AN ABSTRACT OF THE DISSERTATION OF

Sara Jone Codding for the degree of Doctor of Philosophy in Biochemistry & Biophysics presented on December 13, 2016.

Title: Functional and Structural Analyses of Three Distinct proteins by Biochemical and Biophysical Techniques

Abstract approved:_____

Colin P. Johnson

The focus of this dissertation was to understand protein function and structure on a molecular level. To do this successfully, a variety of biochemical and biophysical techniques were employed. A wide variety of techniques were applied to study the proteins of interest in this dissertation to probe protein-protein interactions, protein function, protein structure, and protein regulation. In addition, the use of bioinformatics has placed the similarities and differences of these proteins compared to homologs in perspective within their protein families.

There are three chapters of original research presented here, the second chapter is published, the third contains my contribution to a work that is published, and the fourth will be submitted for publication. The second chapter of this thesis describes the studies on the protein dysferlin. Mutations in dysferlin are responsible for late onset muscular dystrophies where muscle cells are deficient in resealing the cell membrane upon routine injury. We demonstrate that dysferlin catalyzes the rate-limiting step of membrane fusion and mediates lipid mixing in a calcium sensitive manner. This work was the first in the field to determine a mechanism of action of dysferlin and suggests a molecular basis for why mutations that abrogate dysferlin's activity result in muscular dystrophy.

The third chapter includes my contribution to the complete structural solution of the enzyme ValA involved in the synthesis of the crop protectant validamycin A. This was the first structure of a sedoheputlose-7-phoshpate cyclase, and the overall fold

reflects that of similar sugar cyclase proteins. But, the sequence and structural analysis of ValA with respect to other related proteins revealed the first novel theory regarding substrate specificity of ValA. It is postulated that due to sequence alignment and structural data, ValA selectively binds the β-anomer of its substrate which pre-assigns stereochemistry of the enzymatic product.

In the fourth chapter of this dissertation we design several variants of the protein merlin involved in cell contact growth inhibition. Whereas a complete structural solution was never obtained for merlin in this study, we used *in vitro* techniques to test the role of merlin's central domain in the conformational regulation of the protein. This was the first study on merlin, or the homologous proteins ezrin, radixin, and moesin, that focused on the role of the central domain. Our results confirm that this domain is not merely structural, but mediates key interactions that modulate merlin's degree of intramolecular association. That merlin's activity is central to cell contact growth inhibition and that this activity is modulated by merlin's degree of intramolecular association, it is clear now that the central domain plays a key role in the activity of merlin.

In the final chapter and fifth chapter I conclude this dissertation.

Functional and Structural Analyses of Three Distinct Proteins by Biochemical and Biophysical Techniques

by

Sara Jone Codding

A DISSERTATION

submitted to

Oregon State University

In partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented December 13, 2016

Commencement June 2017

Doctor of Philosophy dissertation of Sara Jone Codding presented on December 13, 2016.

APPROVED:

Major Professor, representing Biochemistry & Biophysics

Head of the Department of Biochemistry & Biophysics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorized release of my dissertation to any reader upon request.

Sara Jone Codding, Author

ACKNOWLEDGEMENTS

First, I thank my advisor, Colin Johnson, for his mentorship over the past four years. The opportunity to continue scientific research in your lab has provided me the foundation to grow as a student and individual so that I now approach challenging problems with critical rationale. Your ability to hold unique space for all your students, your calm disposition, and dedication to science has provided a pleasant learning environment for me, and will for students to come. I thank you for your patience as I have developed as a scientist and for encouraging me to work towards my future goals. Finally, I thank you for providing the first lens through which I began to find rationale, or lack there of, to explain molecular biology nuances, as first a chemist, and second as a biophysicist.

I thank Russell Carpenter for his patience and personal dedication to my dissertation work. His contribution to my development as a scientist is broad and has certainly been beneficial. I thank Dale Tronrud for his unwavering criticisms and realistic perspectives of all the things. As well as for his kind introduction to Art Robbins, a truly fabulous individual.

My gratitude to my good friends and lab mates: Chelsea Holman-Wolk for your help, advice, sense of humor, and lighthearted (or hammer-down) presence in the Johnson lab, Murugesh Padmanarayana for our friendly conversations, and Naomi J. Marty, for always reminding me that life is short. I have learned so much from you all and thank you for your support both in and out of the laboratory.

More immediately, my related blood family, while they may not understand the nuances of my research, their foundation of support has never wavered. My parents, Jim and Cindy who both feed my curiosities to understand how things work which reinforced my dexterity to pipette like a champ. My grandmother Lois, who never fails to tell every person she meets about my doctoral work. I am certain she is bragging for herself, Mary, and Jerome. My siblings, who both have come to visit me in Corvallis, and still act humored by my scientific puns. I also thank my extended family, Lynn and Kasey, for

their loving support. Also, I am quite blessed to have two very qualified dissertation supporters and inspectors, Minnie and Tiger, and I am ever thankful for the supportive words of Robert James Smith.

I would like to acknowledge the NIH NIDCD K99/R00 and NIH NIDCD R01 grants to Dr. Johnson for their financial support.

Lastly, and most importantly, I would like to thank Erin, who was my strength when I was weak, my voice of reason, and of course for that uncanny ability to see potential where I cannot. It would be an understatement to say that this dissertation wouldn't have been possible without you.

CONTRIBUTION OF AUTHORS

Dr. Naomi Marty contributed to the cloning and purification of dysferlin and SNARE constructs utilized in Chapter 2. Dr. Nazish Abdulla contributed in cloning full-length soluble dysferlin in our pMCSG9 vector in Chapter 2. Dr. Russell Carpenter who was responsible for the cloning and initial purification of the protein merlin in Chapter 4. Dr. Donnie Berkholz who was responsible for the homology model of merlin and Nicolas O. Thomas for the site directed mutagenesis work performed in Chapter 4. Colin P. Johnson was involved in the design and analysis of experiments in Chapter 2, and the writing of all thesis chapters. P. Andrew Karplus was involved in the design and analysis of experiments in Chapter 3 and 4, as well as the writing of Chapter 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

Table Page

Title: Functional and Structural Analyses of Three Distinct Proteins by Biochemical and Biophysical Techniques

Chapter 1

Dissertation Overview

Introduction to the Study of Protein Structure and Function

Proteins mediate virtually every process in a cell, exhibiting an endless array of diversity in structure and function. The importance of proteins in cellular processes was clear at the onset of biochemical studies, and it was as early as 1926 that most enzymes were found to be proteinaceous, a discovery that earned James Sumner the Nobel Prize in 1946. It was just a few years before Sumner was awarded this prize that DNA was heralded as the genetic material. That DNA displaced protein as the heritable molecule paved the way to subsequent research that frames how we now know proteins, as the "workhorses" of the cell.

On first glance protein diversity appears deceptively simple. Each organism utilizes the same 20 amino acid palette to generate each individual protein required for biological processes. But, with as many permutations as nature has allowed, proteins perform functions as distinct as enzymatic catalysis to organizing cortical actin to give rise to microvilli. The precise nature of protein sequence in protein function is clear when a single mutation can lead to a catastrophic defect or disease. And so it becomes an essential biochemical pursuit to understand how proteins function, as they respond to and carry out the biological requirements of an organism on the molecular level.

The understanding of protein function has advanced molecular treatments of disease and most often it has been via drugable enzymatic targets to regulate a desired physiological process. This has bolstered an industry built on molecular treatments for chronic disease. However, often times it is diseases themselves that make it clear a protein is crucial for a specific cellular function. But, to understand the molecular processes of disease, functional studies are necessary to assign a role to that protein. In addition, it is clear that evolutionary pressure has provided years of trial and error in designing molecular tools that can perform complex chemical reactions more efficiently, and in physiologically benign conditions than that of traditional chemistry. Such that enzymatic catalysis has been used to make compounds for molecular treatments and

nature will continue to be a source of active complex biological molecules for chemical scaffolds in drug manufacturing.

Regardless of the end goal, many research tools exist to test protein function in a variety of settings. Whereas much can be learned about a protein from performing carefully designed functional assays, it is essential to utilize both biochemical and biophysical techniques to glean a more complete understanding of a protein's structural-functional relationships. The overall goal of this dissertation, therefore, is to use a variety of techniques to provide new insights into the function, structure, or regulation of three distinct proteins. Many of these insights offer important advances to our understanding of these individual proteins while utilizing biochemical or biophysical approaches that can be applied in broader applications. In the rest of this chapter, I will briefly introduce some fundamental techniques utilized to gain insight into protein structure-function relationships utilized in subsequent chapters. This will then lead into an overview of each individual protein studied in this dissertation to provide the necessary background for understanding the questions that are addressed in each of the three research chapters. For convenience, the end of this chapter provides a quick summary of the topics of these chapters.

Introduction to Protein Structure and Structure Determination

Protein Structure. Proteins are long polymers of amino acids and the specific sequence of amino acids in a protein is encoded by a single gene. It's this sequence that gives rise to protein structure. In a finished polypeptide there are from 500 to 10,000 single bonds, each with up to 360 rotational degrees of freedom (1). But, protein structures assume only one or a very small number of interchanging structures. Protein structures are largely constrained because of particular dihedral angles that the protein backbone can adopt. This is has been attributed to the double bond-like character of the peptide bond, ω. But the dihedral angles adjacent to ω, ϕ and Ψ, adopt defined values depending on local secondary structure, as determined by Ramachanandran (2). As

depicted in the classic Ramachandran plot, the most commonly observed secondary structures are α-helices and β-sheets, each which adopt a well-defined, distinct range of allowed ϕ and Ψ values. The α-helix is an important structural aspect discussed in Chapter 4. A protein will have secondary structures along a peptide chain that will associate to give rise to tertiary structure through a variety of driving forces involving entropic, ionic, hydrophobic and hydrogen bond interactions in a process that is poorly defined nor well understood. As a complete protein structure folds, there forms distinct interactional sites or catalytic sites that are key to that protein's function in a cell.

Protein structure determination. That proteins assume only a small number of interchanging structures, it has been possible to obtain crystals of proteins; growing crystals requires the ability to assemble an ordered repeating lattice. In fact, protein crystals have been obtained for many proteins. In the event that a protein crystal, is subjected to X-rays and diffracts at a resolution that discerns the distance between adjacent residues, a complete density surface of the protein can be reconstructed from the resultant diffraction pattern to yield a model of a protein's structure. Examination of protein structure can yield powerful insights into protein function. In fact, the number of proteins with experimentally determined structures has been steadily rising as crystallographic techniques improve. But, a limited variety of autonomously folding protein structures have been identified by crystallography, suggesting that proteins encoded by the genome reflect similarities because of genetic heritability, like members of a family (3). Thus, much information can be gleaned from analyzing a protein within context of other genetically related proteins. The structural-functional inferences from comparing related proteins are explored in both Chapters 3 and 4.

Protein structure-functional insights from analyzing familial similarities. While crystallographic models have been obtained for many proteins, it is still only a small minority of the many proteins that have been identified and even purified. This is not because of the fact that it has been only fifty years since the first protein structure; often times technical problems associated with the crystallographic method have prevented a wider variety of structural determinations. Thus, often it becomes necessary to examine a

protein that has eluded crystallization in the context of well studied and/or structurally determined related proteins to begin to draw inferences from familial similarities when testing protein structure-function relationships.

On the most basic level, protein sequences can be analyzed within a group of proteins by computer algorithms to identify similarities. The residues of two or more protein sequences are arranged so that regions of similarity may be identified. Similarities beyond random are often observed and are due to structural or functional relationships as a consequence of close evolutionary relation (4). Primary sequence alignments can be guided by previously determined structural information to add additional dimensions to the alignment by aligning residues that are structurally equivalent (5). The degree of similarity between amino acids occupying a given position in an alignment can be interpreted as how well that particular position is conserved among a protein family. Divergence observed in sequence alignment are often interpreted as point mutations that occurred to give rise to a new protein within a different genome and perhaps with new functionality. Sequence alignments have been utilized in both Chapters 3 and 4 to gain insight about proteins that eluded crystallization, to guide chemical modifications, or to compare similarities between proteins in the same family. In both chapters, structurally determined proteins in that family guided sequence alignment.

Chemical Methods to Study Protein Structure-Function Relationships

In the absence of structural solution and when protein function or mechanism theories are to be tested, the molecular actions of a protein are studied by a diverse collection of techniques. Several of the approaches commonly utilized are types of chemical modifications, that may include mutagenesis, fluorescent probes coupled covalently to proteins, and immunochemical probes. But, these experimental approaches can be as varied as the proteins themselves and if implementation of the technique is

designed to answer a specific experimental question, functional insights on a protein can be obtained.

Site directed mutagenesis. To demonstrate that a particular amino acid is involved in the function of a protein, site-directed mutation is a chemical modification that is typically utilized. This technique converts the amino acid of interest, into one of the other 19 amino acids, and results are assessed by the effect of the modification on some property of the protein (6,7). Using this approach often times necessitates structural guidance to determine which amino acid(s) can be modified, and if a structure is unavailable, sequence analysis among related proteins with structural solution can be insightful. Furthermore, it is possible to test roles of domains by the deletion of many amino acids from the sequence of a protein. This approach was utilized to design protein constructs and to test domain function in Chapter 4.

Fluorescence spectroscopy. The use of fluorescence spectroscopy has been a valuable tool for probing protein function as it is has a relatively high sensitivity and low background due to the quantum yield and specific frequencies of fluorescent dyes. It has been used in a wide variety of applications from studying protein localization in cells to studying protein dynamics and kinetics *in vitro*. In cells, applications of fluorescence may involve a variety of fluorophores of different compositions: antibodies coupled to fluorogenic probes, dyes, and introduction of fluorogenic fusion protein constructs into a cell. *In vitro* experiments typically employ endogenous fluorogenic amino acids in a protein (typically tryptophan) or the introduction of fluorophore at site-specific locations using chemical modification.

Regardless, advancements in fluorescence probe synthesis has allowed for application of fluorescent probes in many settings. Probes designed to label proteins via chemical modification are often times electrophilic reagents that are directed to nucleophilic amino acid side chains. Commonly, those amino acids are cysteines, the amino- or carboxy- termini (8). In addition, fluorescence techniques can be used to examine many types of molecular interactions to analyze protein-protein, protein-nucleic

acid, ligand-receptor, and protein-lipid interactions. These techniques can yield dynamic information on protein conformational changes as well as biophysical information such as binding constants.

Whether an endogenous fluorophore is utilized or a synthetic one that has been introduced to a system of interest, absorption of incident light by a fluorophore leads to an excited molecular state. Decay to the ground state can occur by emission of the light at a wavelength lower than the incident light, loss of energy to vibrational/rotational/heat transfer, or transfer of energy to an alternate molecule that serves to accept the energy from an excited one (9).

Fluorescence resonance energy transfer. Fluorescence resonance energy transfer (FRET) is an application of fluorescence spectroscopy that involves transfer of energy to an alternate molecule that serves to accept the energy from an excited one. Energy is transferred from an excited fluorophore, or donor, to an acceptor fluorophore. The energy transfer efficiency between the two fluorophores is dependent on their molecular composition and falls off at $1/r^6$ distance between the two probes. Thus, there is a limitation on the distance between the donor and acceptor molecule for energy transfer to occur. Typically, the distance between the two probes for energy transfer to be 50% efficient is called the Förster radius, and depending on the FRET pair may range from 1-10nm (10).

FRET experiments are designed to measure a change in distance between the donor and acceptor probes. Generally, this distance is interpreted as a conformational change, a binding event, or a change in molecular orientation or dynamics (10). In general, the donor and acceptor probes are covalently attached to a protein at a specific location(s) either by introduction of a tryptophan by site directed mutagenesis, endogenous tryptophan, or chemical modification by coupling a probe to a cysteine. Both the donor and acceptor may be on the same protein to test conformation change, or they may be on two different putative associating proteins to test intermolecular interactions. In certain applications, synthetic lipids with fluorescent head groups can be

imbedded in a lipid bilayer to test protein-lipid interactions. We utilized FRET extensively to test a variety of functional properties Chapter 2.

Fluorescence Anisotropy. Fluorescence anisotropy is an applied fluorescence technique that is typically used to measure *in vitro* binding events between two proteins. In general, a concentration of fluorescently labeled protein in solution containing an endogenous fluorophore, or a covalently introduced one, is excited with polarized light. In solution, the labeled proteins are randomly oriented with respect to each other. In the time between excitation and emission, the likelihood of molecules in solution to move with respect to their orientation at the time of excitation is dependent on the viscosity of the solution, the temperature, and size of the labeled protein. In the event that a binding event occurs between a labeled protein in solution and a putative unlabeled binding partner, a complex will form if the two associate, increasing the size of the rotating complex containing the fluorophore. This increase in size because of a binding event slows down the rotation of the fluorophore, increasing the probability that the incident light will be remitted polarized (11). The ratio of emitted polarized light to total light emitted by the fluorophore is anisotropy. Anisotropy increases as a function of the amount of unlabeled binding partner added to the solution of labeled proteins if the two associate. Thus, anisotropy experiments can yield binding affinity, a property measured in Chapter 2.

Testing the Functional Properties of The Protein Dysferlin

Cell wounding is a repair or die event. Mammalian cells are prone to membrane disruption. Unlike fungi, plants, or bacteria, they mammalian cells lack a cell wall or cuticle to help maintain cell integrity in the event of membrane stress. Studies have highlighted that cell repair is crucial to physiological processes such as shedding pore forming bacterial toxins (Figure 1.1) (12). Plasma membrane integrity is also compromised if tissue suffer infarction. Generally, oxygen starvation ultimately leads to depletion cellular energy that supply membrane pumps to maintain ionic homeostasis, mammalian cells die if homeostasis cannot be regained (13). Mechanically active tissue

also experiences membrane stress, such as the epithelial cells that line arteries in response to changes in pressure, or skeletal muscle in response to a force loaded contraction (14). In response to plasma membrane tears, mammalian cells have evolved a repair response that is triggered by the influx of extracellular calcium to promote vesicle exocytosis to close a cellular wound. The influx of calcium recruits vesicles to the wound site that fuse together and then to the membrane through exocytosis to form a patch. This patch hypothesis is supported by cell wounding studies in fibroblasts and sea urchin oocytes (Figure 1.2) (12,15,16).

Loss of dysferlin causes muscular dystrophy and loss of membrane repair in skeletal muscle. In skeletal muscle, membrane repair is dependent on calcium much like in fibroblasts and sea urchin oocytes, but while calcium is necessary for repair, it is not sufficient. Skeletal muscle requires the concerted action of both calcium and the protein dysferlin for membrane repair. Mice without functional dysferlin fail to repair the sarcolemma in a calcium dependent manner. These mice also display misstrafficked vesicles clustered in muscle cells beneath wound sites (17). In 1993, mutations in the protein dysferlin were linked to the muscular dystrophies Miyoshi Myopathy (MM) and Limb Girdle Muscular Dystrophy type 2, which are characterized as muscle wasting diseases (LGMD2) (18).

Dysferlin is a large protein with seven C-terminal C2 domains and a single pass transmembrane domain (Figure 1.3). Dysferlin localizes to the plasma membrane and intracellular vesicles in skeletal muscle (19,20). Despite the twenty-three years since dysferlin's discovery, its role in membrane repair has not been tested and its basic molecular properties have not been measured. But, much has been inferred about dysferlin from other C2 domain containing proteins such as synaptotagmin, Doc2, and the other ferlin proteins (21,22). These C2 domain-containing proteins have been extensively studied *in vitro*. Generally, C2 domain proteins share a fold that is composed of two parallel beta sheets connected by loops. These loops contain acidic residues that chelate calcium (Figure 1.4). Upon, calcium chelation, C2 domains typically associate

with lipid bilayers via electrostatic interactions between calcium with negatively charged lipid head-groups (23).

In vivo work has suggested that dysferlin with its multiple C2 domain may work as the calcium sensor in cell membrane repair (24). In support of this hypothesis misstrafficked vesicles are observed in dysferlin deficient cells and muscle membrane repair requires calcium. This has led to the model that dysferlin senses calcium in response to a membrane tear and serves as a protein scaffold or vesicle recruiter to promote membrane repair in skeletal muscle. The model that had developed is similar to that of the patch hypothesis, however, including other key skeletal muscle proteins shown to be near wound sites such as annexins and MG53 (Figure 1.5) (25).

Dysferlin may play a larger role in vesicle trafficking events. Other work has shown that dysferlin plays a role in membrane trafficking events such as cytokine release and acid sphingomyelinase secretion (26,27). Collectively, organismal studies and cellular studies support a wide role for dysferlin in calcium sensitive membrane trafficking events but the mechanism by which dysferlin operates is unclear.

Membrane fusion requires, calcium, SNAREs, and a SNARE effector protein. Membrane trafficking, fusion, and exocytosis, including membrane repair, require minimal membrane machinery called SNAREs (Soluble NSF attachment protein) (15,28). C2 domain containing proteins such as synaptotagmin have been shown to bind SNAREs and lipids to facilitate membrane fusion in a calcium sensitive manner (Figure 1.6) (29). Synaptotagmin is a tandem C2 domain protein that is involved in synaptic exocytosis and is well -studied protein, with clearly defined calcium and lipid binding affinities. Research on synaptotagmin has shown that its function in membrane fusion is to serve as a calcium sensor and SNARE effector protein, as SNAREs are calcium insensitive. Due to dysferlin's similar structure to synaptotagmin, and that membrane fusion and cell repair is SNARE dependent, it is hypothesized that dysferlin could act in a similar fashion to synaptotagmin to promote membrane fusion. However, the basic biochemical properties of synaptotagmin that facilitate its function had not been tested nor measured for dysferlin.

The work performed by Abdulla et al (2014) established the first *in vitro* characterization of dysferlin's calcium and lipid binding properties. It became apparent that dysferlin was a ideal candidate to mediate calcium triggered membrane fusion events in skeletal muscle (30). Furthermore, in the Johnson et al. (2010) study it was found that the dysferlin homolog, otoferlin, mediated SNARE dependent lipid mixing in a calcium sensitive manner (31). It held logical that dysferlin could be capable of a similar action when *in vivo* and *in vitro* evidence were both taken into consideration. In Chapter 2, I set out to determine if dysferlin could act a SNARE effector to facilitate calcium dependent membrane fusion using fluorescent techniques. My studies answered distinct functional questions about dysferlin's role in membrane repair that, when tested in a minimal *in vitro* system, yielded quantitative results. That a function of dysferlin has been determined, a biochemical foundation has been set to perform creative mechanistic and mutational studies to follow up this work.

The Structure and Sequence Analysis of the Enzyme ValA

Sequence analysis and a structural pursuit of ValA. Proteins involved in secondary metabolite synthesis in bacteria and fungi generate compounds that are ancillary to survival, and often help organisms fill a niche or maintain one. Typically, secondary metabolites serve as competitive weapons that are used against other organisms, metal transporting agents, compounds that promote symbiosis, sexual hormones, or differentiation effectors. In addition, secondary metabolites have been used in medical applications as pharmaceuticals or in agricultural applications as pesticides. For example, both penicillin and streptomycin, which are used to treat bacterial infections in humans, are produced in nature and serve survival functions for the organisms producing them (32).

That secondary metabolites are often utilized in the medical or agricultural industries, study of proteins involved in their synthesis is often a biochemical pursuit. In addition, it is advantageous to find additional proteins that may perform similar catalysis.

Typically this is accomplished by searching databases of deposited genomes for proteins that display similar sequences. This process called 'genome mining' can help to discover homologous proteins in gene cassettes of bacteria or fungi. The newly identified gene is isolated and expressed in a heterologous host so biochemical studies to determine its structure-function relationships can be performed.

The protein ValA is an enzyme that is well studied because of its role in the metabolic pathway of validamycin, a crop protectant (33). Through genome mining and bioinformatics techniques, additional proteins with similarities to ValA were identified (34). The genes that encode these proteins were expressed in a heterologous host and characterized biochemically. It was found that, like ValA, they use the same substrate, the seven carbon sugar sedoheptulose-7-phosphate, to generate stereochemically distinct products. The proteins were named after their respective products 2-epi-valionone synthase (EVS) and desmethyl-4-deoxygadusol synthase (DDGS) (35).
Whereas biochemical studies with EVS and DDGS suggested plausible catalytic mechanisms, it was unclear how ValA, EVS, and DDGS could utilize the same substrate to generate distinct products. In Chapter 3, we explore the sequence similarities of ValA, EVS, and EEVS, in the context of other structurally determined proteins in their family. In addition, we solve the structure of ValA and utilize both structural and primary sequence analysis to gain insights into this family of proteins.

Protein design, Mutational Studies, and The Structural Pursuit of the Protein Merlin

The protein merlin is a moesin-ezrin-radixin (ERM) like protein. Merlin is involved in cell contact growth inhibition and somatic mutations in merlin are linked to neurofibromatosis, which typically results in an overgrowth of Schwann cells in the brain (36). *In vivo* studies have shown that loss of merlin is lethal in mice and flies, but leads to tumors in cultured mammalian cells that still express ERM proteins (37,38). Thus, there appears to be distinct functions performed by merlin that are not rescued by other

ERM proteins. A variety of experiments have shown that merlin expression is enriched in mouse lung, intestine, spleen, kidney, spinal cord, and brain (39). But, during fetal development merlin is especially enriched in the heart, the nervous and skeletal systems (40). Cellular studies have localized merlin in cultured cells to the ruffling membranes (41).

In vertebrates, ezrin, radixin, and moesin show high sequence identity and non-vertebrates show clear homologues with high (56-58%) percent identity in sea urchins, flies and worms. The N-terminal functional domain is highly conserved followed by a central α-helical domain that forms a coiled-coil. The C-terminal 80 residue hydrophilic domain shows high sequence identity, with the terminal 26 residues with >78% identity from mammals to nematodes. Merlin shows a relatively high sequence identity in the N-terminal domain to other ERM proteins but lower sequence identity in the central and C-terminal domains (42).

Biochemical and cell biological studies have demonstrated that ERM proteins are conformationally regulated. It was shown that for ezrin, the N-terminal domain binds very tightly to its C-terminal domain and thereby masks binding sites for EBP50 and F-actin (43). Similarly, the radixin N-terminal domain associates with its C-terminal domain, which prevents RhoGDI from binding radixin's N-terminus (44). In addition, transfecting cells to express ERM protein C terminal fragments generates long surface extensions in cells, but this is suppressed by co-transfection of N- and C- terminal fragments (45).

A co-crystal structure of the N- and C-terminal domains of moesin revealed a tightly packed globular tri-lobed N-terminal domain with a meandering C-terminal domain that wanders across the N-terminus, burying a surface area of 2700Å revealing multiple interactions between the N- and C-terminus (46). This interaction blocks the binding site on the N-terminus which localizes ERM proteins to the cortical membrane through interactions with the protein EBP50. This suggests that the intramolecular association of ERM proteins render them inactive. This was recently shown for merlin,

which up until 2012, was thought to be active in its closed state (47). We discuss this further in Chapter 4.

