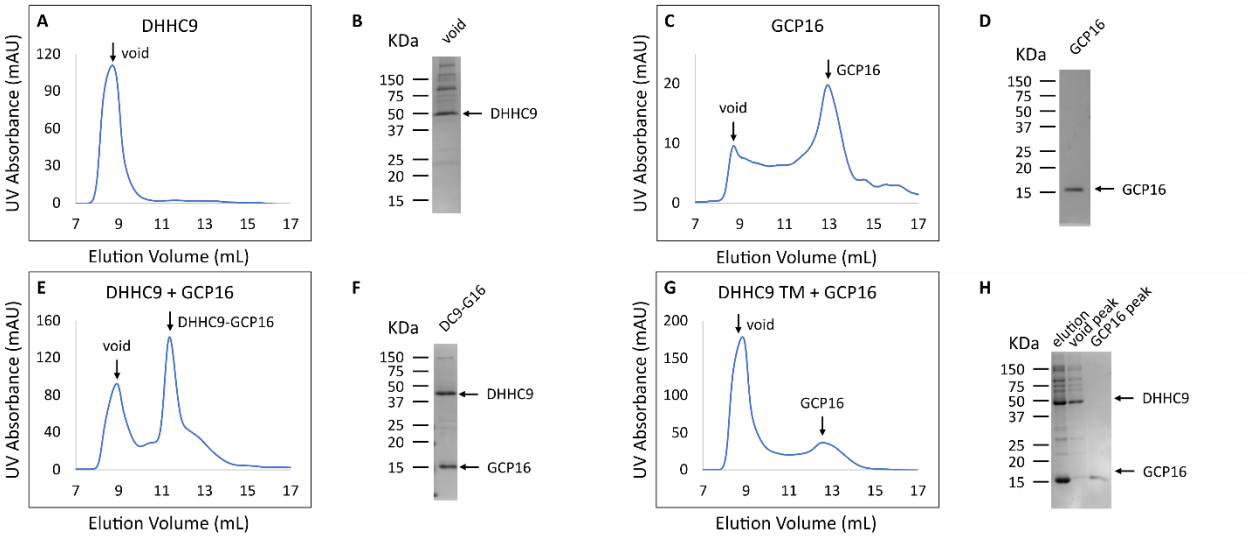

Figure 2.6. DHHC9 CTD Cysteines are Required for Complex Formation with GCP16. Sf9 cells were infected with recombinant baculovirus encoding the indicated variations of the following: DHHC9-strep, DHHC9 with the conserved triple cysteines mutated to serine (DHHC9 TM), or GCP16-strep. Protein was purified via strep-affinity purification and SEC as in the experimental procedures. SEC profiles and SDS-PAGE Coomassie analyses for (A-B) DHHC9, (C-D) GCP16, (E-F) the DHHC9-GCP16 complex, and DHHC9-TM co-purified with GCP16. Arrows in the SEC profiles correspond to the labeled lanes for the corresponding SDS-PAGE.

DHHC9 Enzymatic Activity For H-Ras Correlates with Protein Folding

Protein S-acylation by DHHC proteins occurs via a two-step mechanism (45, 49). The DHHC protein uses acyl-CoA to form an acyl-enzyme intermediate; then, upon substrate binding, the acyl group is transferred to the target cysteine on the protein substrate. Researchers previously demonstrated that GCP16 increases the equilibrium levels of both autoacylated DHHC9 and its substrate H-Ras *in vitro* (69). However, it remains unclear whether GCP16 directly affects catalysis, or whether GCP16 simply increases the amount of folded and active DHHC9. Therefore, I purified DHHC9 with or without GCP16 and classified whether the protein was folded or aggregated based on SEC profiles. Immediately following purification and SEC, I assayed PAT activity using H-Ras and [^{3}H]-Palmitoyl-CoA. The folded DHHC9-GCP16 complex exhibited radiolabeling for both H-Ras and autoacylated DHHC9 (Fig. 2.7 lane 2). The observed activity was attributed to the catalytic activity of DHHC9, as catalytically inactive DHHS9 protein failed to show any PAT activity (Fig. 2.7 lane 4). In contrast, the aggregated DHHC9-GCP16 did not exhibit detectable PAT activity (Fig. 2.7 lane 3). Likewise, aggregated DHHC9 purified in the absence of GCP16 exhibited no detectable activity (Fig. 2.7 lane 1). Similarly, DHHC9 TM was unable to form a complex with GCP16 and resulted in only aggregated protein. The aggregated DHHC9 TM exhibited no detectable activity for H-Ras (Fig. 2.7 lane 5). Taken together, these experiments suggest that DHHC9 requires GCP16 for its enzymatic activity. DHHC9 co-purified with GCP16 results in enrichment of a folded DHHC9 state, which is enzymatically active. However, when GCP16 is absent or when DHHC9 is unable to form a complex with GCP16 (as is the case with DHHC9 TM), essentially all DHHC9 protein is aggregated and exhibits no detectable PAT activity for H-Ras.

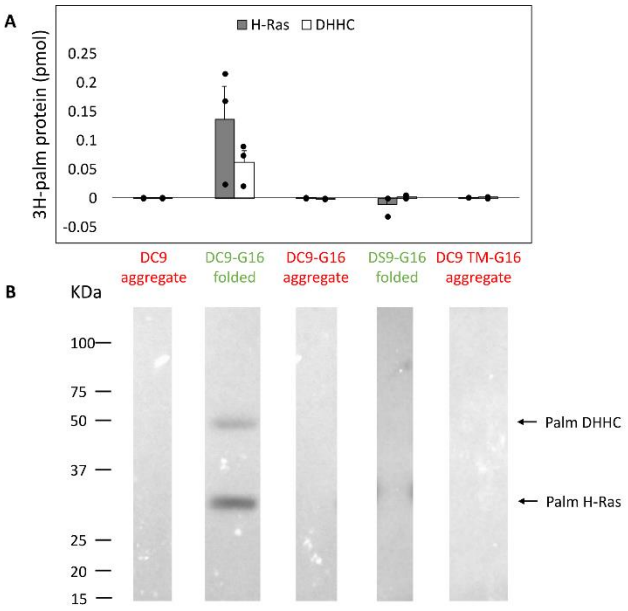

Figure 2.7. DHHC9 Requires Proper Protein Folding for H-Ras Activity. DHHC9 constructs were purified and classified into aggregated (red) or folded (green) states based on their respective SEC profiles. Purified DHHC9 was assayed for activity using H-Ras and ^{3}H-labeled palmitoyl CoA. Radiolabeled protein was quantified via liquid scintillation spectroscopy and detected via fluorography. The Y-axis represents picomoles of ^{3}H-labeled protein after subtracting background radiation. Grey bars represent the quantity of labeled H-Ras, and white bars represent the quantity of labeled DHHC protein. Error bars represent the standard error of the mean for n≥2 independent protein preparations. Arrows indicate bands corresponding to radiolabeled DHHC9 and H-Ras.

DHHC9 Disease Mutations Result in Reduced Protein Stability

DHHC9 protein pathogenic variants include missense mutations R148W, P150S, and R96W, and a nonsense mutation terminating at R298 (93). An earlier study reported that DHHC9 (R148W) and DHHC9 (P150S) possessed reduced steady state levels of autopalmitoylated DHHC9 (98). Given the importance of GCP16 in stabilizing DHHC9 protein folding with a corresponding effect on enzyme activity, I asked how the missense and nonsense mutations affected complex formation with GCP16. Compared to wildtype DHHC9, DHHC9 with R148W, P150S, R96W, or R298X mutations exhibited a greater extent of protein aggregation and reduced monodispersity as monitored by FSEC (Fig. 2.8 A-E, blue traces, F). Co-expression with GCP16 resulted in improved protein behavior compared to the respective DHHC9 constructs alone for all the mutants except R96W, although with a diminished effect compared to wildtype (Fig. 2.8 A-E, orange traces, F). These results suggest the DHHC9 mutations associated with XLID result in decreased protein stability and reflect reduced formation of a DHHC9-GCP16 complex.

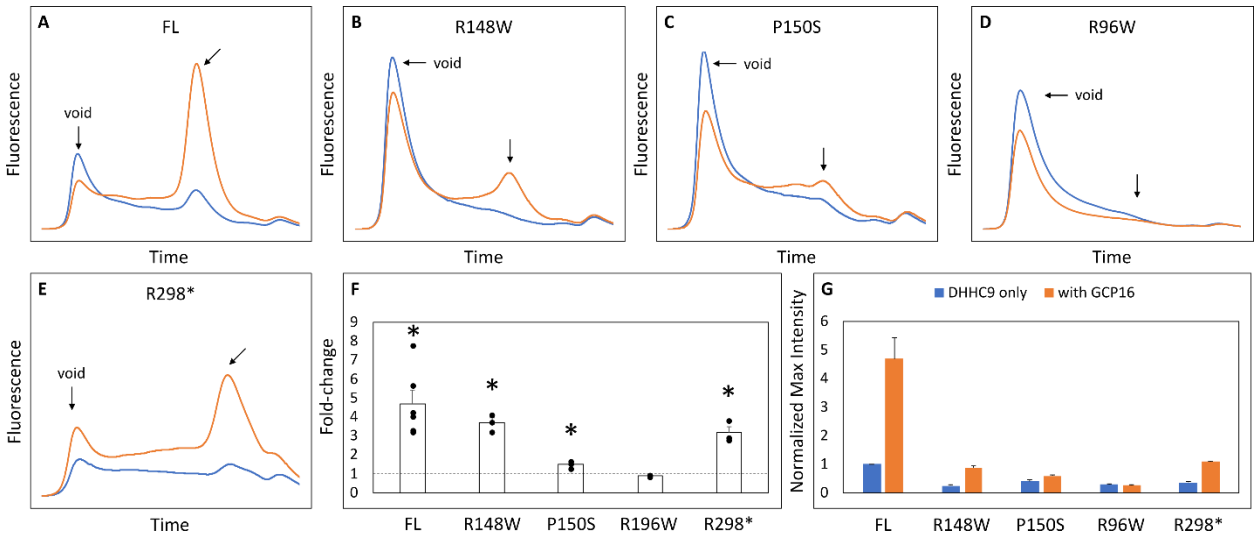

Figure 2.8. DHHC9 Disease Mutations Exhibit Reduced Protein Stability. (A-E) FSEC profiles for full-length DHHC9 (FL), the indicated point mutations, and DHHC9 truncation at R298 (R298*). HEK cells were transfected with vector to express DHHC9 without (blue) and with GCP16 (orange). Bar charts showing (F) average fold-change and (G) normalized max intensity upon GCP16 co-expression for the indicated constructs for n≥3 experiments. Max intensities were normalized to DHHC9 FL as described in the experimental procedures. The dashed line indicates a fold-change of 1. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

GCP16 and GOLGA7B Stabilize DHHCs in a Subtype-Specific Manner

Our experiments suggest that the CCM in DHHC9 plays an important role in complex formation with GCP16. This motif is present in other DHHCs closely related to DHHC9, namely, DHHC14, 18, 5, and 8, but it is not present in distantly related DHHCs, such as DHHC3 and DHHC20 (Fig. 2.9 A). DHHC9 was the first human DHHC identified to require a protein partner for its activation, based on its homology to yeast Erf2/Er4 (69). Recent studies suggest that GCP16 and GOLGA7B, a protein with ~75% amino acid sequence identity to GCP16, function as accessory proteins for additional members of the DHHC protein family. Woodley and Collins reported that DHHC5 interacts with GOLGA7B, facilitating DHHC5 localization at the plasma membrane and enabling its interactions with components of desmosomes to regulate cell adhesion (72). Ko et al. identified ZDHHC5 and GOLGA7 (GCP16) in a screen for genes involved in an unconventional nonapoptotic cell death pathway triggered by the synthetic small molecule oxime, CIL56 (70). They went on to show that DHHC5 and GCP16 form a mutually stabilizing protein complex localized at the plasma membrane. Complex formation is dependent upon C-terminal cysteines in the conserved CCM motif, consistent with my results for DHHC9-GCP16 complex formation. To determine the potential of other DHHC proteins to form complexes with GCP16 and/or GOLGA7B, I tested whether co-expression of GCP16 or GOLGA7B affects the protein behavior of a set of DHHC proteins using FSEC. Like DHHC9, all DHHCs containing the CCM exhibited improved protein expression and monodispersity when co-expressed with GCP16 (Fig. 2.9 B-E, H, I). However, representative DHHCs without the CCM showed no significant difference whether GCP16 was co-expressed (Fig. 2.9 F-I). This supports that the GCP16-mediated stabilization is specific to certain DHHCs that possess the CCM. Interestingly, DHHC protein co-expression with GOLGA7B showed a similar, but distinct, result (Fig. 2.10). While GOLGA7B co-expression significantly improved DHHC5 and DHHC8 protein (Fig. 2.10 D-E, H, I), it had no significant stabilizing effect on DHHC9, 14, 18, 3, and 20 (Fig. 2.10 A-C, F-I). Taken together, my FSEC experiments suggest GCP16 and GOLGA7B stabilize DHHC proteins in a

subtype-specific manner.

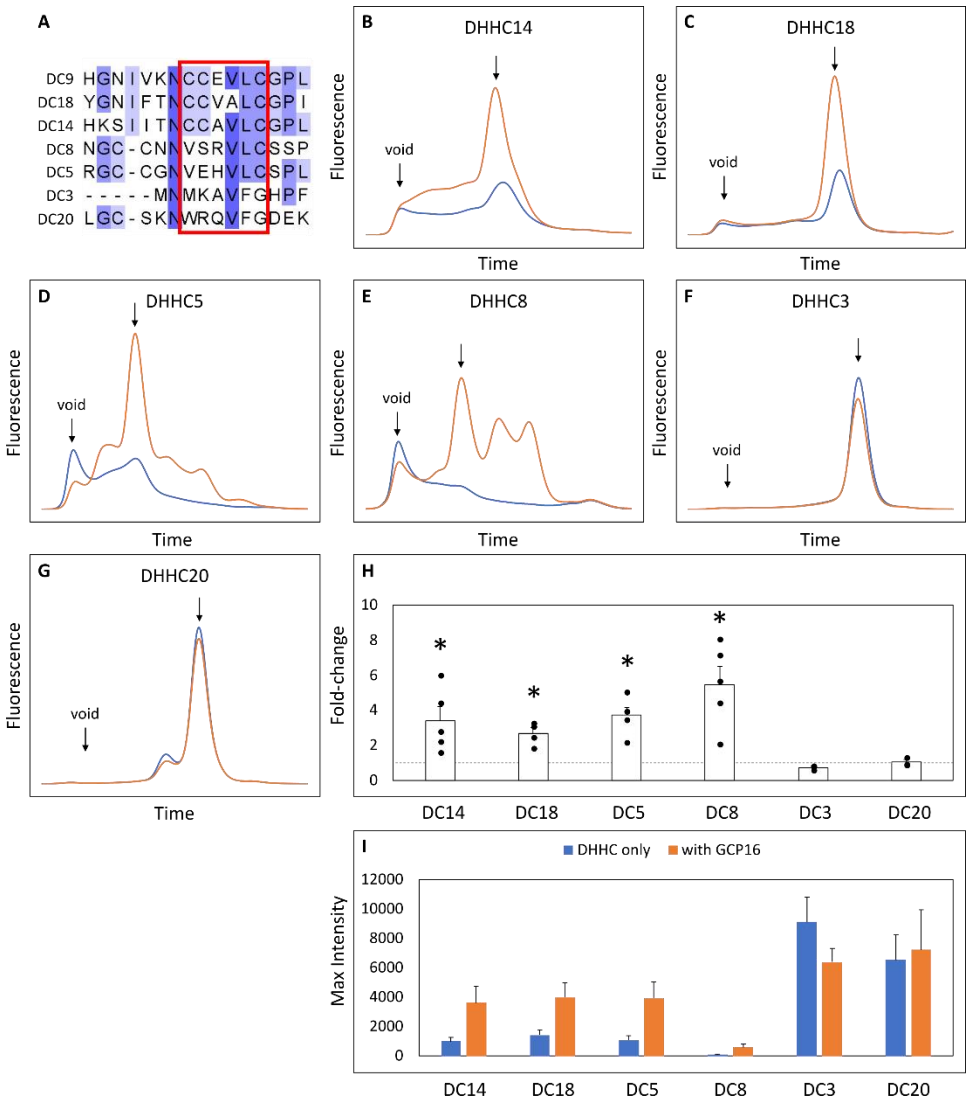

Figure 2.9. GCP16 Co-expression Stabilizes DHHC Proteins with the Conserved C-terminal Cysteine Motif. (A) Amino acid sequence alignment of select DHHC proteins at the conserved CTD cysteine motif. The conserved cysteine motif is boxed in red. (B-G) FSEC analysis of crude HEK lysates for the indicated constructs without (blue) or with GCP16 (orange). Bar charts showing (H) fold-change upon and (I) average max intensity for the indicated constructs upon GCP16 co-expression at n≥3 experiments. Error bars represent the standard error of the mean. Asterisks indicate a significance for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

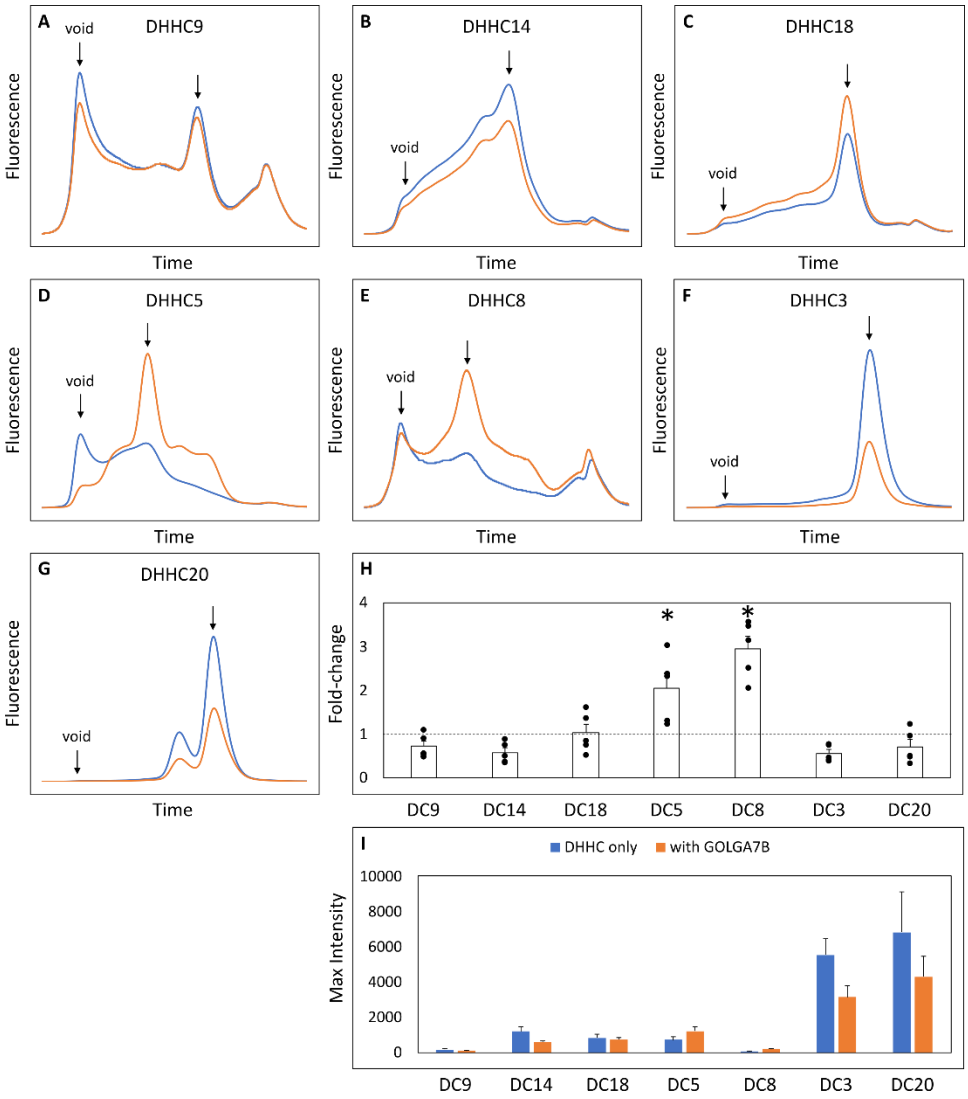

Figure 2.10. GOLGA7B Co-expression Stabilizes DHHC Proteins. (A-G) FSEC profiles for the indicated constructs. HEK cells were transfected with DHHC9 without (blue) or with GOLGA7B (orange). Bar charts showing (H) fold-change upon and (I) average max intensity for the indicated constructs upon GOLGA7B co-expression at n=5 experiments. Error bars represent the standard error of the mean. Asterisks indicate a significant increase with GOLGA7B for p-value <0.05 determined by two-tailed t-test against the null hypothesis.

To better understand the effect of GCP16 on the DHHC9 subfamily, I purified DHHC14 and DHHC18 and assessed the relationship between protein folding and PAT activity (Fig. 2.11). As was the case with DHHC9, DHHC14 purified without GCP16 resulted in nearly complete protein aggregation (Fig. 2.11 A-B) and exhibited no detectable PAT activity (Fig. 2.11 E lane 1). On the other hand, DHHC14 co-purified with GCP16 was monodisperse (Fig. 2.11 C-D) and the folded complex in the included volume exhibited enzymatic activity for H-Ras and DHHC14 autoacylation (Fig. 2.11 E lane 3). Surprisingly, the DHHC14-GCP16 aggregate detected in the void volume exhibited PAT activity, albeit reduced relative to the folded complex (Fig. 2.11 E lane 2). Similarly, DHHC18 purified by itself was aggregated and enzymatically inactive (Fig. 2.11 F-G, J lane 1). Like DHHC14-GCP16, the DHHC18-GCP16 complex purified as two species with the folded complex having more activity than that found in the aggregate (Fig. 2.11 H-I, J lane 2). These results suggest that the stabilizing effect of GCP16 spans the DHHC9 subfamily. To my knowledge, this is the first demonstration of DHHC14 and DHHC18 enzyme activity *in vitro*.

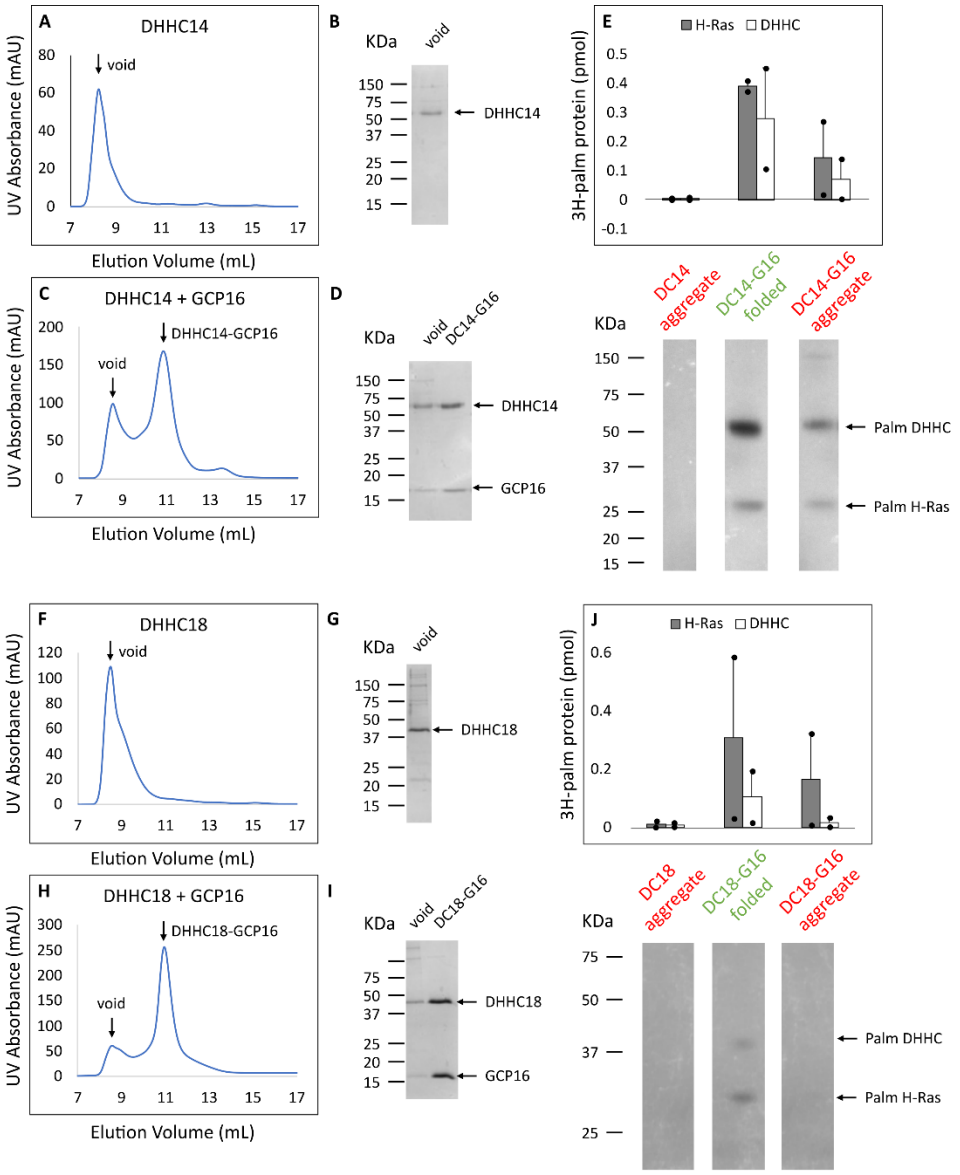

Figure 2.11. DHHC14 and DHHC18 Each Forms a Complex with GCP16 to Confer PAT Activity. Sf9 cells were infected with recombinant baculovirus encoding strep-tagged DHHC alone or DHHC+GCP16. Protein was purified via strep-affinity purification and SEC. SEC profiles and SDS-PAGE Coomassie for (A-B) DHHC14, (C-D) the DHHC14 co-purified with GCP16, (F-G) DHHC18, and (H-I) DHHC18 co-purified with GCP16 are shown. Arrows in the SEC profiles correspond to the labeled lanes for the corresponding SDS-PAGE analysis. (E and J) Purified DHHC14/DHHC18 was classified into aggregated (red) or folded (green) states based on the SEC profile. DHHC14/DHHC18 was assayed for activity using H-Ras and ^{3}H-palmitatoyl CoA.

Radiolabeled protein was quantified via liquid scintillation spectroscopy and detected via fluorography. The y-axis represents picomoles of ^{3}H-labeled protein after subtracting background radiation. Gray bars represent the quantity of labeled H-Ras, and white bars represent the quantity of labeled DHHC protein. Error bars represent the standard error of the mean for n=2 independent protein preparations. Arrows indicate bands corresponding to radiolabeled DHHC14 and H-Ras.