Biochemical studies have revealed that merlin's N-terminal domain associates more strongly with EBP50 than it's own C-terminus suggesting that merlin's intramolecular association is not as tight as ezrin (48). In chapter 4 we explore the sequence similarities between the EBP50 binding epitope to the N-terminus of ERM proteins and the C-termini of the ERMs. This combined with available structural information on the N-termini of ERMs and a co-crystal structure of the self associated N and C- termini of moesin, has led us to design mutations in merlin to alter it's self association.

The role of the central α-domain of ERM. A recent crystal structure of *Sf*moesin has structurally confirmed that the central α-helical domain does form an anti-parallel coiled-coil between the N- and C-terminal domains (49). Due to sequence similarities, the other ERM proteins likely also form an anti-parallel coiled-coil as well. This anti-parallel coiled-coil buries a large surface area of the N-terminus of moesin and, thus, likely plays a role in the intramolecular regulation of ERM proteins. Despite the conformational activation model proposed for ERM proteins suggests a binary switch from open to closed toggled by phosphorylation of the C-terminus, a recent study has shown that merlin in fact behaves differently than it's ERM relatives. In Chapter 4 we explore merlin's degree of intramolecular regulation and the role of the α-domain in this regulation by designing truncations of the central domain.

Contents of The Dissertation

This dissertation contains four remaining chapters. Chapter 2 has been published and Chapter 3 contains my contribution to the work corresponding to the recent collaboration between Taifo Mahmud and P. Andrew Karplus. Chapter 4 is a work that was carried out in Dr. P. Andrew Karplus' laboratory and will be submitted for publication. In Chapter 5 I conclude this dissertation.

Chapter 2: "Dysferlin binds SNAREs (Soluble N-Ethylmaleimide-sensitive Factor (NSF) Attachment Protein Receptors) and stimulates membrane fusion in a calcium sensitive manner" Sara J. Codding, Naomi Marty, Nazish Abdulla, and Colin P. Johnson. Published in *Journal of Biological Chemistry*, 291, 14575-14584, 2016.

This chapter was the logical follow up to the cannon of *in vivo* literature reported on dysferlin and the *in vitro* work performed in the Johnson lab that established calcium and lipid binding data for the multi-domain protein. In this chapter, we follow up this *in vitro* work to establish the precedence with biochemical data that dysferlin can act as a calcium sensor in SNARE dependent membrane fusion. I had initially set out to collect thermodynamic data quantifying a binding event between dysferlin and SNAREs using isothermal titration calorimetry. However, I turned my sights to the comparatively protein sipping technique of fluorescence. Nevertheless, the biochemical data in this chapter provides in cell and *in vitro* evidence on two important areas of dysferlin research: 1. dysferlin binds t-SNAREs and accelerates the rate limiting step in their assembly in a calcium enhanced manner and 2. that dysferlin mediates SNARE dependent lipid mixing in a calcium sensitive manner. This work was the first in the field to determine a mechanism of action for dysferlin and suggests a molecular basis for why mutations that abrogate dysferlin's activity results in muscular dystrophy. I finish this chapter with areas of future research, particularly those involving *in vitro* follow up studies on dysferlin and SNARE interactions.

Chapter 3: In Pursuit of a Crystal Structure to Gain Understanding of the Sedoheptulose-7-phosphate Cyclase Family. Sara J. Codding

This chapter was the result of the collaboration between the Mamud lab and the Karplus lab here at Oregon State University. In this chapter, I analyze the sequences of a family of proteins known as the sedoheptulose-7-phophate cyclases that is composed of three distinct types of enzymes. Interestingly, each enzyme in this family utilizes the

same substrate to generate a different product. Due to this fact, and that the enzymatic products are precursors to many known natural products, the pursuit of structures of these proteins are interesting on two levels: 1. To understand a class of proteins from an evolutionary perspective and 2. To gain understanding of enzymes that could be used for semisynthetic compound production. We sought to crystallize representative enzymes of each of the three types of known sedoheptulose-7-phophate cyclases. Promising initial crystal leads of each enzyme were identified, and we were successful at optimization of one, ValA, that led to a complete X-ray diffraction pattern and identification of the catalytic metal. The structure was solved utilizing new computational and modeling approaches. The structural solution solidified the initial observations laid out in this chapter from analysis of primary sequence alignment. The two key results from this study are 1. the sedoheptulose 7-phosphate cyclases bind different anomers of the same sugar substrate to generate stereochemically distinct products and 2. that structural studies are often times useful when used in conjunction with biochemical studies.

Chapter 4: "Role of the α-domain in the Conformational Regulation of Merlin" Sara J. Codding, Russell Carpenter, Donnie Berkholz, Nicolas O. Thomas, Anthony Bretscher, P. Andrew Karplus.

This chapter was the follow up of the work on merlin performed both *in vitro* and in cell by the Bretscher lab that determined merlin was inactive in the phosphorylated closed state. In addition, these studies showed that merlin's intramolecular interactions were graded in a rheostat like manner rather than a binary switch. In this chapter, we follow up this work to determine the role of merlin's central α-domain in merlin's intramolecular interactions. During this time, I had initially set out to crystallize truncation mutants of merlin where in the central α-domain was systematically shortened to generate more globular constructs of the protein. However, high quality crystals remained elusive. Nevertheless, we collected biochemical data on the constructs developed and this chapter reiterates that merlin is not a binary switch and wild type

protein is partially open. In addition, the length and composition of the α-domain contributes to the degree of which merlin is open. This work was the first in the field to directly study the α-domain of merlin and showed that this domain is more than merely structural. I finish this chapter with areas of future research, particularly those involving in vitro follow up studies on merlin.

Chapter 5: "Conclusions Drawn From Biochemical and Biophysical Analyses of Proteins" by Sara J. Codding.

In this chapter I lay out the conclusions from this collective dissertation work.

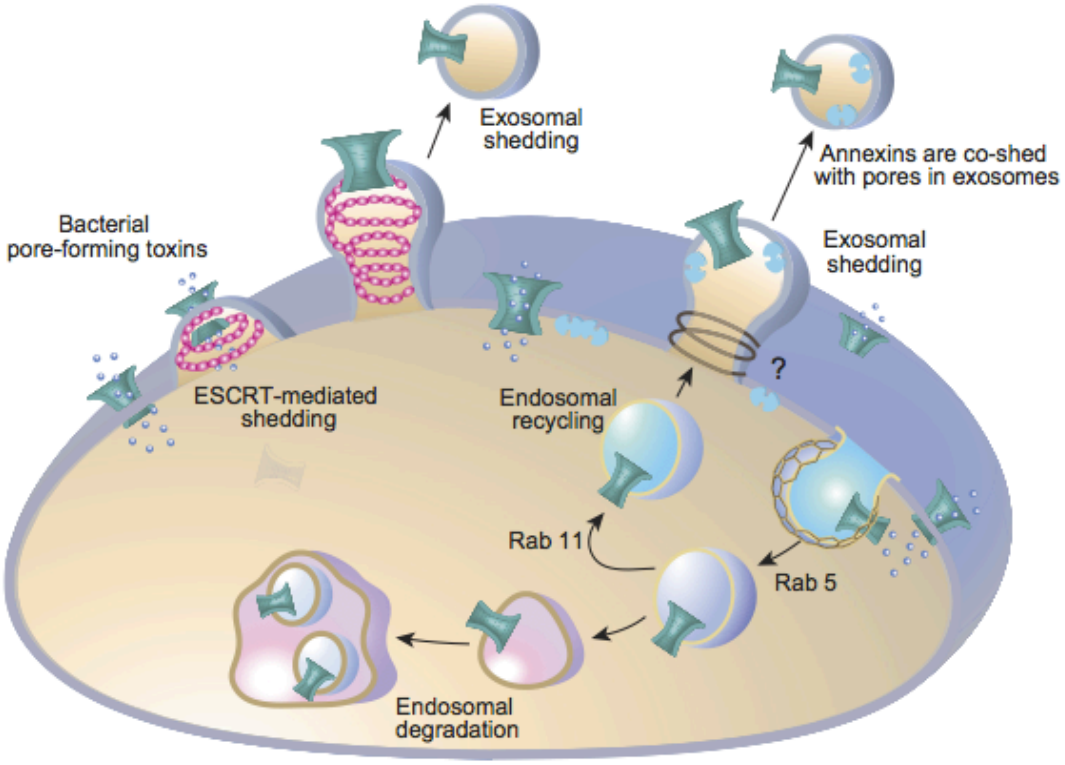


Figure 1.1. Shedding pore forming bacterial toxins via exosomal and endosomal pathways

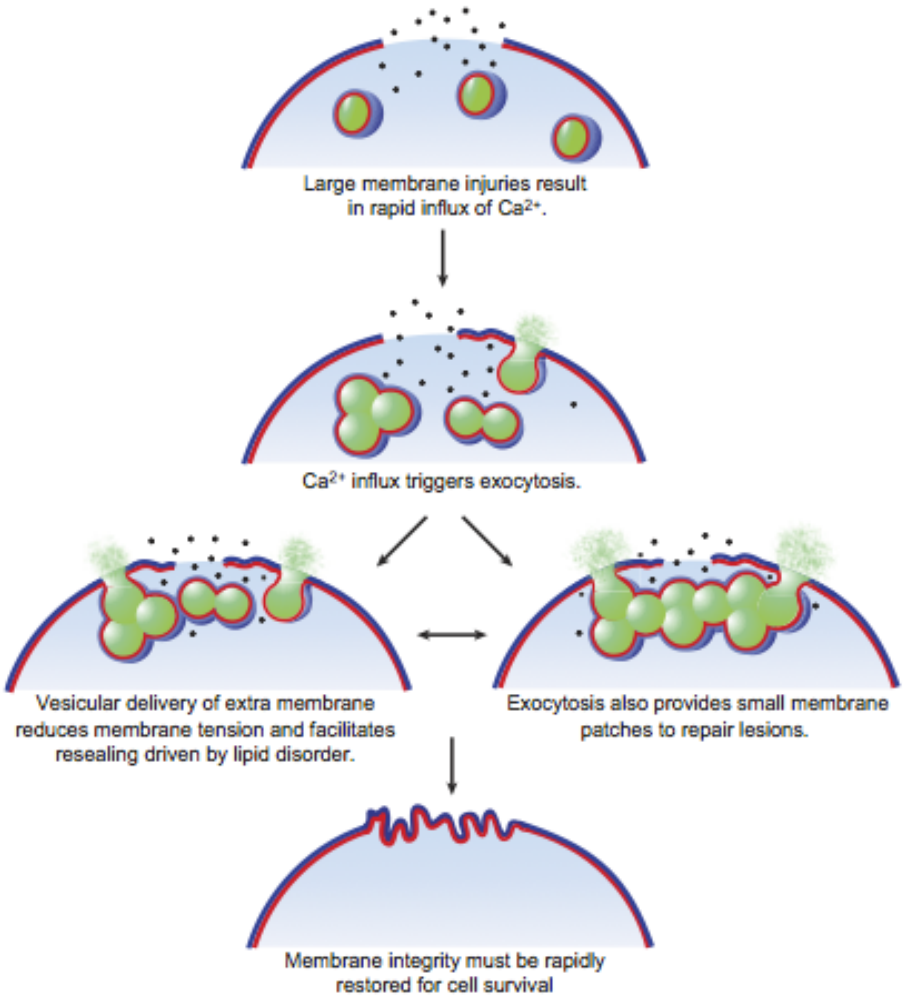


Figure 1.2. Membrane patch hypothesis. Membrane repair occurs via a patch if tears are very large (one micron in diameter). The calcium influx due to injury triggers homo- and heterotypic exocytosis to provide a patch at the membrane disruption site.

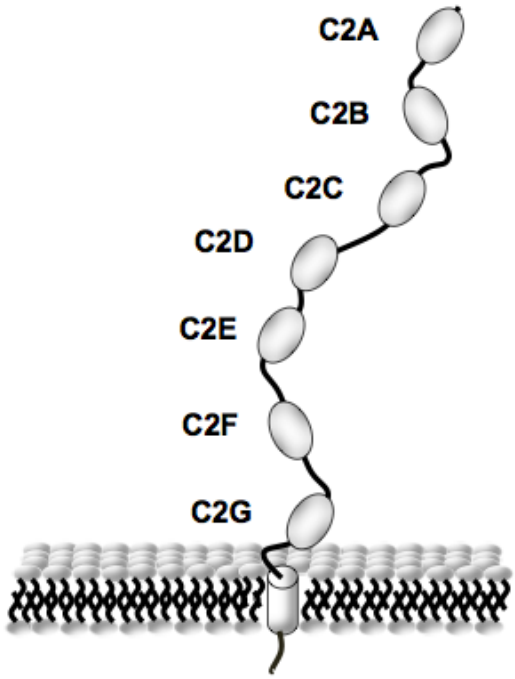
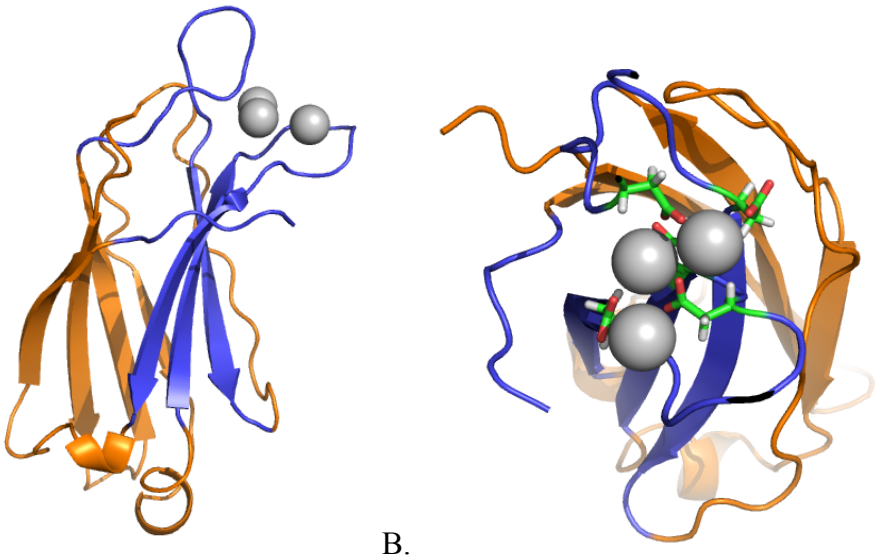


Figure 1.3. Schematic representation of the protein dysferlin. Dysferlin's seven N-terminal cytosolic C2 domain are depicted by ovals followed by a single pass C-terminal transmembrane domain (cylinder) imbedded in a lipid bilayer.



A. B.

Figure 1.4. Depiction of a canonical C2 domain fold. A. C2 domain of synaptotagmin1A C2A pymol 1BYN shown with as a fold with two β-sheets connected by loops that chelate calcium B. Close view of calcium chelation environment with aspartic acid residues shown as sticks, with atoms in the following colors: oxygen (red), nitrogen (blue), hydrogen (grey), carbon (green).

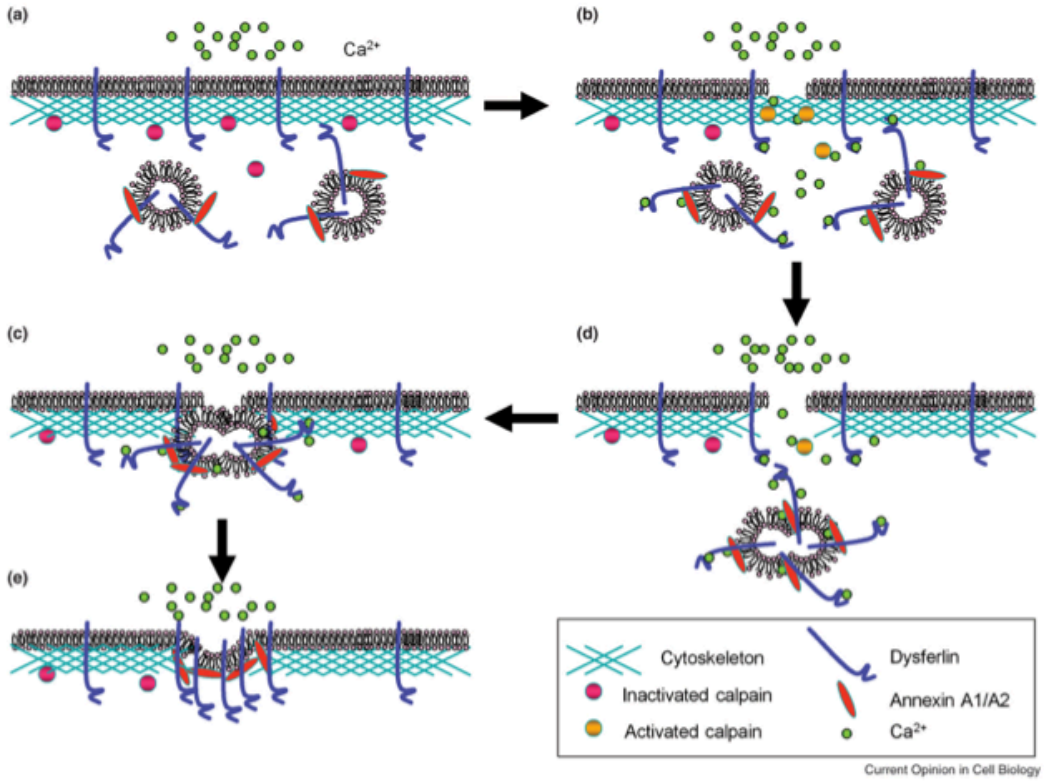
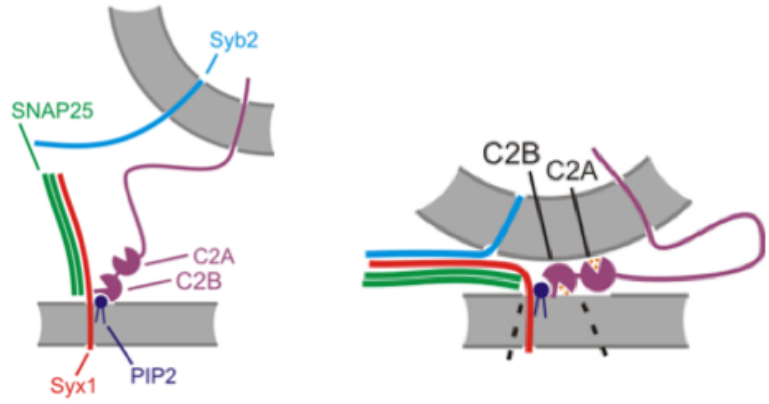


Figure 1.5. Skeletal membrane repair hypothesis. A. Dysferlin is found on the plasma membrane and intracellular vesicles. B. Dysferlin binds calcium that enters the cell from a membrane tear. C. Dysferlin facilitates homotypic fusion and D. heterotypic fusion to form the final patch in E.



A. B.

Figure 1.6 Synaptotagmin binds lipids and interacts with SNAREs to promote membrane fusion. A. Two opposing membranes, intracellular vesicle (curved-grey) are brought into close proximity by synaptotagmin C2 domain interacting with the plasma membrane (grey-flat). The SNAREs, VAMP2 (Syb2) on the vesicular membrane, and SNAP25 and syntaxin1 are brought into close proximity. B. Calcium influx triggers synaptotagmin to bind lipids through both C2 domains and the SNARE proteins are assembled into a fusion competent ternary complex.

Chapter 2

Dysferlin Binds SNAREs and Stimulates Membrane Fusion in a Calcium Sensitive Manner

Sara J. Codding, Naomi Marty, Nazish Abdullah, and
Colin P. Johnson

Published in the Journal of Biological Chemistry (2016), 291, 14575-14584

Abstract

Resealing of tears in the sarcolemma of myofibers is a necessary step in the repair of muscle tissue. Recent work suggests a critical role for dysferlin in the membrane repair process, and that mutations in dysferlin are responsible for limb girdle muscular dystrophy 2B and Miyoshi myopathy. Beyond membrane repair, dysferlin has been linked to SNARE mediated exocytotic events including cytokine release and acid sphingomyelinase secretion. However, it is unclear whether dysferlin regulates SNARE mediated membrane fusion. In this study we demonstrate a direct interaction between dysferlin and the SNARE proteins syntaxin 4 and SNAP-23. In addition, analysis of FRET and in vitro reconstituted lipid mixing assays indicate that dysferlin accelerates syntaxin 4/SNAP-23 heterodimer formation and SNARE mediated lipid mixing in a calcium sensitive manner. These results support a function for dysferlin as a calcium sensing SNARE effector for membrane fusion events.

Introduction

Limb-girdle muscular dystrophy and Miyoshi myopathy are muscle wasting diseases linked to mutations in the protein dysferlin (19,50-54). Dysferlin is a 238 kDa membrane protein composed of seven N-terminal C2 domains and a single pass C-terminal transmembrane domain (55,56). Early studies established a role for dysferlin in calcium triggered sarcolemma repair with dysferlin knockout cells displaying dysfunctional resealing of plasma membrane lesions (17,57).

More recent studies have determined that dysferlin contributes to cytokine secretion, lysosome exocytosis, acid sphingomyelinase secretion, and phagocytosis (27,58-60). These reports suggest a wider role for dysferlin in calcium sensitive membrane trafficking events at the cell membrane. However, the exact functions and mechanisms by which dysferlin operates remain unclear.

One proposed function for dysferlin is as a calcium sensitive scaffold for the recruitment of other proteins involved in membrane trafficking. Support for this comes from studies that have reported that the C2 domains of dysferlin bind annexins and caveolin (61-63). Dysferlin may also act as a regulator of calcium influx through interaction with T-tubule dihydropyridine receptors (64-66). However several studies have reported that deficiencies in dysferlin result in attenuated exocytosis and the accumulation of unfused vesicles at membrane lesions, suggesting a function in membrane fusion events that occur during repair of cell membrane wounds (17,59). In agreement with this, dysferlin has been reported to bind calcium, the cell membrane lipids phosphatidylserine and phosphatidylinositol bisphosphate, and to co-localize with the SNARE protein syntaxin 4 (30,67-69). Knockdown of dysferlin also results in a reduction in lysosome exocytosis and delayed release of acid sphingomyelinase, in agreement with dysferlin's proposed role in membrane fusion (27,58).

Direct evidence establishing the function of dysferlin in membrane trafficking is lacking, and despite the requirement of SNAREs for membrane repair and exocytosis, no study has directly tested whether dysferlin functions as a SNARE effector for membrane fusion (16). This has been due in part to the complexity of membrane trafficking

processes, which can lead to ambiguity in the interpretation of results. In this study we use cell-based techniques and well defined reconstituted systems to characterize the relationship between dysferlin and SNARE proteins and establish a function for dysferlin in membrane fusion events.

Materials and Methods

Molecular biology and generation of recombinant protein constructs. cDNA encoding human full length dysferlin, accession number AF075575, was used as a template for cloning and was a gift from Dr. Kate Bushby (Newcastle University). The single and multi-C2 domain (C2ABC, C2DEFG) dysferlin constructs were described previously (30,68). A soluble dysferlin construct encoding the entire cytosolic region (residues 1-1965) in pMCSG9 was generated via ligation independent cloning with forward primer TACTTCCAATCCAATGCAATGCTGAGGGTCTTCATCCTC and reverse primer TTATCCACTTCCAATGCTATTCTGCTACACAGGGCCAC. Rat SNARE constructs were a gift from Edwin Chapman (University of Wisconsin, Madison) and were expressed in either pGEX2T, pGEX6p3, or pET28a. Rat full length syntaxin 4 in pGEX2T encoding residues 1-298 was subjected to site directed mutagenesis to change a cysteine in the transmembrane domain at residue 279 to an alanine via forward primer GAAAAAGGTCATTGCCATCGCGGTTTCTGTCACTGTTCTCATC and reverse primer GATGAGAACAGTGACAGAAACCGCGATGGCAATGACCTTTTTC and is denoted as syntaxin 4 C279A. To generate a soluble syntaxin 4 with a single cysteine, the syntaxin 4 cytosolic region composed of residues 1-274 was amplified out of the pGEX2T vector with forward primer CGGGGCGGATCCCGCGACAGGACCCATGAGTTG and reverse primer CCGGCGGAATTCTCAGACCTTTTTCTTCCTCGCCTTCTTCTG. The amplicon was restriction cloned into pET28a with BamHI and EcoRI. To generate a syntaxin 4 H3 construct with a single cysteine residue position number 275, the syntaxin 4 H3 domain encoding residues 163-274 was amplified out of the pGEX2T plasmid with forward primer CGGGGCGGATCCACCAATGCTGGAATGGTGTCTGACG and reverse

primer CCGGCGGAATTCTCAGCA GACCTTTTTCTTCCTCGCCTTCTTCTGATTC. The amplicon was restriction cloned into pET28a with BamHI and EcoRI.
To generate a SNAP-23 construct with one remaining endogenous cysteine at residue 87, denoted SN23 C79,80,83,85, rat SNAP-23 was subjected to two sequential rounds of site directed mutagenesis, first changing cysteine residues 79 and 80 to alanine utilizing forward primer AACAGAACTCAACAAGGCTGCTGGCCTCTGCGTCTGCCCTTG and reverse primer CAAGGGCAGACGCAGAGGCCAGCAGCCTTGTTGAGTTCTGTT, followed by a second round changing cysteine residues 83 and 85 to alanine utilizing forward primer CAAGGCTGCTGGCCTCGCCGTCGCCCCTTGTAATAGGACC and reverse primer GGTCCTATTACAAGGGGCGACGGCGAGGCCAGCAGCCTTG. The wild type SNAP-23 and cysteine mutant SNAP-23 C79,80,83,85A were amplified out of pGEX2T via forward primer CGGGGCGGATCCATGGATGATCTATCACCAGAAGAAATT CAGCTTC and reverse primer CCGGCGGAATTCTCAGCTGTCAATGAGTTTCTTTGCTCTTGTATTGG and restriction cloned into pGEX6p3 with BamHI and EcoRI (NEB). All site directed mutagenesis was carried out utilizing the QuikChange II Mutagenesis kit from Agilent. All constructs, generated or received, were sequenced verified by Genscript or The Center for Genome Research and Biocomputing at Oregon State University.