Discussion

In this study, I sought to understand the function of GCP16 in the DHHC9-GCP16 complex. I found that co-expression of GCP16 with DHHC9 improved DHHC9 protein levels and homogeneity. Through truncations, mutagenesis, and FSEC screening, I identified a conserved C-terminal cysteine motif in DHHC9 that is required for the GCP16-mediated stabilization effect. I used the SEC profiles of purified proteins to show the CCM is required for formation of the DHHC9-GCP16 complex. Without GCP16, purified DHHC9 aggregates, and only folded DHHC9-GCP16 complex formation is enzymatically active for H-Ras *in vitro*. Furthermore, I showed that disease mutations associated with XLID result in decreased protein stability and diminished DHHC9-GCP16 complex formation.

DHHC9 and GCP16 were discovered based on homology to yeast palmitoyltransferase Erf2-Er4 (35, 69). DHHC9 and GCP16 share 31% and 17% sequence identity with Erf2 and Er4, respectively. While the exact sequence identity between the pair is limited, they possess functional similarities. For example, both form heteromeric DHHC protein complexes, localize to the Golgi, and are putative Ras PATs in their respective organisms (35, 69). Furthermore, DHHC9-GCP16 can substitute for loss of Erf2 in *S. cerevisiae* (98). Mitchell et al. showed, in the absence of Er4, Erf2 steady-state levels and its half-life are significantly lowered; and, one function of Erf4 is to stabilize Erf2 by impeding its ERAD-mediated degradation (68). Similarly, my data supports that one role for GCP16 is to provide local stability for DHHC9, suggesting this function for the partner protein may be evolutionarily conserved. However, the role of the accessory protein likely extends beyond DHHC protein stabilization. For example in yeast, stabilization of Erf2 does not suppress loss of Erf4 *in vivo,* and Erf2 exhibits distinct enzyme kinetics with and without Erf4 *in vitro* (68). Specifically, the absence of Erf4 does not abolish Erf2 autopalmitoylation, suggesting that residues of Er4 do not participate directly in this step. However, the absence of Erf4 does increase the rate of hydrolysis of the thioester on the acyl-enzyme intermediate, causing Erf2 to undergo a futile cycle of autopalmitoylation and hydrolysis.

Erf4 is required for Erf2 to transfer the palmitoyl group to the protein substrate, with the exact mechanism remaining unclear. Given their structural and functional similarities, I speculate that GCP16 plays some role in catalysis. Unfortunately, despite extensive optimization of purification conditions, protein engineering, and extraction using native lipid nanodiscs (99, 100), I have been unsuccessful in purifying DHHC9 alone that is properly folded with any detectable enzyme activity. Without having a baseline activity for a GCP16-less DHHC9, it is difficult to attribute whether any observed lack of activity is due to a catalytic requirement for GCP16 or a lack of folded DHHC enzyme.

We asked whether these cysteines are involved in direct interaction with GCP16 and/or whether they are palmitoylated. I attempted to detect direct interaction by cross-linking purified DHH9-GCP16 with disuccinimidyl sulfoxide and performing nano LC-MS/MS in orbitrap fusion; however, the results were inconclusive as my sequence coverage did not include the DHHC9 C-terminal domain. Previously, global identification of S-acylation sites in mouse forebrain detected modification of DHHC9 on the active site cysteine and a second cysteine within the DHHC-CRD (101). While sites identified from proteomic studies serve as a valuable first step in mapping S-acylation, factors such as low abundance, hydrolysis, or turnover during experimental procedures may paint an incomplete picture. In an effort to evaluate the S-acylation state of the CCM cysteines in the context of GCP16, I collaborated with the Linder lab and used 17-ODYA labeling (102) and Acyl-Biotin Exchange (ABE) (103, 104) in HEK cells overexpressing DHHC9 or DHHC14 with and without GCP16 co-expression. However, though we were able to detect labeling outside of the catalytic domain, we were unable to definitively assign the S-acylation sites in the CCM. Nonetheless, the CCM partially overlaps with previously annotated DHHC motifs which may provide insight into the implications of the CCM cysteines.

Earlier proteomics studies identified the C-terminal $CCX_{7\text{-}13}C^*(S/T)$ sequence among DHHC proteins where cysteines centered around the “GCxxN” motif were palmitoylated, and these cysteines in DHHC5 and DHHC8 were detected to be palmitoylated *in vivo* (14, 101).

Subsequently, Ko et al. showed DHHC5 and GOLGA7 (GCP16) are mutually stabilizing and that DHHC5-GOLGA7 complex formation, as assessed by immunoprecipitation, is dependent on the DHHC5 C-terminal cysteines (70). The catalytically active DHHC5-GOLGA7 complex is required for nonapoptotic cell death induced by the synthetic oxime-containing small molecule caspase-independent lethal 56 (CIL56) (70). Additionally, palmitoylation of the DHHC5 C-terminal cysteines controls its interaction with and its ability to palmitoylate GOLGA7B, which in turn regulates DHHC5 internalization and turnover and its protein interactome (72). Howie et al. observed that the DHHC5 C-terminus between N218-T334, which includes the CCM, is required for palmitoylation of the Na-pump accessory protein phospholemman (PLM) (105). Interestingly, PLM does not bind directly to this region of DHHC5, but rather it associates with DHHC5 through an intermediate, the Na-pump a subunit, which does bind directly to DHHC5 N218-T334 (106). Nonetheless, the DHHC5 region containing the CCM is specifically required for palmitoylation of its substrates PLM and Flotillin-2, while other substrates such as PSD-5 and GRIP1 require the PSD-95/Discs-large/ZO-1 homology (PDZ) binding motif (105). The CCM I identified in the DHHC9 subfamily overlaps with the end of the previously described $CCX_{7\text{-}13}C^*(S/T)$ sequence. Taken together, I reason that the CCM cysteines are likely to be palmitoylated and mediate accessory protein interaction either directly or indirectly by exposing a GCP16/GOLGA7B binding site. Interaction with the accessory protein provides protein stability, and it expands the DHHC protein interactome, which may be a mechanism of substrate recruitment and regulation. How subtype preferences for GCP16/GOLGA7B are governed is to be resolved, though it seems likely to involve additional protein domain coordination.

DHHC9 *in silico* analysis hints at a role for the CCM. A DHHC9 atomic model was generated using AlphaFold, a ‘knowledge-based’ algorithm that utilizes protein-sequence information available in the Protein Data Bank to understand and predict unknown protein structures (107). Superimposing the DHHC9 model onto the acylated-intermediate mimic 2-bromopalmitate-DHHC20 (2-BP-DHHC20) model, obtained from X-ray crystallography (PDB ID:

6BML) (40), suggests that the CCM is part of a helix that corresponds to the α'2 helix in DHHC20 and that the CCM cysteines are oriented with sidechains towards the membrane (Fig. 2.12). In DHHC20, the α'2 helix is an amphipathic helix that lies almost parallel to the membrane; it follows the palmitoyltransferase conserved C-terminus (PaCCT) motif and contacts TM3 and TM4 to provide local stability (40, 41). Additionally, Plain et al. predicted that the DHHC5 C-terminal PaCCT motif, which includes the CCM, forms a juxtamembrane amphipathic α-helix, and they demonstrated this helix binds to the Na-pump α subunit (106). In the DHHC9 model, S-acylation of the CCM would similarly anchor the CCM helix to the membrane, possibly allowing a similar local stability to this region of the protein. Additionally, the 2-BP-DHHC20 model reveals that the active site is surrounded by hydrophobic residues on all sides except the front of the enzyme (40), and previous molecular dynamics simulations suggested that DHHC20 causes a local deformation in the lipid bilayer to expose the catalytic cysteine for acyl-CoA attack (52). Therefore, it seems the active site must be protected in such a way to only expose the nucleophilic cysteine when appropriate for either acyl-CoA or substrate access. An interesting possibility with the heteromeric DHHC complexes is that the accessory protein regulates exposure to the active site. This is supported by the observation that, in the absence of Er4 or its C-terminal domain, Erf2 undergoes a constant futile cycle of autopalmitoylation and hydrolysis (68). In the DHHC9 subfamily, the C-terminus and palmitoylated CCM may position the accessory protein to regulate access to the active site and to protect the acyl-enzyme from undergoing hydrolysis. Only when conditions are favorable, such as when substrate is bound, does the active site become accessible for palmitate transfer. This may explain why Erf2 requires Erf4 for the substrate transfer step. Looking ahead, it will be especially revealing to see the structure of a heteroprotein DHHC complex and a DHHC-substrate protein complex.

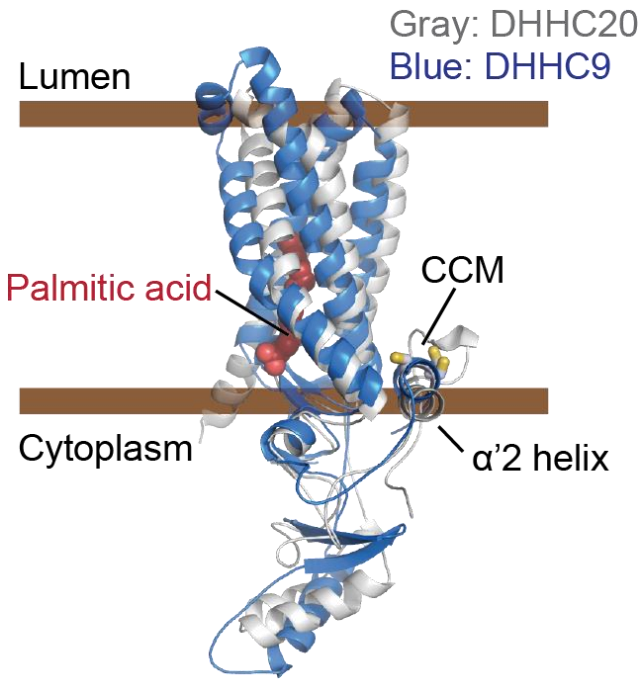

Figure 2.12. In Silico Model of DHHC9 Suggesting that the CCM is Located Where the α'2 Helix Exists in DHHC20. Structural alignment of AlphaFold DHHC9 model (blue, Q9Y397) with the DHHC20 model generated from X-ray crystallography data (gray, PDB ID: 6BML DHHC9 N- and C-terminal regions predicted with lower confidence (residues 1-15 and 295-364) are not shown.). Palmitic acid (sphere representation in red) and α'2 helix in DHHC20, and the CCM (stick representation of cysteine residues) in DHHC9 are indicated. Brown bars indicate predicted membrane boundary between the cytoplasm and the lumen of the Golgi apparatus.

The Ras protein family members are well-known drivers of many cancers and understanding Ras protein regulation and its protein interactors may reveal future drug targets. DHHC9 was discovered based on its homology to the yeast Ras PAT (69), and more recent studies support its role in Ras protein regulation. Shimell et al. showed DHHC9 and N-Ras likely function through the same pathway to promote dendritic cell growth in neuronal cell culture (81). In human fibrosarcoma HT1080 cells, overexpressing *ZDHHC9* promotes, while repressing it diminishes, Ras PM localization (82). *ZDHHC9* silencing reduces oncogenic H-Ras-mediated cell transformation and colony formation in murine NIH 3T3 cells (82). Chai et al. reported microRNA-134 regulates DHHC9-dependent H-Ras localization in SST^+ neurons that is dependent upon the palmitoylated cysteines (83). Using *ZDHHC9* knockout mice, Liu et al. showed Ras palmitoylation is decreased in bone marrow cells (84). Furthermore, ZDHHC9 inactivation inhibits oncogenic N-Ras cellular transformation in hematopoietic cells and slows down the development of myeloid leukemia in N-Ras-induced chronic myelomonocytic leukemia mice models (84). While DHHC9 is the putative Ras PAT, the existence of additional human Ras PATs has been speculated (69). DHHC9 is not present in all tissues. DHHC9 mRNA expression is particularly enriched in the brain, skeletal muscle, lung, and liver, but absent in thymus, spleen, and peripheral blood leukocytes (69). Additionally, Rocks et al. observed that short hairpin RNA-mediated knockdown of *ZDHHC9* showed no detectable effect of Ras palmitoylation in HeLa cells (108). It is probable that another DHHC protein is palmitoylating Ras where DHHC9 is absent. DHHC14 and DHHC18 may be candidates. DHHC14 and DHHC18 localize to the ER and Golgi, respectively (2), in line with stops during the Ras acylation pathway (21). Additionally, Fukata et al. found that overexpression of H-Ras with DHHC18 enhanced H-Ras palmitoylation compared to other DHHCs tested (109). In my study, I found DHHC9, 14, and 18 require complex formation with GCP16 for activity for H-Ras *in vitro*. It is to be determined whether this requirement extends *in vivo*. GCP16 has been found to be expressed ubiquitously (80), possibly allowing it to function

as a coordinator for Ras-palmitoylation with other DHHCs where DHHC9 is absent.

In conclusion, I identified a C-terminal conserved cysteine motif in DHHC9 that is present in a subset of DHHC proteins that includes DHHC9, 14, 18, 5, and 8. GCP16/GOLGA7B co-expression improved DHHC protein stability with subtype specificity, where GCP16 potentiated all subfamily members while GOLGA7B potentiated only DHHC5 and DHHC8. In the context of detergent purification, I showed the CCM is required for DHHC9-GCP16 complex formation and enzymatic activity for H-Ras, and the GCP16 requirement extends to DHHC14 and DHHC18. Additionally, I reported the first DHHC14 and DHHC18 purification with SEC profile and enzymatic activity *in vitro*. My research supports a broader role for GCP16 and GOLGA7B in the function of human DHHC proteins.

Experimental Procedures

Construction of Expression Plasmids

Amino acid sequences were obtained from the UniProt database for the following human proteins: DHHC9 (Q9Y397), DHHC14 (Q8IZN3), DHHC18 (Q9NUE0), DHHC5 (Q8VDZ4), DHHC8 (Q9ULC8), DHHC3 (Q9NYG2), DHHC20 (Q5W0Z9), and GCP16 (Q7F5G4). Genes were synthesized based on protein sequences (GenScript, Piscataway, NJ). The plasmid encoding GOLGA7B was obtained from the Harvard Medical School Plasmid repository. All DHHC genes used were PCR amplified to exclude the start methionine and to include flanking BamHI/XhoI sites. Similarly, *GOLGA7* and *GOLGA7B* were amplified to exclude the start methionine and to include AgeI/NotI sites. All genes were subcloned via standard molecular biology techniques into their respective expression vectors. Any mutations were generated via quick change PCR.

For FSEC, I used a dual mammalian expression vector modified from the pIRES-EGFP RK6 vector provided by M. Mayer, National Institutes of Health, Bethesda, MD. The vector encoded a start methionine followed by mNeonGreen (A0A1S4NYF2), a BamHI/XhoI insertion site, an EMCV internal ribosomal entry site (IRES), an AgeI/NotI insertion site, and a stop codon. DHHC genes were incorporated using BamHI/XhoI, and *golga7* or *golga7b* genes were incorporated using AgeI/NotI.

For protein purification, genes were subcloned into a modified pFastBac baculovirus expression vector (Thermo Fisher Scientific, Waltham, MA). DHHC9, DHHC9 TM (C283S, C284S, and C288S), and GCP16 were encoded to express a C-terminal strep-tag. DHHC14 and DHHC18 were encoded to express an N-terminal strep-tag.

Fluorescence-Detection Size-Exclusion Chromatography

HEK cells at 80-95% confluency were transfected with 2.5 µg of expression plasmid using jetPRIME transfection reagent per the manufacturer's instructions. Cells were grown at 37°C for

48 hours. The media was aspirated, cells were suspended and washed in 2 mL of cold PBS, and samples were lysed in 150 µL of lysis buffer (1x PBS, 1% n-Dodecyl-β-D-Maltopyranoside (DDM), 1x Roche protease inhibitor) for 30 minutes at 4°C with rotation. Whole-cell lysates were cleared by centrifugation in a table-top centrifuge at 15k rpm for 10 minutes at 4°C. The supernatant was further cleared by ultracentrifugation in a TLA 100.3 rotor at 70K rpm for 20 minutes. An aliquot (50µL) was applied onto a Superdex 200 Increase 10/300 GL column pre-equilibrated with running buffer (1x PBS, 0.5mM DDM). The eluent from the SEC column was passed through a fluorometer set to excitation, 480 nm, and emission, 508 nm.

For quantitation, peak intensities in arbitrary fluorescence units were obtained at the expected DHHC peak for every profile. For DHHC9 constructs, values were normalized to either DHHC9 or DHHC9 N2C1, depending on the set of constructs tested for that given day. Data sets across different days were then normalized to DHHC9 globally. For experiments that included other DHHCs, i.e., DHHC14, 18, 5, 8, 3, and 20, raw intensity values were used without normalization. For statistical analysis, I used two-tailed t-tests assuming unequal variances against the null-hypothesis that GCP16/GOLGA7B co-expression has no significant effect on DHHC protein signal.

Expression and Purification of DHHC Proteins and DHHC-GCP16 Complexes.

All constructs were expressed using the Invitrogen Bac-to-Bac® baculovirus-insect cell expression system. Sf9 cells (1-2L) were infected at 2.5 – 4.0 x 10^6 cells/mL with P2 virus (10^7-10^8 PFU/mL, 30mL/L). Cells were incubated at 27°C for 24 hours and moved to 18°C for an additional 48 hours. All purification steps were done on ice or with 4°C buffer unless stated otherwise. Cells were harvested by centrifugation at 2,040 x g and washed with 200 mL of PBS. Cells were solubilized in buffer A (PBS, 15% glycerol, 0.5mM TCEP, 0.5ug/mL leupeptin, 2ug/mL aprotinin, 0.5ug/mL pepstatin A, and 0.5mM phenylmethylsulfonyl fluoride) containing 1% DDM w/v at a ratio of 250 mg DDM/g of cells. Solubilization was performed at 4°C with stirring for one

hour. Large debris were removed by centrifugation at 12,000 x g. The supernatant was further clarified by ultracentrifugation at 185,000 x g for 45 minutes to remove remaining insolubilized material. To the supernatant, 2 mL of equilibrated StrepTactin Sepharose High Performance resin (GE Healthcare, Marlborough, MA) was incubated with stirring for one-hour. The resin was collected by centrifugation and transferred to a gravity column (Bio-Rad). The resin was washed with 10 resin bed volumes of wash buffer (buffer A with 0.5 mM DDM) and eluted with elution buffer (buffer A with 2.5 mM desthiobiotin, 0.5 mM DDM). Protein was concentrated using Amicon ultra concentrators. Size exclusion chromatography was done using a Superdex 200 (GE Healthcare) in SEC buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.5 mM TCEP, 15% glycerol, 0.5 mM DDM).

PAT Assays

PAT assays were performed immediately following purification and SEC as described previously (69). Protein concentrations were determined by NanoDrop (Thermo Scientific) using predicted extinction coefficients. Enzyme was diluted to 250 nM in enzyme dilution buffer (50 mM MES pH 6.4, 100 mM NaCl, 10% glycerol, 0.05% DDM, 0.5 mM TCEP). H-Ras (80 µM, purified from insect cells infected with recombinant Baculovirus) was diluted to 10 µM in enzyme dilution buffer. For a 50 µL reaction, the final concentration was 50 nM enzyme, 2 µM H-Ras, 1 µM palm-CoA. Each reaction was allowed to reach equilibrium by incubation at room temperature for 10 minutes. Reactions were quenched in SDS loading dye containing 25 mM TCEP. Samples were split and run on two SDS-PAGE gels in parallel. For fluorography, one gel was exposed to enhance solution (1M sodium salicylate, 15% methanol), dried under vacuum for 2 hours, and exposed to film for 4-15 days. For liquid scintillation spectroscopy, the gel was stained with Coomassie, H-Ras bands were extracted, and a section of gel around the expected DHHC protein size was extracted. Extracted gel pieces were dissolved in 500 µL Soluene overnight at 37°C. Ultima Gold scintillation fluid (5 mL) was added, and disintegrations per minute (DPM) was

measured using a scintillation counter after allowing for stabilization for 8-24 hours. For each experiment, I used DHHC20 as a positive control and H-Ras without enzyme to account for nonenzymatic acylation. Background DPM was subtracted from each sample before converting counts to picomoles based on the specific activity of palmitoyl-CoA.

Chapter 3: Towards a Structural Basis for the DHHC9-GCP16 Complex

Introduction

My goal was to determine the high-resolution structures of DHHC9, GCP16 and the DHHC9-GCP16 complex. The structures would provide insight into the overall architectures of DHHC9 and GCP16, and it would allow visualization of how they come together to form a functional heterodimeric DHHC enzyme. Comparison of the DHHC9-GCP16 complex with existing structures of hDHHC20 and zfDHHS15 would reveal universal DHHC structural motifs, as well as highlight subfamily specific differences. Furthermore, high-resolution structures would enable structure-guided biochemical experiments to better investigate the mechanism of catalysis for heteromeric DHHC protein complexes. Moreover, visualizing protein-protein interaction surfaces and the critical residues involved would provide a platform for the development of specific inhibitors or potential drugs to combat DHHC9-associated diseases.

DHHC9 is an integral membrane protein and GCP16 is a peripheral membrane protein, which have unique challenges over soluble proteins. Soluble proteins mostly contain hydrophilic surfaces which make soluble proteins generally stable in solution. However, in addition to hydrophilic surfaces, membrane proteins contain hydrophobic and amphiphilic domains that help embed them into biological membranes. This results in two major challenges for membrane proteins. First, it is more difficult to express membrane proteins in high quantities compared to soluble proteins, as there is a greater limitation to the amount of membrane area and membrane insertion machineries. Second, special considerations are required to extract and stabilize membrane proteins while minimizing irreversible protein unfolding and aggregation. The most widely used class of reagents for membrane protein extraction and stabilization are detergents, amphipathic molecules containing a hydrophilic headgroup (usually polar, sometimes charged) and a hydrophobic tail (110). Detergents act by inserting into biological membranes, disrupting membrane integrity, and, at concentrations above the

detergent critical micelle concentration (CMC), extracting membrane-embedded components into detergent micelles. However, detergents are imperfect at replicating biological membranes, due to factors such as detergent molecule exchange rate, varying levels of chemical harshness, and the removal of native lipids (110). As a result, for any given membrane protein purification, a series of detergents are tested empirically to determine the best one for optimal protein extract quantity, solubility, activity, homogeneity, and stability (111). In summary, for a given membrane protein structure target, it is often required to optimize the construct, expression, and purification conditions before any structural determination technique.

To date, X-ray crystallography is the most used experimental technique for determining the structure of macromolecules. Out of ~185,000 structures deposit in the Protein Data Bank, over 161,000 have been solved by X-ray crystallography (112). Briefly, the goal of X-ray crystallography is to obtain a 3D molecular crystal from a protein crystal (113). The protein crystal is exposed to an X-ray source and the X-rays are scattered onto an electron detector, forming a diffraction pattern. The diffraction provides information on the amplitudes of the scattered X-rays, and a follow-up experiment determines the phases of the scattered X-rays. With the amplitudes and phases of the scattered X-rays, an electron-density map can be generated and used for modeling atomic coordinates.

In recent years, cryogenic electron microscopy (cryo-EM) single-particle analysis (SPA) has become a powerful technique for determining the structure of macromolecules. Cryo-EM SPA entails plunge-freezing protein specimens in a vitreous environment to trap proteins in its natural state in (ideally) random orientation. With a transmission electron microscope, an electron beam is passed through the sample and onto a detector, forming a two-dimensional projection. By taking projections of the sample from many different angles and computationally averaging them, a three-dimensional map is generated that is used for structural modeling. Although cryo-EM techniques have been used for over four decades, only recent technological advances in direct electron detectors, computer algorithms that correct for beam-induced

movement and specimen drift, and software for image analysis have enabled high-resolution structure determination (114).

This chapter presents my progress towards obtaining the structures of DHHC9, GCP16, and DHHC9-GCP16. For both DHHC9 and GCP16, I sought to identity a structure candidate by screening orthologs and protein engineering. I optimized expression and purification conditions to obtain sufficient quantities of pure protein for crystallization studies. I report my exploration of DHHC9-GCP16 fusion proteins, before focusing on a DHHC9 and GCP16 co-expression approach. DHHC9-GCP16 complex was the emphasis of extensive crystallization trials, though I was unable to obtain well-ordered crystals. Lastly, I present my progress towards solving the structure by cryo-EM.

Results

The Search for a DHHC9 Structural Candidate

To solve the structure of DHHC9, I sought to find a DHHC9 candidate that was most amenable to structural studies. Previous work in the Linder Lab suggested that human DHHC9 (huDHHC9) exhibits low expression and is technically challenging to purify (69). To identify a more stable DHHC9, I used a bioinformatic approach to generate a set of DHHC9 orthologs, with the goal of screening for the best protein behavior. I used the huDHHC9 amino acid sequence as a query for the Uniprot Protein Basic Local Alignment Search Tool (BLAST) (115). Hits were identified based on percent sequence identity. To avoid overlap with the closely related DHHC14 sequence (59% sequence identity to DHHC9 in humans), a percent identity threshold of 70% was selected. The data was filtered by species, and alternative isoforms from the same species were removed. Lastly, pair-wise identity was compared to ensure that each ortholog was similar to human DHHC9, but distinct from one another by at least 8% i.e., two orthologs can both be 85% similar to huDHHC9, but they are >8% difference from each other.

The final set consisted of 12 orthologs (Table 3.1). Additionally, for each ortholog, the catalytic DHHC cysteine was mutated to serine (referred to as DHHS) to generate the catalytically inactive mutant, with the purpose to decrease protein heterogeneity due to palmitoylation of this residue.

1: D9_frog	100.00	82.84	83.09	84.64	82.23	81.23	77.56	81.74	73.04	76.81	76.22	74.78
2: D9_coral_snake	82.84	100.00	86.90	83.05	85.09	83.29	78.40	82.11	69.10	73.22	73.45	72.83
3: D9_chameleon	83.09	86.90	100.00	84.42	85.34	84.10	78.34	82.38	71.79	74.00	73.37	74.49
4: D9_chicken	84.64	83.05	84.42	100.00	85.26	85.30	85.95	89.61	73.04	75.94	73.56	76.47
5: D9_human	82.23	85.09	85.34	85.26	100.00	93.41	82.04	87.61	72.57	74.71	75.44	75.52
6: D9_opossum	81.23	83.29	84.10	85.30	93.41	100.00	81.08	85.88	72.01	74.34	73.70	74.34
7: D9_sea_turtle	77.56	78.40	78.34	85.95	82.04	81.08	100.00	85.09	65.23	69.21	67.56	69.23
8: D9_alligator	81.74	82.11	82.38	89.61	87.61	85.88	85.09	100.00	70.74	74.43	73.45	74.78
9: D9_arowana	73.04	69.10	71.79	73.04	72.57	72.01	65.23	70.74	100.00	81.18	78.27	80.81
10: D9_spotted_gar	76.81	73.22	74.00	75.94	74.71	74.34	69.21	74.43	81.18	100.00	84.04	84.45
11: D9_zebrafish	76.22	73.45	73.37	73.56	75.44	73.70	67.56	73.45	78.27	84.04	100.00	85.03
12: D9_pufferfish	74.78	72.83	74.49	76.47	75.52	74.34	69.23	74.78	80.81	84.45	85.03	100.00

Table 3.1. DHHC9 Orthologs Percent Identity Matrix. DHHC9 ortholog sequence alignment and percent identity matrix generated using Clustal 2.1. Codename and corresponding Uniprot ID are listed in the experimental procedures.