Recombinant protein purification. The dysferlin fusion protein constructs were purified using methods described previously (30,68). In brief, the pMCSG9 vector, which encodes a polyhistidine-maltose binding domain tag with a TEV protease cleavage site was used to generate dysferlin fusion proteins in E. coli. The dysferlin single domain proteins were expressed in BL21 cells that were cultured in Luria-Bertani broth in the presence of 100 mg/mL ampicillin at 37 ° C and induced for 3 hours with 0.3 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) after optical density (O.D.) reached A_{600}=0.6. The dysferlin C2ABC, C2DEFG, and 7 C2 domain protein constructs were expressed in Rosetta™ (DE3) pLysS cells (EMD Millipore) that were cultured in terrific broth with 100 mg/mL ampicillin and 34 mg/mL chloramphenicol. Protein expression

was induced at A_{600}=0.9 with 0.3mM IPTG for 14 hours at 18 ° C. All cell pellets were isolated by centrifugation at 4000 RPM at 4 ° C for 20 minutes and then re-suspended in lysis buffer, 50mM Tris base pH 7.5, 150mM sodium chloride, 20mM imidazole, in the presence of 1 mM phenylmethanesulfonylfluoride (PMSF) and 1 mM of each of the following protease inhibitors; leupeptin, pepstatin, aprotinin, (RPI). The cells were lysed by sonication. The soluble fraction was isolated by centrifugation in a Beckman J2-21 centrifuge at 20,000G at 4 ° C for 20 minutes and then incubated with Ni-NTA (ThermoFisher) beads in lysis buffer for one hour with rocking at 4 ° C. The beads were washed with twenty column volumes of lysis buffer and the protein was then eluted with lysis buffer plus 500 mM imidazole. The pure fractions were checked via SDS- PAGE, pooled, and extensively dialyzed into 50 mM Tris base pH 7.5, 100 mM sodium chloride using dialysis tubing 4,000-6,000 MWCO (Spectrum Labs).

The SNARE protein constructs were expressed and purified as described previously (70). The pGEX2T vector, which encodes an N terminal glutathione s-transferase protein domain tag with a thrombin protease cleavage site was used to express the full length wild type syntaxin 4, and syntaxin 4 C279A. pGEX6p3 SNAP-23 constructs were expressed with a N terminal glutathione s-transferase with a precision protease cleavage site. The GST fusion SNARE constructs were expressed in BL21 cells cultured in Luria-Bertani broth in the presence of 100 mg/mL ampicillin at 37 ° C and induced for 3 hours after A_{600}=0.6 with 1 mM IPTG. All cell pellets were isolated as noted above. The pellets were resuspended in 50 mM Hepes pH 7.5, 150 mM sodium chloride. 10% glycerol was also added to the full length syntaxin 4 suspension buffer. Protease inhibitors were subsequently added to the cell suspension buffer and the cells were lysed by sonication. Triton X100 to 1% V/V was added to the full length syntaxin 4 lysates and all lysates were rocked for 30 minutes at 4 ° C to promote protein solublization. The soluble fraction of the lysates was isolated by centrifugation and then incubated with glutathione sepharose high performance beads (GE healthcare) for one hour with rocking at 4 ° C. The beads were washed with 20 column volumes of lysis buffer (plus 0.1% Triton X100 for full length syntaxin 4) and the protein was

eluted in lysis buffer with 200 mM reduced glutathione (GoldBio) and dialyzed extensively into 50 mM Hepes pH 7.5, 100 mM sodium chloride, and 0.1% W/V CHAPS (RPI).

The syntaxin 4 DTM and H3 domain constructs were expressed as a pET28a vector construct in Luria-Bertani broth in the same fashion as the GST fusion SNAREs. Cell pellets were isolated and lysed, as described above. The protein was purified similarly to full length syntaxin 4, with a lysis buffer of 50 mM Hepes pH 7.5, 20mM imidazole, 150 mM sodium chloride and a protein solublization step after the addition of Triton X100 to 0.1% V/V concentration. The soluble fraction was isolated as above by centrifugation and then incubated with pre-equilibrated Ni-NTA beads for one hour with rocking in a cold room. The beads were washed with 20 column volumes of lysis buffer plus 0.1% Triton X100 and the protein was eluted in lysis buffer with 500 mM imidazole. The protein fractions were checked for purity via SDS-PAGE and dialyzed extensively into 50 mM Hepes pH 7.5, 100 mM sodium chloride, and 0.1% W/V CHAPS (RPI).

The VAMP 2 pET28a construct was expressed, cell pellets collected, and the soluble fraction isolated, as described for syntaxin 4 and SNAP-23. The lysis buffer consisted of 50 mM Hepes pH 7.5, 150 mM sodium chloride, 20 mM imidazole, 10% glycerol, 2 mM beta-mercaptoethanol. The soluble lysates were incubated for one hour with Ni-NTA beads. The beads were then washed with 10 column volumes of lysis buffer followed by 10 column volumes of wash buffer 1 (50 mM Tris base pH 7.5, 300 mM sodium chloride, 20 mM imidazole, 2 mM beta-mercaptoethanol) and 10 column volumes of wash buffer 2 (50mM Tris base pH 7.5, 300mM sodium chloride, 40mM imidazole). The protein was eluted in 50 mM Tris base pH 7.5, 150 mM sodium chloride, 500 mM imidazole, and 1.5% CHAPS. All SNAREs were checked by SDS-PAGE, pooled as above, and the proteins were extensively dialyzed into 50 mM Tris base pH 7.5, 100 mM sodium chloride or 50mM Hepes pH 7.5, 100mM sodium chloride (plus 0.1% CHAPS for syntaxin 4 DTM, syntaxin H3 domain, and full length VAMP 2) if the protein was to be utilized in a fluorescence labeled assay.

Enzymatic removal of fusion protein tag glutathione s-transferase was achieved by the addition of 10U of thrombin (Sigma) per milligram of protein and incubation at 4 ° C overnight (syntaxin 4) or the addition of 10U of precision protease per milligram of protein and incubation at 4 ° C (SNAP-23) overnight, with rocking. Thrombin was removed with p-Aminobenzamidine-Agarose (Sigma) beads. Cleaved protein was then added to glutathione sepharose beads. The beads were rocked for 1 hour in at 4 ° C and the flow through was collected.

Proteins were concentrated if necessary with a VivaSpin Turbo 10K MWCO spin column and subsequently used for studies. The protein concentrations were measured using a nanodrop ND-1000 and calculated using extinction coefficient based on protein sequence.

Fluorescence anisotropy. Syntaxin 4 DTM, SNAP-23 C79,80,83,85A, and syntaxin 4 H3 proteins were labeled with fluorescein-5-maleimide (FITC, ThermoFisher). FITC was dissolved into sterile DMSO (Sigma) at a concentration of 10 mg/mL and 50 mL was slowly added with gentle agitation to protein with a concentration of ~4 mg/mL in 50 mM Hepes pH 7.5, 100 mM sodium chloride, pH 7.5 (plus 0.1% CHAPS for the syntaxin 4 DTM and H3 domain). The dye was allowed to react with SNAP-23 and syntaxin 4 at 23 ° C for 1 hour isolated from light, while syntaxin 4 DTM and syntaxin 4 H3 domain required 14 or 3 hours of incubation at room temperature respectively, with dye for efficient labeling. The labeled protein was then extensively dialyzed in 4,000-6,000 MWCO tubing into buffer composed of 50 mM Tris base pH 7.5, 100 mM sodium chloride (TBS), or Hepes pH 7.5, 100mM sodium chloride, (with 0.1% CHAPS, 1mM TCEP, for syntaxin 4 DTM and syntaxin 4 H3 domain) with two dialysis buffer exchanges to remove excess dye. The degree of labeling was determined by the manufacturer's protocol, and all absorbances were measured using a nanodrop ND-1000. The degree of labeling (D.O.L.) was calculated as D.O.L = $(A_{max} \times MW)/ ([protein] \times 68{,}000)$. The D.O.L. was determined to be approximately 0.39 labels per syntaxin 4 DTM**,** 0.9 labels per full length syntaxin 4 C279A, 0.6 for SNAP-23 C78,80,83,85A, and 0.25 labels per syntaxin 4 H3 domain. Dysferlin constructs were added to the labeled

SNARE at room temperature in phosphate buffered saline (20mM phosphate pH 7.5, 100mM sodium chloride) with a total reaction volume of 60 uL. Measurements were collected after a 1 minute equilibration time at room temperature with a PTI fluorimeter in a 200 µL quartz cuvette with excitation of l=494 nm, emission of l=518 nm. Anisotropy was calculated using the following equation r = (IVV– IVH) / (IVV + 2IVH) where IVV and IVH correspond to the parallel and perpendicular fluorescence emission intensities. The fraction of protein bound, P, was calculated as $P=(r-r_f)/(r_b-r_f)$ where r is the observed anisotropy, r_f is the anisotropy in the absence of dysferlin, and r_b is the anisotropy of 100% bound SNARE. (rb was set to the maximum anisotropy value at the plateau phase of the titration, where the further increase in dysferlin concentration resulted in little to no increase in the anisotropy value under the concentrations tested). The reported values represent the mean ± propagated error for three or more samples.

SNARE heterodimer assembly assay. The FRET assembly experiments were adapted from a previously described method (29). GST cleaved full-length syntaxin 4 C278A and GST cleaved SNAP-23 C78,80,83,85A were labeled via cysteine reactive FITC or cysteine-reactive Alexa- fluor 546 (Invitrogen) respectively, for 1 hour at 23 ° C in Hepes 50mM pH 7.5, 100mM sodium chloride, as described above. Excess dye was removed by extensive dialysis in TBS. Labeling efficiency was calculated as described above (FITC) and was determined to be approximately 0.8 labels per syntaxin 4, and 0.9 labels per SNAP-23. The labeled syntaxin 4 protein was subsequently reconstituted into liposomes composed of 25% POPS, 75% POPC at approximately 100 copies per liposome as described previously (31,71,72). Briefly, lipids (Avanti Polar Lipids) were evaporated overnight and re-suspended with labeled full length syntaxin 4 C278A (1.2 mg) and brought to 1.5 mL total volume with reconstitution buffer. Two volumes of reconstitution buffer (Hepes 25 mM pH 7.8, 100mM KCl) were added dropwise with light agitation, and the proteoliposome sample was dialyzed overnight in reconstitution buffer with 1 mM DTT. The reconstituted liposomes were subsequently mixed in an eqi-volume of 80% Accudenz solution in a thin walled ultracentrifuge tube, and layered with 2.25 mL 30% Accudenz, followed by 0.750 mL of reconstitution buffer. The

liposomes were then floated on the Accudenz gradient via centrifugation at 41,000 rpm in a SW-41 (Beckman) rotor with adaptors at 4° C. Proteoliposomes were isolated from the 0/30% Accudenz interface.

FRET samples were 100 μL in total volume, and composed of 5 μL syntaxin 4 proteoliposomes, 5 μM SNAP-23, and 1 μM of dysferlin. The assay was run in the absence or presence of calcium **(**500 mM**)** with an excitation wavelength of 494 nm and a FRET emission spectra collected over 500-600nm wavelength. FRET was calculated as the ratio of donor to acceptor emission intensities. The criterion for FRET was a concomitant reduction in donor emission intensity and increase in acceptor emission intensity.

Immunoprecipitation. Cultured C2C12 cells obtained from ATCC were grown to 70% confluence in a T75 flask in high glucose plus pyruvate DMEM (Invitrogen), 10% fetal bovine serum (Invitrogen), and 20 units/mL penicillin-streptomycin (Invitrogen), at 37° C with 5% CO_2. The cells were then changed into a serum deficient media to promote differentiation per ATCC protocol, (high glucose plus pyruvate DMEM, 2% horse serum (Invitrogen), 20 units/mL penicillin-streptomycin) and cultured for 7 days, with serum changes every two days. The media was removed and the cells were washed twice with sterile PBS (50 mM phosphate pH 7.5, 100 mM sodium chloride) and 1 mL of lysis buffer (sterile 50 mM phosphate pH 7.5, 100 mM sodium chloride, and 0.2% Triton X-100 and Roche mini complete protease inhibitors) was added to the flask. Cells were incubated for 30 minutes with rocking at 4° C. The total lysate was removed from the flask and split into two samples, to one sample 3 mM calcium chloride (Sigma) was added, to the other 3 mM EDTA (Sigma) was added. The samples were then centrifuged at 14,000 G for 5 minutes at 4 ° C to remove cell debris. The soluble lysate fraction was subsequently isolated and 50 mL of the Hamlet mouse anti-dysferlin (Abcam) antibody was added to the sample. The samples were rocked at room temperature for 1 hour, followed by the addition of 50 mL fresh PBS washed protein A/G beads. After 1 hour incubation at room temperature with gentle agitation, a magnetic rack was used to pellet the beads, and the supernatant was

removed. The beads were then washed with 3 sample volumes PBS. To dissociate bound proteins from the beads, one sample volume of 2X Lamelli sample buffer was added, the sample was then boiled for 2 minutes and the sample was loaded onto a 12% SDS-PAGE gel. The proteins were then transferred using a wet tank transfer blot box (Invitrogen) in cold Tobin buffer (2.5 mM Tris pH 8.3, 19.2 mM glycine) with 20% methanol onto a PVDF membrane (Immboilon-P Membrane, PVDF, 0.45 mm, Invitrogen) for 90 minutes at 200mA (73). The membrane was blocked with 2 % non-fat milk in PBS plus 0.05 % Tween20 for 1 hour at room temperature and rabbit primary antibodies against SNAP 23 (Sigma Aldrich) or syntaxin 4 at 1:1000 dilution (Sigma Aldrich) were added. After incubating overnight at 4 ° C with rocking, membranes were washed in triplicate for 10 minutes in PBS then incubated with goat anti-rabbit HRP secondary 1:5000 dilution (Sigma Aldrich) in PBS at room temperature for 1 hour. The syntaxin 4 co-IP protocol was performed similarly to as described above, with anti-syntaxin 4 antibody (Abcam). Samples were run on a 8 % gel, blotted, and probed with Hamlet mouse anti-dysferlin (1:500, Abcam) and anti-mouse HRP secondary (1:5000, Sigma Aldrich). Blots were developed with SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) per manufacturer's protocol and imaged with a Kodak Digital Science Image Station 440CF. Quantification of co-immunoprecipitation was conducted by comparison of western blot band intensities using Image J, with the total lysate sample set to 1 and the co-immunoprecipitation band as a fraction of the total lysate. Efficency of the immunoprecipitation for dysferlin is shown in (Figure 1). Statistical significance corresponding to $P>0.1$ are indicated by *. Negative control immunoprecipitations were conducted similarly using mouse IgG (Sigma) (Figure 2).

Duolink. C2C12 cells grown to 70% confluence in full serum medium as described above and were treated with 2 mL 0.25% trypsin (Life Technologies) for 30 minutes followed by the addition of 2 mL of complete media to neutralize trypsin, and 200 μL of cell suspension was subsequently plated onto square ethanol sterilized cover slips (Fisher) in a sterile six well plate (Sigma Aldrich). Cells were allowed to adhere for

1 hour with incubation, followed by the addition 2 mL of full serum medium. After the cells reached 70% confluence, media was changed to serum deficient medium and cells were differentiated for 7 days. Cells were washed twice in PBS, and then fixed with 4% paraformaldehyde in PBS for 10 minutes at room temperature. Fixed cells were washed twice with PBS, permeabilized with 0.2% TritonX-100 in PBS for 10 minutes, and then blocked with 3 drops duolink blocking buffer for 1 hour at 37 ° C. Each sample was then incubated with two primary antibodies in parallel, overnight, in a total volume of 250 µL duolink diluent, at 4 ° C (Hamlet mouse anti-dysferlin 1:250 (Abcam), rabbit anti-syntaxin 4 1:500 (Sigma Aldrich), rabbit anti-SNAP-23 1:500 (Sigma Aldrich), rabbit anti- histone H3 1:500 (Sigma Aldrich), rabbit anti-TFAM 1:500 (Bethyl Labs). Samples were subsequently washed twice with PBS and incubated with 8 µL duolink anti-mouse minus and 8 µL anti-rabbit plus proximity ligation probes for one hour at 37 ° C. Cells were removed from 37 ° C and washed twice with PBS, and two unique oligonucleotides complimentary to the individual PLA probes provided by Sigma were added in the presence of a ligation enzyme per manufacturers' protocol. The samples were incubated for 30 minutes at 37 ° C. The samples were subsequently washed with twice with PBS and subjected to rolling circular amplification for 100 minutes at 37 ° C in the presence of labeled complimentary oligonucleotide probes per duolink instructions. Samples were washed twice in PBS and incubated with Hoechst stain at a 1:5000 dilution for 30 minutes. After incubation the samples were washed and mounted on glass slides, (Fisher) with fluoromount (Life Technologies). Images were captured using a Zeiss Axiovert S100TV fluorescent microscope and Metamorph software version 6.3r7. At least six images per sample were analyzed and each sample was repeated in duplicate. The number of puncta-per-nuclei was recorded. A paired t-test was performed on the data and p-values are reported. Each antibody used in our study was tested for specificity using standard immunofluorescence. The highest dilution of the antibody that gave specific, detectable immunofluorescence, was used for Duolink measurements so as to minimize non-specific antibody interactions.

Reconstituted fusion assays. Reconstitution of proteoliposomes was conducted as previously reported and as noted above (31,71,72). VAMP 2 proteoliposomes were composed of 72% POPC, 25% POPS, 1.5% rhodamine-PE (acceptor), and 1.5% NBD-PE (donor). Syntaxin 4 proteoliposomes were composed of 75 mol% POPC, and 25 mol% POPS. Proteoliposomes contained ~100 copies of VAMP 2 or syntaxin 4 proteins per proteoliposome. All fusion assays were conducted using a PTI fluorimeter with a 200 μL quartz cuvette. Samples contained a total volume of 75 μL with 30 μL purified syntaxin 4 proteoliposomes, 15 μL purified VAMP 2 proteoliposomes, 0 or 15 μM SNAP-23, and varying concentrations of dysferlin protein in TBS. Samples were mixed in a sequential order, with syntaxin 4 proteoliposomes first incubated overnight with SNAP-23 and dysferlin at 4 ° C, followed by the addition of VAMP proteoliposomes just prior to measurements. For samples containing calcium, the free calcium concentration was 500 μM. Measurements were collected over 60 minutes and the dequenching of NBD was monitored. At the end of each experiment, 0.5% wt/vol n-dodecyl-maltoside was added to determine the maximal dequenched NBD signal. Raw fluorescence was normalized to obtain the percentage of maximum fluorescence as described previously (70,74).

Results

Dysferlin binds syntaxin 4 and SNAP-23. Dysferlin has been reported to co-localize with syntaxin 4, a ubiquitously expressed SNARE protein that contributes to lysosomal exocytosis (67). To determine whether dysferlin interacts with syntaxin 4, as well as the cognate SNARE SNAP-23, dysferlin immunoprecipitated samples from serum starved differentiated C2C12 cells were probed by western blot. As shown in Figure 3A, syntaxin 4 and SNAP-23 co-immunoprecipitated with dysferlin in a calcium independent manner. By contrast, actin did not co-immunoprecipitate.

Subsequent analysis of reciprocal co-immunoprecipitation samples generated with an anti-syntaxin 4 antibody also supported an association between syntaxin 4 and dysferlin in cell lysate (Figure 3B). To ascertain whether dysferlin and the SNAREs

syntaxin 4 and SNAP-23 reside in close proximity within cells, we utilized a proximity ligation assay (PLA) on serum starved differentiated C2C12 cells (75). Fixed cells incubated with anti-dysferlin and either anti-syntaxin 4 or anti- SNAP-23 resulted in 9.32 ±3.86, or 1.55 ±0.56 fluorescent puncta per cell nuclei respectively (Figure 3D, E). By contrast, fixed cells incubated with anti-dysferlin and anti-histone h3 or anti-TFAM, resulted in 0.39 ±0.24 and 0.39 ±0.21 puncta per nuclei. Based upon analysis of western blot and PLA data, we conclude that dysferlin associates with syntaxin 4 and SNAP-23.

We next tested whether the co- immunoprecipitation and PLA results were due to a direct interaction between dysferlin and SNAREs by conducting fluorescence anisotropy measurements on a fluorescein-5-maleimide (FITC) labeled syntaxin 4 single cysteine construct (syntaxin 4 DTM a.a. 1-274) in solution. When a construct comprised of the entire cytoplasmic region of dysferlin lacking only the transmembrane domain was added to samples containing labeled syntaxin 4, a concentration dependent increase in anisotropy was observed (Figure 4B). This observed increase in anisotropy was independent of the presence of calcium with an apparent Kd of 0.094 ± 0.020 mM.

Syntaxin is composed of a single-span C-terminal transmembrane helix, a single SNARE domain (H3 domain), and an N- terminal alpha helical regulatory domain (Habc domain) (76-78). Previous reports have established that the C2 domains of the dysferlin homologue otoferlin bind specifically to the H3 domain of syntaxin 1 (79). To determine if the C2 domains of dysferlin interact with the H3 SNARE domain of syntaxin 4, anistropy measurements of a FITC labeled truncated form of syntaxin 4 lacking the Habc and transmembrane regions (a.a. 163-274 with residue 275 changed to a cysteine) were collected. Addition of the cytoplasmic region of dysferlin increased the anisotropy value of the H3 domain of syntaxin 4 in a calcium insensitive dose-dependent manner, with an apparent Kd of 0.022 ±0.001 μM. Similarly, dysferlin constructs composed of the first three N-terminal (C2ABC) or four C-terminal (C2DEFG) C2 domains of dysferlin also increased anisotropy values when added to labeled syntaxin H3 domain, with apparent Kd values of 0.38 ±0.04 μM for C2ABC and 0.32 ±0.02 μM for C2DEFG (Figure 4C). To identify the specific C2

domains of dysferlin that mediate binding, titrations were conducted between the syntaxin 4 H3 domain and each individual C2 domain of dysferlin. As with the larger multi-domain constructs, addition of the individual C2 domains increased the anisotropy values of the FITC labeled H3 domain of syntaxin 4 in a calcium independent manner. In all cases, the increase in SNARE anisotropy upon addition of dysferlin constructs show cooperativity. Such cooperative interactions have been noted for other protein-protein interactions; however the basis for the observed cooperative interactions between C2 domains and SNAREs is not well understood. Apparent Kd values for the individual domains ranged from 2.5 to 10.4 μM, indicating that the isolated domains bind with lower affinity than the multi-domain constructs. The apparent Kd value for a two-domain construct (C2AB) was also determined (Table 1). Anisotropy values did not change significantly with the addition of either recombinant glutathione S-transferase (GST) or maltose binding protein (MBP) to FITC labeled syntaxin 4 samples, which served as negative controls. In addition, no changes in anisotropy were observed when titrations were conducted in the presence of 8M urea which disrupts protein- protein interactions. An increase in anisotropy was also observed when the dysferlin C2ABC or C2DEFG constructs were titrated against a FITC labeled SNAP-23 single cysteine mutant (SNAP-23 C79,80,82,85A), with apparent Kd values of 0.69 ± 0.05 and 0.39 ± 0.01 μM respectively (Figure 4D). A summary of apparent Kd values is listed in Table 1. We conclude that the C2 domains of dysferlin bind both syntaxin 4 and SNAP-23.

Dysferlin stimulates the assembly of SNARE heterodimers. Assembly of syntaxin 4 with the cognate SNARE SNAP-23 is thought to be a necessary prerequisite for the formation of membrane fusion competent SNARE complexes (31,74,76-78,80). This assembly proceeds through interaction between the SNARE motifs of syntaxin 4 and SNAP-23, and is thought to be inhibited by the Habc domain of syntaxin 4 (29,81,82). Having established that dysferlin binds SNAP-23 and the SNARE H3 domain of syntaxin 4, we sought to determine if dysferlin facilitated assembly of the syntaxin 4/SNAP-23 heterodimer. To monitor SNARE heterodimer assembly, we modified a method used previously to detect SNARE interaction by measuring FRET

values between FITC labeled full length syntaxin 4 C278A proteoliposomes and a soluble form of SNAP-23 C79,80,83,85A labeled with an Alexa-flour 546 C5 maleimide, as depicted in Figure 5A (29). Relative to samples lacking dysferlin, the addition of either the C2ABC or C2DEFG constructs enhanced the FRET value as determined by a decrease in donor fluorescence and concomitant increase in acceptor fluorescence (Figure 5B, C). The observed dysferlin dependent increase in FRET was significantly enhanced by the presence of 500 μM calcium (Figure 5C).

Dysferlin stimulates SNARE mediated lipid mixing. Having established that dysferlin facilitates SNARE heterodimer formation we sought to determine what effect dysferlin has on SNARE mediated membrane fusion using an in vitro reconstituted lipid mixing assay (31,71,72,74,80). In this assay, VAMP 2 liposomes containing a FRET pair are mixed with wild-type full length syntaxin 4 containing liposomes that do not harbor fluorophores. When a soluble form of SNAP-23 is added to the liposome mixture, the three SNAREs assemble into a ternary complex that drives fusion between the liposomes, resulting in lipid mixing, dilution of the FRET pair, and dequenching of the NBD fluorophore (Figure 6A). Previous studies using this assay have determined that syntaxin 4/SNAP-23 heterodimer assembly is both a prerequisite and rate limiting step in SNARE mediated lipid mixing between the two populations of proteoliposomes, regardless of whether the proteoliposomes are aggregated (31,80). To determine if dysferlin can accelerate SNARE assembly and lipid mixing, we monitored the dequenching of the NBD signal in proteoliposome samples in the presence or absence of the full length cytoplasmic region of dysferlin. In the absence of dysferlin, a low basal level of fusion was observed (Figure 6B). In the absence of calcium, the addition of dysferlin did not enhance lipid mixing. However, in the presence of calcium, dysferlin enhanced both the rate and final extent of fusion (Figure 6B). Consistent with the requirement of SNARE assembly, the observed stimulatory effect of dysferlin was SNARE dependent, as membrane fusion was not observed in the absence of either SNAP-23, syntaxin 4, or VAMP 2 regardless of the presence of dysferlin or calcium (Figure 6C). Dysferlin also stimulated lipid mixing in assays conducted with dithionite

treated VAMP 2 liposomes, suggesting that fusion extends beyond the hemifused state and that the inner leaflets of the SNARE liposomes mix (31). We next tested shortened forms of dysferlin for the ability to stimulate membrane fusion, and found that truncated constructs composed of either the first three (C2ABC) or last four (C2DEFG) C2 domains stimulated SNARE mediated membrane fusion in a calcium enhanced manner (Figure 6D). In summary, these results suggest that dysferlin promotes lipid mixing between SNARE proteoliposomes in a calcium sensitive manner.

To ensure that the lipid mixing events involved ternary SNARE complex formation, aliquots from the lipid mixing assays were analyzed by western blot (Figure 6E). SNARE proteoliposome samples containing calcium and the C2ABC or C2DEFG constructs that were allowed to react for 1 hour were found to contain high molecular weight SDS resistant syntaxin 4 bands, consistent with the product of SNARE complex formation (Figure 6E, left panel) (83,84). Western blot analysis of SNARE proteoliposomes lacking dysferlin constructs showed little-to-no high molecular weight complexes, consistent with the low rate and extent of NBD dequenching observed in the lipid mixing assays (Figure 6E, left panel). When dysferlin containing aliquots were boiled for 10 minutes, the high molecular syntaxin 4 weight bands were no longer apparent, consistent with previous studies which have demonstrated that SNARE complexes are SDS resistant, and require boiling to dissociate (Figure 6E, right panel).