To screen for the protein behavior of the DHHC9 orthologs, I used fluorescence-detection size-exclusion chromatography (FSEC) (96). Each ortholog was cloned into Kawate Lab vector pCGFP-EU2 and expressed in human embryonic kidney (HEK) cells. Cells were solubilized in dodecyl-maltoside (DDM)-containing buffer, and cleared lysates were applied to a gel-filtration column coupled to a fluorometer to monitor fluorescence (Fig 3.1). The DHHC9 peak was determined based on comparison with frog DHHC20 (frDHHC20) as a positive control, as my and the Linder Lab's prior data with DHHC20 supports it is folded and possesses enzymatic activity (data not shown). Additionally, a GFP only sample was used to assign any signal due to free GFP. The FSEC profile for huDHHS9 showed an expected peak at ~24 min, with an additional shoulder at ~21.5 min that may indicate another oligomeric state (Fig 3.1 A). Notably, huDHHS9 showed a prominent void peak at ~17 min and lower molecular weight species at ~29 min, characteristic of protein aggregation and degradation, respectively (Fig 3.1 A). Many of the orthologs exhibited an FSEC profile that indicated a heterogeneous DHHS9 protein with low abundance, protein aggregation, and degradation including opossum DHHS9, chicken DHHS9, sea turtle DHHS9, alligator DHHS9, spotted gar DHHS9, pufferfish DHHS9, and arowana DHHS9 (Fig1.1 B, D, E, G, I, J, K, L). Fortunately, frog DHHS9 (frDHHS9), coral snake DHHS9 (csDHHS9), chameleon DHHS9 (chDHHS9), and zebrafish DHHS9 (zfDHHS9) (referred to collectively as first-pass orthologs) showed increased expression, less aggregation, and increased monodispersity (Fig 3.1 C, F, H, J). Strikingly, compared to a measured intensity of 800 arbitrary fluorescence units for huDHHS9, the first pass orthologs showed intensities of between 2500 to 4500 (Fig 1.1). These first pass orthologs identified in the initial HEK cell screens were prepared for expression and protein behavior tests in Sf9 insect cells.

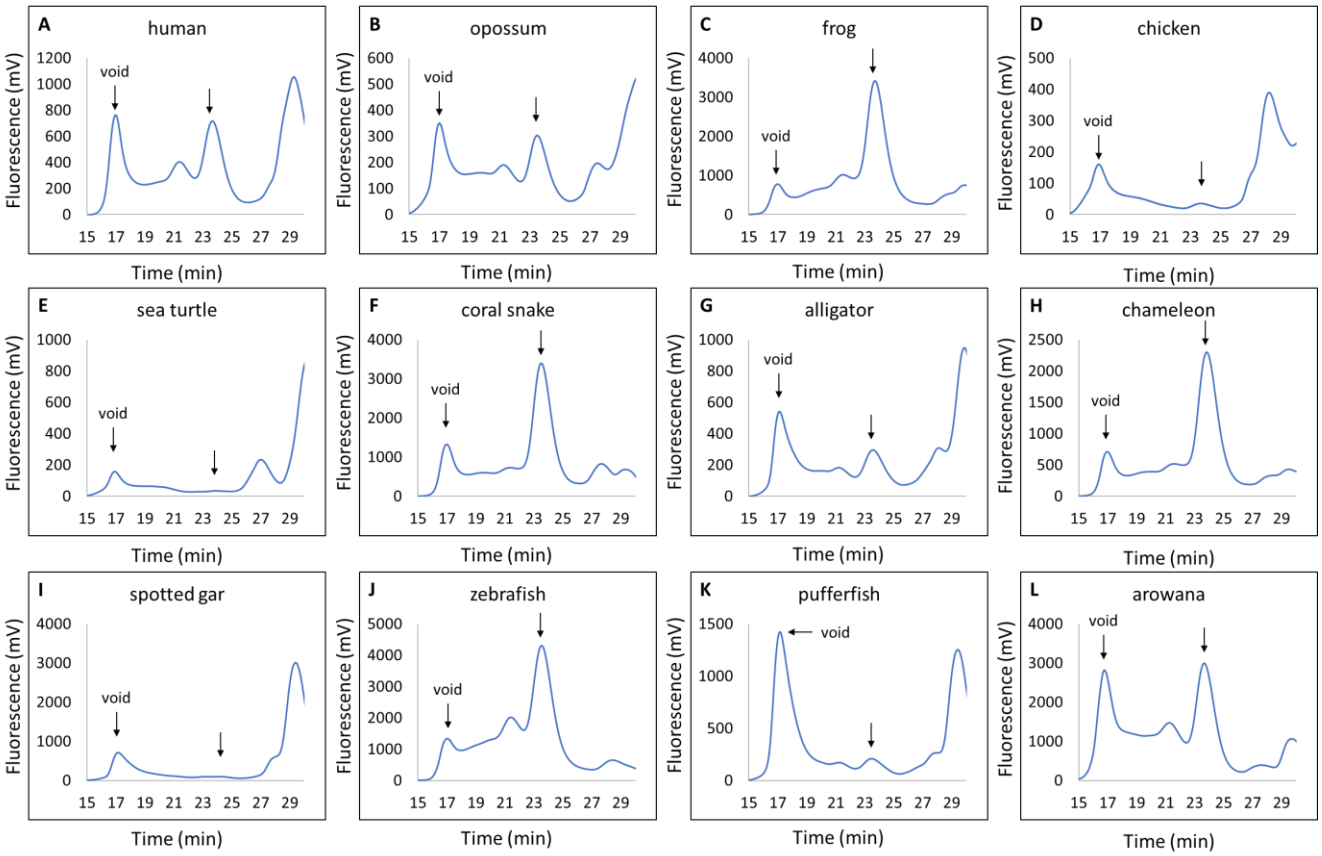

Figure 3.1. DHHC9 Ortholog Screen. (A-L) Representative FSEC profiles of DHHC9 orthologs. HEK cells were transfected with vector to express a C-terminal GFP fusion with DHHC9 protein sequence derived from the indicated organisms. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the DHHC9 retention time.

I assessed the expression and protein behavior of each first pass ortholog after expression in insect cell culture. Protein was expressed in Sf9 cells infected with baculovirus encoding C-terminal GFP-strep-tagged frDHHS9, csDHHS9, chDHHS9, or zfDHHS9. The protein was purified from membranes by triton X-100 solubilization and strep-affinity purification as described in the experimental procedures. GFP was removed by incubation with thrombin protease prior to size-exclusion chromatography (SEC) (Fig 3.2). Unexpectedly, the SEC profiles of every ortholog purified showed extensive aggregation, degradation as indicated by free GFP and lower molecular weight species, and the absence of the expected DHHC9 peak (Fig 3.2 A-D).

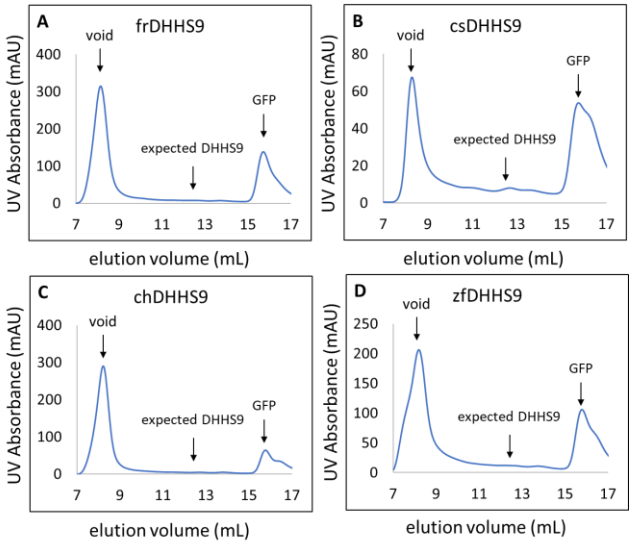

Figure 3.2. Purification of DHHS9 First Pass Orthologs. SEC profiles of the indicated first pass DHHS9 orthologs (A-D). Protein was expressed in Sf9 cells infected with baculovirus encoding C-terminal GFP-strep-tagged frDHHS9, csDHHS9, chDHHS9, or zfDHHS9. The protein was purified by membrane preparation, triton X-100 solubilization, strep-affinity purification, and size-exclusion chromatography (SEC) as in the experimental procedures. Arrows indicate the void, expected DHHC9 peak, and GFP peak.

To identify the source of DHHC9 instability, frDHHS9 was semi-arbitrarily chosen to test for factors specific to the purification protocol that may contribute to protein aggregation. The aggregation occurred regardless of whether the membrane preparation was freeze-thawed or the protein was concentrated, treated with thrombin, or treated with EndoH (data not shown). To determine if aggregation occurred at solubilization, a detergent screen was conducted (Fig 3.3). Different detergents can have varying degrees of 'harshness' based on the polarity of the headgroup and acyl chain length (110). A set of detergents were selected based on those that have been successful in our lab and for other membrane proteins (110). frDHHS9 was expressed in HEK cells, and lysates were prepared by solubilization in either triton X-100, DDM, decyl-maltoside (C10M), nonyl-glucoside (C9G), or lauryldimethylamine oxide (LDAO). Solubilized samples were analyzed by FSEC in a DDM-containing running buffer. As observed in prior screens, DDM solubilization resulted in a distinct DHHC9 peak at ~24 min (Fig 3.3 blue trace). Furthermore, the FSEC profiles of frDHHS9 in HEK cells (Fig 3.1 C) compared to that in Sf9 cells (Fig 3.3 blue trace) were almost identical, supporting that conclusions based on FSEC in one expression system could be applicable to both. Solubilization with C10M resulted in a comparable profile, though with about a 30% decrease in yield compared to solubilization with DDM (Fig 3.3 gray trace). C9G, LDAO, and, most critically, triton X-100 resulted in predominantly aggregated and high molecular weight species (Fig 3.3 yellow, light blue, and orange traces). This result explained the predominant aggregation observed during purification using triton X-100 for solubilization. Moving forward, I focused on frDHHS9 as the candidate towards protein crystallization trials.

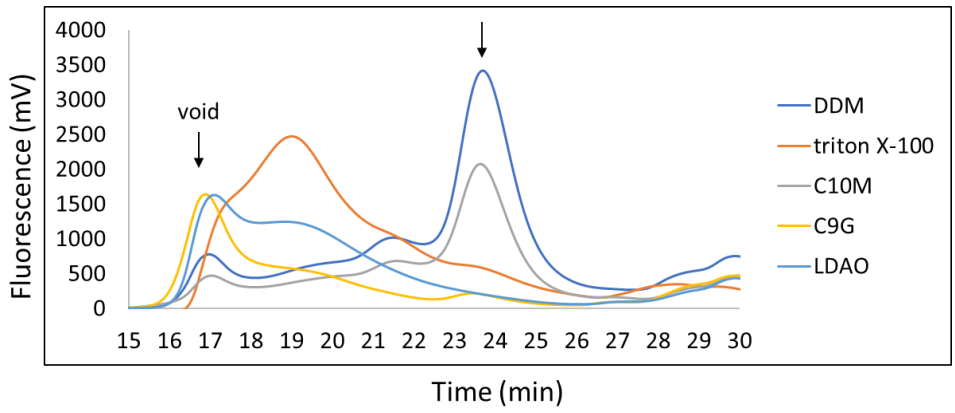

Figure 3.3. frDHHS9 Detergent Screen. Overlay of FSEC profiles of frDHHS9 solubilized with the indicated detergents including dodecyl maltoside (DDM), triton X-100, decyl-maltoside (C10M), nonyl-glucoside (C9G), and lauryldimethyl amine oxide (LDAO). HEK cells were transfected with vector to express frDHHS9 with a C-terminal GFP. Cells were solubilized in detergent-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. The running buffer contained DDM. Arrows indicate the void retention time and the DHHC9 retention time.

Protein Engineering frDHHS9 for Improved Expression and Stability

I aimed to improve the expression and monodispersity of frDHHS9 to improve the likelihood of obtaining protein crystals. Using information from bioinformatic databases (115, 116) including sequence alignments, amino acid conservation, and secondary structure prediction, I generated frDHHS9 deletion constructs to remove potential aggregation-prone flexible regions (Fig 3.4). Truncations were made at regions of lower conservation and, when possible, at glycine residues to minimize deleterious effects. The protein behavior of the deletion constructs was assessed by expression in Sf9 cells and FSEC (Fig 3.5). From the FSEC profiles, the effects of the deletions can be categorized into three groups: N-terminal, C-terminal, and combinatorial deletions. Compared to full-length (FL) frDHHS9 (Fig 3.5 A), N-terminal deletions N1 and N2 showed marked improvements in the intensity of the DHHC9 peak and monodispersity; both N1 and N2 showed an increase in intensity by over 170% and 250%, respectively (Fig 3.5 B and C). By comparison, C-terminal deletion constructs showed milder effects (Fig 3.5 D-F). Construct C1, which removed a long amino acid stretch with no predicted structure, showed a decrease in intensity by ~60% with an overall profile like that of FL (Fig 3.5 D). Construct C2 showed reduced aggregation and increased monodispersity with a comparable level of intensity to FL (Fig 3.5 E). Construct C3 showed a 50% increase in intensity and improved monodispersity (Fig 3.5 F). Interestingly, combinatorial deletions suggest that there is an additive effect when truncating both termini, with an apparent greater contribution from the N-terminal deletion (Fig 3.5 G-L). Overall, construct N2C2 showed the highest increase in intensity at over 300% increase, compared to FL, and a high degree of monodispersity (Fig 3.5 K). Compared to human DHHS9 in Sf9 cells, frDHHS9 N2C2 is over 40-fold higher in intensity (data not shown). frDHHS9 N2C2 was selected as the primary structure candidate.

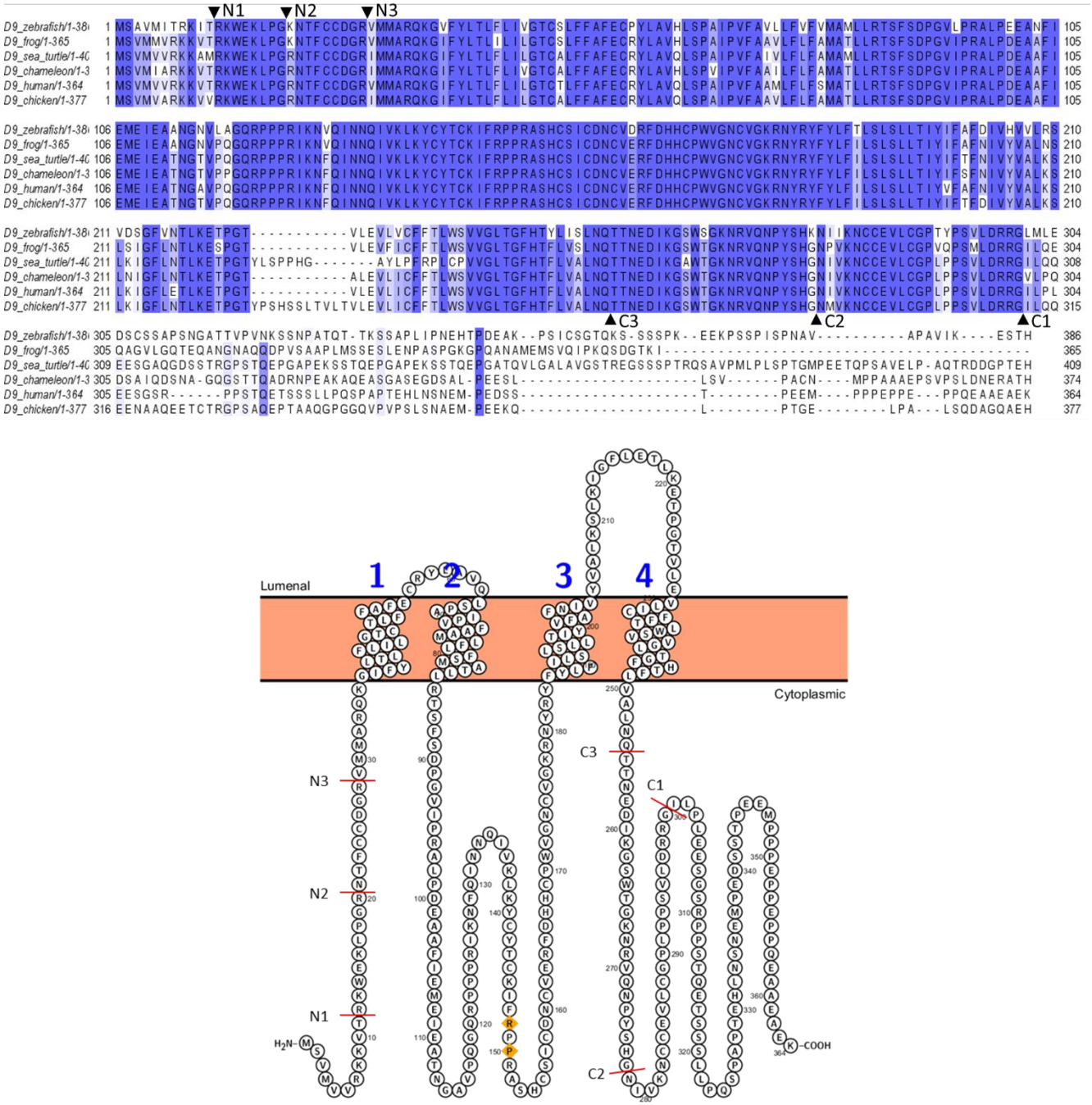

Figure 3.4. Designing DHHC9 Deletion Constructs. (A) Amino acid sequence of human DHHC9 compared to various orthologs. Amino acids highlighted in dark blue are completely conserved, in light blue are somewhat conserved, and in white are not conserved. Arrows indicate where the designated deletions were made. (B) Toplogy of DHHC9 based on the

Protter database. Red lines denote where the indicated constructs start and end.

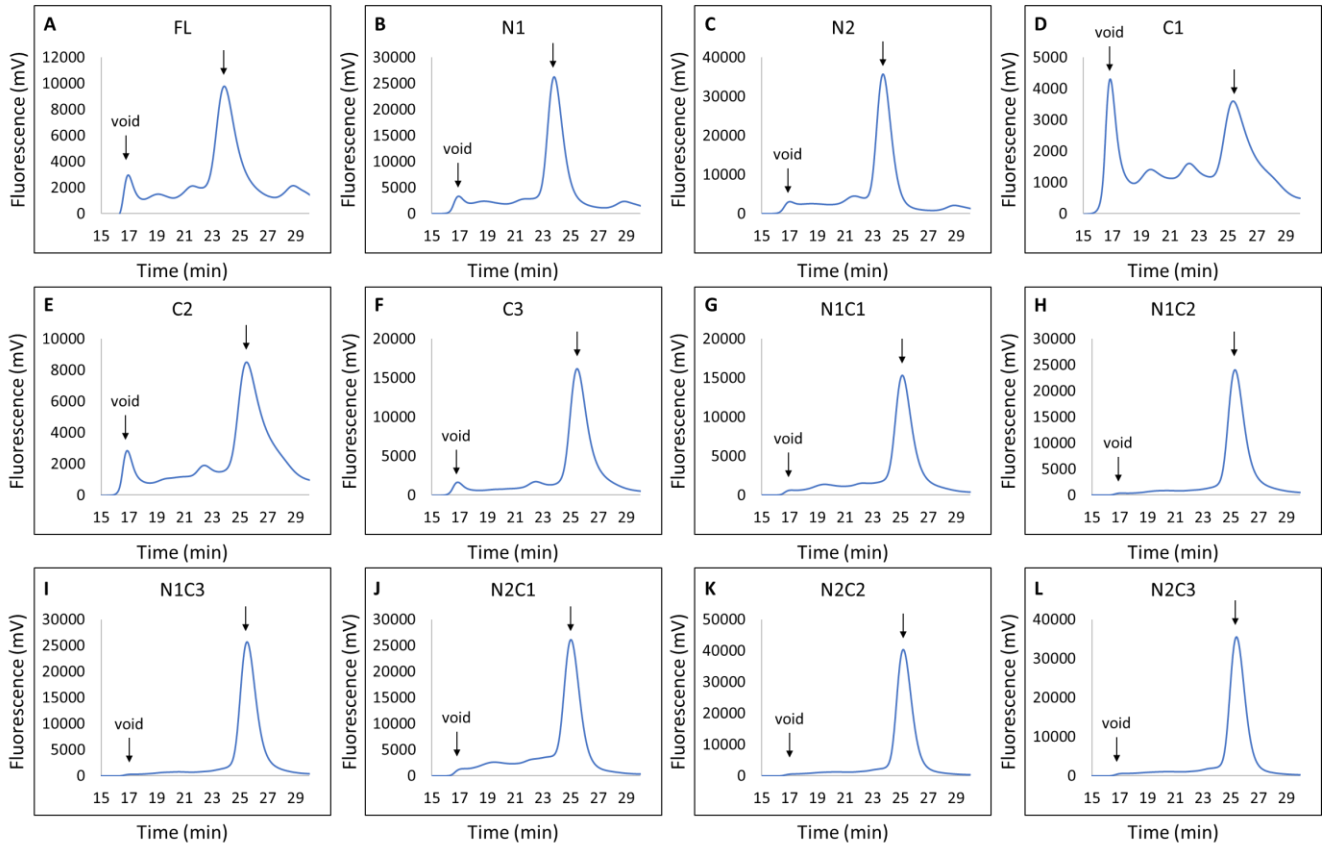

Figure 3.5. Screening frDHHS9 Deletion Constructs. Representative FSEC profiles of frDHHS9 constructs (A-L). Sf9 cells were infected with recombinant baculovirus particles and incubated for 48 hours at 27C. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the DHHC9 retention time.

frDHHS9 Protein Purification and Crystallization Challenges

After identifying frDHHS9 N2C2, I purified the protein in large scale with the goal of advancing through protein crystallization trials. Once I determined the optimal expression conditions using protein expression trials in Sf9 cells (data not shown), I expressed and purified frDHHS9 N2C2-GFP via strep-affinity chromatography and SEC (Fig 3.6). Prior to SEC, GFP was removed by incubation with thrombin protease. SEC analysis showed a clear peak corresponding to frDHHS9 N2C2 centered at ~12.5 mL (Fig 3.6 A). However, SDS-PAGE Coomassie analysis of the DHHC9 peak fractions after SEC revealed additional cleavage products with a significant product at 17 kDa (Fig 3.5 B). With help from Wendy Greentree of the Linder Lab, I identified a non-canonical thrombin protease site centered around R96 that was not detected using previous protease cut prediction software. To prevent non-specific cleavage by thrombin protease, mutations were made to disrupt the thrombin recognition site. frDHHS9 N2C2-GFP constructs were generated where the essential arginine in the thrombin recognition site was mutated to either lysine or alanine. FSEC showed R96K exhibited higher expression and monodispersity compared to R96A (Fig 3.6 C). Therefore, frDHHS9 N2C2 R96K-GFP was expressed in Sf9 cells, purified via strep-affinity purification, and analyzed by SEC (Fig 3.6 D). After thrombin digestion, frDHHS9 N2C2 R96K exhibited a broader peak on SEC compared to the parent construct (Fig 3.6A). More importantly, mutation of the thrombin recognition site did not completely abolish the non-specific cleavage of frDHHS9 (Fig 3.6E). Subsequent GFP fusion proteins were designed with TEV protease rather than thrombin protease.

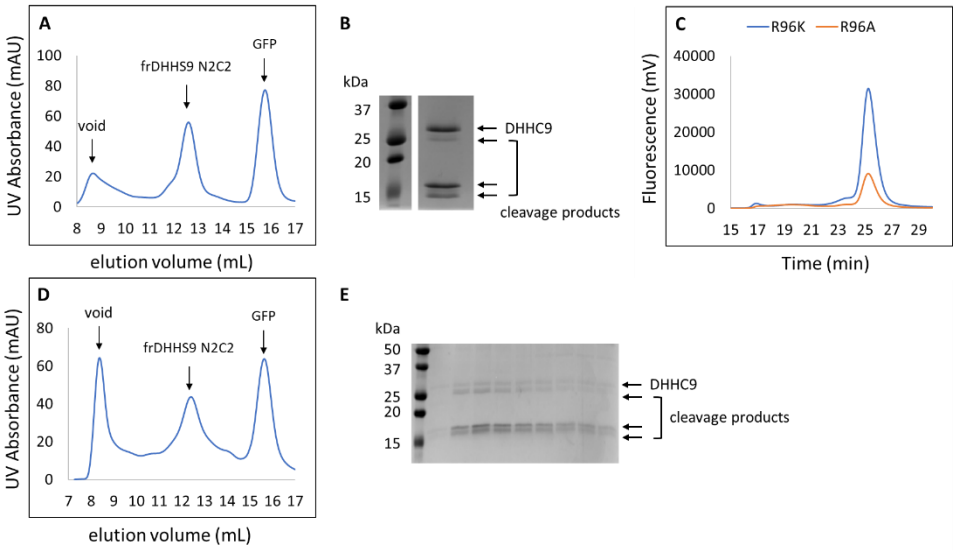

Figure 3.6. frDHHS9 N2C2 is Non-Specifically Cleaved by Thrombin Protease. Sf9 cells were infected with recombinant baculovirus encoding C-terminally strep-GFP-tagged frDHHS9 N2C2 with or without mutations R96K or R96A. Protein was purified via strep-affinity purification and SEC as in the methods. (A) SEC profile for thrombin-digested frDHHS9 N2C2. (B) SDS-PAGE Coomassie analysis of pooled frDHHS9 peak fractions from (a). Black arrows indicate the expected DHHS9 size and cleavage products. (C) FSEC profiles for CGFP frDHHS9 N2C2 R96 Mutants. (D) SEC profile for thrombin-digested frDHHS9 N2C2 R96K. (E) SDS-PAGE Coomassie analysis of frDHHS9 peak fractions from (d). Black arrows indicate the expected DHHS9 size and cleavage products.

With the insights I have obtained up to this point, I sought to purify frDHHS9 N2C2 to advance through protein crystallization trials. frDHHS9 was expressed and purified from Sf9 cells by strep-affinity chromatography and analyzed by SEC. The SEC profile showed frDHHS9 eluted at the expected elution volume for folded DHHC9, and SDS-PAGE Coomassie analysis showed sufficient purity for crystallization (Fig 3.7 A-B). One challenge that remained was low yield due to inefficient binding to the streptactin resin (Fig 3.7E). Nonetheless, for every 1-2L culture of Sf9 cells, the amount of frDHHS9 N2C2 remaining after SEC was sufficient for 1-2 96-well crystallization screens at a relatively low protein concentration between 1-2mg/mL. I repeatedly purified frDHHS9 N2C2 and advanced it through the Kawate Lab FK and other crystallization screens that included a range of polyethylene glycols (PEGs) (10-40%), pH (4.5-9.5), and distinct salts/additives known to facilitate membrane protein crystallization. Despite extensive screening, I was only able to obtain amorphous quasi crystals of frDHHS9 N2C2 that were not of sufficient quality for good X-ray diffraction (Fig 3.8).

Additionally, I tested whether human DHHS9 N2C2 would purify with sufficient quality for crystallization, based on my observations with frDHHS9. In contrast to frDHHS9 N2C2, purification and SEC analysis of huDHHS9 N2C2 prior to GFP cleavage showed primarily a single aggregation peak corresponding to GFP-huDHHS9 N2C2 (Fig 3.7 C-D). This result suggested that huDHHS9 N2C2 was not suitable for crystallization trials.

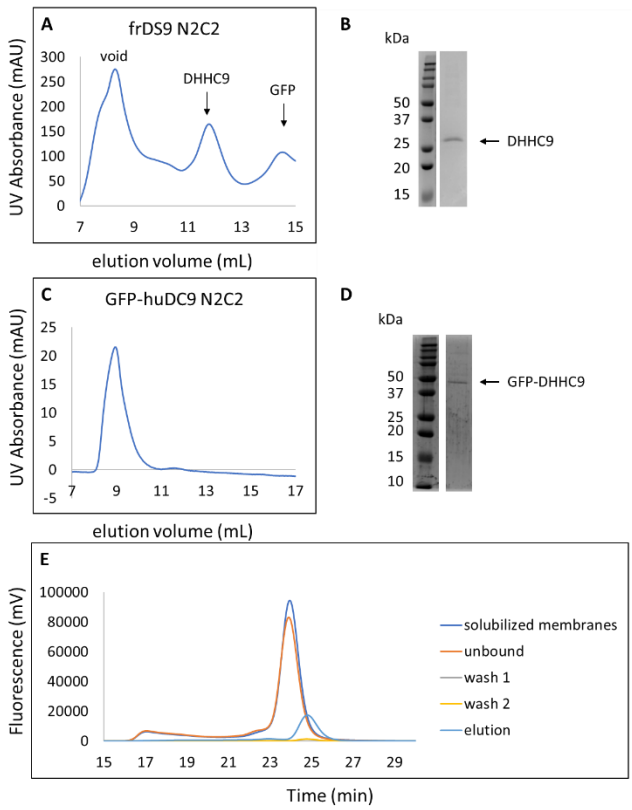

Figure 3.7. Purification of frog and human DHHC9 N2C2 Constructs. Sf9 cells were infected with recombinant baculovirus encoding the frDHHS9 or huDHHS9. Protein was purified via strep-affinity chromatography and SEC as in the experimental procedures. (A) SEC profile for GFP-frDHHS9 N2C2 post treatment with TEV protease. (B) SDS-PAGE Coomassie analysis from the DHHC9 fraction in (a). (C) SEC profile for GFP-huDHHS9 N2C2 and (D) the corresponding SDS-PAGE Coomassie analysis of the main peak after SEC. (E) Overlay of FSEC traces from frDHHS9 N2C2 purification in (a) with the indicated purification fractions after normalization by volume.

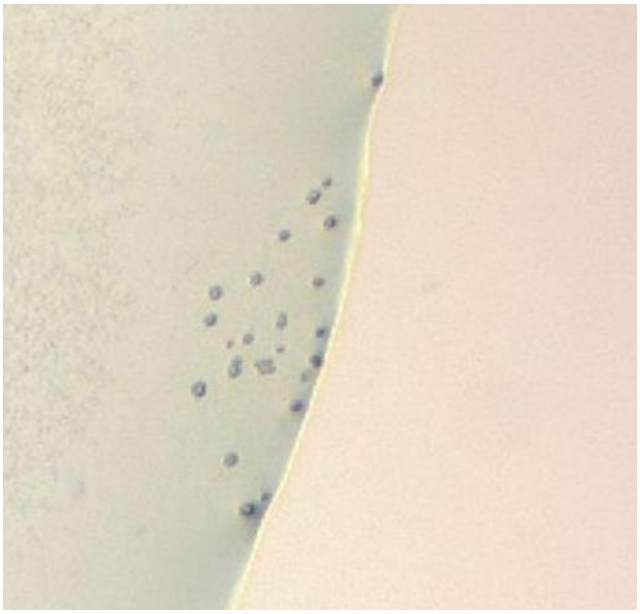

Figure 3.8. frDHHS9 N2C2 Forms Amorphous Quasi Crystals. Representative amorphous crystals after ~14 days in 96-well screens set-up using 1-2 mg/mL.

Around this time, my work with human DHHC9 and GCP16 prompted a reassessment of the DHHC9 structure candidate. My experiments investigating the effect of GCP16 co-expression on DHHC9 stability showed that GCP16 co-expression had little to no effect on DHHC9 N2C2 stability (discussed in Chapter 2). Therefore, I reasoned that an extended C-terminus is required for GCP16 interaction. Since my goal is to better understand the structural basis of DHHC9-GCP16 interaction and function, I re-examined my extended C-terminal constructs frDHHS9 N2, N2C1, and I compared them to construct N2C2 using FSEC in Sf9 cells (Fig 3.9). To further optimize expression and to try to alleviate my observed issues with resin binding (Fig 3.7 E), I tested the protein behavior of N- and C-terminal GFP tags. The FSEC profiles of frDHHS9 N2 and N2C1 (Fig 3.9 A-B) showed the protein was comparable in magnitude to frDHHS9 N2C2 (Fig 3.9 C). Interestingly, C-terminal GFP fusion (Fig 3.9 A-C) consistently showed high expression compared to N-terminal GFP fusion (Fig 3.9 A- C blue traces). Between the two extended C-terminus constructs, frDHHS9 N2 was more heterogenous compared to frDHHS9 N2C1, as indicated by an increased protein signal in the void and higher molecular weight species at 22-23 min in the former (Fig 3.9 A-B). Though C-terminal GFP fusion showed increased expression, it was previously noted that frDHHS9 N2C2-GFP showed inefficient binding to affinity resin (Fig 3.7 E). Therefore, in parallel, both N- and C-terminal frDHHS9 N2C1 were expressed in Sf9 cells and purified by affinity chromatography and SEC (Fig 3.10). Sf9 cells expressing C-terminal GFP-tagged frDHHS9 N2C1 showed higher protein expression compared to N-terminal GFP-tagged as indicated by the intensity of GFP fluorescence 72 hours post infection with recombinant baculovirus (Fig 3.10 A). However, even though the C-terminal construct expressed better, no protein visibly bound to affinity resin compared to the N-terminally tagged construct (Fig 3.10 B). The poor binding efficiency observed qualitatively with frDHHS9 N2C1-GFP was validated by FSEC (data not shown). Therefore, N-terminally GFP-tagged frDHHS9 was applied to an SEC column prior to removal of GFP. Injection of GFP-frDHHS9 N2C1 resulted in increased pressures in the instrument until

the column pressure was maxed out and the run had to be aborted (Fig 3.10 C). I reasoned that NGFP-frDHHS9 N2C1 contains an exposed aggregation-prone region in the C-terminus; so, I tried to purify NGFP-frDHHS9 N2, with the idea that a full-length C-terminus may be more amenable to purification. NGFP-frDHHS9 N2 was expressed in Sf9 cells and purified by affinity chromatography and SEC without removal of GFP (Fig 3.10D). The SEC profile revealed NGFP-frDHHS9 N2 elutes in a broad peak that overlaps with the void volume (Fig 3.10D). Additionally, SDS-PAGE Coomassie analysis of NGFP-frDHHS9 elution and post SEC fractions showed the protein to be completely aggregated during purification (Fig 3.10 E). To potentially alleviate aggregation during detergent solubilization, styrene maleic acid copolymers (SMAs) were extensively investigated for their potential to solubilize various frDHHS9 constructs with the idea of retaining the protein in their native lipid environments. However, no combinations of SMAs tested (2:1, 3:1, 1.4:1) were able to show sufficient yields of folded protein after SEC (data not shown). In the end, no extended frDHHS9 construct was able to be purified or advanced through protein crystallization trials with approaches at the time.

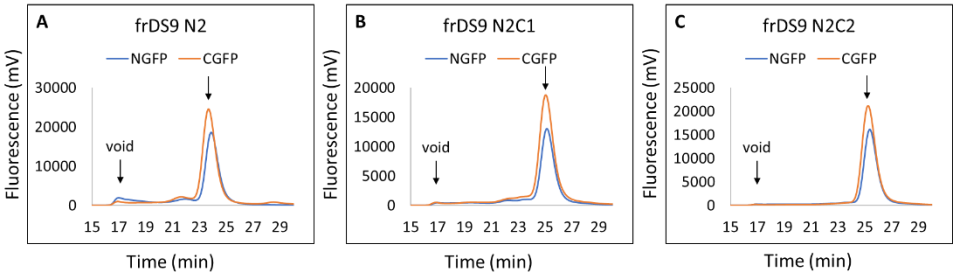

Figure 3.9. CGFP-tagged DHHC9 Consistently Shows a Modest Improvement in Expression. (A-C) FSEC profiles for the indicated frDHHC9 constructs with N- and C-terminal GFP fusion shown in blue and orange, respectively. Sf9 cells were infected with recombinant baculovirus particles and incubated for 48 hours at 27C. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the DHHC9 retention time.

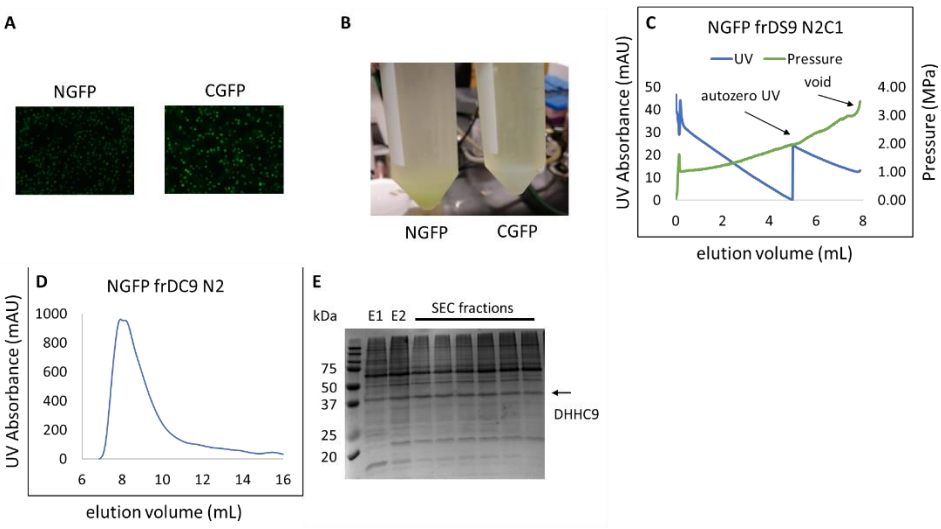

Figure 3.10. frDHHS9 Extended C-terminus Constructs Aggregate When Applied to SEC. Sf9 cells were infected with recombinant baculovirus encoding the indicated frDHHS9 constructs. Protein was purified via immobilized metal affinity chromatography (IMAC) purification and SEC as in the experimental procedures. (A) Light microscopy image (GFP filter) of Sf9 cells 72 hours after infection with baculovirus encoding N- and C-GFP-tagged frDHHS9 N2C1 constructs. (B) Image showing the extent of frDHHS9 N2C1-GFP fusion protein bound after 2-hour incubation with Ni-NTA resin. (C) Incomplete SEC profile for IMAC-purified and uncleaved GFP-frDHHS9 N2C1. (D) SEC profile for IMAC-purified and uncleaved GFP-frDC9 N2. (E) SDS-PAGE Coomassie analysis showing elution fractions and fractions corresponding to the aggregation peak from (d).

Identification of a GCP16 Structure Candidate

GCP16 is a 16 kDa peripheral membrane protein that is membrane-bound via two acylated cysteines at amino acids 69 and 72 (80). Initially, I asked whether GCP16 could be expressed in a bacterial expression system, where the S-acylation machinery is absent, to determine whether GCP16 would purify like a soluble protein; however, these experiments yielded low amounts of impure GCP16 (data not shown). Thus, I applied an approach similar to my approach with DHHC9 to identify a GCP16 structure candidate. I used the human GCP16 amino acid sequence as a query for Uniprot Protein BLAST (115). Hits were identified based on percent sequence identity, and candidates were refined based on species, alternative isoforms, and pair-wise identity. GCP16 is highly conserved across a broad range of metazoans; after refinement, the final set consisted of six distinct GCP16 orthologs (Table 3.2).

1: pufferfish	100.00	86.01	79.26	83.21	81.75	81.02
2: zebrafish	86.01	100.00	82.22	80.88	80.15	80.88
3: african_frog	79.26	82.22	100.00	82.96	85.19	85.93
4: chameleon	83.21	80.88	82.96	100.00	89.78	88.32
5: human	81.75	80.15	85.19	89.78	100.00	94.89
6: zebra_finch	81.02	80.88	85.93	88.32	94.89	100.00

Table 3.2. GCP16 Orthologs Percent Identity Matrix. GCP16 ortholog sequence alignment and percent identity matrix generated using Clustal 2.1. Codename and corresponding Uniprot ID are listed in the experimental procedures.

For each GCP16 ortholog, protein behavior was assessed by expression in Sf9 cells and FSEC analysis (Fig 3.11). All GCP16-GFP constructs showed well-behaved protein corresponding to a retention time of ~27 min (Fig 3.11). Zebra finch GCP16 (zfGCP16), chameleon GCP16 (chGCP16), and human GCP16 (huGCP16) showed higher homogeneity (Fig 3.11 A, B, F), while African frog GCP16 (afGCP16), pufferfish GCP16 (pfGCP16), and zebrafish (zeGCP16) showed an additional shoulder species at 24 min (Fig 3.11 C, D, E). One possibility is that FSEC may be skewed towards the behavior of GFP (27 kDa) due to the relatively small size of GCP16 (16 kDa). Therefore, I reassessed the protein behavior of each GCP16 after removal of GFP. From Sf9 cells, I expressed and purified each GCP16 ortholog by strep-affinity purification and SEC (Fig 3.12). SDS-PAGE Coomassie analysis showed each elution fraction contained a band corresponding to the size of GCP16-GCP16 (42 kDa), and, upon treatment with thrombin protease, the expected bands for GFP (27 kDa) and GCP16 (16 kDa) were present (Fig 1.12 A-B). SEC analysis showed zfGCP16 and huGCP16 primarily elutes at the expected retention volume at ~13.5 mL (Fig 3.12 C and H). Though the expected GCP16 peak was identified in the profiles of all orthologs, chGCP16, afGCP16, pfGCP16, and zeGCP16 showed multiple protein peaks and a high signal at the void volume, supporting that these orthologs are heterogeneous and prone to aggregation (Fig 3.12 D-G). zfGCP16 showed the highest degree of homogeneity and it exhibited sufficient levels of expression; therefore, zfGCP16 was chosen as the GCP16 structure candidate.

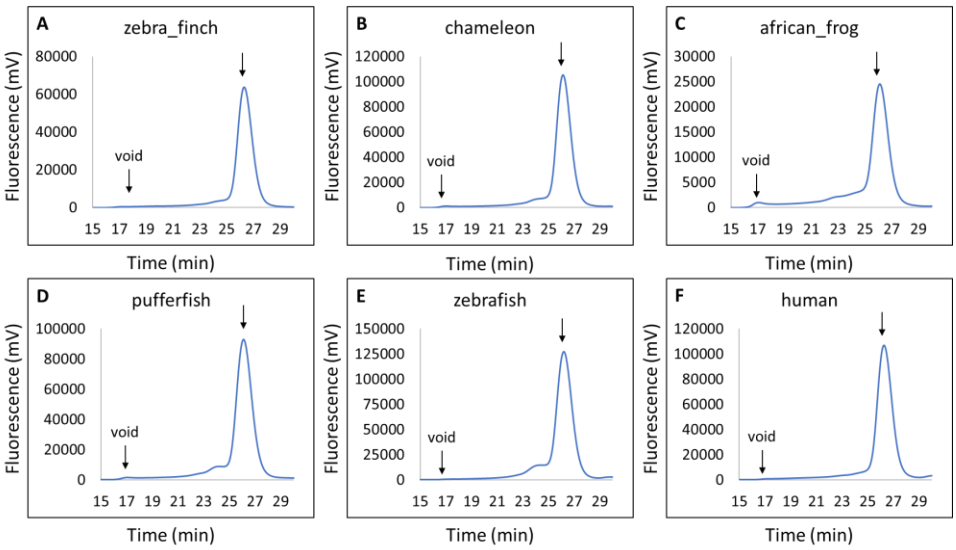

Figure 3.11. GCP16 Ortholog Screen. (A-F) Representative FSEC profiles of GCP16 orthologs. Protein was expressed in Sf9 cells infected with recombinant baculovirus encoding C-terminal GFP-tagged GCP16. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the GCP16 retention time.

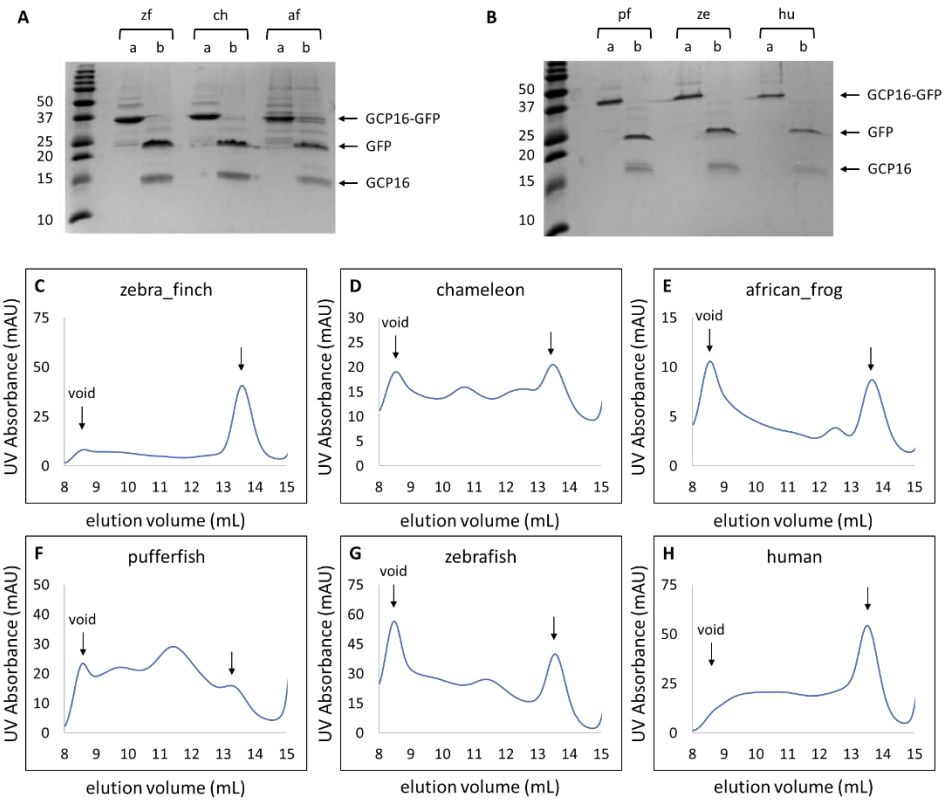

Figure 3.12. Purification of GCP16 Orthologs. Sf9 cell cultures were infected with recombinant baculovirus particles for a given ortholog and incubated for 48 hours at 27C. Protein was purified via membrane preparation, triton x-100 solubilization, and strep-affinity chromatography before subjecting to SEC. Black arrows indicate the expected elution volume for GCP16. (A-B) SDS-PAGE Coomassie analysis of elution fractions (lanes a) and elution fractions after incubation with thrombin protease (lanes b) for each ortholog after purification. (C-H) SEC chromatograms for GCP16 orthologs.

zfGCP16 Purification and Crystallization

zfGCP16 was the structure candidate to move forward with crystallization trials. I expressed zfGCP16 in Sf9 cells, and I purified the protein by strep-affinity chromatography and SEC (Fig 3.13). While a significant portion of GCP16 was lost to aggregation during purification and GFP cleavage (Fig 3.13 A), the SEC fractions corresponding to the GCP16 showed sufficiently pure zfGCP16 in adequate quantities for crystallization trials (Fig 3.13B). With repeated purifications, I screened zfGCP16 using the Kawate Lab FK and other crystallization screens that included a range of PEGs (10-40%), pH (4.5-9.5), and distinct salts/additives known to facilitate membrane protein crystallization. Between 7-21 days, I observed amorphous crystals in conditions that were generally low pH (4.6), contained PEG monomethyl ether, and chloride-containing salts (Fig 3.13 C). However, these types of protein crystals do not diffract sufficiently for structure determination. I performed additive screening, but I only ever obtained similar crystals to those above.

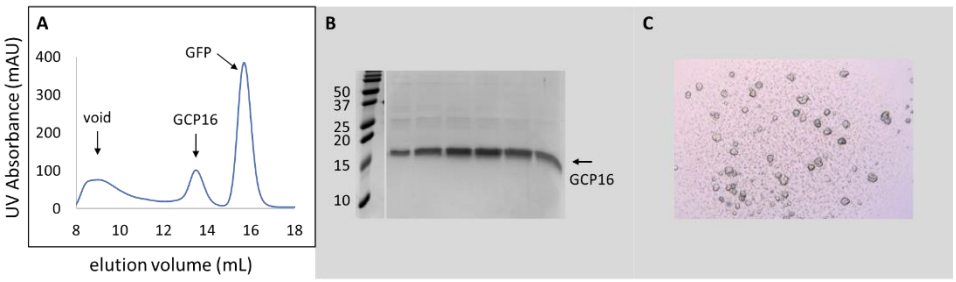

Figure 3.13. Purification and Crystallization of zfGCP16. Sf9 cell cultures were infected with recombinant baculovirus for zfGCP16 and incubated for 48 hours at 27C. Protein was purified via membrane preparation, triton x-100 solubilization, strep-affinity chromatography, and SEC as in the methods. (A) SEC profile for zfGCP16. (B) SDS PAGE Coomassie analysis of GCP16 fractions after SEC. (C) Representative amorphous crystals from crystallization trials after 2-12 weeks of growth using zfGCP16 at 2-4mg/mL.

Protein cysteine residues can affect protein crystallization due to potential non-specific disulfide bond formation and/or contributing to protein heterogeneity through post-translational modifications. GCP16 contains four cysteine residues at amino acid positions 24, 69, 72, and 81. Additionally, two of which, cysteine 69 and 72, have been showed to be acylated and affect membrane localization (80). To try to find another zfGCP16 construct for crystallization, I performed cysteine mutation analysis by generating C-terminal GFP-tagged zfGCP16 with several cysteine to serine mutations: C24S, C69S, C72S, C81S, and the C69, 72S double mutant (DM). Each construct was expressed in Sf9 cells and analyzed by FSEC (Fig 3.14). Compared to wildtype (Fig 3.14 B), C24S showed an increase in monodispersity but with less total expression (Fig 3.14C). Interestingly, C69S and C72S mutations resulted in multiple peaks on FSEC (Fig 3.14 D-E); when both residues are mutated in the DM construct (Fig 3.14 G), a single peak is observed again. C81S mutation showed a slightly more heterogeneous FSEC profile to wildtype (Fig 3.14 F). Based on these results, zfGCP16 C24S was selected as a potential construct for crystallization.

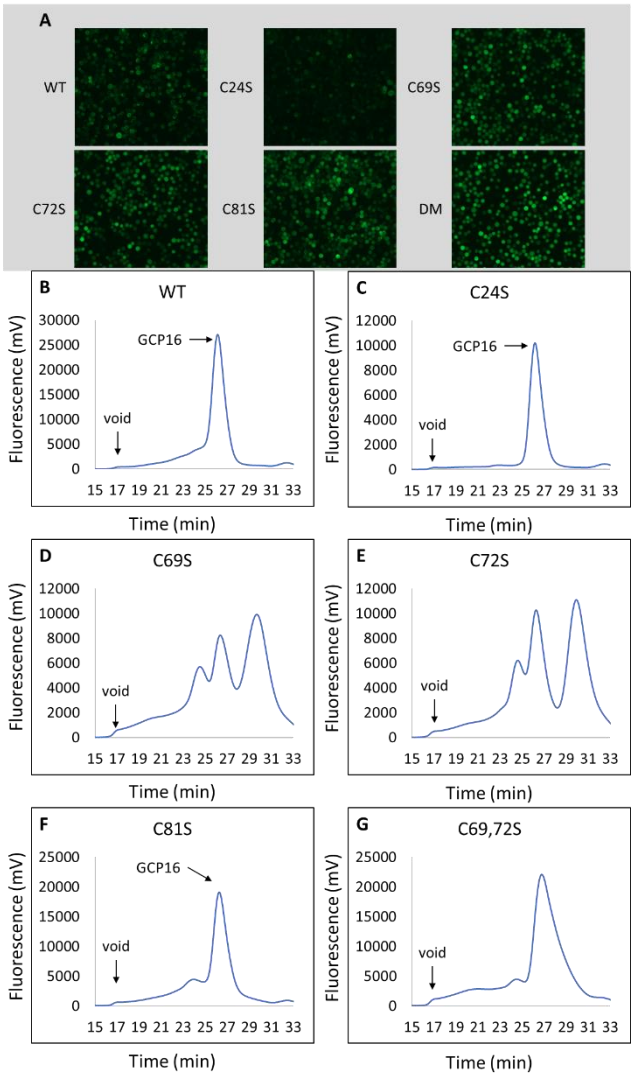

Figure 3.14. zfGCP16 Cysteine Mutation Analysis. Sf9 cells were infected with recombinant baculovirus encoding the indicated zfGCP16 constructs. (A) Light microscopy images (GFP filter) of Sf9 cells 48 hours after infection with baculovirus encoding C-GFP-tagged zfGCP16 constructs with either wildtype sequence (WT), or mutations C24S, C69S, C72S, C81S, or both C69, 72S (DM). (B-G) FSEC profiles for the indicated zfGCP16 constructs. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the primary GCP16 retention time.

To determine the protein behavior of zfGCP16 C24S after GFP cleavage, this construct was expressed and purified from Sf9 cells as in the experimental procedures. The SEC profile for zfGCP16 C24S showed two peaks at the expected retention volume for GCP16 at ~13.5 mL and multiple small degradation products between 18 to 24 mL (Fig 3.15A). SDS-PAGE Coomassie analysis revealed that the additional GCP16 peak corresponded to incomplete cleavage of GFP by thrombin protease (Fig 3.15 B). This suggests the C24S mutation may either interfere with GFP cleavage or GFP cleavage may cause destabilization of the protein. This construct was not pursued any further for crystallization.

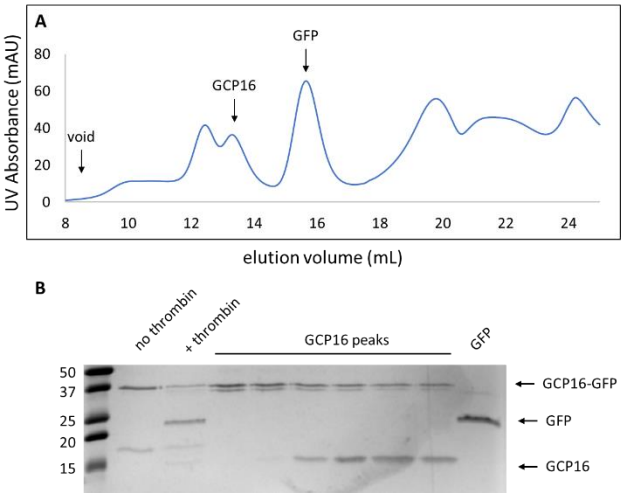

Figure 3.15. Purification of zfGCP16 C24S. Sf9 cell cultures were infected with recombinant baculovirus for zfGCP16 C24S and incubated for 48 hours at 27C. Protein was purified via membrane preparation, triton x-100 solubilization, strep-affinity chromatography, and SEC as in the methods. (A) SEC profile for zfGCP16 C24S. (B) SDS PAGE Coomassie analysis of elution pre- and post- thrombin treatment and GCP16 and GFP fractions after SEC.