Discussion

Dysferlin has been implicated in lysosome exocytosis, cell membrane repair, and vesicle-vesicle fusion (25,27,58,59). In this study we used in vitro approaches to test dysferlin's effects on membrane fusion. Analysis of co-immunoprecipitation samples indicated that dysferlin associates with syntaxin 4 and SNAP-23, two ubiquitously expressed SNAREs previously implicated in lysosome exocytosis. Using a FRET based assay, dysferlin was found to facilitate assembly of syntaxin 4/SNAP-23 heterodimers. Using a reconstituted membrane fusion assay we found that dysferlin promoted SNARE

mediated lipid mixing in a calcium sensitive manner. These results suggest dysferlin could act as a calcium sensor for membrane repair, cytokine secretion, and lysosome exocytosis (27,58-60,85). The results of our studies may also explain why dysferlin deficiencies result in attenuated rates of exocytosis and an increase in unfused vesicles (17). Our findings that regions in both the N- and C-terminus of the protein engage SNAREs and stimulate fusion may explain cell studies that have implicated both the C2A domain and the C-terminal C2 domains in dysferlin function (26,58,86,87).

Previous studies have determined that the dysferlin homologue otoferlin also stimulates SNARE mediated membrane fusion and binds plasma membrane lipids (31,68,88,89). Similarly, defects in Fer1, the C. elegans ferlin homologue, results in defective calcium dependent membranous organelle fusion with the sperm plasma membrane during spermatogenesis (90). That multiple ferlin proteins in both vertebrate and invertebrate species have been determined to promote calcium dependent membrane fusion suggests a common function among members of this family. However, members of the ferlin family appear to have distinct binding partners and calcium sensitivities, and ferlins are not believed to be completely functionally redundant (91). We speculate ferlin proteins share a conserved function as calcium sensing membrane fusion proteins that have diversified to regulate different exocytotic events.

Future Directions

One of the most well studied physiological process involving SNAREs is synaptic vesicle exocytosis for synaptic communication (28) and researchers have gone to great lengths to identify the molecular steps that underlie this this process. In so much as assigning names to the steps, some which have yet to be molecularly identified, have been argued to be priming, docking, then fusion (92). While the post fusion conformation state of SNAREs is well understood as the neuronal SNARE ternary complex has been crystallized, the conformation changes in SNAREs during steps leading to fusion is not well defined. In the steps prior to fusion, syntaxin is thought to reside in a closed state, with the Habc inhibitory domain folded over onto the H3 domain

(SNARE motif). This closed conformation limits the formation of a t-SNARE heterodimer, a rate-limiting step to fusion. Our results indicate that t-SNARE dimer formation is stimulated by dysferlin in a calcium enhanced manner, however it remains to be shown if dysferlin can drive t-SNARE complex formation by mediating the transition from the closed to open conformation of syntaxin.

In an EPR study published in 2011, the protein Munc13-1 was shown to facilitate the opening of syntaxin 1A (29). To monitor if dysferlin facilitates the opening of syntaxin 4 in a similar manner, a FRET assay that showed a change in intramolecular association of the syntaxin protein could be used. Standard fluorescent labeling chemistry to perform FRET is relatively facile but inherent limitations exist. For example, to demonstrate intramolecular interactions the coupling chemistry must be distinct to directionally incorporate two separate fluorophores on one protein. The caveats of commercially available fluorophores prevent incorporation at discrete locations with 100% certainty; both N and C terminal probes can react off target with amino acid side chains. To overcome this limitation the incorporation fluorescent non-canonical amino acids is one viable option. In such, this technique can incorporate a synthetic fluorogenic amino acid by utilizing an orthogonal synthetase and paired orthogonal tRNA at a site selective location of a growing polypeptide chain utilizing endogenous ribosomal machinery (93). This approach would add a fluorophore at a site-specific location side stepping the limitations of commercially available coupling chemistry. Recently, it has been shown that the non-canonical fluorescent amino acid acridon-2-ylalanine can incorporate *in vivo* and FRET with many commercially available maleimide fluorophores (94). In fact, acridon-2-ylalanine has been used to monitor dynamic conformational changes in the HP35 subunit of villin (95). Thus, the use of non-canonical amino acids is a suitable option to monitor intramolecular FRET with the main hurdle being choosing a permissive site for non-canonical amino acid incorporation.

Many studies have reported different functions and conflicting views on which dysferlin domains are crucial for rescuing the phenotype (86,87). In addition, the role of

dysferlin in healthy cells is not yet clear. *In vitro* studies are hampered by the complexity of the cell and the *in vitro* work outlined here takes advantage of answering one clear question in a reductive system. A question that remains to be answered, on dysferlin and ferlins in general, is the diving force behind the key feature that sets them apart from other C2 domain containing proteins: why are there so many C2 domains and why are they spaced so far apart? Unlike synaptotagmins that have a consistent seven amino acids between their C2 domains, the ferlins have larger distances between C2 domains and additional domains between C2 domains that are unique to the ferlins. Two such domains, the Fer and DysF domains, lie between the C2 domains, perhaps adding a yet untested aspect to the protein's function. These domains were identified by PFAM and remain uncharacterized (56).

Of the three Fer domains in dysferlin, the most N terminal, FerI lies between the C2B and C2C domains. The endocytic rate of dysferlin appears to be depend on the C2B-FerI-C2C region of the protein, and endocytosis of dysferlin proceeds through a clathrin independent pathway with syntaxin 4 and Rab5 (67,96). It is unclear if these domains work in coopertively or if FerI adds a quaternary structure to yield a C2B-FerI-C2C fold distinct from the domains separate folds. Further, two more Fer domains exist in dysferlin, FerA and FerB, with no existing biochemical data and the central DysF domain between C2C and C2D, has no assigned function despite being conserved across all ferlins. While this region has not been structurally defined as yet, mutations in DysF have been linked to dysferlinopathies suggesting an important role in the protein function (97).

It would be of interest to assess whether the DysF and Fer domains fold autonomously using circular dichroism and to pursue crystallization trials if possible. These domains can be tested for any additive function that they may lend to multi-domain constructs, in calcium, lipid, and SNARE biding assays by generating multi-domain constructs that lack these domains. Finally, atomic force microscopy or cryoEM could be used to determine the quaternary fold difference between C2ABC containing and lacking FerI.

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author Contributions

CPJ conceived and coordinated the study. CPJ and SJC wrote the paper. SJC and CPJ, performed and analyzed the experiments. NA and NM provided technical assistance. Authors reviewed the results and approved the final version of the manuscript.

Table 2.1. SNARE binding constants

SNARE	C2 domain	Kd (μM)
Syntaxin 4 H3	C2A	4.9 ± 0.1
Syntaxin 4 H3	C2B	2.9 ± 0.1
Syntaxin 4 H3	C2C	3.4 ± 0.1
Syntaxin 4 H3	C2D	4.9 ± 0.1
Syntaxin 4 H3	C2E	10.4 ± 0.2
Syntaxin 4 H3	C2F	5.5 ± 0.2
Syntaxin 4 H3	C2G	2.5 ± 0.1
Syntaxin 4 H3	C2AB	2.9 ± 0.1
Syntaxin 4 H3	C2ABC	0.38 ± 0.04
Syntaxin 4 H3	C2DEFG	0.32 ± 0.02
Syntaxin 4 H3	C2ABCDEFG	0.022 ± 0.001
Syntaxin 4 ΔTM	C2ABCDEFG	0.094± 0.002
SNAP-23	C2ABC	0.69 ± 0.05
SNAP-23	C2DEFG	0.39 ± 0.01

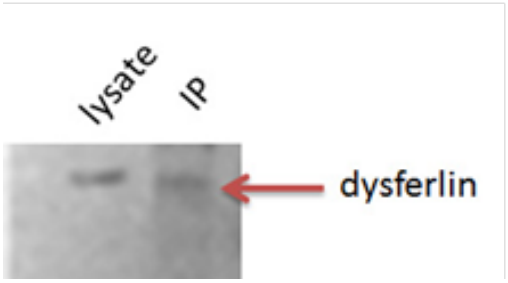


Figure 2.1. Representative western blot showing the efficiency of immunoprecipitation. Left lane shows the amount of dysferlin in C2C12 lysate before immunoprecipitation. Right lane shows the amount of dysferlin in the immunoprecipitation sample after pull-down. Dysferlin positive bands of the correct molecular weight were detected in both the cell lysate and dysferlin-immunoprecipitated samples, and quantitative analysis indicate approximately 43-58% of the dysferlin in the cell lysate is immunoprecipitated under the conditions used (N=2).

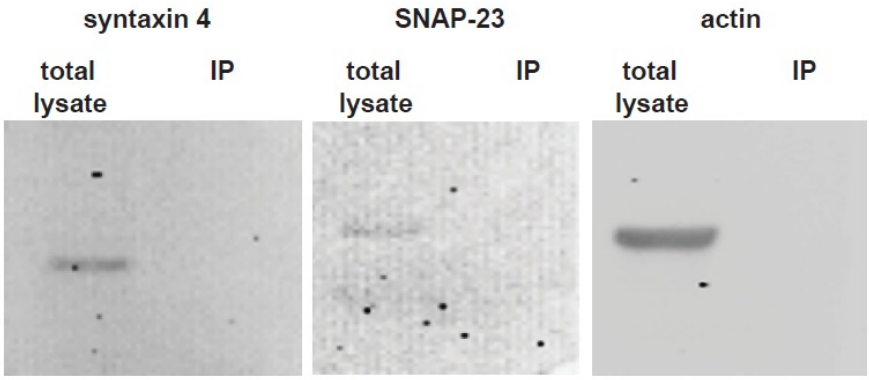


Figure 2.2 IgG control immunoprecipitation results. Control IgG corresponding to the host species of the primary dysferlin antibody (mouse) did not immunoprecipitate syntaxin 4, SNAP-23, or actin. Total lysate corresponds to C2C12 lysate before immunoprecipitation, IP corresponds to IgG immunoprecipitated sample.

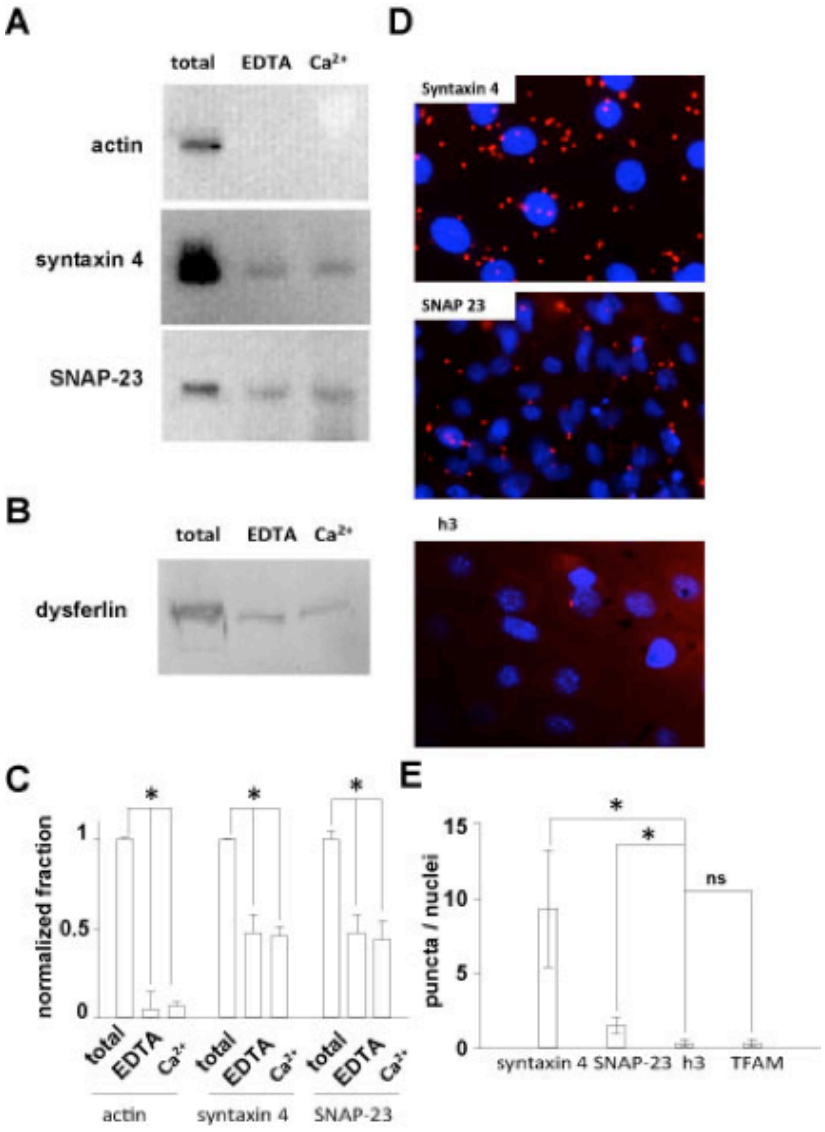


Figure 2.3. Dysferlin interacts with syntaxin 4 and SNAP-23. (A) Western blot of dysferlin immunoprecipitation samples from differentiated C2C12 cells probed for syntaxin 4, SNAP-23, or actin. Syntaxin 4 and SNAP-23, but not actin co-immunoprecipitate. Total refers to the total soluble cell lysate; EDTA and Ca^{2+} refer to immunoprecipitation samples in the presence of EDTA or calcium. (B) Western blot of syntaxin 4 immunoprecipitated samples from differentiated C2C12 cells probed for dysferlin. (C) Quantitation of western blot results. N=3.(D) Dysferlin co- localizes with syntaxin 4 and SNAP-23, but not histone h3 as determined by in situ proximity ligation. (E) Quantitation of proximity ligation results. Error bars represent std, N= 12, *P<0.05, ns (not significant).

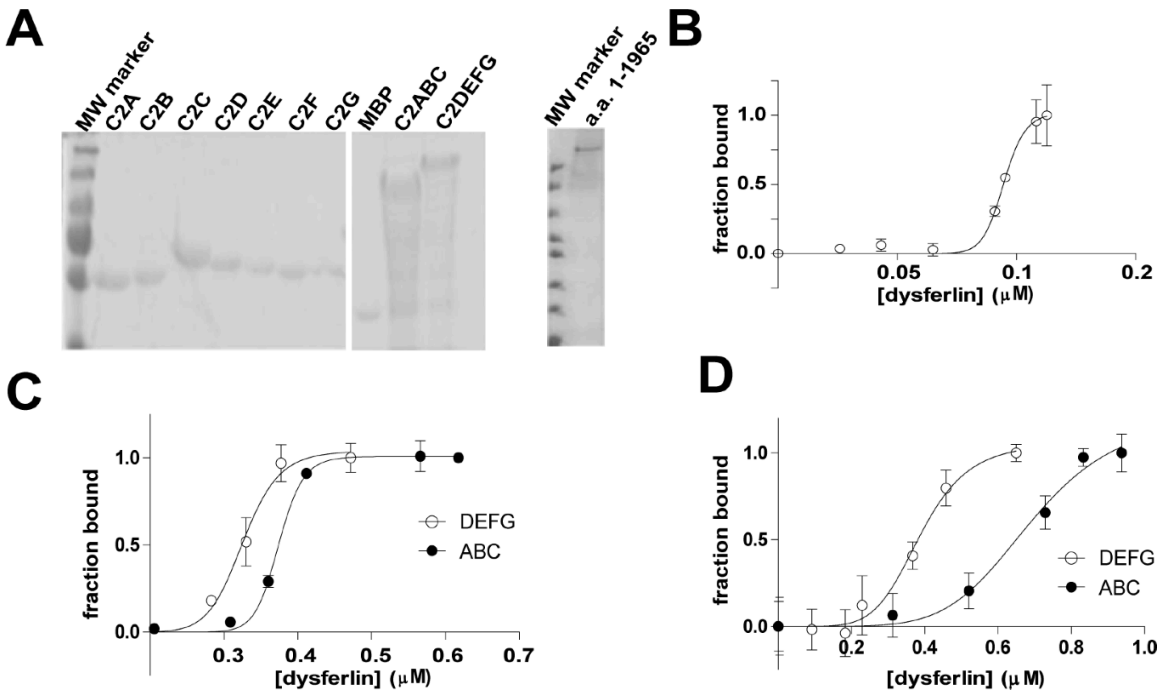


Figure 2.4. Dysferlin directly binds syntaxin 4 and SNAP-23 as determined by fluorescence anisotropy. (A) Representative SDS gel showing purity of recombinant dysferlin proteins. Purified C2 domains are shown on the left, with the dysferlin construct possessing the entire cytoplasmic region of the protein (lacking only the transmembrane domain) is shown on the right. (B) Dose-response curve for the addition of the full-length cytoplasmic region of dysferlin (a.a. 1-1965) to a sample containing FITC labeled syntaxin 4 lacking a transmembrane domain. (C) Dose-response curves for C2ABC and C2DEFG region of dysferlin show a progressive increase in binding for the H3 domain of syntaxin 4. (D) Titration curves for FITC labeled SNAP-23 mixed with either the C2ABC or C2DEFG constructs of dysferlin. Measurements were collected at 23° C. Error bars represent std, N= 3.

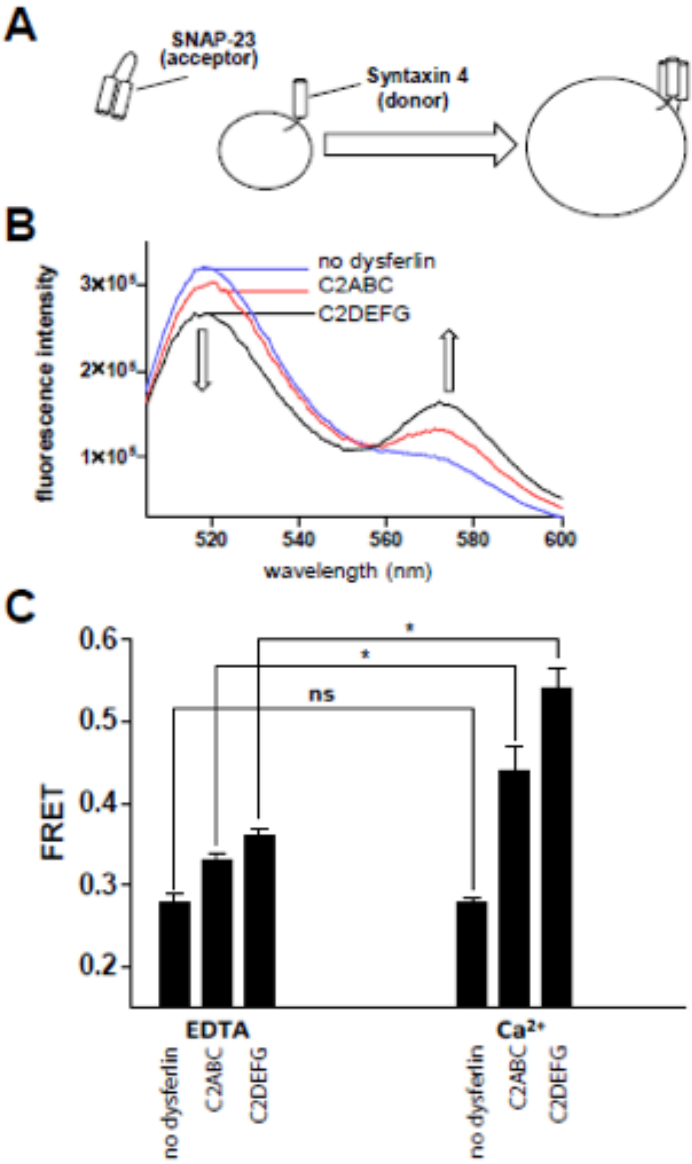


Figure 2.5. Dysferlin accelerates assembly of syntaxin 4/SNAP-23 heterodimers. (A) Schematic of the assay used in this study. A soluble form of SNAP-23 labeled with an acceptor fluorophore is mixed with donor labeled syntaxin 4 reconstituted into POPS/POPC liposomes. Heterodimer formation results in FRET, with decreased donor intensity and increased acceptor intensity. (B) Representative fluorescence spectra of samples containing C2ABC, C2DEFG, or no dysferlin in the presence of 0.5 mM calcium. Arrows denote change in donor and acceptor fluorescence. (C) Quantitation of FRET efficiency after incubation with dysferlin at 23°C for 1 hour. EDTA and calcium concentrations were 500 μM. Error bars represent std, N= 3, *P<0.05, ns not significant.

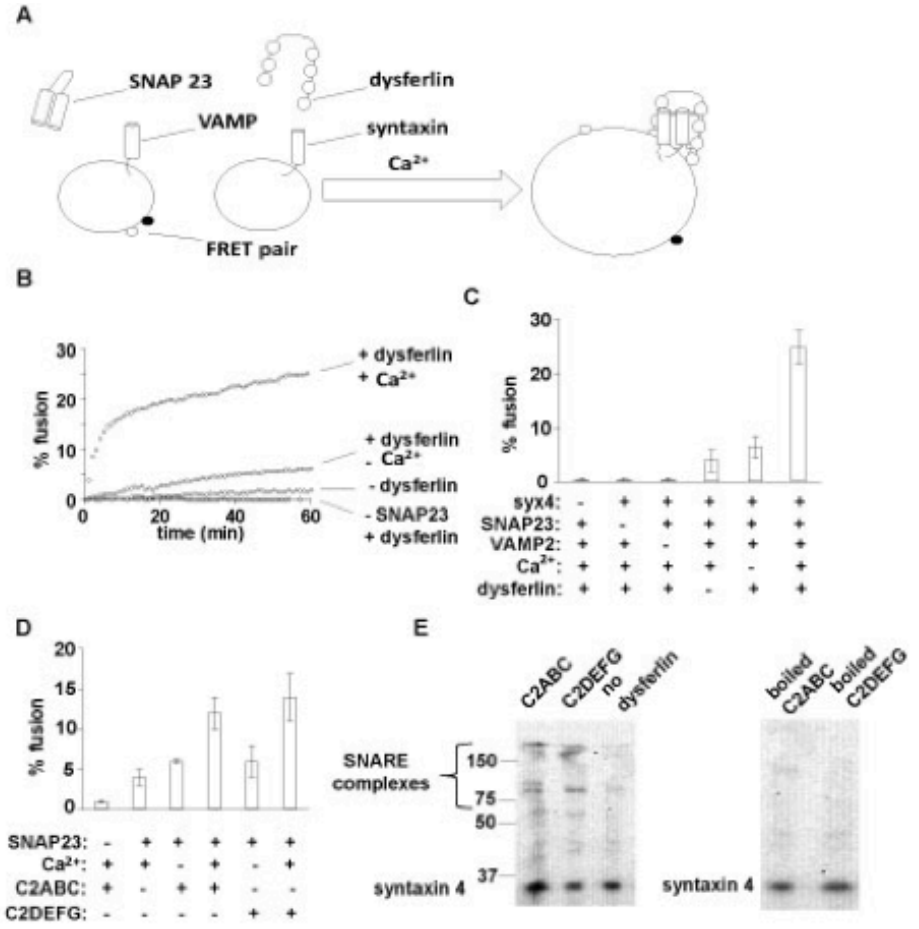


Figure 2.6. Dysferlin stimulates SNARE mediated membrane fusion. (A) Schematic of the in vitro fusion assay. Lipid mixing between SNARE bearing proteoliposomes results in dilution of NBD and rhodamine lipid-FRET pairs (denoted as black and white circles) and dequenching of NBD donor fluorescence. (B) Representative lipid mixing traces in the presence or absence of dysferlin, calcium, EDTA and SNAP-23 at 37 ° C. The concentration of SNAP-23 and the cytoplasmic region of dysferlin was 15 and 2 μM respectively. (C) Quantitation of the effects of the cytoplasmic region of dysferlin on membrane fusion. (D) Representative fusion assays conducted in the presence of 10 μM of C2ABC or C2DEFG demonstrate the ability of truncated forms of dysferlin domain to stimulate fusion between SNARE bearing proteoliposomes. Error bars represent std, N = 3, EDTA and calcium concentrations were 500 μM. (E) Western blot of lipid mixing assay products probed for syntaxin 4 after reacting for 1 hour at 37°C with calcium and either C2ABC, C2DEFG, or no dysferlin. Syntaxin 4, as well as prominent high molecular weight SDS resistant complexes positive for syntaxin 4 was detected in dysferlin containing samples (left panel). Boiling of dysferlin containing lipid mixing assay products for 10 minutes eliminated the syntaxin 4 positive high molecular weight bands. Molecular weights in kDa are indicated.

Chapter 3

In Pursuit of a Crystal Structure to Gain Understanding of the Sedoheptulose 7-phosphate Cyclase Family

By Sara J. Codding

Abstract

Sedoheptulose 7-phosphate cyclases (SH7PC) are enzymes that utilize the seven carbon pentose phosphate pathway intermediate sedoheptulose 7-phosphate to generate cyclic precursors incorporated into many therapeutic bioactive natural products. The enzyme ValA from *Streptomyces hygroscopicus* which makes the 2-epi-5-epi-valiolone product is one of the three different types of SH7PCs that have been identified; the 2-epi-5-epi-valiolone synthases, the des-methyl-4-deoxygaudsol syntheses, and the 2-epi-valiolone synthases, that each generate a different cyclic product from the same substrate. These proteins are distant relatives to the well-characterized and structurally determined dehydroquinate synthases (DHQS) from the shikimate pathway. DHQS and SH7PCs both bind a seven carbon sugar substrate and are part of the more recently recognized superfamily the sugar phosphate cyclases (SPCs). We report that despite the low overall sequence identity between the SH7PCs and dehydroquinate synthases, as well as among the SH7PCs, the majority of active site residues identified by sequence alignment are conserved or semi-conserved between these proteins lending little insight into the biochemistry that allows each SH7PC enzyme to utilize the same substrate yet generate a distinct product. Comparative analysis of the primary sequence alignment of the SPCs suggests that the active site residue differences are predominately localized to areas where the SH7PCs substrate differs from that of DHQS. This suggests that SH7PCs' active site chemistry is likely conserved and similar to that of DHQS. Due to this, we postulate that product stereochemistry is a fait accompli, in so much as it is not likely the result of catalysis but due to selective binding of either the α or β pyranose anomer of the substrate by each type of SH7PCs. We utilize the biophysical technique of X-ray crystallography and report the purification, crystallization, data collection, and metal scanning data of the SH7PC ValA. The ValA structural solution supported the observations based on sequence comparisons among the SPCs.