Exploring DHHC9-GCP16 Fusion Proteins

Purified DHHC9 alone resulted in very prevalent aggregation, challenging my approaches up to this point. At the time, my early work with DHHC9 and GCP16 co-expression suggested GCP16 has a stabilizing effect on DHHC9, I explored whether a DHHC9-GCP16 fusion protein would be a way to purify the complex while avoiding any issues related to differing protein molar ratios between preparations. I used FSEC to compare the protein behavior of several human DHHS9-GCP16 constructs to the corresponding DHHS9 only constructs. HEK cells were transfected with vector to express N-terminally GFP-tagged DHHS9 constructs with or without C-terminal GCP16 fusion. Cells were solubilized in DDM-containing buffer, and cleared lysates were analyzed by FSEC (Fig 3.16). Compared to huDHHS9 full-length (Fig 3.16 A blue trace), huDHHS9-GCP16 fusion showed 4-fold higher expression, increased monodispersity, and decreased protein aggregation (Fig 3.16A orange trace). Similarly, fusion proteins for DHHS9 N2 and N2C1 showed drastic improvement in protein expression and homogeneity compared to the DHHS9 only counterpart (Fig 3.16 D-E). DHHS9 N2-GCP16 exhibited the highest expression overall (Fig 3.16 G), while DHHS9 N2C1-GCP16 showed the most improvement compared to the corresponding DHHS9 N2C1 construct. The huDHHS9 C1-GCP16 fusion protein showed a slight decrease in protein aggregation but the effect is not as pronounced, possibly due to low signal from this construct in general (Fig 3.16B). Interestingly, GCP16 fusion for DHHS9 constructs C2 and N2C2 resulted in drastically reduced DHHS9 signal (Fig 3.16 C, F). This may be due to GCP16 fusion to construct C2 interferes with DHHC9 protein folding and causing aggregation; this was reflected in the larger protein pellets observed during ultracentrifugation (data not shown). Together, these results show that GCP16 fusion improves DHHS9 protein levels and monodispersity for certain DHHC9 constructs.

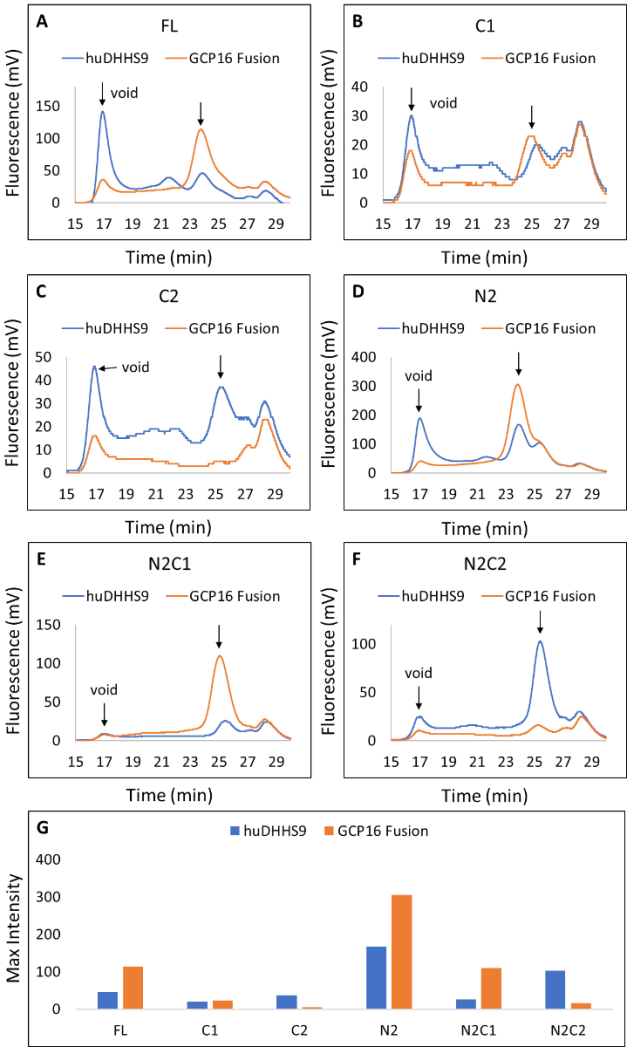

Figure 3.16. huDHHS9-GCP16 Fusion Protein Screen. (A-F) FSEC profiles for the indicated huDHHS9 constructs (blue traces) and the fusion protein counterpart (orange traces). HEK cells were transfected with vector to express the indicated N-terminal GFP-DHHC9 without or with C-terminal GCP16 fusion proteins. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. Arrows indicate the void retention time and the DHHC9 retention time. (G) Bar chart showing the max fluorescence intensity at the DHHC9 peak for constructs a-f.

Because frDHHC9 and zfGCP16 were more well-behaved compared to the human proteins, I used the result from the human DHHC9-GCP16 fusion protein screen to generate a frDHHS9 N2-zfGCP16 fusion protein with the goal of large-scale purification and protein crystallization. I verified that co-expression of zfGCP16 with frDHHS9 showed improved protein behavior compared to frDHHS9 alone, though to a lesser degree than the human orthologs (data not shown). Both N- and C-terminal GFP-tagged frDHHS9 N2-zfGCP16 were expressed in Sf9 cells and protein behavior was analyzed by FSEC (Fig 3.17 A). Though C-terminal GFP-tagged frDHHS9 N2-zfGCP16 showed higher GFP fluorescence in infect cells compared to the N-terminal GFP-tagged construct (Fig 3.17 B), approximately equal amounts were extracted in analyzed lysates (Fig 3.17 A). Thus, both N- and C-terminal GFP-tagged frDHHS9 N2-zfGCP16 were expressed in mid-scale (200 mL cultures) and purified by his-IMAC purification. To determine the efficiency of resin binding between the two constructs, purification fractions were analyzed by FSEC (Fig 3.17 C-D). Analysis of the GFP-frDHHS9 N2-zfGCP16 purification fractions showed less than 15% of available protein in the solubilized membrane fraction (Fig 3.17 C blue trace) is recovered in the final elution fraction (Fig 3.17 C yellow trace). Much of the protein remains unbound or was lost in the wash fraction (Fig 3.17 C orange and gray traces). Analysis of the C-terminal GFP-tagged construct showed over 90% of the available protein remains unbound (Fig 3.17 D blue and orange traces), and essentially no protein is detected in the final elution (Fig 3.17 D yellow trace). From purification fractions, the presence of N-terminal GFP-tagged frDHHS9 N2-zfGCP16 was verified by SDS-PAGE Coomassie and western blot (Fig 3.17 E-F), and this construct was the focus of the following large-scale purifications.

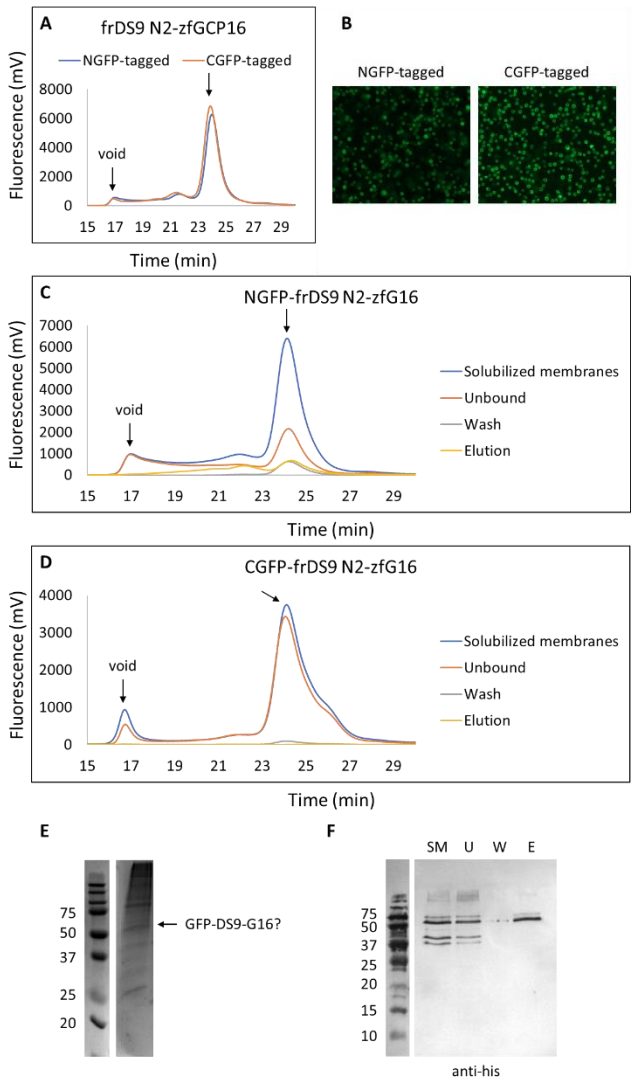

Figure 3.17. Initial Screening of frDHHS9 N2-zfG16 Fusion Protein. Representative (A) FSEC and (B) Sf9 cell expression profiles for N- and C-GFP tagged fusion protein. Sf9 cells were infected with recombinant baculovirus particles for the indicated constructs and incubated for 48 hours. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. (C) FSEC profiles of purification fractions for his-tagged NGFP- and (D) CGFP-tagged fusion proteins purified as describes in the experimental procedures. Fractions injected were normalized by volume. (E) SDS-PAGE Coomassie analysis of the elution fraction and (F) western blot of solubilized membrane,

unbound, wash, and elution fractions for the NGFP fusion protein. Arrows indicate the expected protein band.

GFP-frDHHS9 N2-zfGCP16 was expressed in Sf9 cells and purified by his-IMAC affinity chromatography for the purpose of using the protein in crystallization trials. However, SEC analysis of DDM-solubilized GFP-frDHHS9 N2-zfGCP16 showed extensive protein aggregation and insufficient protein detected in the expected DHHC9 elution volume (Fig 3.18 A). Since the prior FSEC profiles showed GFP-frDHHS9 N2-zfGCP16 was relatively monodispersed, I reasoned that aggregation was occurring at some point during the purification process. FSEC is done using lysates, which may still contain components that stabilize DHHC9 such as lipids and cholesterol; therefore, I hypothesized that the purification process may be removing such components, causing DHHC9 protein aggregation. To artificially create a more native environment, I purified GFP-frDHHS9 N2-zfGCP16 using buffers supplemented with brain lipids, as DHHC9 is enriched in the brain (69). SEC analysis showed GFP-frDHHS9 N2-zfGCP16 purification with supplemented brain lipids still resulted in extensive protein aggregation, as indicated by a majority of the protein eluting in the void peak and the increase in SEC column pressure (Fig 3.18 B). A similar result seen upon repeating the purification with buffers supplemented with cholesterol hemisuccinate (Fig 3.18 C). Lastly, I attempted to using SMA solubilization to extract and preserve GFP-frDHHS9 N2-zfGCP16 in its native environment (Fig 3.18 D). SEC analysis showed SMA-purified GFP-frDHHS9 N2-zfGCP16 showed a minor peak at the expected elution volume and extensive aggregation indicated by the prominent void peak and increased column pressure (Fig 3.18 D). Though SDS-PAGE Coomassie analysis showed relatively pure protein after SEC (Fig 3.18 E), difficulties with GFP removal and protein yields convinced me not to pursue this construct further.

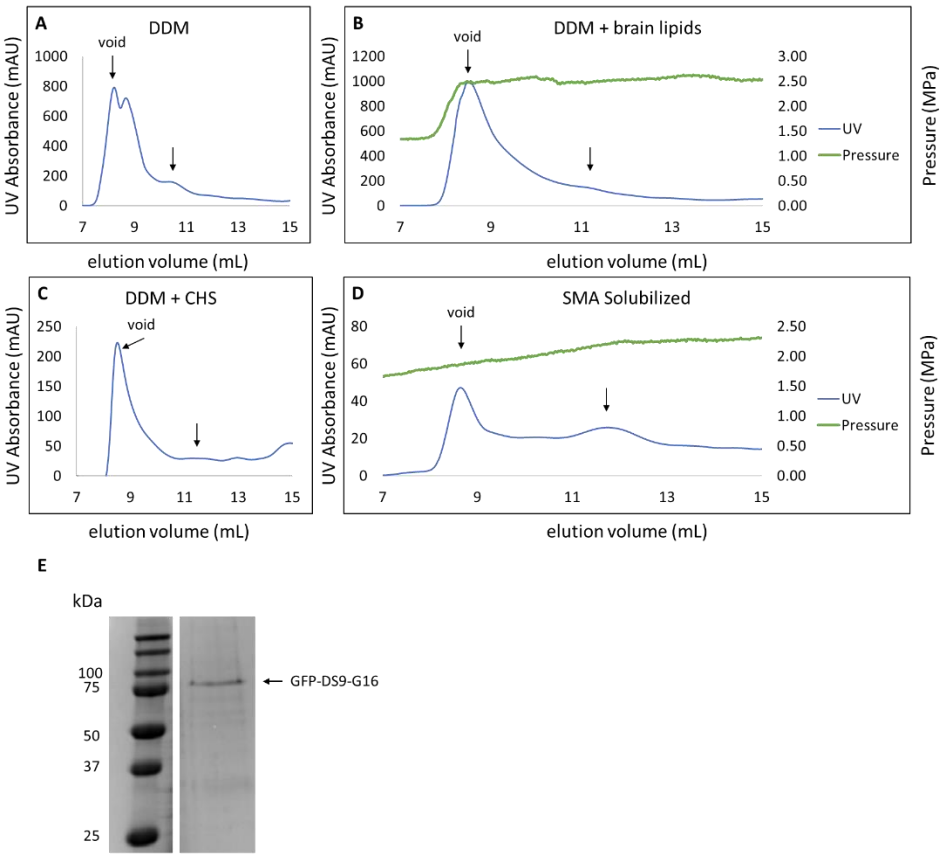

Figure 3.18. Purification of frDHHS9 N2-zfG16 Fusion Protein. SEC profiles for uncleaved his-NGFP frDHHS9 N2-zfGCP16 purified via IMAC and SEC as in the experimental procedures in SEC buffer containing (A) DDM, (B) DDM + brain lipids, and (C) DDM + cholesterol hemisuccinate (CHS). (D) SEC profile for his-NGFP frDHHS9 N2-zfGCP16 purified via IMAC and SEC following solubilization with SMA 2:1 polyscope. (E) SDS-PAGE Coomassie analysis of SMA-purified protein in (d).

In my last study with DHHC9-GCP16 fusion proteins, I revisited the human orthologs. frDHHC9 and zfGCP16 share 83% and 95% sequence identity, respectively, to their human counterparts. Though I observed zfGCP16-mediated stabilization of frDHHC9, the effect was less pronounced than human DHHC9 and GCP16 (data not shown). Therefore, I tested the potential for human DHHC9-GCP16 fusions for crystallization trials by rescreening the fusion proteins in Sf9 cells using FSEC (Fig 3.19). Excitingly, GFP-huDHHS9 N2C1-GCP16 showed a monodispersed peak that was approximately 4-fold higher than GFP-frDHHS9 N2-zfGCP16 (Fig 3.19 A gray and yellow traces). huDHHS9 N2C1-GCP16 was expressed in Sf9 cells and purified by strep-affinity chromatography and SEC. Based on prior qualitative observations, the purification went smoothly compared to frDHHS9 N2-zfGCP16: SDS-PAGE Coomassie showed relatively pure protein at elution (Fig 3.19 B), no obvious precipitation was observed after storage overnight at 4°C, and protein concentration prior to SEC took substantially less time (2 hours compared to 6+ hours). However, once the protein was concentrated above 3mg/mL, very visible precipitation occurred within minutes of incubation on ice (Fig 3.19 C). Even after ultracentrifugation, once applied to SEC, the protein sample continued to aggregate and maxed the column pressure at the void volume (data not shown). Reconstitution of proteins into lipid nanodiscs was previously shown to stabilize proteins in our laboratory such as P2x7, Pannexin, and E-NTPDase 2. To stabilize DHHC9-GCP16 fusion protein prior to concentration and SEC, I reconstituted the protein into saposin A/brain lipid nanodiscs (Fig 3.19 D). However, despite my positive control working for E-NTPDase 2 (data not shown), no DHHC9 peak was observed in any of the reconstituted samples, suggesting DHHC9-GCP16 fusion protein is unable to reconstitute under these conditions. Based on these experimental results, a DHHC9-GCP16 fusion protein approach was no longer pursued.

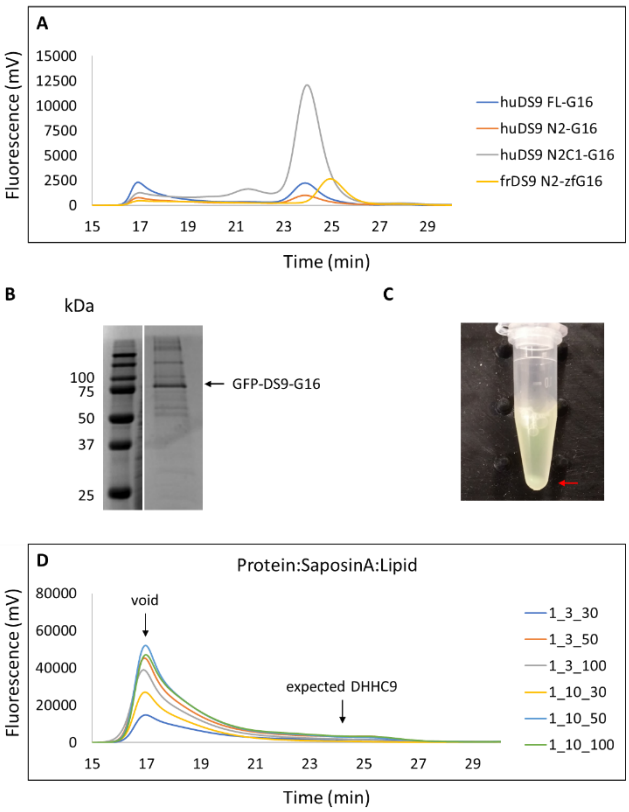

Figure 3.19. Purification of human DHHS9 N2C1-GCP16 Fusion Protein. (A) FSEC profiles human DHHS9-GCP16 fusion proteins. Sf9 cells were infected with recombinant baculovirus particles encoding the indicated his-N-GFP DHHC9 constructs fused to GCP16 and grown for 48 hours at 27C. Cells were solubilized in DDM-containing buffer, and the cleared lysate was analyzed via SEC coupled to a fluorometer to monitor fluorescence at excitation and emission wavelengths 480 nm and 508 nm, respectively. GFP-DHHS9 N2C1-GCP16 was purified via his-IMAC purification as in the experimental procedures. (B) SDS-PAGE Coomassie analysis of purified GFP-DHHS9 N2C1-GCP16. (C) Image of concentrated protein sample after leaving on ice for ~20 minutes before SEC. The red arrow indicates visible precipitation. (D) FSEC profiles for purified fusion protein reconstituted with saposin A into liposomes at the indicated molar ratios.

DHHC9-GCP16 Complex Protein Purification and Crystallization

After years of working on DHHC9 and over eighty purifications of various DHHC9, GCP16, and fusion constructs, the first positive SEC profile of DHHC9-GCP16 complex was obtained. Protein was expressed from Sf9 cells co-infected with baculovirus encoding his-DHHC9 and GCP16-strep and protein was purified by dual-affinity chromatography and SEC as in the experimental procedures (Fig 3.20). SDS-PAGE Coomassie analysis of purification fractions showed his-DHHC9 and GCP16-strep co-eluted after elution from Ni-NTA (Fig 3.20 B lane “E1”) and after elution from streptactin resin (Fig 3.20 B lane “E2”). The SEC profile showed a clear signal at the expected DHHC9-GCP16 peak at ~12 mL and a GCP16 peak at ~14 mL (Fig 3.20 and after SEC (Fig 3.20 B lane “SEC). Additionally, both proteins were visualized in the DHHC9-GCP16 peak after SEC (Fig 3.20 B lane “SEC”). With this initial positive result, there were two major obstacles to overcome. The first obstacle was that protein yield using dual-affinity purification was too low at ~30µg of protein per 400 mL culture of Sf9 cells, and the maximum concentration was 0.6mg/mL. This may be due to loss of protein during additional resin binding steps, as indicated by unbound protein fractions (Fig 3.20 B lanes “U1” and “U2”). To add to the yield issue, DHHC9-GCP16 must be solubilized in DDM as solubilization with triton X-100 was shown to aggregate DHHC9 earlier; this made obtaining more protein through scaling up culture volume not economically viable for our lab. The second obstacle was that protein concentration needed to be higher. A concentration of 10mg/mL is an empirical standard for crystallization, though 1mg/mL is the minimum.

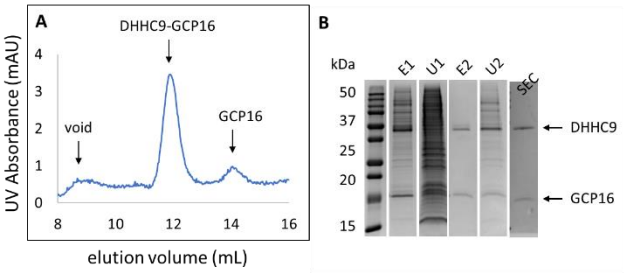

Figure 3.20. Dual-Affinity Purification of the DHHC9-GCP16 complex. (A) SEC profile for his-DHHC9 GCP16-strep complex purified using tandem his-IMAC affinity purification followed by strep-affinity purification as in the experimental procedures. Arrows indicate the void, DHHC9-GCP16 peak, and the GCP16 peak. (B) SDS-PAGE Coomassie analysis for purification fractions showing first elution off Ni-NTA (E1) and the unbound fraction (U1), and elution off streptactin resin (E2) and its unbound fraction (U2), and the DHHC9-GCP16 peak fraction after SEC (SEC).

I first aimed to improve the yield of DHHC9-GCP16. Poor resin-binding efficiency had been a challenge for previous DHHC9 purifications. The affinity-tag location had been tested for every new major construct, and different constructs have a preferred tag location for better yield, e.g., full-length DHHC9 is better with a C-terminal tag while construct N2C2 is better with N-terminal tag or a C-terminal GFP-tag. To determine the optimal tagging position for DHHC9, his-DHHC9 and DHHC9-his were co-expressed with GCP16-strep, purified by dual-affinity chromatography, and visualized by SDS-PAGE Coomassie (Fig 3.21 A). Both N- and C-terminal his-tagged DHHC9 had comparable yields, though it was still too low for crystallization trials. One explanation is that the dual-affinity purification approach, the low resin-binding efficiency for each protein is compounded, resulting in a significant proportion of the protein remaining in the unbound fraction (Fig 3.21 A lanes 3 and 6). Therefore, I tested whether single-step his-IMAC affinity-purification from cells expressing his-DHHC9 and GCP16-strep would preserve yield with sufficient purity (Fig 3.21 B-C). This approach resulted in sufficiently higher yields at 2mg of protein at elution per liter of Sf9 culture. However, the protein showed a higher extent of heterogeneity and impurity, and DHHC9 was enriched over GCP16 (Fig 3.21 B-C) resulting in inconsistent protein preparations. Thus, I switched DHHC9 from a his-tag to a strep-tag and repeated the purification by single-step strep-affinity purification (Fig 3.21 D-E). SEC analysis of strep-DHHC9 and GCP16-strep showed a monodispersed peak for the DHHC9-GCP16 complex around 11.5 mL and a GCP16 peak around 14 mL (Fig 3.21 D). SDS-PAGE Coomassie analysis showed both proteins co-eluted and were pure. Subsequent experiments showed N- and C-terminal strep-tag DHHC9 resulted in (1.7mg protein)/(L of Sf9 cells) and (2.8mg protein)/(L of Sf9 cells), respectively. This yield and purity were sufficient for crystallization trials.

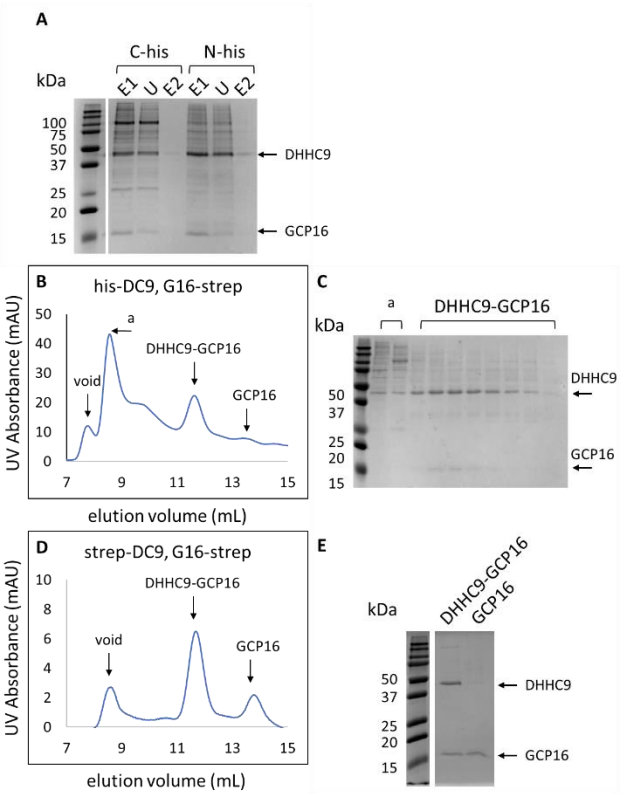

Figure 3.21. Single-Strep Strep-Affinity Purification Yields the most DHHC9-GCP16 complex. Sf9 cells were infected with recombinant baculovirus encoding his-DHHC9 and GCP16-strep, DHHC9-his and GCP16-strep, or strep-DHHC9 and GCP16-strep. Purification was performed as described in the experimental procedures. (A) SDS-PAGE Coomassie analysis of purification fractions from purifications done in parallel with N- vs C-terminally his-tagged DHHC9 showing elution from Ni-NTA (E1), unbound (U), and elution from streptactin resin (E2) fractions. (B) SEC profile and (C) SDS-PAGE Coomassie analysis of peak fractions for his-DHHC9 and GCP16-strep purified via single-step nickel-affinity purification. (D) SEC profile and (E) SDS-PAGE Coomassie of analysis of the complex peak for strep-DHHC9 and GCP16-strep purified via single-step strep-affinity purification.