Introduction

Sedoheptulose 7-phosphate cyclases (SH7PC) are members of the recently identified sugar phosphate cyclase superfamily (SPC) of proteins and use the pentose phosphate pathway (PPP) sugar sedoheptulose 7-phosphate (SH7P) as their substrate to form C_7-cyclitols; cycloalkanes that contain three or more hydroxyl groups (Figure 3.1) (35). Enzymes in this family utilize NAD^+ and a metal ion, either cobalt or zinc as prosthetic groups, to perform catalysis (34,35,98-100). The majority of the research interest in SH7PCs is largely due to their ability to synthesize cyclitol products that are incorporated into natural products (101). The natural products that are synthesized down stream of the SH7PCs are secondary metabolites, pseudooligosaccharides, such as acarbose for diabetes treatment, the crop protectant validamycin A, the antitumor agent cetoniacytone A, and the sunscreen mycosporine-like amino acids (Figure 3.2) (102-108). Interestingly, the natural products noted above are molecules that are generated from the same core cyclic precursor, sedoheptulose 7-phosphate (Figure 3.1), by one of three enzymes present in some bacteria and fungi that allow SH7P to be used in secondary metabolism. Of the identified enzymes involved, there are three types; 2-epi-5-epi valiolone synthases (EEVS), the recently characterized 2-epi-valiolone synthases (EVS) and the des-methyl-4-deoxygaudisol synthases (DDGS) (35,105,109).

Half of the six biochemically characterized phylogenetic groups in the sugar phosphate superfamily are SH7P dependent (DDGS, EEVS, EVS) (Figure 3.3) (35). Despite their large representation in the SPC superfamily and the myriad of bioactive natural products that arise from SH7PCs, they remain structurally uncharacterized. In fact, our current understanding of enzymatic catalysis of the SPCs is largely based on two other well-studied SPCs, dehydroquinate synthase (DHQS Figure 3.3, green) and 2-deoxy-scyllo-inoso synthase (DOIS). The DHQS enzymes like SH7PCs cyclize a PPP seven carbon sugar but, DHQS acts on the substrate 3-deoxy-D-arabinoheptulosonate 7-phosphate (DHAP) as part of the shikimate pathway for synthesis of aromatic amino acids in bacteria, fungi, and plants (Figure 3.4) (110,111). Due to its necessary function, DHQS is a target of antimicrobial drug development which in turn has made DHQS

largely studied (98). As for the DOIS type enzymes, they act on the substrate glucose 6-phosphate and while they are not as well studied as DHQS, they are involved in the biosynthesis of aminoglycoside antibiotics including neomycin, kanamycin, gentamicin, and ribostamycin (Figure 3.4) (111). In addition, structural pursuits of DOIS have been successful; in so much as structures of DOIS type enzymes have been solved both in the presence and absence of inhibitors (99).

Of structurally determined SPCs, both DOIS and DHQS share the same fold. The enzymes assemble as functional homodimers. For reference we have included a representative, high resolution structure of a DHQS, pdb code 1DHS, (Figure 3.5) (98). Each monomer is composed of two domains, an N-terminal α/β-sandwich (Figure 3.5, red) that binds NAD^+(Figure 3.5, sticks). The carboxy-terminal α-helical domain (Figure 3.5, blue) harbors most of the residues involved in catalysis and metal binding. The active site resides in a cleft between the N and C terminal domains.

As mentioned above, DHQS is largely studied and the extensive research has proposed a chemical mechanism based on biochemical studies and substrate analog studies, both which are consistent with a high resolution crystal structure of the DHQS, pdb code 1DQS (Figure 3-boxed) (110) (100,112,113). The reaction is a complex multistep mechanism that involves alcohol oxidation, phosphate β-elimination, carbonyl reduction, ring opening and intramolecular aldol condensation (Figure 3.6) and all steps are catalyzed by DHQS with Zn^{2+} and NAD^+ as cofactors (112,113). It appears that the DOIS crystal structure shares high structural similarity to DHQS, and thus it is proposed to have a similar mechanism (114). Further, on the simple basis of sequence comparisons of SH7PCs to both DHQS and DOIS, it is suggested that SH7PCs also adopt a similar mechanism to DHQS (35,98). But, this analysis would be supported by structural data to help elucidate how the active sites produce distinct products from the same substrate, especially for EEVS and EVS whose products vary only by differing stereochemistry. Finally, as SH7PCs are crucial in the biosynthesis of natural products, developing an understanding of their structure-function relationships may provide a door to new therapeutics through synthetic biology and semisynthetic production (115).

Here, we report the, crystallization, data collection and processing, and metal scanning data, to support the structural solution of the EEVS enzyme ValA from *Streptomyces hygroscopicus* 5008, that is crux in the biosynthetic pathway of validamycin A, a crop protectant (109). As the first SH7PC structure reported, we took extensive steps to validate the active site metal as zinc and were pleased to collect anomalous scattering to specifically locate the heavy atom in the structure (116). We have provided sequence comparisons of ValA, with representative EVS and DDGS sequences, and two structurally determined DHQS proteins and one DOIS protein. Interestingly, although there was compiled structural data reported in Keen et al, and a comparative structural analysis ValA to DHQS, we conclude that active site views support the conclusion here that the SH7PCs EEVS (ValA) and EVS bind different anomers to generate stereochemically distinct products.

Materials and Methods

Expression, purification, and crystallization of ValA. The expression of recombinant ValA was done as previously described (117). For purification, all done at 4 °C, cell pellets from 100 mL cultures were each resuspended in ~5 ml of 40 mM HEPES, 300 mM NaCl, pH 8.0 (buffer A) with 10 mM imidazole, sonicated (13 watts, 4 x 1 min), and centrifuged (14,500 rpm, 30 min). The supernatant was loaded onto a Ni-NTA resin column (5 ml resin; 0.8 ml/min). After washing with 100 ml buffer A with 20 mM imidazole, the protein was eluted using a 200 ml gradient from 20 to 500 mM imidazole in buffer A. Fractions (~6 ml each) containing protein were combined and dialyzed overnight against 2 L 10 mM Tris-HCl, 300 mM NaCl, 5 mM imidazole, pH 8.0. A second phase of purification was done similarly using a Talon column (~40 ml run at 0.3 ml/min) in buffer B (20 mM Tris-HCl, 300 mM NaCl, pH 8.0) with 5 mM imidazole for column equilibration, 10 mM imidazole for washing, and a 200 ml gradient from 10 mM to 200 mM imidazole for elution. Fractions (~4 ml each) containing pure ValA as judged by SDS gel electrophoresis were combined and dialyzed against 2 L 10 mM Tris-HCl,

pH 7.5 (3 x 3 h). The protein solution was concentrated by ultrafiltration (10 K cutoff membrane) to 10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

The crystals of ValA were optimized from the initial commercial screen hit to grow robust diffracting crystals at 4 °C in hanging drops to full size after 14 days from 4 μL protein and 1 μL of a 0.6 M succinic acid, pH 6.5 reservoir solution. The resulting crystals were rod-like with dimensions of about 50 x 50 x 200 μm^3 (Figure 3.7) at 4°C and diffracted to 1.9Å.

X-ray diffraction data collection. For diffraction data collection (at –170 °C), crystals were briefly passed through a solution containing 20% PEG 400 and then flash frozen in loops by plunging into liquid nitrogen. Data were collected from two crystals using λ = 1.0 Å X-rays and $\Delta\varphi$ = 1° steps at beamline 5.0.2 at the Advanced Light Source (Berkeley, CA). From both crystals, 120 2.0 s images were collected at detector distance d = 250 mm, and from the second, an additional 200 2.0 s images were collected at d = 350 mm. All these images were integrated using XDS (118) and merged using the CCP4 suite of programs (119,120) to obtain the data set used for the initial structure solution and refinement. The merged data set was usable out to 2.1 Å using $CC_{1/2}$ ~0.2 as the cutoff criteria, and a random 5% of reflections were marked for cross-validation. In addition, a third crystal was used for a fluorescence scan and to collect a data set at beamline 5.0.2 using λ = 1.282 Å X-rays to maximize the anomalous signal from the bound zinc. This data set included two sets of 60 $\Delta\varphi$ = 1°, 4.0 s images offset by $\Delta\varphi$ = 90° to collect the bijovet pairs, and yielded data useful to 3.5 Å resolution (data not shown).

Structure determination. With a diffraction pattern of ValA in hand we first performed molecular replacement (MR) using Phaser to try and solve the structure (121). Initial phases for MR are typtically the starting point from the closest structurally determined homolog of ValA. In this case 3OKF shares a 28% sequence identity to ValA by ClustalW. After MR was performed, the phases were marginally improved by early refinements using Phenix (122). After which, all density map calculations and model building was done in Coot (123). Our initial MR solution is correct and was verified by the bijvoet pair data collected at the zinc (the catalytic metal) frequency.

Few SPC structures are deposited in the protein data bank (19 DHQS structures and 2 DOIS structures). In general, molecular replacement solutions are successful if above 35% sequence identity, which has made finding a structural solution to the diffraction data non-trivial. Follow up work by Keen et al. solved the phase problem by molecular replacement using MR-Rosetta with default settings (124).

Structural comparisons and sequence analyses. Secondary structure assignments were made using DSSP (125,126) and structure-based alignments were generated using the Dali server and ClustalW (127,128).

Results and Discussion

Sequence analysis of SPC family revels active site residue conservation. When representative proteins of the SPC family are aligned in a ClustalW multiple sequence alignment (Figure 3.8) it is clear that there is little overall sequence identity (Table 3.1). The greatest being between the two DDGS enzymes isolated from different species, *Np*DDGS from *Nostoc punctiforme* and *Av*DDGS from *Anabaena variabilis,* both discovered via a phylogenetic analysis of SPCs in deposited genomes (34). The three structurally determined proteins included in this alignment, the *Aspergillus nidulans* DHQS pdb code 1DQS (1.8 Å plus substrate), the *Vibrio cholerae* DHQS pdb code 3OKF (2.5 Å), and the *Bacillus circulans* DOIS pbd code 2D2X (2.3 Å) also show little overall sequence identity both, amongst each other and with the SH7PCs (Table 3.1) (98,99). Yet, strikingly, unlike the two DDGS enzymes, the two DHQS proteins share substantially fewer residues. In addition, between the different types of enzymes in the SH7PC family, the greatest percent identity is shared between ValA and two the aforementioned DDGSs. Due to similarity of reaction products it could be hypothesized that EVS and EEVS (ValA) would share greater similarity than EEVS and DDGS. However, it remains that this alignment is not extensive and that the SPCs included here display a low overall percent shared identity.

While there is low total sequence identity among the different SPCs, it is clear that the metal and NAD^+ binding, and the majority of active site residues are conserved

(Figure 3.8, n= NAD^+ binding, m=metal binding, *=active site residue). In fact, the all metal binding residues observed in 1DQS, 3OKF, and 2D2X, are conserved among ValA, *Actinosynnema mirum* EVS (*Am*EVS), and the two DDGS enzymes by sequence alignment. It appears all SPCs may share residues equivalent to 1DQS, Lys152, Glu197 and His287 involved in catalytic metal binding (Figure 3.8 and Table 3.2) (125,128,129). This supports a role for a catalytic metal in SH7PCs. The same is true for NAD^+ (bold denoted n in Figure 3.8) such that all ten identified residues involved in NAD^+ binding are conserved.

When the sequence alignment is annotated for putative active site residues of SH7PCs compared to those determined by structure in proteins 1DQS, 3OKF, and 2D2X, some differences arise (Table 3.2). Of the ten active site residues found in 1DQS, 3OKF, and 2D2X, six of those residues are conserved among all SPCs analyzed in our alignment; those residues are equivalent to Arg130, Arg146, Asn162, Leu197, Glu260, His271 in 1DQS. However, there remain four residues that are not conserved across the SPC family. In 1DQS, residue 250 is a lysine that is conserved in 3OKF and 2D2X, but this aligns with a methionine in ValA, *Av*DDGS, and *Np*DDGS. The lysine residue of 1DQS at position 250 interacts with the carbonyl oxygen of carbon 1 of its substrate DHQ (Figure 3.4). But, SH7P lacks this carbonyl oxygen, thus the methionine found in SH7PCs *Sh*EEVS (ValA), *Av*DDGS, and *Np*DDGS perhaps confers some substrate specificity. In that, both EEVS and DDGS enzymes generate products with the same stereochemistry at carbon 5 of their respective products (Figure 3.1). But, in *Am*EVS, the sequence alignment shows a conserved lysine as in DHQS and DIOS at this position and, the stereochemistry of the EVS product is inverted compared to that of the EEVS and DDGS products (Figure 3.1). Further supporting the hypothesis that this residue could confer substrate specificity.

The 1DQS residue 264 is an arginine and it is conserved in *Am*EVS and ValA, but not in *Av*DDGS and *Np*DDGS. But upon careful analysis, both *Av*DDGS and *Np*DDGS harbor a neighboring residue in the alignment is in fact arginine (Figure 3.8 orange arrow, italic, bold). It could be that this arginine, rather than the aspartic acid identified by

sequence alignment is key in the active site of *Av*DDGS and *Np*DDGS (Table 3.2). It is here where structural studies could be highly informative.

The 1DQS Asn268 residue is not well conserved among the SPCs. It can be argued that in the representative DOIS pdb code 2D2X, and in ValA, that they harbor semi-conserved residues at this location. But it is unclear as of yet how DDGS enzymes compensate for the sequence identified alanine at this position (Table 3.2).

Of those residues that have been shown to be critical for 1DQS enzymatic activity, Arg130, Lys152, Arg264, His275, the SH7PCs showed conservation of these residues by sequence alignment except for at the catalytic histidine 275 (130). In 1DQS, is it proposed that His275 acts as a base to facilitate ring opening of the substrate, but ValA and the DDGSs have a proline at this site. It is unclear if there exists a residue that can act as a catalytic Lewis base in place of His275 for ValA and DDGS. We anticipate that a structural solution may be helpful in identifying one such residue in the active site that is not identified by sequence alignment.

When the primary sequence alignment of the SPCs noted above is analyzed with DSSP for secondary structure assignment the secondary structure boundaries are shared by all structurally determined SPCs (Figure 8). This combined with the observation that by alignment, the NAD^+, metal binding, and six of ten active site residues are conserved, supports the hypothesis that all SPCs share the same fold.

The alignment generated active site residues that were analyzed here are summarized in Table 3.2.

Crystallization. We have found that often, the bottleneck to protein structure determination is in the pursuit to obtain the first crystals. Thus, we have utilized a streamlined approach that both conserves protein and screens 384 different crystal cocktail conditions. Our micro-crystallization trials were carried out with commercial crystallization cocktails; Emerald Wizard 1&2, Emerald Wizard 3&4, Hampton Index, and Hampton Crystal Screen 2. The cocktails were dispensed in 1:1, 1:2, 1:4 crystal cocktail: protein ratios with the Crystal Phoenix robotic liquid handling system by Art Robbins Instruments in microbatch sub microliter sitting drops. Thus, what would have

required many hours of manual pipetting and ten times the volume of protein was cut to four hours and utilized less than 300μL of protein.

Using this approach, we performed crystal trials of *Am*EVS, ValA, *Av*DDGS, and *Np*DDGS and obtained initial hits for all four of the proteins in this study. *Am*EVS has produced two promising crystal leads, one condition has reproduced well and formed visibly single crystals in 5-7 days at 4 °C after optimization (Figure 3.7 A). Of the DDGS proteins, *Av*DDGS has two promising crystal leads, one diffracting to 1.7 Å directly out of the microplate (Figure 3.7 B), but the diffraction pattern was indicative of a more than one crystal (ie: not a single diffraction pattern). As of yet, *Np*DDGS has only produced small needles that are not of diffraction quality (not pictured). It may be useful to search for crystallization conditions of the EVS and DDGS proteins above by lowering the concentration of the protein to support ordered growth of crystalline solid as we observed protein precipitation in many of the conditions. Often the addition of salts and organic additives are commonly used in conjunction with standard commercial screens to produce new crystal leads (131). However, the initial hits from the micro-crystallization trials did produce a crystal of ValA that we were then able to optimize. In the future it may be advantageous to seed the EVS and DDGS crystals for nucleation with ValA crystals in a cross seeding experiment to generate additional hits in commercial screens (131).

The initial hits from ValA were grown out of 0.8M succinic acid at pH 7.0. But, while crystals did form, we also noted precipitation of protein in the initial conditions. Thus, we found that reduction of the succinic acid concentration from 0.8M to 0.6M, as well as reduction of the pH from 7.0 to 6.5, produced visibly larger crystals in the presence of notably less precipitate. Crystals grown in the 1:1 protein to cocktail ratio produced the largest crystals. However, our drops were 1 μL total volume and we noted that the crystals grew fast and growth attenuated quickly. In attempts to slow the crystal growth we increased the drop size to 4 μL total, while retaining the protein:cocktail ratio. We did this under the premise that a larger drop takes a greater amount of time to reach of equilibration with the reservoir, slowing the equilibration time, and thereby promoting

crystal growth. We ultimately found crystal conditions that were reproducible and produced crystals of diffraction quality using our optimization method (Figure 3.7 C).

Cryoprotectant trials, diffraction collection, and metal scanning. To prevent the formation of ice during flash freezing of protein crystals, cryoprotectants are used. To determine the appropriate cryoprotectant for the ValA crystals, we tested their stability in mother liquor plus common cryoprotectants. We found that including 20% PEG 400 in the mother liquor prevented ice formation and preserved crystal stability. Thus, several crystals were flash frozen in liquid nitrogen post cyroprotectant soak and cryo-diffraction patterns were collected.

It quickly became apparent that although the ValA crystals diffracted to high resolution, the diffraction spots were weakly intense. Thus, we first collected data from the best diffracting crystal at a large detector distance to capture the high-resolution reflections with relatively long exposure times to increase the signal to noise. After this, a second data set was collected from a different crystal to obtain the low resolution reflections with reduced exposure times. The two data sets were then combined to overcome the compounded issue of low intensity reflections at high resolution with overexposed reflections at low resolution and high intensity.

To obtain information on the identity of the metal in the active site of ValA, we utilized a tunable X-ray source to collect fluorescence scans at the metal absorption edge of cobalt, nickel, and zinc. The identity of the metal was confirmed to by collection at 1.2873 Å and the scattering coefficients f`=9662eV midpoint and f"=9670eV maximum were observed. These are of characteristic anomalous zinc k-shell fluorescence spectra (Figure 3.10).

With this in hand we then collected a full multi-wavelength anomalous diffraction (MAD) data set at the zinc absorption edge (132). This allowed us to maximize the anomalous signal from the bound zinc determined above. This data set included the bijovet pairs.

Initial structural solutions by standard molecular replacement. Of the SPC structures deposited in the protein data bank (19 DHQS structures and 2 DOIS structures)

3OKF has the highest sequence identity to each of the SH7PCs in this study (Table 3.1). In general, molecular replacement solutions are successful if the structure that is in need of phasing shares at least 35% sequence identity with a known structure. Due to the fact that the calculated percent sequence identity between ValA and 3OKF is below the 35% threshold, the use of standard molecular replacement to find the initial phases of ValA was non-trival. This lends gravitas to our reported difficulty with traditional molecular replacement to obtain a structural solution to ValA. However, we when did utilize standard molecular replacement approaches and the phases from the solved structure 3OKF as a starting point, we observed a promising result. The initial phases that resulted from this standard approach gave high R/R_{fee} the best calculated at 45/50 from Phenix MR replacement. But, this result was corroborated by the anomalous data from zinc. The modeled active site from this MR replacement aligned the metal binding residues with the density determined from the MAD data, suggesting that the initial phases are an acceptable approximation (Figure 3.11). However, attempts to build additional residues using buster were not successful.

Final structural solution to ValA. The final solution of ValA was completed by Keen using the computational package MR_Rosetta (133). This computational approach combines the tools of standard molecular replacement with the structure-modeling program Rosetta to improve crystallographic molecular replacement when difficulties, as noted above, are encountered. Mr_Rosetta can also aid in model-building, density modification, and refinement. Using the approach as described in Diamo et al., the application of MR_Rosetta was successful to solve the ValA crystal structure (134).

Structural solution supports results from primary sequence alignment. The MR_Rosetta solution of the ValA structure obtained by Keen supports the observations made above from primary sequence alignment. The structural solution shows that the overall fold of SPCs structurally identified in both DHQS and DOIS is retained by ValA. This was predicted by sequence alignment and the DDPS assignment of secondary structure to that alignment (Figure 3.12).

The structural solution also supports that which was observed by primary sequence alignment of SPCs, the ValA active site shows a conservation of Zn^{2+} and NAD^{+} binding residues as identified in Table 3.2. In addition, the structural solution also shows that the neighboring arginine residue noted in the alignment (orange arrows Figure 3.8) of *Av*DDGS and *Np*DDGs, are likely equivalent to Arg264 in 1DQS. This is due to the fact that the ValA residue, Arg277 that is equivalent to Arg264 in 1DQS was in fact far from the active site. By structural overlay of ValA to 1DQS it was found that the neighboring ValA residue, Arg278 was found to be to close in conformation to Arg264. This arginine 278 in ValA aligns with the identified 'neighboring arginine' in DDGS enzymes (Figure 3.8, orange arrow italic, bold). The ValA structural solution suggests that the shifted residue is likely to support the substrate binding pocket of the SH7PC substrate where it differs, at carbon 2, from that of 1DQS. This, plus the fact that the product of EVS is stereochemically inverted compared to that EEVS, and DDGS, suggests that an induced fit binding model may lend a form of substrate preference at the active sites of EVS vs EEVS/DDGS and supports the observations made above. However, in regards to DDGS it remains to be confirmed by structural or functional studies.

A side-by-side structural comparison and overlay of ValA and 1DQS by did not reveal the catalytic Lewis base residue in ValA (116). Which active residues may provide a putative alternative for a Lewis base in the EEVS or DDGS remains unclear. We speculate that it may be useful to solve a structure of ValA in the presence of a substrate analog to help understand the divergence.

Finally, as noted above, the 1DQS residue 250 is conserved in 3OKF and 2D2X, but this aligns with a methionine in ValA, *Av*DDGS, and *Np*DDGS. As this is a residue that interacts with the carbonyl oxygen of carbon 1 of the 1DQS substrate, we speculated that the non-conserved residue at this site may confer some substrate specificity to *Sh*EEVS (ValA), *Av*DDGS, and *Np*DDGS. This, combined with the fact that both EEVS and DDGS enzymes generate products with the same stereochemistry at carbon 5 of their respective products, the observation becomes plausible (Figure 3.1). In addition, in

*Am*EVS product the stereochemistry is inverted at carbon 5 compared to the other two SH7PCs. Thus, to determine a role that the aligned methionine may play EEVS and DDGS active sites becomes an intriguing pursuit. This structural solution of ValA did not provide additional insight into the role of the equivalent methionine in the EEVS and DDGS proteins (Figure 3.8). Mutational analysis would be beneficial to determine if this residue is critical to preserve the stereochemistry of the enzymatic products.

In total, the analysis of the primary sequence was recapitulated by the final structural solution of ValA. The primary sequence alignment noted conservation of, NAD^+, metal, and the majority of active site residues, and this was in fact observed in the ValA structure. In addition, the structure supported that the overall fold was conserved among SPCs as well. Due to these facts, observed by alignment and recapitulated by the structural results, it is unlikely that the chemical catalysis of SH7PCs is largely divergent from DHQS.

When analyzing the sequence alignment this fact was initially apparent and we began to hypothesize that that the SH7PCs, EEVS and EVS could bind alternate anomers of the substrate thereby pre-assigning the stereochemistry of anomeric carbon that becomes carbon five in their respective products. Interestingly, the DDGS product can be produced from either anomer of the product due to the proposed reaction mechanism. Interestingly, as noted by primary sequence alignment EEVS and DDGS are more similar to each other in sequence than they are to EVS. In addition, as noted in Table 3.1 both ValA and the DDGS share the conserved neighboring arginine residue that is structurally equivalent to Arg264 in 1DQS, while EVS does not. And, EVS has the conserved asparagine at Asn268 in 1DHQ where ValA harbors an aspartic acid at the equivalent position. These three residues, Arg264, Asn268, Lys250, where EVS and DDGS/EEVS diverge, all reside in the binding pocket surrounding the substrate at carbon 2. Thus, we conclude that the two SPCs EEVS and EVS have distinct binding sites for the substituents of the substrate at the anomeric carbon. This leads us to the hypothesis, that EEVS and EVS bind either the α or β anomer respectively of SH7P rather than use catalysis to generate alternate stereochemistry at carbon five of their respective products.

We view the SH7PCs as enzymes that bind distinct forms of the same substrate to give products with distinct stereochemistries.

Future Directions

This structural study would be largely complemented by biochemical analyses to test the hypotheses of SH7PCs substrate specificity developed above. Traditional biochemical approaches guided by the ValA structure can guide site directed mutational studies on SH7PCs in the hopes of discerning the role of divergent active site residues. The key residues that appear integral in substrate specificity, or catalysis, can be systematically altered to test their role in substrate binding. Mutations are easily introduced and kinetic characterization that has been performed on wild type proteins has previously reported can be performed on mutants (35).

It would be insightful to perform studies on SH7PCs with a substrate analog as proposed in Figure 3.13. This analog is designed to strongly resemble SH7P but to participate in only the first two steps of (putative) catalysis. It is not available commercially and would need to be synthesized as previous substrate analogs for DHQS (112,135,136). In addition, the analog could be characterized for its interaction with all SH7PCs to determine a K_I and thus infer anomeric preference of EEVS, DDGS, and EVS. If this inhibitor is found to be ineffective for any SH7PCs, it is likely that synthesizing a compound with alternate stereochemistry at carbon 2 (*) may be beneficial (Figure 3.13). Further, the alternate chair conformation of the analog may be sought synthetically for biochemical binding studies if necessary.

Pursuit of high quality crystals of EVS and DDGS proteins to round out a complete comparative structural study on the SH7PCs could provide insight on anomeric preference as well. Pursuit of high quality crystals of EVS and DDGS proteins is non-trivial, however a solid leads have already been elucidated. In addition, the use of the substrate analog described above could be used to solve a co-crystal of ValA, furthering the analysis of the SH7PCs structure-function relationships.