Next, I determined the protein concentration tolerance for DHHC9-GCP16. Across multiple independent purification, I observed varying degrees of visible protein precipitates during protein concentration prior to SEC (data not shown). Based on these observations and the stochastic nature of the co-infection approach I used, I postulated there exists a population of free DHHC9, free GCP16, and DHHC9-GCP16 complex present in expressing Sf9 cells. One or more of these populations may be sensitive to protein concentration, resulting in aggregation of itself and aggregation of the 'good' protein as well. To test this idea, DHHC9-strep and GCP16-strep was purified, and the cleared elution fraction was applied to SEC without concentrating (Fig 3.22 A). From the unconcentrated fraction at 0.5mg/mL, SEC showed DHHC9-GCP16 eluted at ~11 mL, in agreement with the expected DHHC9-GCP16 complex peak seen previously (Fig 3.22 A). However, the same elution fraction concentrated to 5mg/mL showed a shifted, broad peak at ~9 mL (Fig 3.22 B), suggesting protein concentration was shifting the protein to a higher molecular weight species or aggregate. To examine this phenomenon more carefully, DHHC9-strep and GCP16-strep were purified again, and samples of unconcentrated (0.5mg/mL), 2mg/mL, and 4mg/mL were applied to SEC (Fig 3.22 C). Concentrating the protein showed a consistent shift towards a higher molecular weight species (Fig 3.22 C). Interestingly, the SEC profiles of DHHC9-GCP16 applied to SEC at 0.5mg/mL (Fig 3.22 C blue trace) and at 4mg/mL (Fig 3.22 C gray trace) had a shoulder but the 2mg/mL (Fig 3.22 C orange trace) was more monodispersed. Additionally, SDS-PAGE Coomassie analysis of the elution and fractions after peak fractions after SEC showed that the molar ratio of GCP16 to DHHC9 increased as the protein concentration increased. From these experiments, DHHC9-GCP16 at a concentration of 2mg/mL was used for protein crystallization trials.

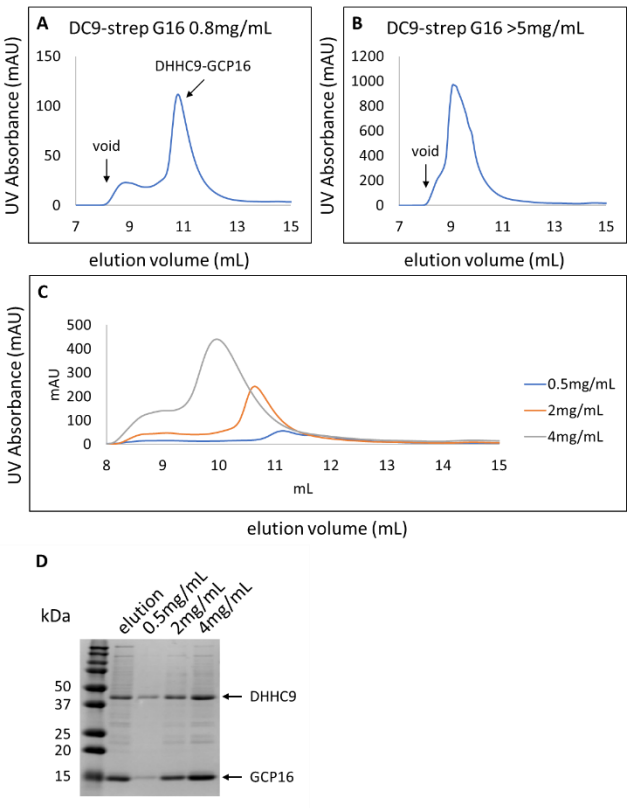

Figure 3.22. Protein Concentration Prior to SEC Affects DHHC9-GCP16 Complex Elution and Protein Molar Ratios. (A-B) SEC profile for purified DHHC9-strep GCP16-strep complex un concentrated at 0.5mg/mL and concentrated to 5mg/mL prior to SEC. (C) Overlay of SEC profiles for strep-DHHC9 GCP16-strep complex concentrated to 0.5, 2.0, and 4.0mg/mL, respectively, prior to SEC. (D) SDS-PAGE Coomassie analysis of elution and peak fractions from (c).

With the yield and protein concentration addressed, I purified DHHC9-GCP16 and proceeded with crystallization trials using the Kawate Lab FK and other crystallization screens that included a range of PEGs (10-40%), pH (4.5-9.5), and distinct salts/additives known to facilitate membrane protein crystallization. Despite extensive screening, I was only able to obtain amorphous quasi crystals of DHHC9-GCP16 that were not of sufficient quality for good X-ray diffraction (Fig 3.23). This was the last attempt to date at crystallization of DHHC9-GCP16.

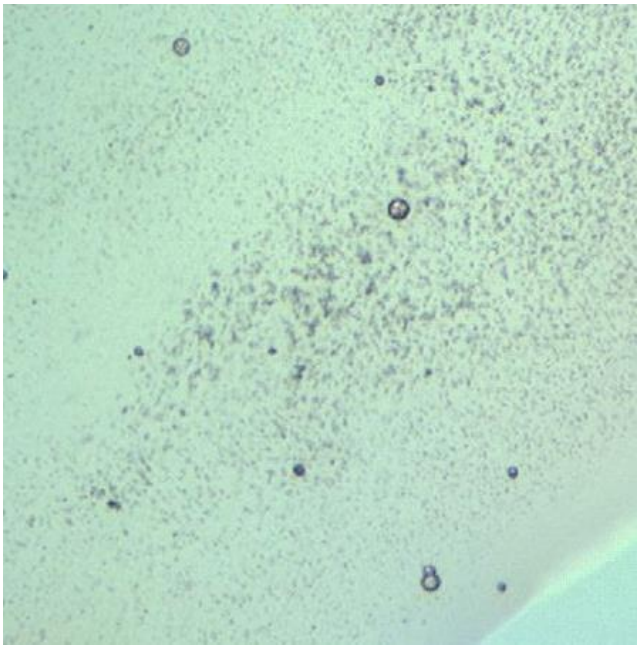

Figure 3.23. DHHC9-GCP16 Complex Forms Amorphous Crystals. Representative amorphous crystals after ~21 days in 96-well screens set-up using 1.3 to 2mg/mL DHHC9-GCP16.

Towards the Structure DHHC9-GCP16 Using Cryogenic Electron Microscopy

In recent years, advances in single-particle cryogenic electron microscopy (cryo-EM) has led to a revolution in structural biology (117–119). The basic principle of cryo-EM is to image biological molecules frozen in a thin layer of vitreous ice, thereby obtaining projections of the protein that can be used to process and reconstruct the protein in two- and three dimensions (117). This approach can be advantageous over traditional X-ray crystallography and nuclear magnetic resonance (NMR) for several reasons: 1. (a) it does not require protein crystals; (b) it is suitable for proteins and complexes of large molecular weight; (c) it reduces radiation damage and better maintains native activity and functional states including posttranslational modifications; and (d) it can capture different conformational states in one experiment (117). However, cryo-EM does have its own challenges in that data processing and 3D reconstruction remains challenging with protein of smaller size (generally less than 150 kDa) and/or protein with numerous flexible domains (117). Since I was unsuccessful in obtain well-diffracting crystals of DHHC9-GCP16, I aimed to use cryo-EM to solve the structure. Based on procedures I adapted from my cryo-EM studies with E-NTPDase 2, Prominin-1, and P2x7, I prepared DHHC9-GCP16 samples. From Sf9 cells, I purified DHHC9-GCP16 in detergent and reconstituted the complex into MSP1D1 lipid nanodiscs containing a 3:1 molar ratio of 1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoethanolamine (POPE) and 1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoglycerol (POPG) (Fig 3.24 A-C). Cryo-EM grids were prepared, and images were acquired using a Talos Arctica. Figure 3.24 D shows a representative micrograph at 63,000x magnification showing the first DHHC9-GCP16 particles observed. Image motion correction was performed using RELION (120), contrast transfer function (CTF) was determined and corrected by Gctf (121), and 2D- and 3D-classification was performed using a combination of RELION and CryoSPARC Software (122). While individual particles could be distinguished and extracted (Fig 3.24 D), 2D classification showed oval shaped classes that did not converge (Fig 3.24 E). This suggested the algorithm could not confidently align features in the 2D

projects. Furthermore, 3D classification showed an egg-shaped blob at below 15-20 angstrom resolution (Fig 3.24 F). Together, this suggested that the DHHC9-GCP16 sample was still too heterogenous at a single-particle level. Moving forward, I aimed to improve the homogeneity of DHHC9-GCP16 particles by testing another expression system and by optimizing nanodisc reconstitution.

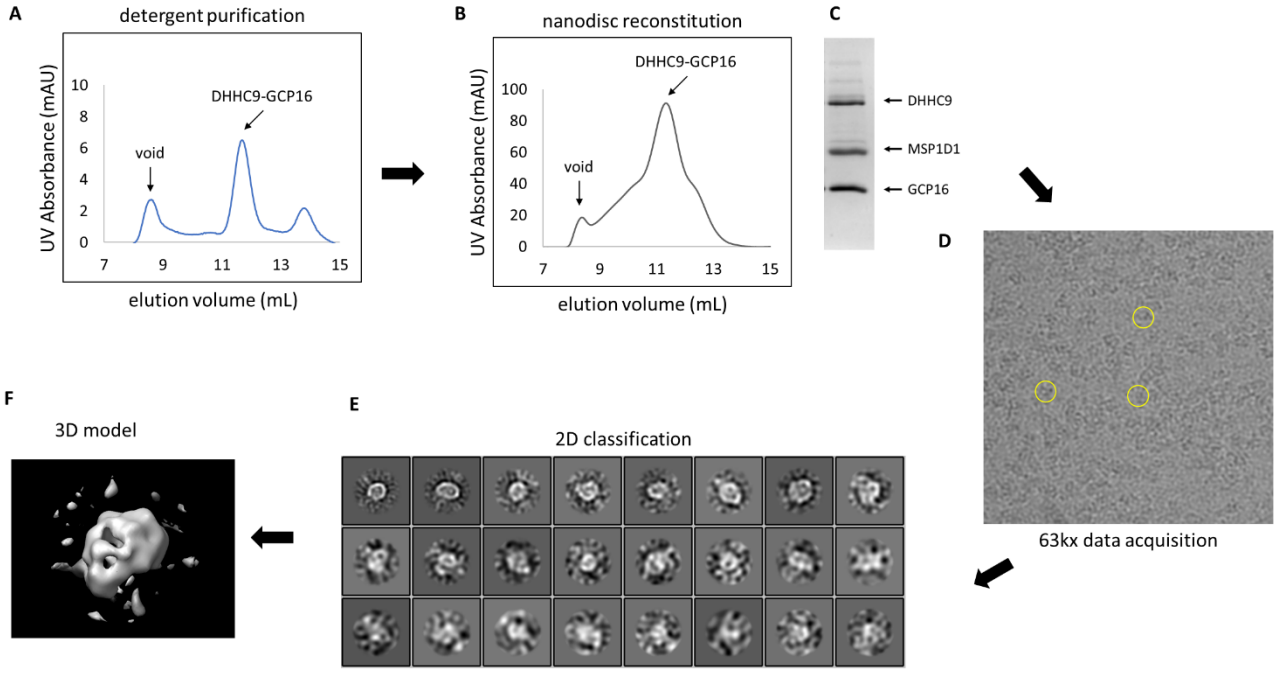

Figure 3.24. Cryo-EM of DHHC9-GCP16 in Nanodiscs Reconstruction #1. Flowchart of DHHC9-GCP16 cryo-EM studies showing (A) SEC profile of detergent-purified DHHC9-GCP16 from Sf9 cells, (B) SEC profile and (C) SDS-PAGE Coomassie analysis of DHHC9-GCP16 reconstituted into MSP1D1 POPE:POPG:Cholesterol nanodiscs, (D) representative micrograph at 63,000x magnification with ~100 angstrom diameter yellow circles highlighting particles, (E) representative 2D classes, and (F) representative 3D map at low resolution ~20 angstroms.

I tested whether expression and purification from mammalian cells could improve DHHC9-GCP16 protein quality. HEK293 cells with N-acetylglucosaminyltransferase knock out (HEK GnTI-) lack the ability to form complex N-glycans, and, thus, have become a powerful tool for high-yield expression of homogeneously glycosylated proteins (123). For example, HEK GnTI- cells have been used in functional and structural studies of membrane proteins including GCPRs, nuclear receptors, ion channels, ATP-binding cassette drug transporters, glutamate receptors (124–130). Additionally, from my work with E-NTPDase 2, CD133, and pannexins, I observed improved protein yields and homogeneity when purifying from HEK GnTI- cells compared to Sf9 cells (data not shown). To determine whether there is a difference in protein quality from HEK GnTI- cells, DHHC9-strep was expressed without and with GCP16-strep and purified by strep-affinity chromatography and SEC (Fig 3.25). Like in Sf9 cells, SEC analysis showed DHHC9 alone was aggregated (Fig 3.25 A-B). By contrast, SEC analysis of DHHC9 co-expressed and purified with GCP16 showed a primary peak at the expected DHHC9-GCP16 peak at ~11.5 mL (Fig 3.25 C) with both proteins co-eluted (Fig 3.25D lane “3”). Interestingly, a larger species that eluted at ~10.5 mL showed enrichment of DHHC9 (Fig 3.25 C, D lane “2”). DHHC9-GCP16 purified from HEK GnTI- cells showed comparable levels of activity for H-Ras *in vitro* compared to DHHC9-GCP16 purified from Sf9 cells (data not shown). Previously, DHHC9-GCP16 purified from Sf9 cells showed a shift towards a higher molecular weight species depending on protein concentration prior to SEC (Fig 3.22 C). This phenomenon may be a cause of protein heterogeneity. To test whether DHHC9-GCP16 purified from HEK GnTI- cells showed sensitivity to protein concentration, FSEC was used to determine whether there was a shift between purified DHHC9-GCP16 unconcentrated (0.7mg/mL) compared to the same DHHC9-GCP16 concentrated to ~3mg/mL (Fig 3.25 E). Both the unconcentrated and concentrated samples showed a primary peak at ~25 (Fig 3.25 blue vs. orange traces), suggesting that DHHC9-GCP16 purified from HEK cells is not affected by protein concentration.

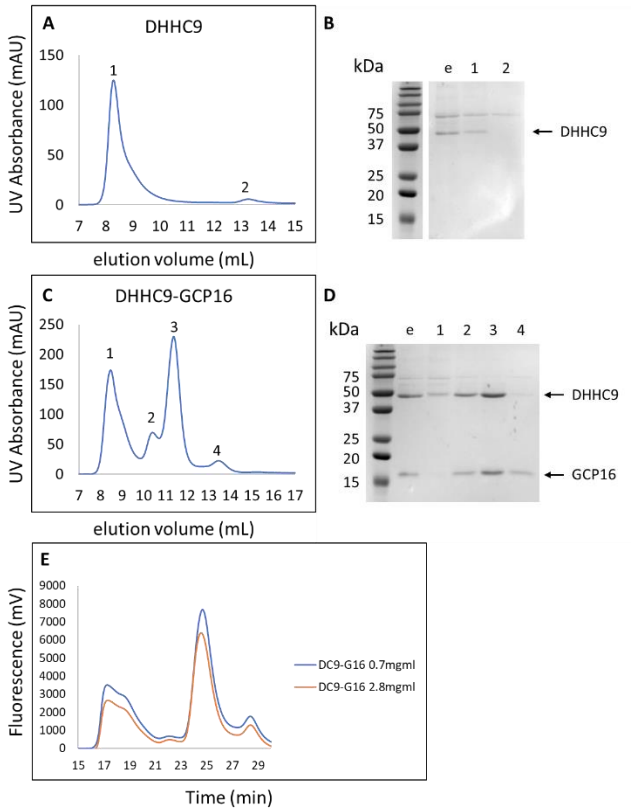

Figure 3.25. Purification of DHHC9 and DHHC9-GCP16 Complex from HEK GnTI- Cells. DHHC9-strep or DHHC9-strep with GCP16-strep was expressed and purified from HEK GnTI- cells infected with recombinant BacMam cells as in the experimental procedures. SEC profiles and SDS-PAGE Coomassie analysis for (A-B) DHHC9 and (C-D) DHHC9 and GCP16. Gel lane "e' corresponds to the affinity-purified elution fraction prior to SEC and the numbered lanes correspond to the indicated peaks in the SEC profiles. (E) FSEC Trp fluorescence profile for purified DHHC9-GCP16 complex unconcentrated (blue trace) and concentrated to 2.8mg/mL (orange trace).

Next, I optimized the reconstitution of DHHC9-GCP16 into lipid nanodiscs. My previous DHHC9-GCP16 reconstitution with MSP1D1 and POPE:POPG (Fig 3.24 B-C) was based on conditions for E-NTPDase 2, a ~60 kDa membrane protein ATPase. To finetune the reconstitution parameters, I screened a matrix of protein-to-scaffold-to-lipid molar ratios that included MSP1D1 scaffolding protein with lipids POPE, POPG, and1,2-Distearoyl-sn-glycero-3-phosphoglycerol (DSPG), as well as cholesterol. POPE:POPG lipids were chosen as a reference point based on prior reconstitution studies in our lab (131), with the goal of modulating membrane fluidity with cholesterol or DSPG due to its saturated lipid chain. To determine the amount of scaffolding protein and lipid to center around, linear extrapolation was performed to obtain the moles of MSP and lipid per empty nanodisc at a given diameter and temperature (132). Purified DHHC9-GCP16 was reconstituted and screened based on the above conditions and analyzed by FSEC (Fig 3.26). DHHC9-GCP16 reconstituted into POPE:POPG nanodiscs with the highest yield with a ratio of 1-to-5-to-244 (Fig 3.26 A gray trace). The presence of incomplete nanodisc (including improperly assembled nanodiscs, empty nanodiscs, and free MSP) peaks at ~28 min and ~32 min supports that MSP and lipids were not limiting factors (Fig 3.26 A). For DSPG nanodiscs, all conditions showed DHHC9-GCP16 reconstitution, though there was a higher excess of incomplete nanodiscs (Fig 3.26 B). DHHC9-GCP16 reconstitution into POPE:POPG:Cholesterol nanodiscs showed the highest yield across all conditions with a ratio of 1-to-5-to-163 (Fig 3.26 C). Since the FSEC screen showed that all three conditions were comparable, multiple DHHC9-GCP16 purifications were performed to reconstitute in amounts sufficient for cryo-EM (Fig 3.26 D-F). Upon scaling up, POPE:POPG and DSPG nanodiscs showed an additional shoulder peak that overlapped with the main DHHC9-GCP16 peak (Fig 3.26 D-E). Fortunately, DHHC9-GCP16 reconstituted into POPE:POPG:Cholesterol nanodiscs showed one primary peak containing DHHC9, GCP16, and MSP1D1 (Fig 3.26 F-G). Therefore, DHHC9-GCP16 reconstituted into POPE:POPG:Cholesterol nanodiscs was used for cryo-EM studies.

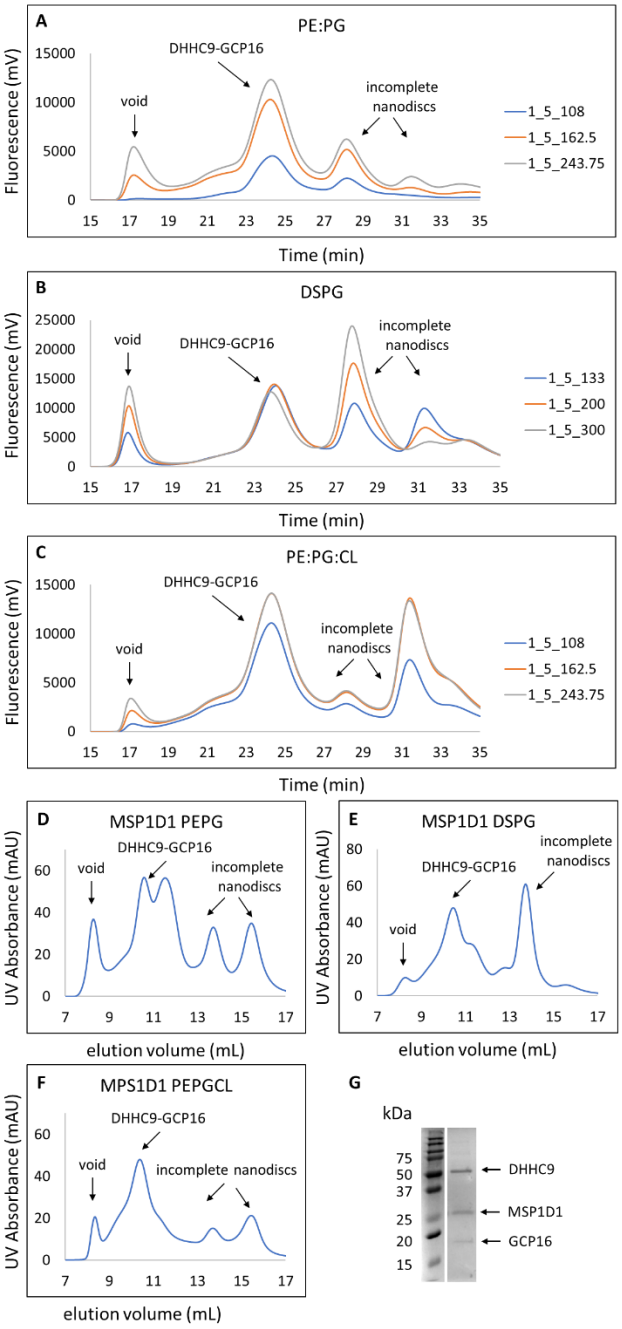

Figure 3.26. Reconstitution of DHHC9-GCP16 into Lipid Nanodiscs. Overlay of FSEC Trp fluorescence profiles for DHHC9-GCP16 reconstitution screen using MSP1D1 scaffolding protein at the indicated protein-to-scaffold-to-lipid molar ratios for (A) 3:1 molar ratio of POPE:POPG, (B) DSPG, and (C) 3:1:1 molar ratio of POPE:POPG:Cholesterol. (D-F) SEC profiles of DHHC9-GCP16 reconstituted into the indicated lipid nanodiscs. (G) Representative

SDS-PAGE Coomassie analysis of the reconstituted DHHC9-GCP16 fraction shown in (f). Abbreviations used include the following: POPE (1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoethanolamine), POPG (1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoglycerol), and DSPG (1,2-Distearoyl-sn-glycero-3-phosphoglycerol).

DHHC9-GCP16 reconstituted into POPE:POPG:Cholesterol nanodiscs was used for cryo-EM studies. Grids were prepared and data was collected using a Talos Arctica Transmission Electron Microscope. Motion correction, CTF estimation, and classification were performed as in the experimental procedures. Particles showed good distribution and sufficient contrast (Fig 3.27 A). Compared to the previous cryo-EM run, 2D classification and iterative refinement showed clearer views at different orientations (Fig 3.27 B-C). Some views showed the top/bottom (Fig 3.27 C top row 4th from the left and bottom row 3rd from the left), while others showed tilted side views. With the refined particle stack, I was able to generate a 3D map (Fig 3.27 D). While the resolution is currently estimated at ~12-15 angstroms, the 3D map resembles the shape of DHHC20 (44), where the nanodisc density is shown at the top, with the catalytic domain protruding towards the bottom (Fig 3.27 D). Additionally, there is an unresolved density located near where the C-terminus should be (Fig 3.27 D green circle). While the current resolution needs to be improved for accurate modeling, this is the first cryo-EM map of a DHHC protein complex.

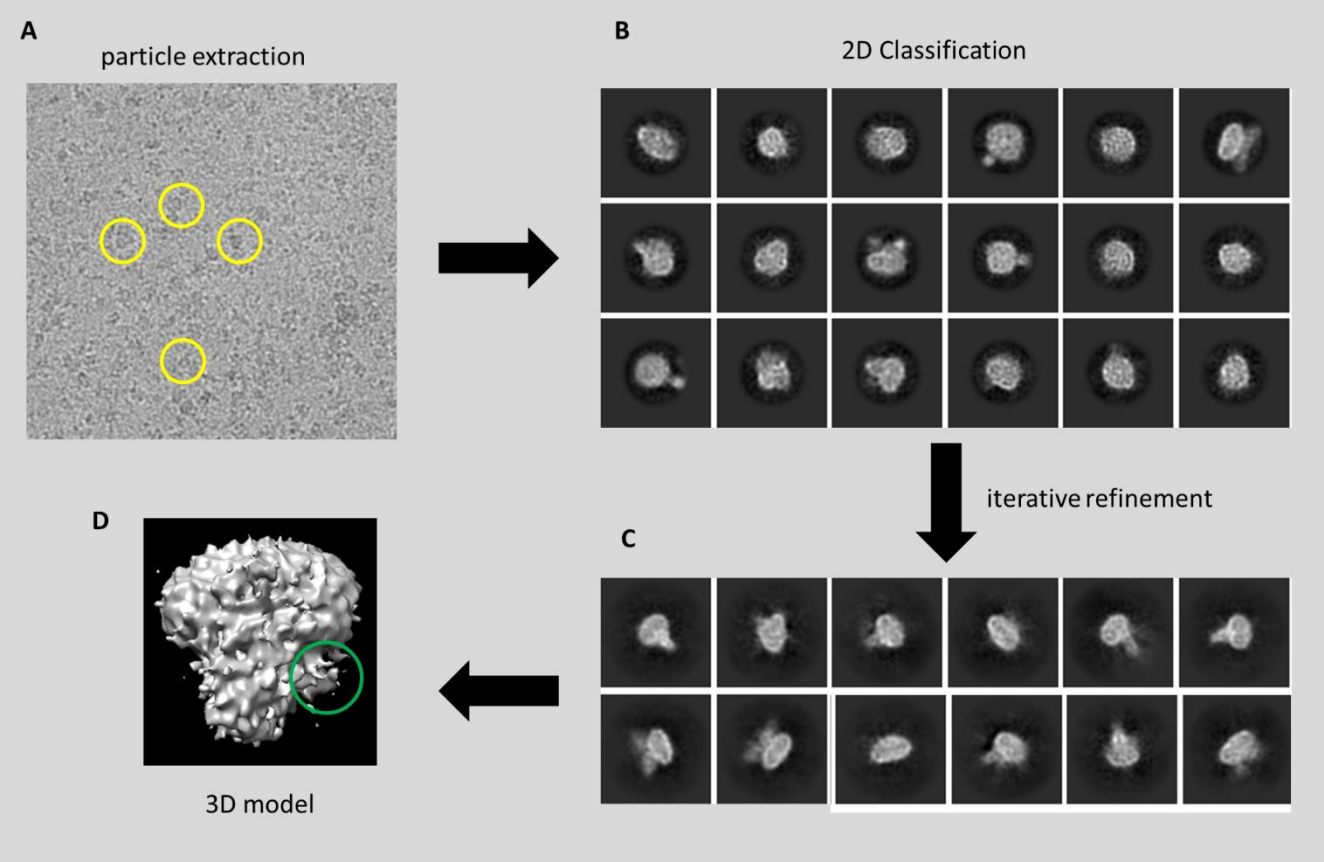

Figure 3.27. Cryo-EM of DHHC9-GCP16 in Nanodiscs Reconstruction #2. DHHC9-GCP16 was purified from HEK GnTI and reconstituted into MSP1D1 POPE:POPG:Cholesterol nanodsics as in the experimental procedures. (A) Representative micrograph at 63,000x magnification with ~100 angstrom diameter yellow circles highlighting particles. (B) 2D classification from an early particle stack compared to (C) after iterative cleanup and refinement in 2D. (D) final 3D map at ~12-15 angstroms with an unresolved density highlighted with a green circle.

Discussion

DHHC9 and GCP16 Structure Constructs and Purification

I started my structural studies by screening DHHC9 and GCP16 orthologs to identify the most well-behaved constructs for purification and crystallization. Compared to other proteins I have worked with, DHHC9 and GCP16 orthologs showed very high evolutionary conservation among their respective orthologs. For example, human DHHC9 and GCP16 showed over ~80% amino acid sequence identity to orthologs in amphibians and birds, despite being relatively distant evolutionarily. One challenge this created is that many orthologs I initially came up with differed from the human sequence by about the same percent sequence identity. Therefore, I used an unbiased, bioinformatic-based approach where I used the human sequences as a query for BLAST. For example, this allowed an initial set of ~100 orthologs for DHHC9, and I was able to filter by pair-wise differences to increase the coverage within the same identity range. Ultimately, this approach resulted in an above average number of positive hits at ~30% (4/12) in the initial DHHC9 screen, compared to an average of 1 out of 15 for previous ortholog screens in the Kawate Lab.

frDHHS9 emerged as the focus of my optimization due to its consistently high expression and above average monodispersity compared to other orthologs. Yeast Erf2 shares 31% sequence identity to frDHHC9. Erf2 contains a C-terminus that targets it for ubiquitin-mediated degradation; removal of the C-terminal 58 amino acids increased the half-life of Erf2 in cells by ~ two-fold, but removal of the N-terminal 119 amino acids had almost no effect (68). Based on the conservation of DHHC9 in general, I initially hypothesized that frDHHS9 would behave similarly to Erf2 in the context of termini deletion. However, the opposite was observed in my deletion construct screen. Deletions of frDHHS9 C-terminus up to ~120 amino acids had very mild effects compared to deletions of 12 or 20 amino acids in the N-terminus (Fig 3.5). Furthermore, N-terminal deletions were the dominant effect, as the combinatorial deletions

showed high similarities to their respective N-terminal deletion. This difference in stabilizing truncations between Erf2 and frDHHC9 may suggest that the DHHC9 termini in metazoans have diverged and that protein degradation may be regulated in a manner distinct from that seen in Erf2. For now, it is unclear whether this involves ubiquitination or the endoplasmic reticulum-associated degradation system. Additionally, for huDHHC9, deletion N2C2 showed improved protein behavior on FSEC, in a manner like the frog construct (data not shown). However, purified huDHHS9 N2C2 was almost entirely aggregated on SEC, while frDHHS9 N2C2 could remain folded (Fig 3.7). The exact amino acid differences between frog and human constructs could be explored to obtain purified huDHHC9 that is folded.

Perhaps my progress in optimizing DHHC9 constructs for crystallization can be more appropriately called “all the ways in which DHHC9 will aggregate.” Though frDHHS9 N2C2 showed very promising behavior in FSEC screens, this construct exhibited low yields likely due to poor binding efficiency to resin. At the time, I evaluated what could be learned from a structure of frDHHS9 N2C2. DHHC9 shares 28% amino acid sequence identity with DHHC20, with most existing between the start of the DHHC domain and ending at the region around deletion C2. The N- and C-termini of DHCH9 are where the proteins diverge, and the C-terminus is where my concurrent experiments with GCP16 were suggesting where GCP16 may be interacting (Chapter 2). A structure of DHHC9 N2C2 may not provide many new insights into DHHC proteins in general, and it may not reveal any information on how DHHC9 interacts with GCP16 if the C-terminus beyond region C2 is required for interaction. Therefore, I focused on stabilizing a construct that contained an extended C-terminus such as construct N2 or N2C1.

Unfortunately, whether DHHC9 was expressed alone (Fig 3.10) or as a fusion protein with GCP16 (Fig 3.17-1.19), all my purification showed extremely prevalent protein aggregation by or at SEC. The most promising was human DHHS9 N2C1-GCP16 fusion, which purified well but aggregated at higher concentrations (Fig 3.19). My efforts at stabilizing the solubilization step with lipid supplements, cholesterol, or SMA nanodiscs did not yield enough protein for

crystallization trials. In the end, I focused on DHHC9-GCP16 complex purification co-expressed and co-purified for crystallization.

Additionally, purification of DHHC9 is further complicated by inefficient resin binding with construct-specific differences. While C-terminal GFP-tagged DHHC9 expressed better in most cases compared to N-terminally GFP-tagged constructs, they showed less than 5% binding of the total protein (Fig 3.17). In a few cases where N-tagged DHHC9 constructs without GFP were purified, almost no protein was obtained at elution; however, western blot confirmed that the protein remained in the unbound fraction (data not shown).

Secondary structure prediction of DHHC9 using JPred 4 software (133) predicts three beta strands with hydrophobic characteristic between residues 11 – 31. This region contains two cysteine residues, Cys24 and Cys25. The SwissPalm database predicts these to be S-acylated. Prior, I sent DHHC9-GCP16 to Cornell Proteomics and Metabolomics Facility for palmitoylation analysis by iodoacetamide replacement and liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) in Orbitrap fusion. The report stated that both Cys24 and Cys25 were found to be modified by iodoacetamide, supporting that that these residues are S-acylated. Based on this information and my observations from purification of construct N2 (Δ1-19), I speculate that the DHHC9 N-terminus contains an ordered structure that—unless GFP is fused to it—is tightly associated to the main body of the protein or to the membrane. Without GFP or a full-length N-terminus, a purification tag is localized in a way to occlude exposure to the aqueous environment. Similarly, the C-terminus likely contains ordered domains that were discussed in Chapter 2.

zfGCP16 was identified as a crystallization candidate for GCP16. One consideration that arose from these experiments is the limitation of FSEC. FSEC was unable to reveal differences between the behavior of GCP16 orthologs (Fig 3.11). This could be because GCP16 is relatively small at 16 kDa, compared to GFP at 27 kDa, which may have skewed the FSEC profile. Another interesting observation came from the zfGCP16 cysteine mutation

analysis (Fig 3.14). Interestingly, light microscopy images using GFP filters of cells overexpressing these constructs showed different localization patterns (Fig 3.14 A). Wildtype and C24S mutation showed diminished GFP signal, with more fluorescence around the edges of cells or in internal compartments (Fig 3.14 A top left and middle panels). Mutation of cysteines 69, 72, or both to serine resulted in more intense and diffuse GFP fluorescence (Fig 3.14 top right, bottom left, and bottom right panels). This may suggest these mutations disrupt GCP16 localization, which agrees with what Ohta et al. observed (80). Intriguingly, C81S mutation also displayed more diffuse GFP fluorescence, which may suggest C81 in zfGCP16 is acylated or contributing to localization in some way. Perhaps the additional peaks observed for C69S or C72S mutations (Fig 3.14 D-E) could reflect oligomerization or different membrane-bound states. However, a thorough analysis of these cysteines and their effects on GCP16 localization is outside the scope of this work.

Crystallization Trials

Often, protein crystallization requires testing many possible crystallization conditions. For a given construct, conditions to vary include protein concentration, temperature, ionic strength, pH, chemical additives, and the method of crystallization, e.g., vapor diffusion, microbatch, etc. For DHHC9, GCP16, and DHHC9-GCP16 complex, I set up hanging drop 96-well crystallization trays using the Kawate Lab FK and other (MemGold I and II, JBScreen) crystallization screens that included a range of polyethylene glycols (PEGs) (10-40%), pH (4.5-9.5), and distinct salts/additives known to facilitate membrane protein crystallization. In the end, I was only able to obtain small, amorphous quasi crystals that are not well-ordered with representative images shown in Figures. 1.8, 1.13, and 1.23. The published structures of DHHC20 and DHHC15 were obtained from protein crystals that formed from protein at ~15-20mg/mL by lipidic cubic phase (LCP) and hanging drop conditions that were optimized by reductive methylation for the highest diffraction data (44). Given DHHC9's sensitivity to

concentration (Fig 3.22), protein concentration was limited to under 2mg/mL, which prevented screening with higher concentrations. Given the near limitless combinations of crystallization conditions to test, my screens were not exhaustive. Future trials could include different detergents, batch seeding, bicelles, amphipols, nanodiscs, LCP, and many more. However, I focused the remainder of my efforts into DHHC9-GCP16 structure determination using cryo-EM.

DHHC9-GCP16 Cryo-EM

During the last year of the project, I sought to obtain the structure of DHHC9-GCP16 by cryo-EM. After several collections and extensive data processing, I generated a 12-15 angstrom 3D density of DHHC9-GCP16 reconstituted into lipid nanodiscs (Fig 3.27). They blob showed the lipid nanodisc ring, with a protrusion that resembles the catalytic domain in DHHC20. Furthermore, an unresolved density is present where the C-terminal α'2-helix is expected. The density is consistent with a monomeric DHHC9, though determining the number of GCP16 subunits requires higher resolution.

Cryo-EM can be limited by protein size. The smaller the object, the lower the signal to noise ratio is from extracted micrographs with the current technology. Only a few proteins less than 100 kDa in size have been solved, and they are known to be well-behaved proteins (134, 135). DHHC9-GCP16 is ~57 kDa. Its size is very likely a limiting factor in achieving higher resolutions. Recently, the structure another 57 kDa acyltransferase, hedgehog acyltransfer, was solved with substrate and product complex at 2.7 and 3.2 angstroms, respectively (136). The authors generated two IgG fab fragments (~50 kDa each) that, when bound to the target protein, increased the effective size to ~150 kDa and enabled the generation of a high-resolution 3D map. Moving forward, a similar approach to increase the size of DHHC9-GCP16 complex may be worth exploring.

Experimental Procedures

Construction of Expression Plasmids

Amino acid sequences were obtained from the UniProt database for the proteins listed in Table 3.3. Genes were synthesized based on protein sequences (GenScript, Piscataway, NJ). The plasmid encoding GOLGA7B was obtained from the Harvard Medical School Plasmid repository. Unless stated otherwise, all genes used were PCR amplified to exclude the start methionine and to incorporate flanking BamHI/XhoI sites. For DHHC9-GCP16 fusion proteins, GCP16 was fused to DHHC9 with a Gly-Ser-Gly linker in between. Standard molecular biology techniques were used to subclone genes into their respective expression vectors. Any mutations were generated via quick change PCR.

Uniprot ID	Scientific Name	Codename
A4IIG3	*Xenopus tropicalis*	frog DHHC9
U3FYR5	*Micrurus fulvius*	coral snake DHHC9
H9GEW2	*Anolis carolinensis*	chameleon DHHC9
A0A1D5PK56	*Gallus gallus*	chicken DHHC9
Q9Y397	*Homo sapiens*	human DHHC9
F7EDC7	*Monodelphis domestica*	opossum DHHC9
M7BVD7	*Chelonia mydas*	sea turtle DHHC9
A0A1U7RXI8	*Alligator sinensis*	alligator DHHC9
A0A0P7WKI1	*Scleropages formosus*	arowana DHHC9
W5NDV0	*Lepisosteus oculatus*	spotted gar DHHC9
F1R013	*Danio rerio*	zebrafish DHHC9
H2TW56	*Takifugu rubripes*	pufferfish DHHC9
Q4SBV8	*Tetraodon nigroviridis*	pufferfish GCP16
Q6IQI1	*Danio rerio*	zebrafish GCP16
Q5XK89	*Xenopus laevis*	african frog GCP16
H9G3X9	*Anolis carolinensis*	chameleon GCP16
Q7Z5G4	*Homo sapiens*	human GCP16
H0Z441	*Taeniopygia guttata*	zebra finch GCP16

Table 3.3. Table of DHHC9 and GCP16 Genes and Codenames.

For FSEC in HEK293 cells, I used Kawate Lab vectors pNGFP-EU2 or pCGFP-EU2 for either N- or C-terminal fusion with eGFP, respectively. The vectors encoded a start methionine. Genes were incorporated using BamHI/XhoI sites.

For protein expression in Sf9 cells, genes were subcloned into a modified pFastBac baculovirus expression vector (Thermo Fisher Scientific, Waltham, MA). Protein was expressed from Kawate Lab vectors with one of the following tags as indicated: N-terminal strep-tag (pNNG FB7), N-terminal his-tag (pNNG FB6b), N-terminal strep-GFP (pNGFP-FB7), N-terminal his-GFP (pNGFP-FB6b), C-terminal strep-tag (pCNG FB7), C-terminal his-tag (pCNG FB6b), C-terminal GFP-strep (pCGFP FB7), or C-terminal GFP-his (pCGFP FB6b).

Fluorescence-Detection Size-Exclusion Chromatography

For FSEC in HEK 293 cells, cells at 80-95% confluency were transfected with 2.5 µg of expression plasmid using jetPRIME transfection reagent per the manufacturer's instructions. Cells were grown at 37°C for 48 hours. For FSEC in Sf9 cells, 2.0-2.5 x 10^6 cells were infected with 50 µL of P2 virus for 48 hours. The media was aspirated, cells were suspended and washed in 2 mL of cold PBS, and samples were lysed in 150 µL of lysis buffer (1x PBS, 1% n-Dodecyl-β-D-Maltopyranoside (DDM), 1x Roche protease inhibitor) for 30 minutes at 4°C with rotation. When solubilization was done with other detergents, 1% w/v was used. Whole-cell lysates were cleared by centrifugation in a table-top centrifuge at 15k rpm for 10 minutes at 4°C. The supernatant was further cleared by ultracentrifugation in a TLA 100.3 rotor at 70K rpm for 20 minutes. An aliquot (50µL) was applied onto a Superdex 200 Increase 10/300 GL column pre-equilibrated with running buffer (1x PBS, 0.5mM DDM). The eluent from the SEC column was passed through a fluorometer set to excitation, 480 nm, and emission, 508 nm. For Trp fluorescence detection, the fluorometer was set to excitation, 280nm, and emission, 350nm).

Expression and Purification of DHHC Proteins and DHHC-GCP16 Complexes.

All constructs were expressed using the Invitrogen Bac-to-Bac® baculovirus-insect cell expression system. Sf9 cells (1-2L) were infected at 2.5 – 4.0 x 10^6 cells/mL with P2 virus (10^7-10^8 PFU/mL, 30mL/L). For DHHC9 and DHHC9-GCP16 fusion proteins, cells were incubated at 27°C for 24 hours and moved to 18°C for an additional 48 hours. For GCP16, cells were incubated at 27°C for 48 hours. All purification steps were done on ice or with 4°C buffer unless stated otherwise. Cells were harvested by centrifugation at 2,040 x g and washed with 200 mL of PBS. Where membrane preparation was performed, cells were lysed by nitrogen cavitation at 750 psi for 20 minutes in a 4635 cell disruption vessel (Parr Instrument). The sample was centrifuged at 10,000 x g for 10 minutes to separate unbroken cells and debris. Membrane fraction was collected by centrifugation at 185,000 x g for 45 minutes. Either homogenized membranes or whole cells were solubilized in buffer A (PBS, 15% glycerol, 0.5mM TCEP, 0.5ug/mL leupeptin, 2ug/mL aprotinin, 0.5ug/mL pepstatin A, and 0.5mM phenylmethylsulfonyl fluoride) containing 1% DDM w/v at a ratio of 250 mg DDM/g of cells (1% triton X-100 was used for GCP16). Solubilization was performed at 4°C with stirring for one hour. Large debris were removed by centrifugation at 12,000 x g. The supernatant was further clarified by ultracentrifugation at 185,000 x g for 45 minutes to remove remaining insolubilized material. To the supernatant, 2 mL of equilibrated StrepTactin Sepharose High Performance resin (GE Healthcare, Marlborough, MA) was incubated with stirring for one-hour. The resin was collected by centrifugation and transferred to a gravity column (Bio-Rad). The resin was washed with 10 resin bed volumes of wash buffer (buffer A with 0.5 mM DDM) and eluted with elution buffer (buffer A with 2.5 mM desthiobiotin, 0.5 mM DDM). Protein was concentrated using Amicon ultra concentrators. Size exclusion chromatography was done using a Superdex 200 (GE Healthcare) in SEC buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.5 mM TCEP, 15% glycerol, 0.5 mM DDM). Other purifications were performed as above with minor adjustments mentioned below.

Where indicated, GFP-tags were removed prior to SEC by incubation with either 1:50-

1:100 thrombin protease or TEV protease overnight at 4°C.

For purification with his-tagged protein, purification was performed up to resin binging as above. The cleared lysate was incubated with 2 mL of equilibrated Ni-NTA per liter of initial cell culture. The resin was collected by centrifugation and transferred to a gravity column (Bio-rad). The resin was washed with 10 resin bed volumes of wash buffer (buffer A with 20mM imidazole, 0.5mM DDM) and eluted with elution buffer (buffer A with 250mM imidazole, 0.5mM DDM). Protein was concentrated using amicon ultra concentrators. Size exclusion chromatography was done using a Superdex 200 (GE Healthcare) in SEC buffer (150mM NaCl, 20mM HEPES pH 7.4, 0.5mM TCEP, 15% glycerol, 0.5mM DDM).

For purification using brain lipids, brain lipids stocks were prepared by aliquoting 5 mL of 10mg/mL brain lipid into a glass vial. The solution was dried using nitrogen gas, suspended and washed in pentane and dried again. 2.5 mL of 50mM Tris pH8, 500mM NaCl, and 15% glycerol was added to the dried lipids. Lipids were dissolved partially via sonication for 10 minutes. Prior to adding to buffer, this stock was vortexed vigorously. DHHC9 was purified as above, except wash, elution, and SEC buffers contained 0.01mg/mL brain lipids.

For purification with cholesterol hemisuccinate (CHS), A solution of 10% DDM and 2% CHS w/v was mixed in mq water and allowed to mix at room temperature overnight. This solution was syringe filtered before using. Purification of DHHC9 was done as above except for solubilization buffer containing 0.2% CHS, and wash, elution, and SEC buffers containing 0.05mg/mL CHS.

For styrene maleic acid (SMA) purification, whole cells were dissolved in 50mM Tris pH8, 500mM NaCl, 15% glycerol, 0.5mM TCEP, protease inhibitors, and 2.5% SMA at a ratio of 40mg cells per mL buffer for 2 hours at RT with stirring. The remainder of the purification was done as above except with overnight binding to Ni-NTA and no detergents in the buffers.

For nanodisc reconstitution, 10-20 µg of detergent-purified protein was used per condition during screening, and 200-400 µgwas used for structural studies. Lipids

cryo-EM prep and data.

Nanodisc Reconstitution

Lipids were prepared by transferring stock lipids in chloroform (Avanti Polar Lipids) at desired w/w ratios into glass vials. A gentle stream of nitrogen was used to evaporate the solvent. Lipids were dissolved in chloroform and dried again. The procedure was repeated an additional time with pentane. To the dried lipids, reconstitution buffer (50mM HEPES pH 7.4, 150mM NaCl) was added to a concentration of 20 mg/mL. The solution was sonicated until it became clear. The lipid stock was freeze-thawed 3x and diluted to 5mg/mL for a working stock.

For the generation of protein into MSP nanodiscs, purified MSP1D1 at 8mg/mL was incubated with lipids at molar ratios of 2 MSP to 120-200 lipids for 30 minutes at room temperature (at 37°C for DSPG). Purified protein at >2 mg/mL was added at a molar ratio of 1 to 2 MSP and incubated for 45 minutes at room temperature. The solution volume was adjusted to 100 µL per 20 µgprotein. Detergent was removed by incubation with biobeads overnight at 4°C. Biobeads were removed by pipetting and syringe filter. For the generation of protein into saposin nanodiscs, the above procedure was performed with adaptations. Protein was incubated with lipids for 45 minutes before the addition of saposin for 30 minutes.

Protein Crystallization

DHHC9, GCP16, and DHHC9-GCP16 were crystallized by hanging drop vapor diffusion method by mixing 1:1 or 1:1.5 (v/v) ratio of protein and reservoir solutions at 4°C or room temperature.

Cryo-EM

DHHC9-GCP16 in nanodiscs (3 uL) at a concentration of 1 mg/mL was applied to glow-discharged C-flat holey carbon grids 1.2/1.3 (Electron Microscopy Services), blotted for 4 seconds at 7 force and 100% humidity and immediately plunge frozen in liquid ethane using an

FEI Mark IV Vitrobot.

Data collection was performed using a Thermo Fisher Talos Arctica operating at 200 keV with a Gatan K3 detector in counting mode with 0.5x binning (super-resolution mode) and a Gatan BioQuantum energy filter with a slit width of 20 eV. EPU software was used for automated data collection with a dose of 1 e^-/angstrom^2 per frame for 50 frames. Data was collected at 63,000x magnification for a physical pixel size of 1.26 angstroms per pixel.

Cryo-EM data image motion correction was performed using RELION (120), contrast transfer function (CTF) was determined and corrected by Gctf (121), and 2D- and 3D-classification was performed using a combination of RELION and CryoSPARC Software (122).

Concluding Remarks

In my thesis work, I sought a better understanding of the DHHC9-GCP16 protein complex. I started by studying the role of GCP16 in the DHHC9-GCP16 complex. One role of GCP16 is to stabilize DHHC9 by preventing its aggregation through formation of a protein complex. I demonstrated that only properly folded DHHC9-GCP16 is enzymatically active *in vitro*. Additionally, mutations in *ZDHHC9* linked to X-linked intellectual disability result in reduced protein stability and DHHC9-GCP16 complex formation. This provides molecular insight at the protein level as to how *ZDHHC9* mutations can contribute the disease phenotype. Through my investigations with DHHC9-GCP16, I identified a C-terminal cysteine motif (CCM) that is conserved among the DHHC9 subfamily (DHHC14, 18, 5, and 8). Co-expression of GCP16 with DHHCs containing the CCM improves DHHC protein stability. Like DHHC9, DHHC14 and DHHC18 require GCP16 for their enzymatic activity *in vitro*. Furthermore, GOLGA7B, an accessory protein with 75% sequence identity to GCP16, improves protein stability of DHHC5 and DHHC8, but not the other members of the DHHC9 subfamily, suggesting selectivity in accessory protein interactions. My study supports a broader role for GCP16 and GOLGA7B in the function of human DHHC proteins. Additionally, though my structural studies, I determined the protein constructs and purification conditions to obtain purified DHHC9-GCP16 in sufficient quantities for protein crystallization and cryo-EM analysis. Currently, the cryo-EM map at ~12-15 angstroms is too low for structure modeling. Future directions could include increasing the size of the DHHC9-GCP16 complex through a combination of fabs, fusion proteins, or substrates. A high-resolution structure will be invaluable in improving our understanding of a unique class of proteins with strong biomedical relevance.

References

1. Chen, B., Sun, Y., Niu, J., Jarugumilli, G. K., and Wu, X. (2018) Protein Lipidation in Cell Signaling and Diseases: Function, Regulation, and Therapeutic Opportunities. *Cell Chem. Biol.* **25**, 817–831
2. Jiang, H., Zhang, X., Chen, X., Aramsangtienchai, P., Tong, Z., and Lin, H. (2018) Protein Lipidation: Occurrence , Mechanisms , Biological Functions , and Enabling Technologies. 10.1021/acs.chemrev.6b00750
3. Cheung See Kit, M., and Martin, B. R. (2019) *Protein Lipidation*, 10.1007/978-1-4939-9532-5_6
4. S. Resh, M. D. (2013) Covalent Lipid Modifications of Proteins. *Curr. Biol.* **23**, 431–435
5. Stoffyn, P., and Folch-Pi, J. (1971) On the type of linkage binding fatty acids present in brain white matter proteolipid apoprotein. *Biochem. Biophys. Res. Commun.* **44**, 157–161
6. Schmidt, M. F., Bracha, M., and Schlesinger, M. J. (1979) Evidence for covalent attachment of fatty acids to Sindbis virus glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1687–1691
7. Schlesinger, M. J., Magee, A. I., and Schmidt, M. F. G. (1980) Fatty acid acylation of proteins in cultured cells. *J. Biol. Chem.* **255**, 10021–10024
8. Wang, Y., Yang, W., Sciences, B., Oschin, S., Cancer, C., Angeles, L., States, U., Angeles, L., and States, U. (2021) Proteome-scale Analysis of Protein S-acylation Comes of Age. *J. Proteome Res.* **20**, 14–26
9. Martin, B. R., Wang, C., Adibekian, A., Tully, S. E., and Cravatt, B. F. (2011) Global profiling of dynamic protein palmitoylation. *Nat. Methods*. **9**, 84–89
10. Chamberlain, L. H., and Shipston, M. J. (2015) The Physiology of Protein S-acylation. *Physiol. Rev.* **95**, 341–376
11. Blanc, M., David, F., Abrami, L., Migliozzi, D., Armand, F., Bürgi, J., and van der Goot, F. G. (2015) SwissPalm: Protein Palmitoylation database [version 1; referees: 3 approved].