Abbreviations. SPC, sugar phosphate cyclase; DHQS, dehydroquinate synthase;

SH7P, sedoheptulose 7-phosphate; DHQ, dehydroquinate; SH7PC, sedoheptulose 7-phosphate cyclase; EV, 2-epi-valiolone; aminoDHQS, amino-dehydroquinate synthase; DOIS, 2-deoxy-scyllo-inososе synthase; DOI - 2-deoxy-scyllo-inososе; G6P, glucose 6-phosphate; EEVS, 2-epi-5-epi-valiolone synthase; DDG, des-methyl-4-dexoygaudsol; EEV, 2-epi-5-epi-valiolone; PPP, pentose phosphate pathway; EVS , 2-epi-valiolone synthase; MR, molecular replacement; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; DDGS, des-methyl-4-deoxygaudsol synthase

Table 3.1. Percent sequence identity calculated by ClustalW primary sequence alignment

	1DHQ	3OKF	2D2X	*Am*EVS	*Sh*EEVS	*Av*DDGS	*Np*DDGS
1DHQ	-	35	17	29	19	19	34
3OKF		-	23	34	28	24	22
2D2X			-	20	16	19	21
*Am*EVS				-	32	28	27
*ValA (Sh*EEVS)					-	35	35
*Av*DDGS						-	72
*Np*DDGS							-

Table 3.2. Key active site residues of SPCs are highly conserved. Residues identified by ClustalW primary sequence alignment predicted to be involved in metal binding (m) or to be located in active site (*). 1DQS is the reference protein to denote the residue numbering depicted here. The residues that change identity after a sequence alignment is performed using the structural solution of ValA are denoted with arrows.

	*	*	m	*	m	*	*	*	*	*	*	*	m	*
Protein	Residue 130	146	152	162	194	197	250	260	264	268	271	275	287	356
1DQS	R	D	K	N	E	K	K	E	R	N	H	H	H	K
3OKF	R	D	K	N	E	K	K	E	R	N	H	H	H	K
2D2X	R	D	K	N	E	K	K	E	**G**	**E**	H	H	H	K
*Am*EVS	R	D	K	N	E	K	K	E	R	N	H	H	H	**K->*R***
ValA (*Sh*EEVS)	R	D	K	N	E	K	**M**	E	**R->*R***	**D**	H	**P**	H	**K->*H***
*Av*DDGS	R	D	K	N	E	K	**M**	E	**D->*R***	**A**	H	**P**	H	**L**
*Np*DDGS	R	D	K	N	E	K	**M**	E	**D->*R***	**A**	H	**P**	H	**L**

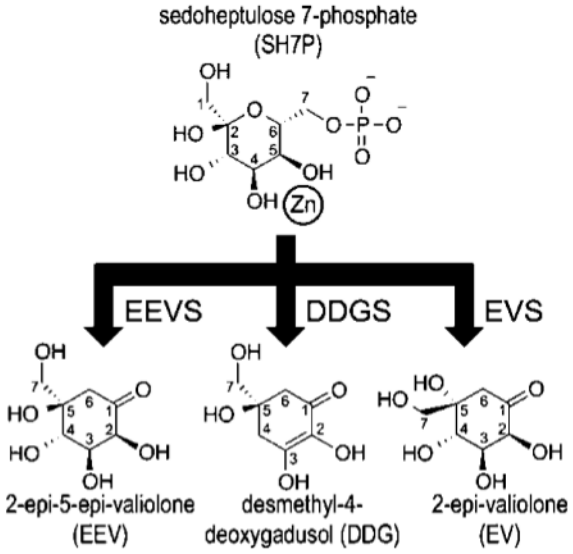
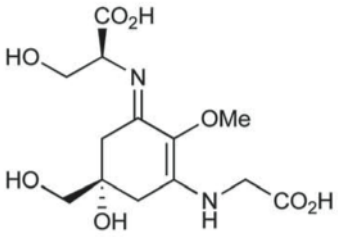


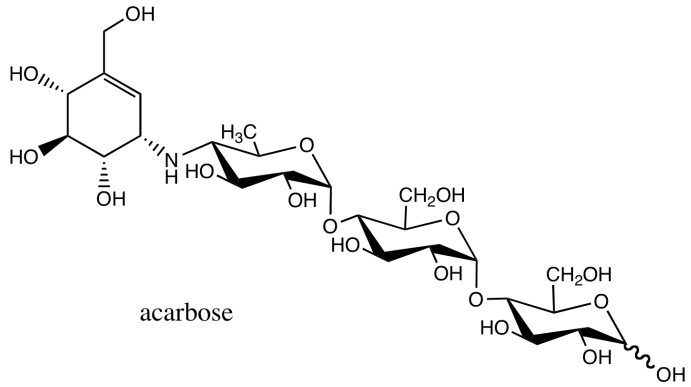
Figure 3.1. Reactions catalyzed by sedoheptulose 7-phosphate cyclases (SH7PC)

A.



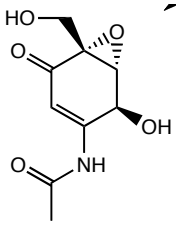
Shinorine

B.



acarbose

C.



cetoniacytone A

Figure 3.2. Clinically relevant SH7PC derived natural products

.

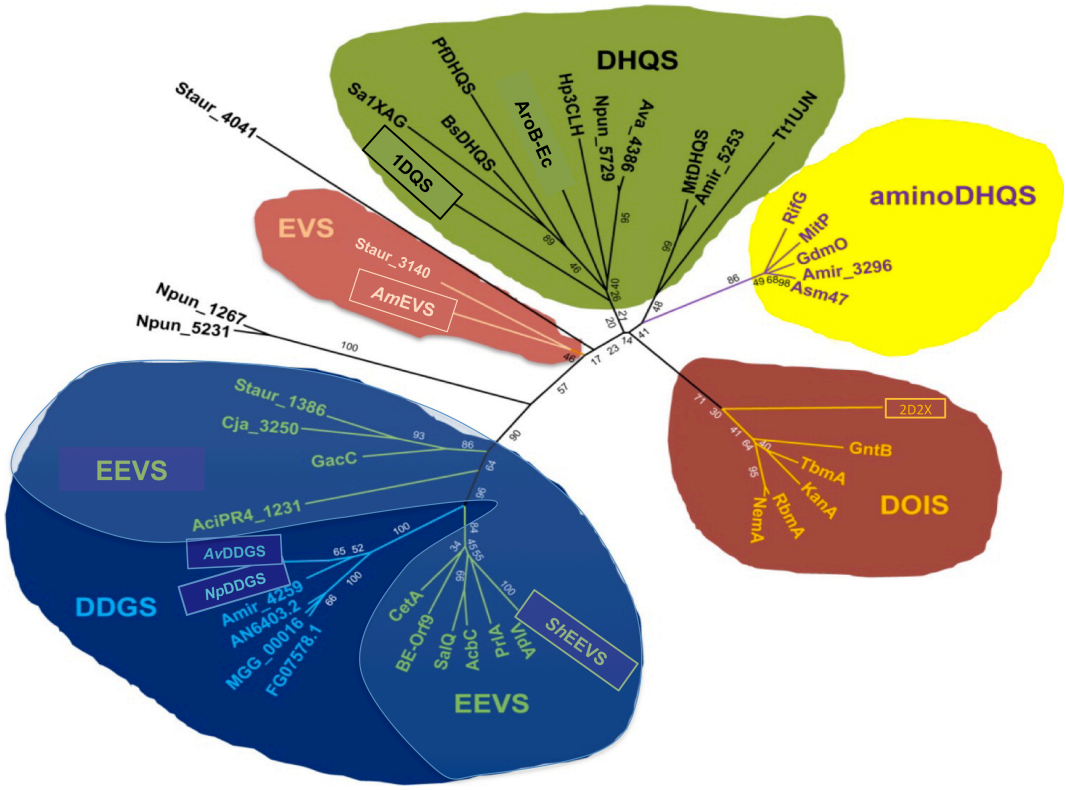


Figure 3.3. SPC Protein Phylogenetic Tree**.** Each SPC type is labeled in caps. The proteins utilized the construction of this tree were labeled after their corresponding gene except those included in this study, which are boxed

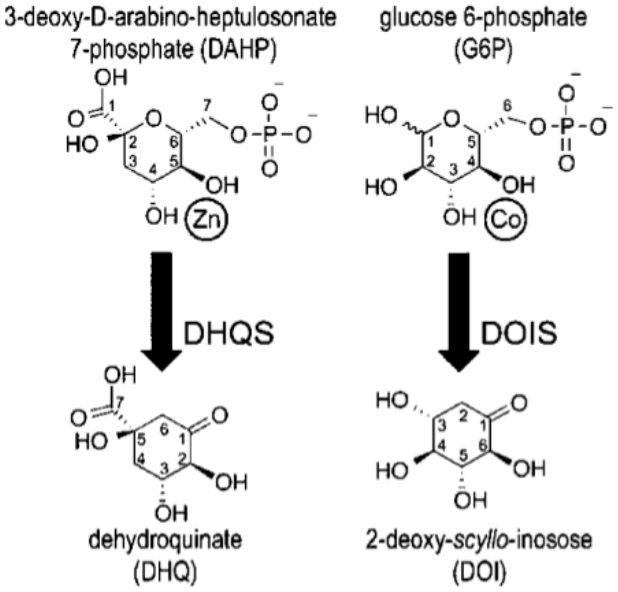


Figure 3.4. DHQS and DOIS catalyzed reaction substrates and products

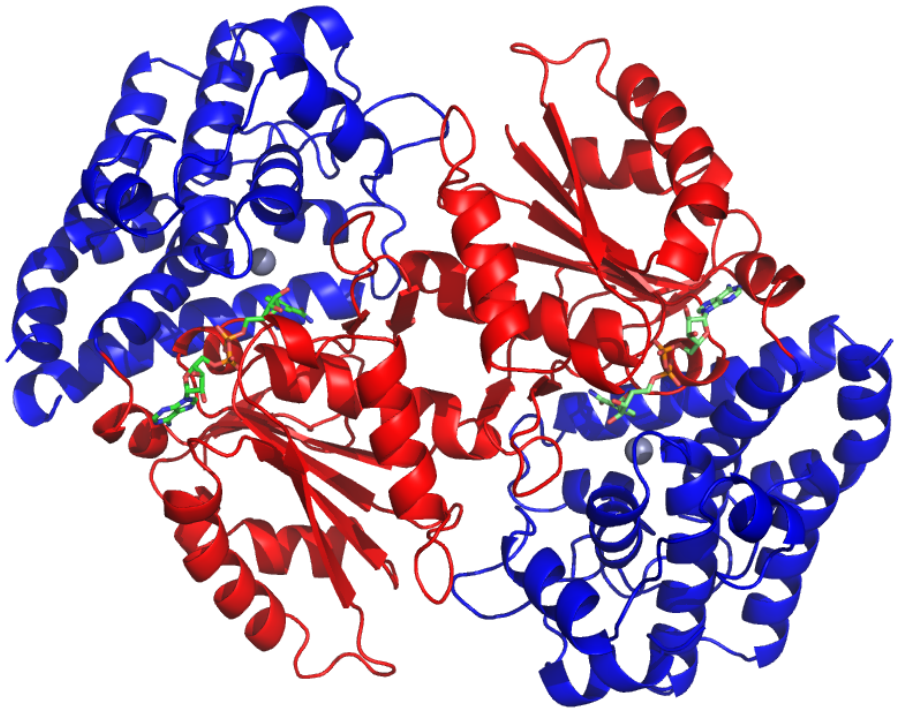


Figure 3.5. Representative DHQS structure. This DHQS shows the homodimer interface and the two distinct domains of each monomer (N-terminus in red or C-terminus in blue,), with space filling model of NAD^+ (greens) and metal cofactor Zn^{2+} (silver) pdb code 1DQS

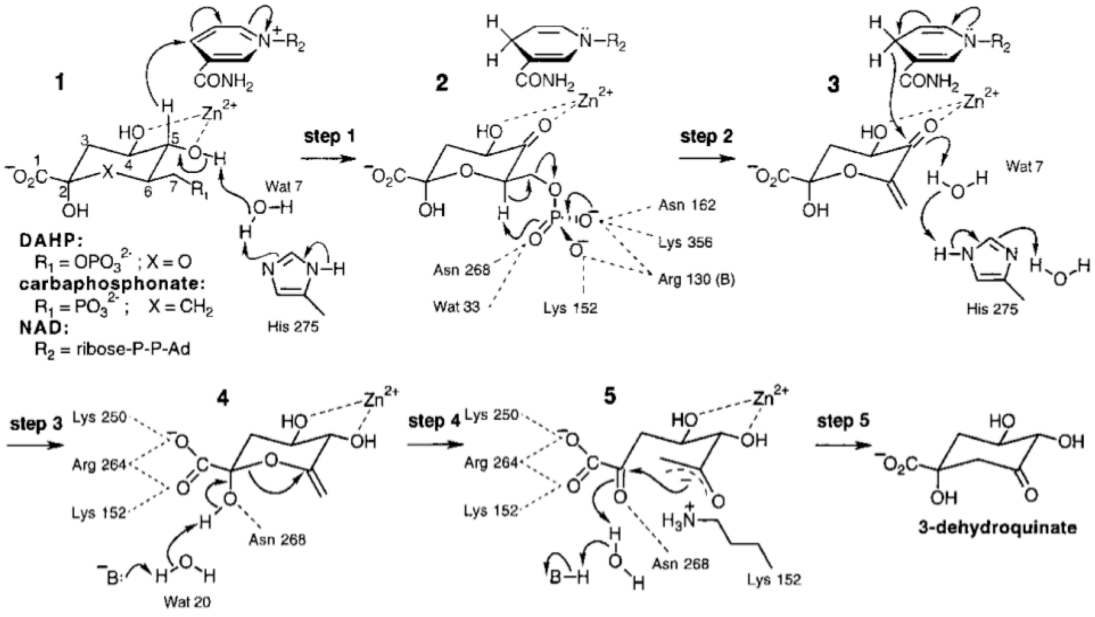


Figure 3.6. Reaction Mechanism of DHQS

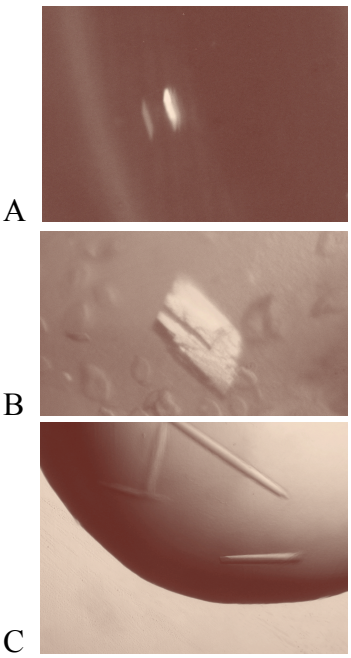


Figure 3.7. Protein crystals of SPCs. A. *Am*EVS primary crystal lead, 50x50x50μm B. the *Av*DDGS primary crystal lead 125x125x25 μm C. EEVS enzyme ValA, 25x25x300μm

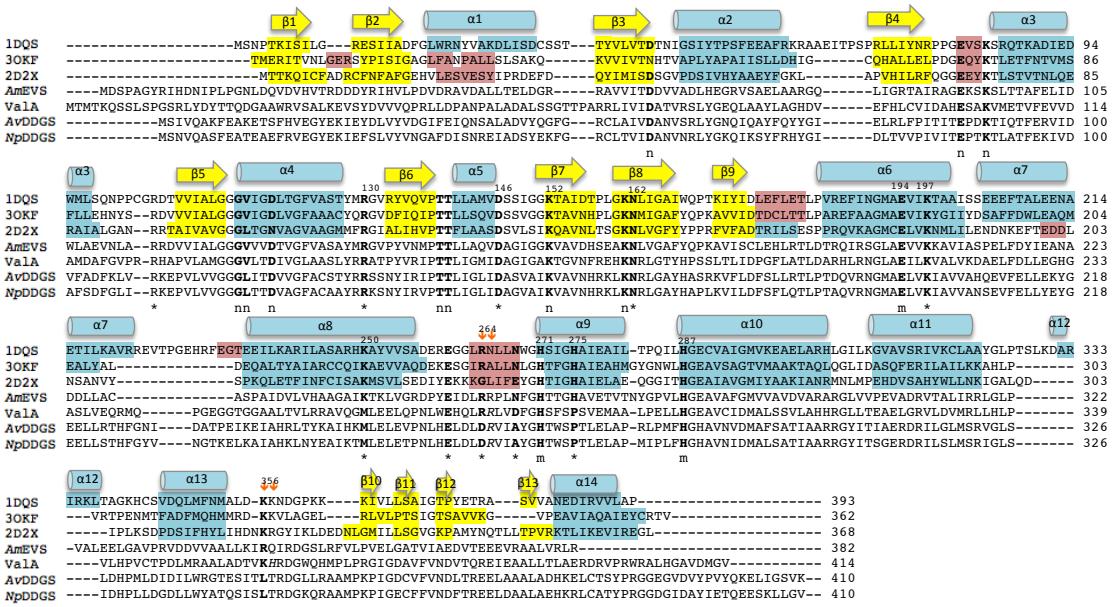


Figure 3.8. Multiple sequence alignment of SPCs using ClustalW. Representative DHQSs, *Aspergillus nidulans* DHQS pdb code 1DQS (1.8 Å plus substrate), and the *Vibrio cholerae* DHQS pdb code 3OKF (2.5 Å). The representative DIOS *Bacillus circulans* DOIS pbd code 2D2X (2.3 Å). The EEV known as ValA from *Steptomyces hygroscopicus, Sh*EEVS (ValA), the EVS enzyme *Actinosynnema mirum Am*EVS, and two DDGS both *Anabaena variabilis Av*DDGS, and *Nostoc punctiforme* (*Np*DDGS). Secondary structure is included for the structurally determined proteins and is color coded yellow=b strand, blue=a helix, pink=3_{10} helix. Secondary structure was determined by DSSP analysis of the solved crystal structures. Numbering of β strand and α helixes as in 1DQS. Bold are confirmed metal (m), NAD^+ (n) or active sites (*) residues determined by crystal structure or inferred by sequence alignment, the only exceptions to the primary sequence alignment determined metal, NAD+ or active site residues of ValA or DDGS are two residues denoted by orange arrow and italicized that were determined as the neighboring residue after ValA was solved. Numbers correspond to 1DQS.

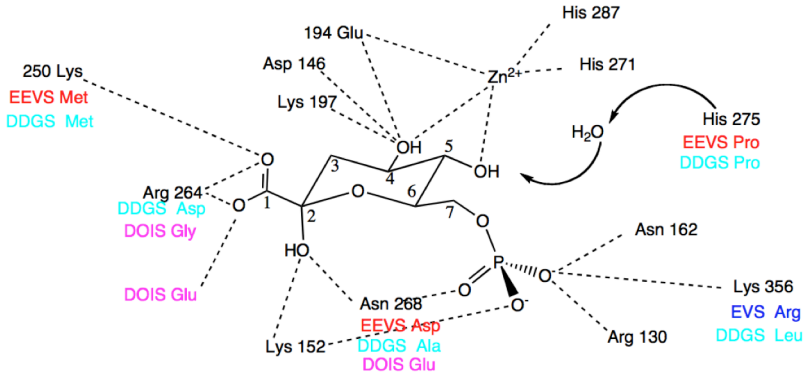


Figure 3.9. Active site residues of DHQS protein overlaid with putative residues involved in the substrate binding in SH7PC. Non-conserved residues determined by sequence alignment corresponding to SH7PCs or DOIS are listed below the DHQS numbered residue, mapped to the DHQS binding site and substrate DAHP

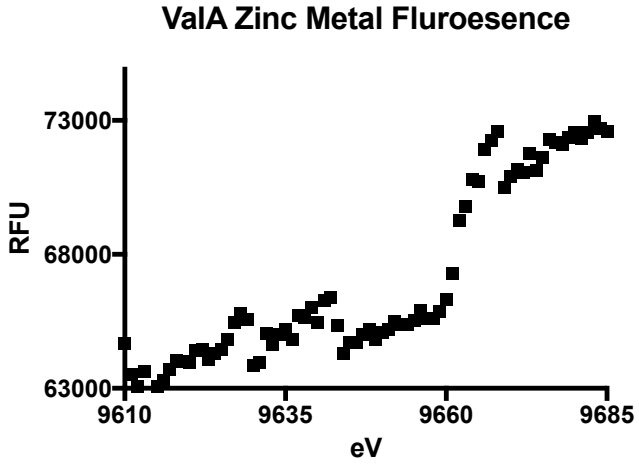


Figure 3.10. Fluorescence scan of the ValA crystal at the zinc absorption edge. X-ray wavelength at 1.2837 Å

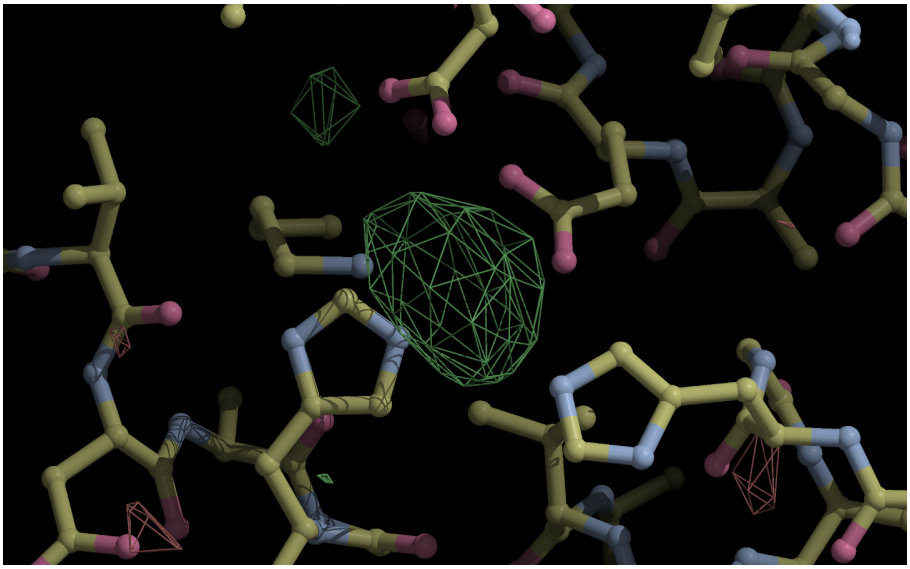


Figure 3.11. Initial structural solution to ValA. Inital structural solution by phenix and built by buster with an active site view and the difference map in green calculated from zinc anomalous diffraction data.

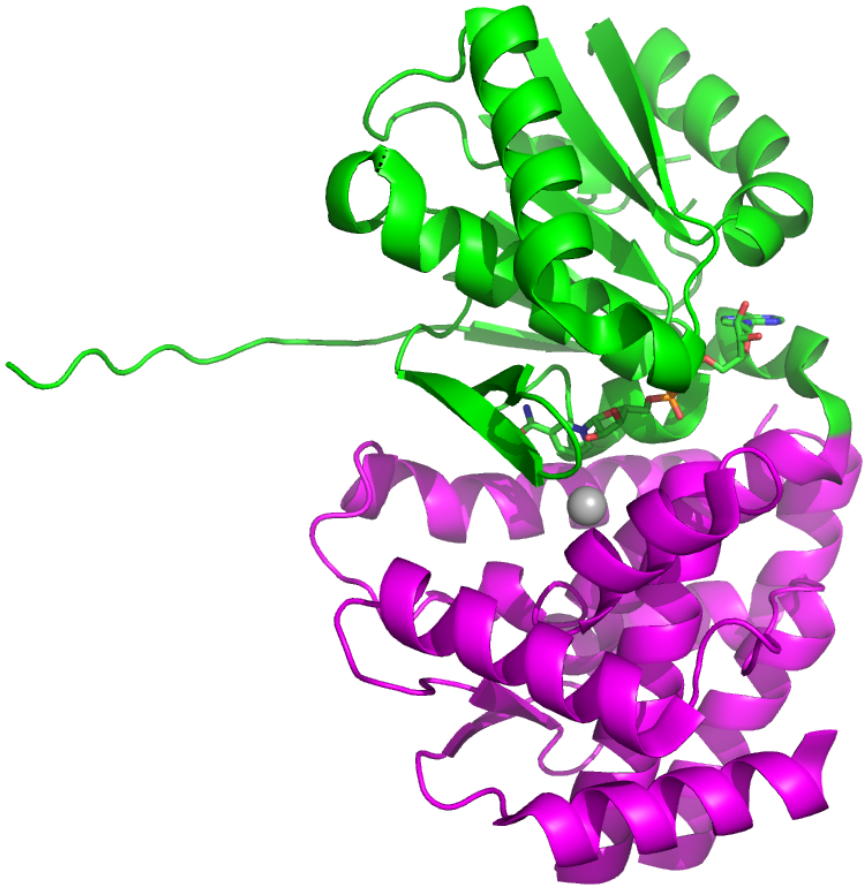
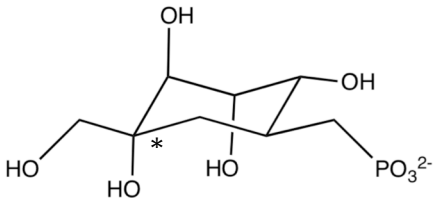


Figure 3.12. Final structural solution of ValA



((1*S*,2*R*,3*R*,4*S*,5*R*)-2,3,4,5-tetrahydroxy-5-(hydroxymethyl)cyclohexyl)methylphosphonate

Figure 3.13. Substrate analog of SH7PC

Chapter 4

Role of the α-domain in the Conformational Regulation of Merlin

By Sara J. Codding, Russell Carpenter, Donnie Berkholtz, Nicolas O. Thomas, Anthony Bretscher, P. Andrew Karplus,

Abstract

Neurofibromatosis 2 (NF2) is an inherited disease associated with mutations in the gene encoding the protein merlin. Merlin is thought to be a key mediator of contact inhibition of cell growth, and its essentiality in mice, drosophila and *C. elegans* proves its general importance in multicellular organisms. Merlin's mechanism of action is not known, but its similarities with ERM (ezrin-radixin-moesin) proteins has provided a conceptual basis for understanding its structure-function relationships. It is well-known that the tumor-suppressor activity of merlin is regulated by the intramolecular association of the N and C terminal parts of the protein, with modulation by phosphorylation at serine 518 (Ser518) blocking tumor-suppressor activity. Contrary to the current paradigm, recent biochemical studies using a novel engineered form of merlin show that it has varied degrees of intramolecular association in a rheostat like fashion, that the open form is responsible for tumor-suppressor activity, and that Ser518 phosphorylation favors the closed form. These developments underscore the crucial need for biochemical studies to gain better understanding of merlin regulation and function, in order to guide mechanistic insight and the development of interventions for NF2. Here we investigate the uncharacterized α-domain of merlin and its role in the rheostat like behavior of the merlin protein. This domain has been identified as an antiparallel coiled-coil by sequence alignment with other ERM proteins. We have designed, expressed, and purified a series of seven variants with a truncated α-domain. Each variant is truncated in the α-domain incrementally by heptad repeats observed in the anti-parallel coiled-coil. Through biochemical techniques we characterize this series of variants and their role in merlin's open-closed spectrum. We report that the α-domain contributes to merlin's rheostat and unfavorable interactions in the α-domain heptad repeats contribute to the degree of openness of the protein. In addition, these characterized variants may be of use to test the of the role α-domain in contact growth inhibition and could also serve as a tool to understand why the length of the α-domain is highly conserved across the ERM proteins. Thus, they may have the potential to reveal fundamentally new aspects of merlin function.