F1000Research. **4**, 261

12. Magee, A. I., Koyama, A. H., Malfer, C., Wen, D., and Schlesinger, M. J. (1984) Release of Fatty Acids From Virus Glycoproteins by Hydroxylamine. *Biochim. Biophys. Acta*. **798**, 156–166
13. Kaufman, J. F., Krangel, M. S., and Strominger, J. L. (1984) Cysteines in the transmembrane region of major histocompatibility complex antigens are fatty acylated via thioester bonds. *J. Biol. Chem.* **259**, 7230–7238
14. Yang, W., Vizio, D. Di, Kirchner, M., Steen, H., and Freeman, M. R. (2010) Proteome Scale Characterization of Human S-Acylated Proteins in Lipid Raft-enriched and Non-raft Membranes*. *Mol. Cell. Proteomics*. **9**, 54–70
15. Muszbek, L., and Laposata, M. (1989) Covalent modification of platelet proteins by palmitate. *Blood*. **74**, 1339–1347
16. Muszbek, L., and Laposata, M. (1993) Myristoylation of proteins in platelets occurs predominantly through thioester linkages. *J. Biol. Chem.* **268**, 8251–8255
17. Muszbek, L., and Laposata, M. (1993) Covalent modification of proteins by arachidonate and eicosapentaenoate in platelets. *J. Biol. Chem.* **268**, 18243–18248
18. Liang, X., Nazarian, A., Erdjument-Bromage, H., Bornmann, W., Tempst, P., and Resh, M. D. (2001) Heterogeneous Fatty Acylation of Src Family Kinases with Polyunsaturated Fatty Acids Regulates Raft Localization and Signal Transduction. *J. Biol. Chem.* **276**, 30987–30994
19. Salaun, C., Greaves, J., and Chamberlain, L. H. (2010) The intracellular dynamic of protein palmitoylation. *J. Cell Biol.* **191**, 1229–1238
20. Goodwin, J. S., Drake, K. R., Rogers, C., Wright, L., Lippincott-Schwartz, J., Philips, M. R., and Kenworthy, A. K. (2005) Depalmitoylated Ras traffics to and from the Golgi complex via a nonvesicular pathway. *J. Cell Biol.* **170**, 261–272
21. Rocks, O., Peyker, A., Kahms, M., Verveer, P. J., Koerner, C., Lumbierres, M.,

Kuhlmann, J., Waldmann, H., Wittinghofer, A., and Bastiaens, P. I. H. (2005) An Acylation Cycle Regulates Localization and Activity of Palmitoylated Ras Isoforms. *Science (80-.).* **307**, 1746–1752

22. Roy, S., Plowman, S., Rotblat, B., Prior, I. A., Muncke, C., Grainger, S., Parton, R. G., Henis, Y. I., Kloog, Y., and Hancock, J. F. (2005) Individual Palmitoyl Residues Serve Distinct Roles in H-Ras Trafficking, Microlocalization, and Signaling. *Mol. Cell. Biol.* **25**, 6722–6733
23. Duncan, J. A., and Gilman, A. G. (1996) Autoacylation of G protein α subunits. *J. Biol. Chem.* **271**, 23594–23600
24. Bizzozero, O. A., McGarry, J. F., and Lees, M. B. (1987) Autoacylation of myelin proteolipid protein with acyl coenzyme A. *J. Biol. Chem.* **262**, 13550–13557
25. Bharadwaj, M., and Bizzozero, O. A. (1995) Myelin P0 Glycoprotein and a Synthetic Peptide Containing the Palmitoylation Site Are Both Autoacylated. *J. Neurochem.* **65**, 1805–1815
26. Maeda, A., Okano, K., Park, P. S. H., Lem, J., Crouch, R. K., Maeda, T., and Palczewski, K. (2010) Palmitoylation stabilizes unliganded rod opsin. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8428–8433
27. Kümmel, D., Walter, J., Heck, M., Heinemann, U., and Veit, M. (2010) Characterization of the self-palmitoylation activity of the transport protein particle component Bet3. *Cell. Mol. Life Sci.* **67**, 2653–2664
28. Broach, J. R., and Deschenes, R. (1990) The Function of Ras Genes in Saccharomyces cerevisiae. *Adv. Cancer Res.* **54**, 79–139
29. Clarke, S. (1992) Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annu. Rev. Biochem.* **61**, 355–386
30. Kato, K., Cox, A. D., Hisaka, M. M., Graham, S. M., Buss, J. E., and Der, C. J. (1992) Isoprenoid addition to Ras protein is the critical modification for its membrane association

and transforming activity. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6403–6407

31. Hancock, J. F. (2003) Ras proteins: Different signals from different locations. *Nat. Rev. Mol. Cell Biol.* **4**, 373–384

32. Bhattacharya, S., Chen, L., Broach, J. R., and Powers, S. (1995) Ras membrane targeting is essential for glucose signaling but not for viability in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2984–2988

33. Bartels, D. J., Mitchell, D. A., Dong, X., and Deschenes, R. J. (1999) Erf2, a Novel Gene Product That Affects the Localization and Palmitoylation of Ras2 in Saccharomyces cerevisiae . *Mol. Cell. Biol.* **19**, 6775–6787

34. Zhao, L., Lobo, S., Dong, X., Ault, A. D., and Deschenes, R. J. (2002) Erf4p and Erf2p form an endoplasmic reticulum-associated complex involved in the plasma membrane localization of yeast Ras proteins. *J. Biol. Chem.* **277**, 49352–49359

35. Lobo, S., Greentree, W. K., Linder, M. E., and Deschenes, R. J. (2002) Identification of a Ras palmitoyltransferase in Saccharomyces cerevisiae. *J. Biol. Chem.* **277**, 41268–41273

36. Roth, A. F., Feng, Y., Chen, L., and Davis, N. G. (2002) The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. *J. Cell Biol.* **159**, 23–28

37. Mitchell, D. A., Vasudevan, A., Linder, M. E., and Deschenes, R. J. (2006) Protein palmitoylation by a family of DHHC protein S-acyltransferases. *J. Lipid Res.* **47**, 1118–1127

38. González Montoro, A., Quiroga, R., and Valdez Taubas, J. (2013) Zinc co-ordination by the DHHC cysteine-rich domain of the palmitoyltransferase Swf1. *Biochem. J.* **454**, 427–435

39. Gottlieb, C. D., Zhang, S., and Linder, M. E. (2015) The cysteine-rich domain of the DHHC3 palmitoyltransferase is palmitoylated and contains tightly bound zinc. *J. Biol. Chem.* **290**, 29259–29269

40. Rana, M. S., Kumar, P., Lee, C. J., Verardi, R., Rajashankar, K. R., and Banerjee, A.

(2018) Fatty acyl recognition and transfer by an integral membrane S-acyltransferase. *Science (80-.).* 10.1126/science.aao6326

41. González Montoro, A., Quiroga, R., Maccioni, H. J. F., and Valdèz Taubas, J. (2009) A novel motif at the C-terminus of palmitoyltransferases is essential for Swf1 and Pfa3 function in vivo. *Biochem. J.* **419**, 301–308
42. Gottlieb, C. D., and Linder, M. E. (2017) Structure and function of DHHC protein S -acyltransferases. 10.1042/BST20160304
43. Zaballa, M. E., and van der Goot, F. G. (2018) The molecular era of protein S-acylation: spotlight on structure, mechanisms, and dynamics. *Crit. Rev. Biochem. Mol. Biol.* **53**, 420–451
44. Rajashankar, K. R., and Banerjee, A. (2018) Fatty acyl recognition and transfer by an integral membrane S -acyltransferase. 10.1126/science.aao6326
45. Jennings, B. C., and Linder, M. E. (2012) DHHC protein S-acyltransferases use similar ping-pong kinetic mechanisms but display different Acyl-CoA specificities. *J. Biol. Chem.* **287**, 7236–7245
46. Greaves, J., Munro, K. R., Davidson, S. C., Riviere, M., Wojno, J., Smith, T. K., Tomkinson, N. C. O., and Chamberlain, L. H. (2017) Molecular basis of fatty acid selectivity in the zDHHC family of S-acyltransferases revealed by click chemistry. *Proc. Natl. Acad. Sci.* **114**, E1365–E1374
47. Stix, R., Lee, C. J., Faraldo-Gómez, J. D., and Banerjee, A. (2020) Structure and Mechanism of DHHC Protein Acyltransferases. *J. Mol. Biol.* **432**, 4983–4998
48. Reddy, K. D., Malipeddi, J., DeForte, S., Pejaver, V., Radivojac, P., Uversky, V. N., and Deschenes, R. J. (2017) Physicochemical sequence characteristics that influence S-palmitoylation propensity. *J. Biomol. Struct. Dyn.* **35**, 2337–2350
49. Mitchell, D. A., Mitchell, G., Ling, Y., Budde, C., and Deschenes, R. J. (2010) Mutational analysis of Saccharomyces cerevisiae Erf2 reveals a two-step reaction mechanism for

protein palmitoylation by DHHC enzymes. *J. Biol. Chem.* **285**, 38104–38114

50. Birktoft, J. J., and Blow, D. M. (1972) Structure of crystalline a-chymotrypsin. *J. Mol. Biol.* **68**, 187–240

51. Baxter, A., Chambers, M., Edfeldt, F., Edman, K., Freeman, A., Johansson, C., King, S., Morley, A., Petersen, J., Rawlins, P., Spadola, L., Thong, B., Poël, H. Van De, and Williams, N. (2011) Non-covalent inhibitors of rhinovirus 3C protease. *Bioorganic Med. Chem. Lett.* **21**, 777–780

52. Stix, R., Song, J., and Banerjee, A. (2020) DHHC20 Palmitoyl-Transferase Reshapes the Membrane to Foster Catalysis. *Biophys. J.* **118**, 980–988

53. Montoro, A. G., Ramirez, S. C., and Taubas, J. V. (2015) The Canonical DHHC Motif Is Not Absolutely Required for the Activity of the Yeast S -acyltransferases Swf1 and Pfa4 * □. **290**, 22448–22459

54. Ohno, Y., Kashio, A., Ogata, R., Ishitomi, A., Yamazaki, Y., and Kihara, A. (2012) Analysis of substrate specificity of human DHHC protein acyltransferases using a yeast expression system. *Mol. Biol. Cell*. **23**, 4543–4551

55. Lemonidis, K., Gorleku, O. A., Sanchez-Perez, M. C., Grefen, C., and Chamberlain, L. H. (2014) The Golgi S-acylation machinery comprises zDHHC enzymes with major differences in substrate affinity and S-acylation activity. *Mol. Biol. Cell*. **25**, 3870–3883

56. Verardi, R., Kim, J.-S., Ghirlando, R., and Banerjee, A. (2017) Structural Basis for Substrate Recognition by the Ankyrin Repeat Domain of Human DHHC17 Palmitoyltransferase. *Structure*. 10.1016/j.str.2017.06.018

57. Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., Phinney, B. S., Yates, J. R., and Davis, N. G. (2006) Global Analysis of Protein Palmitoylation in Yeast. *Cell*. **125**, 1003–1013

58. Fukata, Y., Iwanaga, T., and Fukata, M. (2006) Systematic screening for palmitoyl transferase activity of the DHHC protein family in mammalian cells. **40**, 177–182

59. Greaves, J., and Chamberlain, L. H. (2011) DHHC palmitoyl transferases: Substrate interactions and (patho)physiology. *Trends Biochem. Sci.* **36**, 245–253

60. Lemonidis, K., Sanchez-perez, M. C., and Chamberlain, L. H. (2015) Identification of a Novel Sequence Motif Recognized by the Ankyrin Repeat Domain of zDHHC17 / 13 S - Acyltransferases *. **290**, 21939–21950

61. Thomas, G. M., Hayashi, T., Huganir, R. L., and Linden, D. J. (2013) DHHC8-dependent PICK1 palmitoylation is required for induction of cerebellar long-term synaptic depression. *J. Neurosci.* **33**, 15401–15407

62. Thomas, G. M., Hayashi, T., Chiu, S. L., Chen, C. M., and Huganir, R. L. (2012) Palmitoylation by DHHC5/8 Targets GRIP1 to Dendritic Endosomes to Regulate AMPA-R Trafficking. *Neuron*. **73**, 482–496

63. Sanders, S. S., Hernandez, L. M., Soh, H., Karnam, S., Walikonis, R. S., Tzingounis, A. V., and Thomas, G. M. (2020) The palmitoyl acyltransferase zdhhc14 controls kv1-family potassium channel clustering at the axon initial segment. *Elife*. **9**, 1–32

64. Fredericks, G. J., Hoffmann, F. K. W., Rose, A. H., Osterheld, H. J., Hess, F. M., Mercier, F., and Hoffmann, P. R. (2014) Stable expression and function of the inositol 1,4,5-triphosphate receptor requires palmitoylation by a DHHC6/selenoprotein K complex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16478–16483

65. Fang, C., Deng, L., Keller, C. A., Fukata, M., Fukata, Y., Chen, G., and Lüscher, B. (2006) GODZ-mediated palmitoylation of GABAA receptors is required for normal assembly and function of GABAergic inhibitory synapses. *J. Neurosci.* **26**, 12758–12768

66. Lai, J., and Linder, M. E. (2013) Oligomerization of DHHC protein S-acyltransferases. *J. Biol. Chem.* **288**, 22862–22870

67. Salaun, C., Locatelli, C., Zmuda, F., González, J. C., and Chamberlain, L. H. (2020) Accessory proteins of the zDHHC family of S-acylation enzymes. *J. Cell Sci.* **133**, 1–9

68. Mitchell, D. A., Hamel, L. D., Ishizuka, K., Mitchell, G., Schaefer, L. M., and Deschenes,

R. J. (2012) The Erf4 subunit of the yeast Ras palmitoyl acyltransferase is required for stability of the Acyl-Erf2 intermediate and palmitoyl transfer to a Ras2 substrate. *J. Biol. Chem.* **287**, 34337–34348

69. Swarthout, J. T., Lobo, S., Farh, L., Croke, M. R., Greentree, W. K., Deschenes, R. J., and Linder, M. E. (2005) DHHC9 and GCP16 Constitute a Human Protein Fatty Acyltransferase with Specificity for H- and N-Ras *. **280**, 31141–31148

70. Ko, P. J., Woodrow, C., Dubreuil, M. M., Martin, B. R., Skouta, R., Bassik, M. C., and Dixon, S. J. (2019) A ZDHHC5-GOLGA7 Protein Acyltransferase Complex Promotes Nonapoptotic Cell Death. *Cell Chem. Biol.* **26**, 1716-1724.e9

71. Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Brian, K., Wühr, M., Chick, J., Zhai, B., Kolippakkam, D., Mintseris, J., Obar, R. A., Harris, T., Artavanis-tsakonas, S., Sowa, M. E., Paulo, J. A., Harper, J. W., and Gygi, S. P. (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*. **162**, 425–440

72. Woodley, K. T., and Collins, M. O. (2019) S-acylated Golga7b stabilises DHHC5 at the plasma membrane to regulate cell adhesion. *EMBO Rep.*

73. Huang, K., Sanders, S. S., Kang, R., Carroll, J. B., Sutton, L., Wan, J., Singaraja, R., Young, F. B., Liu, L., El-Husseini, A., Davis, N. G., and Hayden, M. R. (2011) Wild-type HTT modulates the enzymatic activity of the neuronal palmitoyl transferase HIP14. *Hum. Mol. Genet.* **20**, 3356–3365

74. Tabrizi, S. J., Flower, M. D., Ross, C. A., and Wild, E. J. (2020) Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. *Nat. Rev. Neurol.* **16**, 529–546

75. Yanai, A., Huang, K., Kang, R., Singaraja, R. R., Arstikaitis, P., Gan, L., Orban, P. C., Mullard, A., Cowan, C. M., Raymond, L. A., Drisdel, R. C., Green, W. N., Ravikumar, B.,

Rubinsztein, D. C., El-Husseini, A., and Hayden, M. R. (2006) Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function. *Nat. Neurosci.* **9**, 824–31

76. Young, F. B., Butland, S. L., Sanders, S. S., Sutton, L. M., and Hayden, M. R. (2012) Putting proteins in their place: Palmitoylation in Huntington disease and other neuropsychiatric diseases. *Prog. Neurobiol.* **97**, 220–238

77. Yeste-Velasco, M., Linder, M. E., and Lu, Y. J. (2015) Protein S-palmitoylation and cancer. *Biochim. Biophys. Acta - Rev. Cancer*. **1856**, 107–120

78. Mesquita, F. S., Abrami, L., Sergeeva, O., Turelli, P., Qing, E., Kunz, B., Raclot, C., Paz Montoya, J., Abriata, L. A., Gallagher, T., Dal Peraro, M., Trono, D., D'Angelo, G., and van der Goot, F. G. (2021) S-acylation controls SARS-CoV-2 membrane lipid organization and enhances infectivity. *Dev. Cell*. **56**, 2790-2807.e8

79. De, I., and Sadhukhan, S. (2018) Emerging Roles of DHHC-mediated Protein S-palmitoylation in Physiological and Pathophysiological Context. *Eur. J. Cell Biol.* 10.1016/j.ejcb.2018.03.005

80. Ohta, E., Misumi, Y., Sohda, M., Fujiwara, T., Yano, A., and Ikehara, Y. (2003) Identification and Characterization of GCP16, A Novel Acylated Golgi Protein That Interacts with GCP170. *J. Biol. Chem.* **278**, 51957–51967

81. Shimell, J. J., Shah, B. S., Cain, S. M., Thouta, S., Kuhlmann, N., Tatarnikov, I., Jovellar, D. B., Brigidi, G. S., Kass, J., Milnerwood, A. J., Snutch, T. P., and Bamji, S. X. (2019) The X-Linked Intellectual Disability Gene Zdhhc9 Is Essential for Dendrite Outgrowth and Inhibitory Synapse Formation. *Cell Rep.* **29**, 2422-2437.e8

82. Young, E., Zheng, Z.-Y., Wilkins, A. D., Jeong, H.-T., Li, M., Lichtarge, O., and Chang, E. C. (2014) Regulation of Ras Localization and Cell Transformation by Evolutionarily Conserved Palmitoyltransferases. *Mol. Cell. Biol.* **34**, 374–385

83. Chai, S., Cambronne, X. a, Eichhorn, S. W., and Goodman, R. H. (2013) MicroRNA-134 activity in somatostatin interneurons regulates H-Ras localization by repressing the

palmitoylation enzyme, DHHC9. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.1317528110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1317528110

84. Liu, P., Jiao, B., Zhang, R., Zhao, H., Zhang, C., Wu, M., Li, D., Zhao, X., Qiu, Q., Li, J., and Ren, R. (2016) Palmitoylacyltransferase Zdhhc9 inactivation mitigates leukemogenic potential of oncogenic Nras. *Leukemia*. **30**, 1225–1228

85. Tian, L., McClafferty, H., Jeffries, O., and Shipston, M. J. (2010) Multiple palmitoyltransferases are required for palmitoylation-dependent regulation of large conductance calcium- and voltage-activated potassium channels. *J. Biol. Chem.* **285**, 23954–23962

86. Adachi, N., Hess, D. T., McLaughlin, P., and Stamler, J. S. (2016) S-Palmitoylation of a Novel Site in the β2 -Adrenergic Receptor Associated with a Novel Intracellular Itinerary. *J. Biol. Chem.* **291**, 20232–20246

87. McMichael, T. M., Zhang, L., Chemudupati, M., Hach, J. C., Kenney, A. D., Hang, H. C., and Yount, J. S. (2017) The palmitoyltransferase ZDHHC20 enhances interferon-induced transmembrane protein 3 (IFITM3) palmitoylation and antiviral activity. *J. Biol. Chem.* **292**, 21517–21526

88. Yang, Y., Hsu, J. M., Sun, L., Chan, L. C., Li, C. W., Hsu, J. L., Wei, Y., Xia, W., Hou, J., Qiu, Y., and Hung, M. C. (2019) Palmitoylation stabilizes PD-L1 to promote breast tumor growth. *Cell Res.* **29**, 83–86

89. Zhang, Z., Li, X., Yang, F., Chen, C., Liu, P., Ren, Y., Sun, P., Wang, Z., You, Y., Zeng, Y. X., and Li, X. (2021) DHHC9-mediated GLUT1 S-palmitoylation promotes glioblastoma glycolysis and tumorigenesis. *Nat. Commun.* **12**, 1–12

90. Raymond, F. L., Tarpey, P. S., Edkins, S., Tofts, C., O'Meara, S., Teague, J., Butler, A., Stevens, C., Barthorpe, S., Buck, G., Cole, J., Dicks, E., Gray, K., Halliday, K., Hills, K., Hinton, J., Jones, D., Menzies, A., Perry, J., Raine, K., Shepherd, R., Small, A., Varian,

J., Widaa, S., Mallya, U., Moon, J., Luo, Y., Shaw, M., Boyle, J., Kerr, B., Turner, G., Quarrell, O., Cole, T., Easton, D. F., Wooster, R., Bobrow, M., Schwartz, C. E., Gecz, J., Stratton, M. R., and Futreal, P. A. (2007) Mutations in ZDHHC9, which encodes a palmitoyltransferase of NRAS and HRAS, cause X-linked mental retardation associated with a Marfanoid habitus. *Am. J. Hum. Genet.* **80**, 982–987

91. Baker, K., Astle, D. E., Scerif, G., Barnes, J., Smith, J., Moffat, G., Gillard, J., Baldeweg, T., and Raymond, F. L. (2015) Epilepsy, cognitive deficits and neuroanatomy in males with ZDHHC 9 mutations . *Ann. Clin. Transl. Neurol.* **2**, 559–569

92. Han, J. Y., Lee, I. G., Shin, S., Kim, M., Jang, J. H., and Park, J. (2017) The first patient with sporadic X-linked intellectual disability with de novo ZDHHC9 mutation identified by targeted next-generation sequencing. *Eur. J. Med. Genet.* **60**, 499–503

93. Schirwani, S., Wakeling, E., Smith, K., and Balasubramanian, M. (2018) Expanding the molecular basis and phenotypic spectrum of ZDHHC9-associated X-linked intellectual disability. *Am. J. Med. Genet. Part A*. **176**, 1238–1244

94. Zhang, M. M., Wu, P. Y. J., Kelly, F. D., Nurse, P., and Hang, H. C. (2013) Quantitative Control of Protein S-Palmitoylation Regulates Meiotic Entry in Fission Yeast. *PLoS Biol.* 10.1371/journal.pbio.1001597

95. Yeste-Velasco, M., Linder, M. E., and Lu, Y. J. (2015) Protein S-palmitoylation and cancer. *Biochim. Biophys. Acta - Rev. Cancer*. **1856**, 107–120

96. Kawate, T., and Gouaux, E. (2006) Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. 10.1016/j.str.2006.01.013

97. Malgapo, M. I. P., and Linder, M. E. (2021) Substrate recruitment by zDHHC protein acyltransferases. *Open Biol.* 10.1098/rsob.210026

98. Mitchell, D. A., Hamel, L. D., Reddy, K. D., Farh, L., Rettew, L. M., Sanchez, P. R., and Deschenes, R. J. (2014) Mutations in the X-linked intellectual disability gene, zDHHC9,

alter autopalmitoylation activity by distinct mechanisms. *J. Biol. Chem.* **289**, 18582–18592

99. Lee, S. C., Knowles, T. J., Postis, V. L. G., Jamshad, M., Parslow, R. A., Lin, Y., Goldman, A., Sridhar, P., Overduin, M., Muench, S. P., and Dafforn, T. R. (2016) A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat. Protoc.* **11**, 1149–1162

100. Dörr, J. M., Scheidelaar, S., Koorengevel, M. C., Dominguez, J. J., Schäfer, M., van Walree, C. A., and Killian, J. A. (2016) The styrene–maleic acid copolymer: a versatile tool in membrane research. *Eur. Biophys. J.* **45**, 3–21

101. Collins, M. O., Woodley, K. T., and Choudhary, J. S. (2017) Global , site-specific analysis of neuronal protein S-acylation. *Sci. Rep.* **7**, 1–14

102. Cravatt, B. R. M. and B. F. (2014) Large-Scale Profiling of Protein Palmitoylation in Mammalian Cells. *Bone*. **23**, 1–7

103. Renaldo C. Drisdel, and William N. Green (2004) Labelling and quantifying sites of protein palmitoylation. *Biotechniques*. **36**, 276–285

104. Wan, J., Roth, A. F., Bailey, A. O., and Davis, N. G. (2007) Palmitoylated proteins: Purification and identification. *Nat. Protoc.* **2**, 1573–1584

105. Howie, J., Reilly, L., Fraser, N. J., Vlachaki Walker, J. M., Wypijewski, K. J., Ashford, M. L. J., Calaghan, S. C., McClafferty, H., Tian, L., Shipston, M. J., Boguslavskyi, A., Shattock, M. J., and Fuller, W. (2014) Substrate recognition by the cell surface palmitoyl transferase DHHC5. *Proc. Natl. Acad. Sci.* **111**, 17534–17539

106. Plain, F., Howie, J., Kennedy, J., Brown, E., Shattock, M. J., Fraser, N. J., and Fuller, W. (2020) Control of protein palmitoylation by regulating substrate recruitment to a zDHHC-protein acyltransferase. *Commun. Biol.* **3**, 1–10

107. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J.,

Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589

108. Rocks, O., Gerauer, M., Vartak, N., Koch, S., Huang, Z. P., Pechlivanis, M., Kuhlmann, J., Brunsveld, L., Chandra, A., Ellinger, B., Waldmann, H., and Bastiaens, P. I. H. (2010) The palmitoylation machinery is a spatially organizing system for peripheral membrane proteins. *Cell*. **141**, 458–471

109. Fukata, M., Fukata, Y., Adesnik, H., Nicoll, R. A., and Bredt, D. S. (2004) Identification of PSD-95 palmitoylating enzymes. *Neuron*. **44**, 987–996

110. Stetsenko, A., and Guskov, A. (2017) An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals*. **7**, 197

111. Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G., and Vogel, H. (2011) Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* **29**, 335–340

112. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242

113. Rupp, B. (2009) *Biomolecular crystallography: Principles, practice, and application to structural biology*

114. Lyumkis, D. (2019) Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197

115. Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J.,

Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C. S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M. R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K. C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M. L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T. B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and Zhang, J. (2021) UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489

116. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788

117. Benjin, X., and Ling, L. (2020) Developments, applications, and prospects of cryo-electron microscopy. *Protein Sci.* **29**, 872–882

118. Henderson, R. (2015) Overview and future of single particle electron cryomicroscopy. *Arch. Biochem. Biophys.* **581**, 19–24

119. Kalienkova, V., Alvadia, C., Clerico Mosina, V., and Paulino, C. (2020) Single-Particle Cryo-EM of Membrane Proteins in Lipid Nanodiscs. *Methods Mol. Biol.* **2127**, 245–273

120. Scheres, S. H. W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM

structure determination. *J. Struct. Biol.* **180**, 519–530

121. Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12

122. Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017) CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*. **14**, 290–296

123. Goehring, A., Lee, C.-H., Wang, K. H., Michel, J. C., Claxton, D. P., Baconguis, I., Althoff, T., Fischer, S., Garcia, K. C., and Gouaux, E. (2014) Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585

124. Ames, Robert S., James A. Fornwald, Parvathi Nuthulaganti, John J. Trill, James J. Foley, Peter T. Buckley, Thomas A. Kost, Zining Wu, and M. A. R. (2004) BacMam recombinant baculoviruses in G protein–coupled receptor drug discovery. *Recept. Channels*. **10**, 99–107

125. Boudjelal, M., Mason, S. J., Katso, R. M., Fleming, J. M., Parham, J. H., Condreay, J. P., Merrihew, R. V., and Cairns, W. J. (2005) The application of BacMam technology in nuclear receptor drug discovery. *Biotechnol. Annu. Rev.* **11**, 101–125

126. Fonfria, E., Clay, W. C., Levy, D. S., Goodwin, J. A., Roman, S., Smith, G. D., Condreay, J. P., and Michel, A. D. (2008) Cloning and pharmacological characterization of the guinea pig P2X 7 receptor orthologue. *Br. J. Pharmacol.* **153**, 544–556

127. Shukla, S., Schwartz, C., Kapoor, K., Kouanda, A., and Ambudkar, S. V. (2012) Use of baculovirus BacMam vectors for expression of ABC drug transporters in mammalian cells. *Drug Metab. Dispos.* **40**, 304–312

128. Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*. **504**, 107–112

129. Baconguis, I., and Gouaux, E. (2012) Structural plasticity and dynamic selectivity of acid-

sensing ion channelspider toxin complexes. *Nature*. **489**, 400–405

130. Lee, C. H., Lü, W., Michel, J. C., Goehring, A., Du, J., Song, X., and Gouaux, E. (2014) NMDA receptor structures reveal subunit arrangement and pore architecture. *Nature*. **511**, 191–197

131. Karasawa, A., Michalski, K., Mikhelzon, P., and Kawate, T. (2017) The P2X7 receptor forms a dye-permeable pore independent of its intracellular domain but dependent on membrane lipid composition. *Elife*. **6**, 1–22

132. Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M., and Sligar, S. G. (2009) Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol.* **464**, 211–231

133. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394

134. Herzik, M. A., Wu, M., and Lander, G. C. (2019) High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* **10**, 1–9

135. Zhang, Y., Tammaro, R., Peters, P. J., and Ravelli, R. B. G. (2020) Could Egg White Lysozyme be Solved by Single Particle Cryo-EM? *J. Chem. Inf. Model.* **60**, 2605–2613

136. Jiang, Y., Benz, T. L., and Long, S. B. (2021) Substrate and product complexes reveal mechanisms of Hedgehog acylation by HHAT. *Science (80-.).* **372**, 1215–1219