Introduction

Neurofibromatosis 2 (NF2) is an autosomal dominant familial cancer associated with development of tumors of the central nervous system (CNS). This disease affects 1 in 25,000 individuals and current NF2 treatment is limited to invasive surgery. The tumors associated with NF2 commonly involve the Schwann cells of the cranial nerves, and may lead to deafness, blindness, or impaired ambulatory function in effected individuals. Patients with NF2 are heterozygous at the *NF2* gene locus having inherited one defective copy; disease occurs when they lose function of the wild-type allele through somatic mutation and certain CNS cells exhibit uncontrolled growth. NF2 is associated with many different mutations at the *NF2* locus, which result in the loss of function of the encoded gene product, the tumor suppressor protein merlin (137).

Merlin is thought to be a key mediator of contact inhibition of cell growth and its essentiality in mice, drosophila and *C. elegans* proves its general importance in multicellular organisms (138-140). Merlin's mechanism of action is not known, but its sequence, structural, and functional similarity with ERM (ezrin-radixin-moesin) proteins has provided a conceptual basis for understanding its structure-function relationships (42,43). Like other ERM proteins, merlin is composed of an N-terminal FERM (4.1 ERM) domain (residues 1-335), a central α-helical domain (336-505), and a C-terminal tail domain (CTT, residues 506-595). It is well known that the tumor-suppressor activity of merlin is regulated by the intramolecular association of the N-terminal FERM domain and the C-terminal tail, with the 'closed-vs.-open' balance modulated by phosphorylation at serine 518 (Ser518) which blocks tumor-suppressor activity (141-143). Contrary to the dominant view of the last decade, recent work using a newly designed strongly-closed 'AR' merlin mutant has definitively shown that phosphorylation at Ser518 actually closes the merlin structure (47). Among many evidences presented, the merlin's N-terminal FERM domain binds to many ligands, and the ability to do so is modulated by the CTT association with the FERM domain. Thus it is clear that intramolecular regulation of

merlin is integral to its activity.

Merlin, like other ERM proteins, binds to the plasma membrane scaffolding protein EBP50 (144). The ability of merlin to associate with EBP50 varies, depending on the accessibility of the FERM domain (145). Recent studies have shown that of the CTT mutant merlin variants commonly used in cell culture, the phospho-deficient S518A, the phosphomimetic S518D, and the newly described highly closed 'AR' variant, the degree of merlin's openness varies as a function of the CTT composition. This suggests there are key interactions between the CCT and FERM domain that are likely highly regulatory for merlin's function. This suggests that merlin function is controlled in a rheostat like manner and the open-closed balance is likely modulated by various cellular factors (47,49). This new information highlights that biochemical studies on merlin regulation can greatly aid in our understanding of the protein.

With this new this information, we set out to characterize the role the central α-domain in regulating merlin's rheostat. The α-domain is a coiled-coil that folds back onto itself and links the FERM and CTT domains. This domain is highly conserved in both length and composition among the ERM proteins, and has been minimally characterized. In merlin, its importance in tumor suppression is evidenced by the presence of four NF2-linked and five cancer-linked missense mutations in this domain (137). Based on a homology model of human merlin, we have designed, expressed, and purified seven variants (Δ1- Δ7) that have increasing portions of the α-domain deleted.

By studying these deletion variants, we aim to determine how portions of α-domain contribute to the regulation of the FERM domain accessibility. Now, with the recent discovery that the open state of merlin is the one responsible for its growth inhibitory activity, the stage is set for careful characterization of the open-closed rheostat of merlin and to carry out the first studies to reveal the role of the central α-domain in this rheostat. Studies suggest that specific locations in this domain significantly modulate the closed-vs.-open balance, and we propose that in addition to structural roles it must have, interactions in the α-domain play a part in the rheostat control of merlin's open-closed equilibrium. Now characterized, these constructs can then become tools to assess its

broader role perhaps via *in vivo* studies.

Materials and Methods

DNA Constructs (merlin WT, merlin FERM, MPB, MBP-EBP50, merlin Δ1-Δ7). His-SUMO fusion constructs of full-length merlin, merlin FERM (residues 1-313), were generated by the Bretscher lab as described in (*Sher et al., 2012*) (47). Briefly, human merlin accession number NP000259 was PCR amplified out of a holding vector followed by restriction cloning into pE-SUMO (Life Sciences) vector to generate full-length merlin residues 1-595, and the merlin FERM construct consisting of residues 1-313. The seven α-domain truncation variants, Δ1-Δ7, where the α-domain of merlin was shortened in both the B-helix and C-helix, while retaining the hairpin between the two, and are summarized in Table 4.1. The cloning for the Δ1-Δ7 variants was performed by Genescript with the human gene NP000259 as the template. The seven resultant Δ variants were enzymatically restricted out of the supplied Genescript holding vector and ligated into pE-SUMO. The Δ7AR mutant was generated by site directed mutagenesis using Agilent Quikchange II kit via manufacturer's protocol to change the merlin residues A585 to W and R558 to L with pE-SUMO
Δ7 as the template. The maltose binding protein (MPB) and the MBP-fusion construct MBP-EBP50 (ezrin binding protein 50, last 39 amino acids) were supplied by Dr. Anthon y Bretcher and described in (*Garbett et al., 2010, (JCB)*) (146). All constructs were sequence verified by Oregon State University Center for Biocomputing.

Expression and purification of recombinant proteins. The pE-SUMO vector, which encodes an N-terminal 6-His-sumo-domain tag with a Ulp1 protease site was used to express merlin fusion constructs in *E. coli*. The purification of 6His-SUMO-tagged constructs was similar to that previously described (146). In brief, the merlin protein Δ variants were expressed in BL21 Rosetta™(DE3)pLysS cells (EMD Milipore) that were cultured in terrific broth with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol at 37 ° C and protein expression was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) after optical density (O.D.) reached A_{600}=1. The use of

BL21 Rosetta™(DE3)pLysS cells facilitated expression of the variants in *E. coli* as the human gene was not codon optimized prior to protein expression. The protein was expressed for 18 hours at 20 °C. All cell pellets were isolated by centrifugation at 4000 RPM at 4 °C for 20 minutes and then re-suspended in lysis buffer, 20 mM sodium phosphate pH 7.4 (RPI), 300 mM sodium chloride (RPI), 20 mM imidazole (Sigma), 1 mM TCEP tris (2-carboxyethyl) phosphine (Gold Bio), 10% Glycerol (JT Baker) in the presence of Roche Complete Mini protease inhibitor tablets per manufacturers protocol. The cells were lysed by sonication and the soluble fraction was isolated by centrifugation in a Beckman J2-21 centrifuge at 20,000 G at 4 °C for 45 minutes. The soluble lysate was then incubated with 1.5 % w/v streptomycin sulfate with light stirring at 4 °C for 1 hour. Streptomycin sulfate was removed by centrifugation at 10,000 G at 4 °C for 30 minutes. The soluble lysate was decanted and bound onto a column of pre-equilibrated Ni-NTA (ThermoFisher) beads in lysis buffer. The beads were washed with twenty column volumes of lysis buffer and the protein was then eluted with lysis buffer plus 200 mM imidazole. The pure fractions were checked via SDS-PAGE, pooled, and cleaved by addition of ~10 μL per milligram protein of N-terminally labeled 6His-Ulp1 = [1mg/mL] for 1-2.5 hours at 30 °C. The cleaved protein was extensively dialyzed into 20 mM phosphate buffered saline pH 7.5, 150 mM sodium chloride, using dialysis tubing 4,000-6,000 MWCO (Spectrum Labs) To the dialyzed protein Trition X-100 was added to 0.1 %, and the protein was bound to a second pre-equilibrated Ni-NTA column. The flow through was collected in fractions, checked for purity, and pooled as noted above. If necessary, the protein was concentrated utilizing 10 K MWCO centriprep spin columns by Milipore. The final protein concentration was determined by calculated extinction coefficient and UV absorbance at λ=280 by NanoDrop (ND 1000).

The p-GEX vector, which encodes an N-terminal MBP fusion tag was used to express constructs MBP and MBP-EBP50 in *E. coli* strain BL21(DE3) (Novagen) cells. The MBP constructs were cultured in Luria Bertani (LB) broth with ampicillin 100 μg/mL and 1mM IPTG was added after optical density reached A_{600}=0.6 and the culture was incubated for an additional 4 hours at 37 °C. MBP and MBP-tagged proteins were

purified as previously described, but for our purposes, left bound to amylose resin (NEB) and re-suspended as a 50 % resin slurry in phosphate buffered saline, (PBS), 20 mM Phosphate pH 7.4, 150 mM sodium chloride, with 0.1% TritonX-100 (146). Protein purity was checked by 10 % SDS-PAGE gel and final protein concentration of the 50 % slurry was determined by Qubit Fluorometer (Molecular Probes) protein assay per manufacturer's protocol.

In vitro binding assays. Binding assays were performed in a similar fashion to that previously described (47). A 50% slurry of amylose beads bound to MBP or MBP-EBP50 at a concentration of 0.7-0.9 mg/mL were pelleted at 700 G for 2 minutes and the supernatant removed by aspiration. The protein bound amylose resin, 50 mL, was washed twice with 1 mL of binding buffer, (PBS plus 0.1 % TritonX-100) and merlin constructs were added to amylose beads in a 3-4:1 molar ratio MBP:merlin variant. Each reaction was brought to a total of 1 mL by the addition of binding buffer. The beads were incubated for 1 hour at 4 °C with light agitation, then pelleted by centrifugation at 700 G and washed 4 times in binding buffer. The bound proteins were then eluted by boiling in 45μL of 2x Laemmli sample buffer for 2 minutes.

Gel electrophoresis and western blots. Protein eluted from the binding assays were separated by SDS-PAGE 10 % gel and wet tank transferred in Tobin buffer (25 mM Tris, 192 mM glycine) plus 10 % methanol to polyvinylidene fluoride (PVDF) membrane (Immobilon-FL, Millipore) with a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-rad). Blots were blocked with 10 % nonfat dry milk in Tris Buffered Saline (TBS) 20 mM Tris pH 7.5, 150 mM sodium chloride, for one hour at room temperature. The blot was then incubated with 1:400 rabbit polyclonal anti-merlin antibody (SC331 Santa Cruz Biotechnologies) in 3 % nonfat dry milk in Tris 50 mM pH 7.5, 150 mM NaCl plus 0.05 % tween (TBS-T) overnight at 4 °C. The blot was then washed 3 times in TBS-T followed by the addition of 1:20,000 IRDye 680-conjugated secondary antibody goat anti-rabbit (LI-CORE) in TBS-T plus 0.02 % SDS for 1 hour. The blot was washed twice with TBS-T, once with PBS, and imaged using an Odyssey infrared imaging system (LI-CORE Biosciences). Bound merlin constructs detected by immunobloting were

quantified with LI-CORE software with respect to input standards of 5 % or 10 % protein load. The average and standard deviations reported were from 3-5 independent experiments. The averages were plotted as a relative percent normalized to the positive FERM control (100 %). Statistical analysis was by performed as multiple ANOVA test, not all comparisons are plotted. Our negative control, MBP, showed no appreciable Δ variant binding under these conditions, (data not shown).

Crystallization trials of merlin Δ-variant. Merlin Δ1- Δ7 variants were expressed and purified as above, but dialyzed into a final buffer of 50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM TCEP, 10 % glycerol for crystallization trials. Each protein was concentrated to 5 mg/mL by centrifugation if necessary, and dispensed in microcrystallization trays (ARI) via the robotic crystal phoenix liquid handling system at 1:1, 2:1, and 4:1 protein:conditions in sitting drops. For these trials, we utilized the Emerald Wizard 1&2, Emerald Wizard 3&4, Hampton Index, and Hampton Crystal Screen. Each screen contained 96 crystallization cocktails each, for a total of 384 cocktails tested for each truncation variant. The crystallization trials were run in duplicate, one set of trials at room temperature and the other at 4 °C.

Homology model of merlin. The homology model of merlin was modeled using the *Sf*moesin structure and generated by Dr. Donnie Berkholz (49).

Sequence alignments and structural overlay. Alignments were generated using ClustalW (128) and the structural overlay was generated using Pymol by Schrödinger.

Results

Rational design of the Δ7AR mutant. Our objective was to generate a merlin construct that possessed a C-terminal tail that strongly interacted with the FERM domain. Thus, we decided to guide our approach with structural and biochemical data on merlin and ERM proteins, and their strong interaction with the C-terminus of EBP50. Our rationale was, that if the merlin CTT more closely resembled the EBP50 tail, the merlin

construct would have a considerably stronger N- to C- intramolecular interaction. This is based on evidence that the C-terminal tail of EBP50 binds much more tightly and at the same location as merlin's tail to its own FERM domain (47,147,148).

When the FERM domain structures depicting, ezrin, moesin, are overlaid with the radixin:EBP50 tail co-crystal structure, it becomes apparent that there exists a hydrophobic pocket in the FERM domain that buries both the W348 and K351 residues of the EBP50 tail (Figure 4.2 A) (49,149,150). In addition, it is clear that when the ERM tails are aligned with EBP50, the key structurally determined residues of EPB50 that interact with the FERM domain reside near or in that FERM hydrophobic binding pocket (Figure 4.2 B asterisks). But, there are two fundamentally different residues in EBP50 compared to the CTT of the ERM proteins; the EBP50 residues W385 and K351 (Figure 4.2 B orange box) (149). Thus, we designed the merlin Δ7AR construct to more closely resemble the EBP50 C-terminal tail by mutating the merlin residues A585 to W and R558 to L.

Rational design of the Δ variants. Each of the Δ variants were rationalized and structurally guided by the *Sf*mosesin crystal structure. The central α-domain of *Sf*moesin is an antiparallel coiled-coil that is composed of two helices, the B- and C- helices, that are connected by a hairpin turn (Figure 4.3A). Typically, anti-parallel coiled-coils are composed heptad repeats that are denoted **a**bc**d**efg, where the interacting surface between the two coils occurs at the **a** and **d** positions in the repeat (151). To determine if this structural characteristic of coil-coiled motifs was present in other ERM proteins, their sequences were aligned as guided by the structure of *Sf*moesin, in an antiparallel fashion as to depict the self-association of the α-domain (Figure 4.3 B) (49). As is apparent from this alignment, the ERM proteins share the antiparallel coiled-coil heptad repeat motif, and thus we used this as a guide to begin truncating the α-domain of merlin. We designed each merlin variant, Δ1- Δ7, to retain the hairpin turn region between the two helices, but the B-helix and C-helix of the α-domain were each shortened by a heptad repeat from the apex of the coiled-coil where the hairpin turn resides (Δ1 with one heptad repeat deletion from each coil and Δ7 with seven deleted from each coil) (Figure, 4.3,

Table 4.1), to maintain the integrity of the domain while systematically decreasing its size. The heptad deletions are depicted pictorially as shades of blue in the homology model of merlin (Figure 4.4).

Crystallization trials of the Δ1- Δ7 merlin variant. As each variant, Δ1- Δ7, got shorter with each heptad pair deletion, the proteins approached a more globular fold (Figure 4.4). Thus, we anticipated that the probability of merlin to pack well into an ordered crystalline lattice improved as the construct became smaller. In addition, as the α-domain shortened, the probability of the rheostat adopting multiple conformations decreased. This also improved the likelihood that the construct was less dynamic to promote a well-ordered crystal growth. But, our rational design of the merlin variants retained the majority of the α-domain that is thought to interact with the FERM domain burying much of the FERM surface.

Currently, the only solved structure of merlin is the FERM domain (152-154). Thus, these variants become a novel structural pursuit in that additional structural information would be gained on the α-domain:FERM interaction even if the only suitable domain for crystallization contained the largest deletion, variant Δ7. Each Δ1- Δ7 variant expressed and purified well for crystallization trials (Figure 4.5A). We hypothesized that of the Δ1- Δ7 variants, the smallest, Δ7, would be most likely to crystallize for reasons noted above. The Δ7 construct did produce a hit that was optimized, but only grew as needle like crystals that unfortunately diffracted to low resolution, 7.0 Å (Figure 4.5B, C). Additionally, crystal trials for the remaining Δ1- Δ6 variants produced no leads under our conditions. We thus focused our effort on biochemical characterization of the Δ variants.

Biochemical characterization of the Δ1- Δ7 variants imply specific locations in the α-domain significantly modulate the closed-vs.-open balance. We set out to test each variant for their degree of openness. Since the intramolecular interaction between the FERM:C-terminal tail masks the EBP50 binding site, an interaction between EBP50 and our merlin variants served as an assay for the degree of merlin's openness. For these

experiments we used a published protocol in which the EBP50 tail was immobilized on amylose beads as a maltose binding fusion protein and merlin constructs were added to the beads (155). After equilibration and washing, the merlin that had bound to the beads was eluted and quantified by western blot referenced to an internal merlin input standard. A representative blot is depicted in Figure 4.6.

At physiological salt (150 mM NaCl) concentrations, it is evident that the α-domain plays a role in merlin's rheostat. Key (presumably unfavorable) interactions of the coiled-coil heptad repeats that favor the open state appear to be lost in Δ1- Δ3, which is then regained to a similar degree as wild type in Δ 4 (Figure 4.7). In addition, the open state regained in Δ4, is once again lost in Δ 5-Δ6, which returns the variants to a more closed state as in the Δ1-Δ3. But, Δ7 represents an ultra-closed form of merlin, where in the construct is 50-fold more closed than wild type. This result is recapitulated in the Δ7AR mutant suggesting that Δ7 results in the loss of unfavorable interactions so great that the addition of the AR mutation in the C-terminal tail does not significantly change the degree of which Δ7 is closed. This is unlike the result observed when comparing the wild type and full length AR variant of merlin, where the AR mutation shows a 50% decrease in the degree of openness compared to full length (155). These results indicate that the α-domain plays an active role in regulating the merlin rheostat.

Discussion

Despite nearly two decades of research since the discovery that NF2 is caused by the loss of normal merlin function, our understanding of how merlin carries out its cell growth control is still quite limited. Part of the reason for the slow progress in NF2 research is that merlin function is complex and the intensive efforts focused on genetic and *in vivo* approaches have not had been complemented by sufficient biochemical studies that can make the genetic approaches more effective. In addition, while the NF2 community is actively seeking novel compounds to intervene with disease progression**,** the efforts in development of effective treatments are limited by our current knowledge of merlin structure, function, and regulation. In fact, most ideas about merlin's structure,

and its structure-function relations, are the results of inferences derived from the crystal structures and the biochemical studies of ezrin-radixin-moesin (ERM) proteins (46,147). For instance, the crystal structure of the moesin and radixin FERM domains with the C-terminal residues of EBP50 revealed the binding site of EBP50 on the FERM domain that is was then later shown to be relevant for merlin (148,149). Although many such inferences are accurate, it was one such inference made from the ERM proteins combined with a dearth of direct biochemical studies of merlin that allowed the idea to develop that merlin phosphorylation at Ser518 favors the open conformation and that the closed conformation is thus active in growth control (142). This model has now been conclusively shown to be incorrect (47). Also, a few years ago a full-length structure of moesin from Sf9 cells, revealed that the central α-helical domain interacts extensively with the FERM domain (burying ~3800 $Å^2$ surface area), leading to the proposal that the central α-helical domain can also mask binding sites and influence the open-closed equilibrium of ERM proteins and of merlin (49).

Here, it is clear that a precise interplay of molecular interactions indicates that merlin is not a binary switch. Our results show that the α-domain does in fact play a role in regulation of merlin's conformation.

The contribution of the α-domain in determining the level at which merlin's rheostat is dialed towards open or closed is not a direct function of its length. The unfavorable interactions lost in Δ4 open the protein to the same extent as wild type merlin and both the heptads on either side of Δ4, promote a more closed form of the protein. But, when seven heptads of the coiled-coil portion of the α-domain is deleted (Δ7 variant), which predominantly leaves only the portion of the α-domain that interacts with the FERM domain, the protein is closed 98% of the time, in so much that Δ7AR variant could not close the domain to any more appreciable degree. This result is striking when compared to the differences between wild type merlin and 'full-length merlin AR' where the degree of openness falls from 50% to 25% with respect to the AR mutations. With this collective evidence, we propose that there is an interplay of unfavorable vs. favorable

interactions in the α-domain that dial merlin's rheostat and may represent energetic barriers to full extension of the domain.

Future Work

Although many merlin-binding proteins have been identified, it is not clear for most of these proteins whether their interaction with merlin has functional relevance, if they interact with merlin in the open or closed state, and if their interaction with merlin promotes cell proliferation or cell growth inhibition (156). Also, merlin is thought to participate in many cellular processes (reviewed in McClatchey et al. 2005, (140)), almost nothing is known about how the specific cellular contexts in which merlin acts influence its conformational state. Furthermore, the role of the α-domain in merlin's erlin'sin's has yet to be tested. The Δ variants described here now biochemically characterized, have the potential to flesh out the α-domain's role in cell growth inhibition in cell culture or other model systems.

Additional quantitative biochemical and biophysical studies on this series of variants could provide useful to understand the energetic barriers to full extension of merlin. Carefully designed studies could help determine if this is a function of the length of the domain and or the composition of heptads in the α-domain. Finally, testing the ability of α-domain disease-linked missense mutations to alter merlin's rheostat or energetic barrier to full extension would be interesting. Such studies could aid in understanding the roles on merlin's conformational states in NF2 disease. This, combined with experiments suggested above could aid our understanding of how merlin's role in contact growth inhibition is a function of its rheostat.

Table 4.1. Merlin constructs. WT denoting full length wild type protein

Merlin Δ construct	Retained Residues N-terminal	First Deletion B-coil of α-domain	Hairpin	Second Deletion C-coil of α -domain	Retained Residues C-terminal	Total Residues	Total Deleted Residues	MW
FERM	1-313	-	-	-	-	314	281	36689.6
WT	1-595	-	-	-	-	595	0	69690.1
Δ1	1-406	407-413	414-427	428-431	429-595	581	14	67978.1
Δ2	1-399	400-413	414-427	428-441	442-595	567	28	66523.4
Δ3	1-392	393-413	414-427	428-438	439-595	553	42	64910.6
Δ4	1-385	386-413	414-427	428-455	456-595	539	56	63297.8
Δ5	1-378	379-413	414-427	428-462	463-595	525	70	61743.1
Δ6	1-371	372-413	414-427	428-470	471-595	511	84	60085.3
Δ7	1-364	365-413	414-427	428-476	477-595	497	98	58542.6
Δ7AR	1-364	365-413	414-427	428-476	477-595	497	98	58629.6

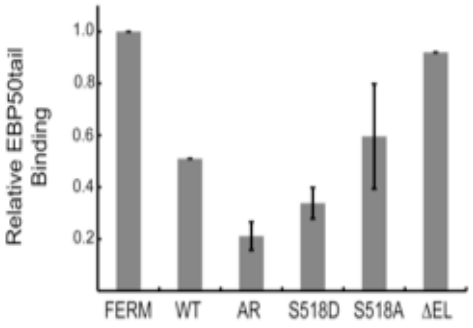
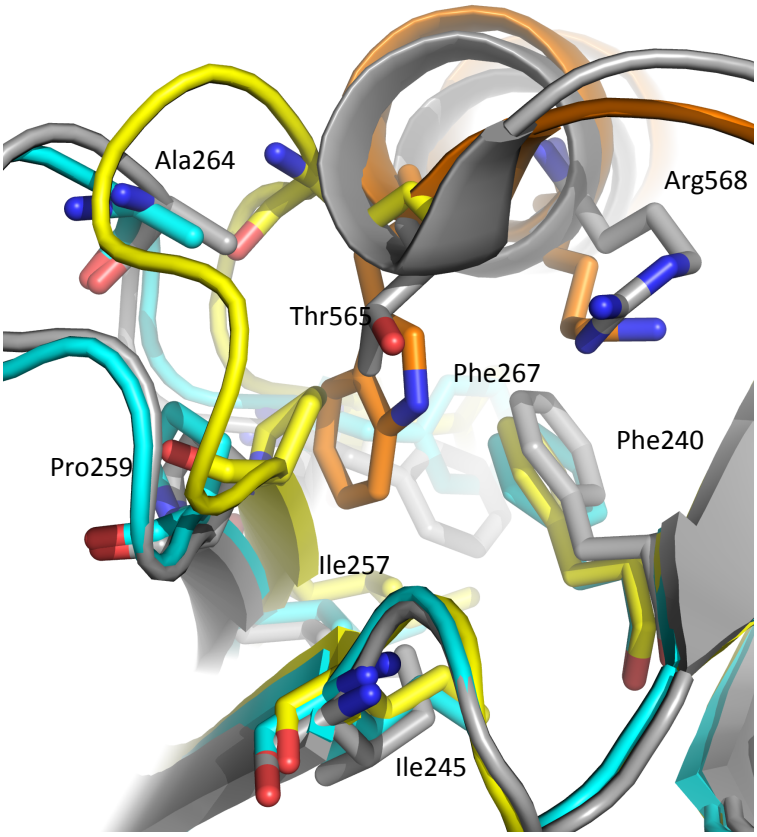


Figure 4.1. CTT merlin variants display a varied degree of openness. Recombinant merlin assayed via EBP50 pull down and percent bound normalized to the FERM domain.

A.

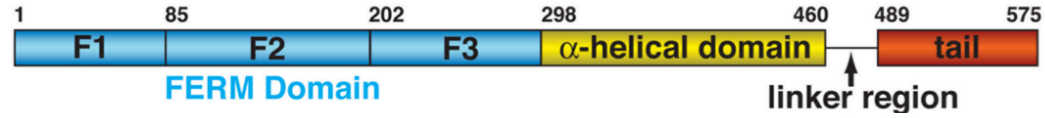


B.

```
ezrin          RDKYKTLRQIRQGNTKQRIDEFEAL 586
radxin         RDKYKTLRQIRQGNTKQRIDEFEAM 583
moesin         RDKYKTLRQIRQGNTKQRIDEFESM 577
merlin         SSKHNTIKKLTLQSAKSRVAFFEEL 595
EBP50          SS-----KRAPQMDWSKKNELFSNL 358
                             *  *   *  *
```

Figure 4.2. A. Overlay of structurally determined FERM domains. Proteins are depicted a ribbons, ezrin in yellow pdb 1NI2, radxin:EBP50 pdb code 2D10 in cyan and orange respectively, and moesin pdb code 2I1K in grey. Residues are numbered for the FERM domains, which share residue numbering, and are highlight the FERM domain hydrophobic pocket that buries the EBP50 c-terminal tryptophan (orange, sticks) and the lysine (orange, sticks). B. Sequence alignment of the c-terminal tails of human proteins ezrin, radxin, moesin, merlin, and EBP50. The residues of merlin that have been mutated A, R, are highlighted in a orange box.

A.



B.

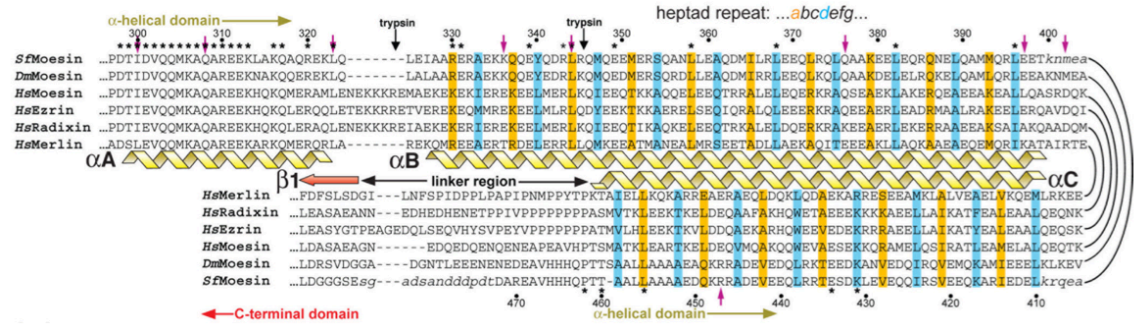


Figure 4.3. Domain topology of *Sf*moesin and sequence alignment of the ERM-merlin -helical domain depicted as a coiled-coil. (a) Domain topology of Sf*moesin*. Residue numbers at the domain boundaries are indicated. (b) Alignment of ERM-Merlin α-helical domains. Alignment is folded back onto itself such that the sequence for coil C runs right to left to indicate its coiled-coil heptad interaction in register with the B helix. Helical regions are indicated by a yellow coil, and the 1 strand of the C-terminal tail by a red arrow. The *a* and *d* positions of the coiled-coil heptad repeat are shown with orange and cyan backgrounds respectively. These positions interact with the other helix of the coil, as shown. Positions in human merlin associated with cancer (purple arrows) are indicated. Sequences used are as follows: human merlin (*Hs*Merlin), SwissProt accession no. **P35240**; human radixin (*Hs*Radixin), **P35241**; human ezrin (*Hs*Ezrin), **P15311**; human moesin (*Hs*Moesin), **P26038**; and *D. melanogaster* moesin (*Dm*Moesin), GenBank accession no. **NP_996392**.

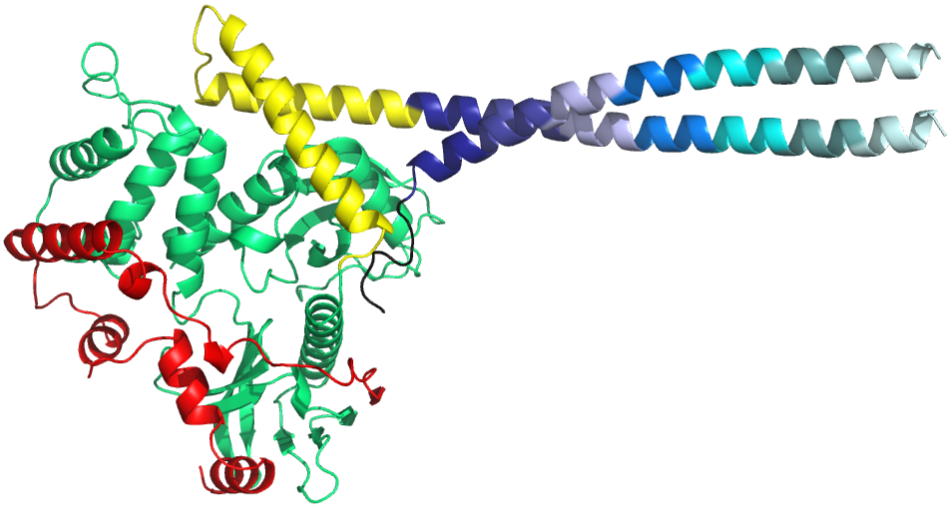
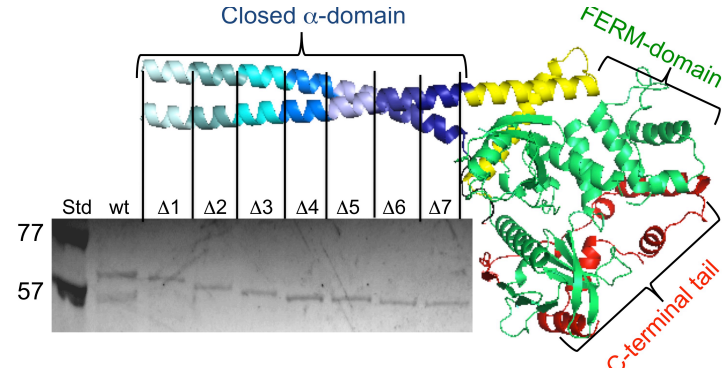


Figure 4.4. Merlin homology model generated from *Sf*moesin structure as a starting point. The N-terminal FERM domain is depicted in green, the
aoesin str as a starting point. The N-terminal FERM domain is depicted in green, the , th e ther helix of the coe blue color pairing on both the B- coil and C-coil correspond to dele ted □paired□ heptad repeats that were incrementally deleted to generate the Δ1-7 constructs. The C-terminal tail is depicted in red.

A.



B.

C.

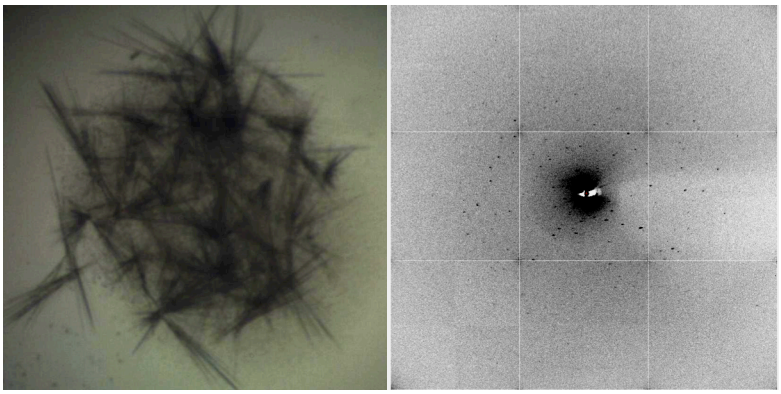


Figure 4.5. Purified Δ variants utilized in this study. A. SDS-PAGE gel showing purity of each Δ1-7 variant as purified in this study. Merlin homology model is overlaid onto the gel with the deleted portion of the a-7 variant as purified in this sΔ variant above the lane in various shades of blue. B. Needle bundles of Δ variant Δ7. C. The diffraction pattern of a Δ7 needle.

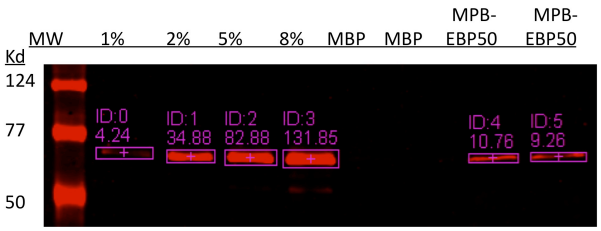


Figure 4.6. Representative blot of pull-down experiment. Δ5 merlin variant MBP-EBP50 pull down assay detected by western blot. Molecular weight standards (MW) and percent input (%) of the total protein subjected to pull down assay, control lane MBP, and binding assay in duplicate MBP-EBP50. The bolt was imaged and quantified on the LI-COR-Odyssey Imaging System.

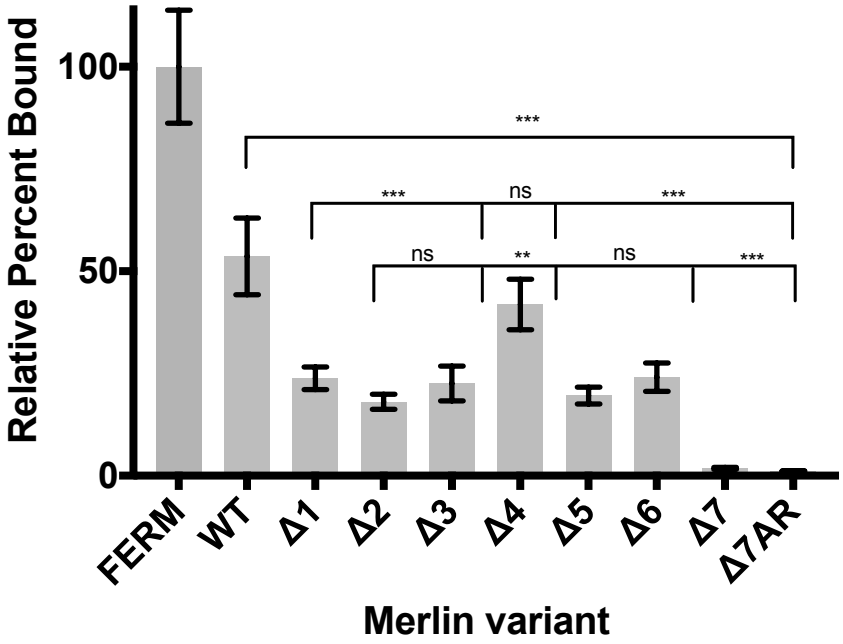


Figure 4.7. Relative percent bound of merlin Δ variants to EBP50. Merlin variants were assayed for their degree of openness by relative binding to immobilized EBP50. Bars represent the percent bound as normalized to the FERM domain. The average values were plotted and statistically analyzed by performing a multiple ANOVA test. Statistical values are denoted ** for P less than or equal to 0.01, and *** for P less than or equal to 0.001, ns=not significant. Top bracket are P values compared to the FERM domain, middle bracketed are P values compared to wild type merlin, and the bottom bracket are P values compared to Δ1.

Chapter 5

Conclusions Drawn From Biochemical and Biophysical Analyses of Proteins

Sara J. Codding

Testing Biochemical and Biophysical Properties

The usefulness of *in vitro* assays. As I developed as a scientist to the point of this dissertation, I have come to appreciate the distilled approach of carefully designed *in vitro* assays. When well crafted and with proper controls, working with proteins in a simplified *in vitro* environment can answer direct fundamental questions that elude cellular and organismal studies. This was the case for both dysferlin and merlin proteins mentioned in this thesis. It is however, in my experience quite beneficial to approach the study of a protein involved in physiological relevant processes to include both *in vitro* and *in vivo* approaches. The impacts of an *in vitro* study may be bolstered with complimentary *in vivo* work. The converse is also true, and complex processes that occur in vivo may be too difficult to dissect with available tools.

The benefits of protein structure determination. Protein structure determination is useful to illuminate a snap shot of dynamic protein structure. That proteins typically adopt one or very few fully folded structures, biophysical approaches to elucidate protein structure have been useful in my research. Specifically, when analyzing newly identified members of a protein family that share a similar biochemical function but are highly divergent, protein structure can illuminate aspects that are hypothesized by sequence analysis. However, I have found that conclusions drawn from structural studies are solidified by complimentary biochemical studies. Future work will incorporate both into my research.

The power of protein design. Protein design utilized in this dissertation was to generate constructs for crystallography and subsequently, the role of a protein domain in protein autoregulation. This field is currently burgeoning and protein tools for studies on dose dependent Ras signaling to proteins that measure membrane voltage. As researcher become better at predicting protein structure and function relationships, applications of protein design will continue to grow in scientific research. The power of protein design applications lies in the nuance. For example, Ras signaling is typically studied in cells with brute force overexpression. Development of a dose dependent activatable Ras has revealed different cellular responses that correlate with dose but not with cell type (157).

In addition, the engineering of optically active protein sensors allow for live cell imaging and measurements in tissue that would have otherwise been resected, fixed, or in the case of membrane voltage, required expensive apparatus and specialized training (158).

Conclusions Drawn From This Work and Other Research on Dysferlin

On the function of dysferlin. While it is clear that dysferlin can act as a SNARE effector to facilitate membrane fusion in a calcium sensitive manner *in vitro,* dysferlin may play a broader role in skeletal muscle (158). In fact, Many dysferlin mutations are linked to MD, some lead to premature stop codons, others lead to fully translated proteins that localize properly but are functionally deficient for unknown reasons. How different mutations in the dysferlin gene can affect distinct muscle groups remains to be elucidated. Interestingly, phenotypic heterogeneity has been observed in familial lineages where identical mutations appear in affected individuals as either LGMD or distal myopathy, but it must be pointed out that phenotypes are consistent among sibships, suggests that other effector genes are key in phenotype distinction (159).

Dysferlin deficiency appears to alter the immune response in damaged muscle, and likely exacerbates the failure of the muscle regenerative process (85). Monocytes deficient in dysferlin show delayed chemotaxis and decreased adhesion, but when matured to macrophages in tissues show increased motility and enhanced phagocytosis. However, it remains to be shown if this result is due to a loss of signal from damaged muscle or if it is due to attenuated receipt of signal in immune cells (60,160). This when combined with the evidence that dysferlin-null mice fail to reseal the muscle cell membrane post injury, develop dilated cardiomyopathy, show deficiencies in angiogenesis, decreased acetylcholine uptake at the neuromuscular junction, and that dysferlin has been shown to associate with the transverse tubules to support DHPR calcium signaling and calcium homeostasis suggests that dysferlin might play a large role in cellular processes involving cell signaling, membrane integrity, fusion, and remodeling (65).

Research supports a role for dysferlin as the calcium sensor for repair of the sarcolemma (17,20,57,161). But, membrane repair in muscle is a multifaceted event and to further compound dysferlin's role in the repair process, it has also been shown to associate with many other proteins involved in membrane patching - including annexins A1 and A2, calveolin-3, calpain-3, MG53, AHNAK, and affixin (26,58,61,64,86,162). When taken together, it raises the question as to where or when in the concerted repair process dysferlin contributes to patching as it is supported to recruit proteins upstream of wound patching (25). If dysferlin can both simultaneously recruit and facilitate fusion remains to be determined. But, that dysferlin acts in other calcium sensitive fusion events such as acid sphingomyelinase secretion and lysosome exocytosis, lends evidence for dysferlin's wider role in membrane trafficking as a calcium sensitive SNARE effector fusion in events (27).

Recently, it has been shown that calpain cleaves dysferlin upon membrane injury, yielding a small 72 kDa C-terminal isoform termed mini-dysferlin$_{C72}$ that was enriched at the site of injury (163). This truncated, but functional mini-dysferlin$_{C72}$ has been suggested as a model upon which to base a therapeutic construct, as the large size of the full-length protein prohibits its utility in gene delivery therapies (87). But it is of note that the cleavage susceptible isoform is not present in abundance in skeletal muscle tissues. In rescue studies mini-dysferlin$_{C72}$ has restored membrane repair but not the MD phenotype. Suggesting that the many domains of dysferlin are not functionally redundant and that dysferlin's role in repair and trafficking is more complex than membrane patching.

On dysferlin's role in other pathways. Aside from its well-studied function in muscle membrane repair, dysferlin has also been implicated in several other disparate pathways. For example, dysferlin may play a part in muscle differentiation and development, as dysferlin expression is linked to mRNA levels of myogenin in cultured myoblasts, a transcription factor important in this process (164). Also, loss of dysferlin has been shown cause misstrafficking of insulin-like growth factor receptor (IGFR) and results in increased endocytosis of IGFR and smaller cultured mytotubes (165). In

endothelial cells, loss of dysferlin resulted in misstrafficking and degradation of platelet endothelial cellular adhesion molecule-1 causing decreased cell adhesion essential for angiogenesis (166). In addition, mutant dysferlin is rapidly endocytosed via a clathrin independent pathway with syntaxin 4 and Rab 5 that is largely influenced by caveolin, and appears to be sensitive to the C2B-FerI-C2C fold (67,96). In sea star oocytes, loss of dysferlin shows marked reduction of endocytosis, however the molecular mechanism is unclear and may be associated with partial knock down via morphlino and loss of any potential compensating proteins through potential upregulation (167).

Due to the expression profile of dysferlin and the numerous mutations in the gene that lead to dysferlinopathies, the extent of clinical disease incidence in other tissues is not well documented and is exacerbated by the search for a well defined role of the protein in healthy tissues. Still, *in vivo* and organismal studies have implicated dysferlin in vesicular trafficking in many cell types. Clear and well designed experiments will aid in our future understanding of dysferlin's role in non-damaged or diseased tissues.

Bibliography

1. Kyte, J. (1995) *Structure in protein chemistry*, Garland Pub., New York

2. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-99

3. Doolittle, R. F. (2010) The roots of bioinformatics in protein evolution. *PLoS Comput Biol* **6**, e1000875

4. Mount, D. W. (2001) *Bioinformatics : sequence and genome analysis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

5. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549

6. Hutchison, C. A., 3rd, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **253**, 6551-6560

7. Zoller, M. J., and Smith, M. (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res* **10**, 6487-6500

8. Stark, G. R. (1970) Recent developments in chemical modification and sequential degradation of proteins. *Adv Protein Chem* **24**, 261-308

9. Lehrer, S. S. (1971) Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry* **10**, 3254-3263

10. Clegg, R. M. (2002) FRET tells us about proximities, distances, orientations and dynamic properties. *J Biotechnol* **82**, 177-179

11. Gadella, T. W. J. (2009) *FRET and FLIM techniques*, 1st ed., Elsevier, Amsterdam ; Boston

12. Cooper, S. T., and McNeil, P. L. (2015) Membrane Repair: Mechanisms and Pathophysiology. *Physiol Rev* **95**, 1205-1240

13. Kalogeris, T., Baines, C. P., Krenz, M., and Korthuis, R. J. (2012) Cell biology of ischemia/reperfusion injury. *International review of cell and molecular biology* **298**, 229-317

14. Friden, J., Sjostrom, M., and Ekblom, B. (1983) Myofibrillar damage following intense eccentric exercise in man. *International journal of sports medicine* **4**, 170-176

15. Steinhardt, R. A., Bi, G., and Alderton, J. M. (1994) Cell membrane resealing by a vesicular mechanism similar to neurotransmitter release. *Science* **263**, 390-393

16. Bi, G. Q., Alderton, J. M., and Steinhardt, R. A. (1995) Calcium-regulated exocytosis is required for cell membrane resealing. *The Journal of cell biology* **131**, 1747-1758

17. Bansal, D., Miyake, K., Vogel, S. S., Groh, S., Chen, C. C., Williamson, R., McNeil, P. L., and Campbell, K. P. (2003) Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature* **423**, 168-172

18. Bashir, R., Britton, S., Strachan, T., Keers, S., Vafiadaki, E., Lako, M., Richard, I., Marchand, S., Bourg, N., Argov, Z., Sadeh, M., Mahjneh, I., Marconi, G., Passos-Bueno, M. R., Moreira Ede, S., Zatz, M., Beckmann, J. S., and Bushby, K. (1998) A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nature genetics* **20**, 37-42

19. Anderson, L. V., Davison, K., Moss, J. A., Young, C., Cullen, M. J., Walsh, J., Johnson, M. A., Bashir, R., Britton, S., Keers, S., Argov, Z., Mahjneh, I., Fougerousse, F., Beckmann, J. S., and Bushby, K. M. (1999) Dysferlin is a plasma membrane protein and is expressed early in human development. *Human molecular genetics* **8**, 855-861

20. McDade, J. R., and Michele, D. E. (2013) Membrane damage-induced vesicle-vesicle fusion of dysferlin-containing vesicles in muscle cells requires microtubules and kinesin. *Human molecular genetics*

21. Earles, C. A., Bai, J., Wang, P., and Chapman, E. R. (2001) The tandem C2 domains of synaptotagmin contain redundant Ca2+ binding sites that cooperate to engage t-SNAREs and trigger exocytosis. *The Journal of cell biology* **154**, 1117-1123

22. Ernst, J. A., and Brunger, A. T. (2003) High resolution structure, stability, and synaptotagmin binding of a truncated neuronal SNARE complex. *The Journal of biological chemistry* **278**, 8630-8636

23. Rizo, J., and Sudhof, T. C. (1998) C2-domains, structure and function of a universal Ca2+-binding domain. *The Journal of biological chemistry* **273**, 15879-15882

24. McNeil, P. L., and Steinhardt, R. A. (2003) Plasma membrane disruption: repair, prevention, adaptation. *Annual review of cell and developmental biology* **19**, 697-731

25. Han, R., and Campbell, K. P. (2007) Dysferlin and muscle membrane repair. *Current opinion in cell biology* **19**, 409-416

26. Matsuda, C., Miyake, K., Kameyama, K., Keduka, E., Takeshima, H., Imamura, T., Araki, N., Nishino, I., and Hayashi, Y. (2012) The C2A domain in dysferlin is important for association with MG53 (TRIM72). *PLoS currents* **4**, e5035add5038caff5034

27. Defour, A., Van der Meulen, J. H., Bhat, R., Bigot, A., Bashir, R., Nagaraju, K., and Jaiswal, J. K. (2014) Dysferlin regulates cell membrane repair by facilitating injury-

triggered acid sphingomyelinase secretion. *Cell Death Dis* **5**, e1306

28. Jahn, R., and Scheller, R. H. (2006) SNAREs--engines for membrane fusion. *Nature reviews. Molecular cell biology* **7**, 631-643

29. Ma, C., Li, W., Xu, Y., and Rizo, J. (2011) Munc13 mediates the transition from the closed syntaxin-Munc18 complex to the SNARE complex. *Nature structural & molecular biology* **18**, 542-549

30. Abdullah, N., Padmanarayana, M., Marty, N. J., and Johnson, C. P. (2014) Quantitation of the calcium and membrane binding properties of the c2 domains of dysferlin. *Biophysical journal* **106**, 382-389

31. Johnson, C. P., and Chapman, E. R. (2010) Otoferlin is a calcium sensor that directly regulates SNARE-mediated membrane fusion. *The Journal of cell biology* **191**, 187-197

32. Demain, A. L., and Fang, A. (2000) The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* **69**, 1-39

33. Bai, L., Li, L., Xu, H., Minagawa, K., Yu, Y., Zhang, Y., Zhou, X., Floss, H. G., Mahmud, T., and Deng, Z. (2006) Functional analysis of the validamycin biosynthetic gene cluster and engineered production of validoxylamine A. *Chem Biol* **13**, 387-397

34. Wu, X., Flatt, P. M., Schlorke, O., Zeeck, A., Dairi, T., and Mahmud, T. (2007) A comparative analysis of the sugar phosphate cyclase superfamily involved in primary and secondary metabolism. *Chembiochem* **8**, 239-248

35. Asamizu, S., Xie, P., Brumsted, C. J., Flatt, P. M., and Mahmud, T. Evolutionary divergence of sedoheptulose 7-phosphate cyclases leads to several distinct cyclic products. *J Am Chem Soc* **134**, 12219-12229

36. Gusella, J. F., Ramesh, V., MacCollin, M., and Jacoby, L. B. (1999) Merlin: the neurofibromatosis 2 tumor suppressor. *Biochim Biophys Acta* **1423**, M29-36

37. Fehon, R. G., Oren, T., LaJeunesse, D. R., Melby, T. E., and McCartney, B. M. (1997) Isolation of mutations in the Drosophila homologues of the human Neurofibromatosis 2 and yeast CDC42 genes using a simple and efficient reverse-genetic method. *Genetics* **146**, 245-252

38. McClatchey, A. I., Saotome, I., Ramesh, V., Gusella, J. F., and Jacks, T. (1997) The Nf2 tumor suppressor gene product is essential for extraembryonic development immediately prior to gastrulation. *Genes Dev* **11**, 1253-1265

39. Claudio, J. O., Lutchman, M., and Rouleau, G. A. (1995) Widespread but cell type-specific expression of the mouse neurofibromatosis type 2 gene. *Neuroreport* **6**, 1942-1946

40. Huynh, D. P., Tran, T. M., Nechiporuk, T., and Pulst, S. M. (1996) Expression of neurofibromatosis 2 transcript and gene product during mouse fetal development. *Cell Growth Differ* **7**, 1551-1561

41. Gonzalez-Agosti, C., Xu, L., Pinney, D., Beauchamp, R., Hobbs, W., Gusella, J., and Ramesh, V. (1996) The merlin tumor suppressor localizes preferentially in membrane ruffles. *Oncogene* **13**, 1239-1247

42. Bretscher, A., Chambers, D., Nguyen, R., and Reczek, D. (2000) ERM-Merlin and EBP50 protein families in plasma membrane organization and function. *Annu Rev Cell Dev Biol* **16**, 113-143

43. Gary, R., and Bretscher, A. (1995) Ezrin self-association involves binding of an N-terminal domain to a normally masked C-terminal domain that includes the F-actin binding site. *Mol Biol Cell* **6**, 1061-1075

44. Takahashi, K., Sasaki, T., Mammoto, A., Takaishi, K., Kameyama, T., Tsukita, S., and Takai, Y. (1997) Direct interaction of the Rho GDP dissociation inhibitor with ezrin/radixin/moesin initiates the activation of the Rho small G protein. *J Biol Chem* **272**, 23371-23375

45. Martin, M., Andreoli, C., Sahuquet, A., Montcourrier, P., Algrain, M., and Mangeat, P. (1995) Ezrin NH2-terminal domain inhibits the cell extension activity of the COOH-terminal domain. *J Cell Biol* **128**, 1081-1093

46. Pearson, M. A., Reczek, D., Bretscher, A., and Karplus, P. A. (2000) Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. *Cell* **101**, 259-270

47. Sher, I., Hanemann, C.O., Karplus, P.A., Bretscher, A.P. (2012-in press) The tumor suppressor merlin controls growth in its open state and is converted by phosphorylation to less active more closed state. *Dev Cell*

48. Morrison, H., Sherman, L. S., Legg, J., Banine, F., Isacke, C., Haipek, C. A., Gutmann, D. H., Ponta, H., and Herrlich, P. (2001) The NF2 tumor suppressor gene product, merlin, mediates contact inhibition of growth through interactions with CD44. *Genes Dev* **15**, 968-980

49. Li, Q., Nance, M. R., Kulikauskas, R., Nyberg, K., Fehon, R., Karplus, P. A., Bretscher, A., and Tesmer, J. J. (2007) Self-masking in an intact ERM-merlin protein: an active role for the central alpha-helical domain. *J Mol Biol* **365**, 1446-1459

50. Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., Angelini, C., Serrano, C., Urtizberea, J. A., Hentati, F., Hamida, M. B., Bohlega, S., Culper, E. J., Amato, A. A., Bossie, K., Oeltjen, J., Bejaoui, K., McKenna-Yasek, D., Hosler, B. A., Schurr, E., Arahata, K., de Jong, P. J., and Brown, R. H., Jr. (1998) Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nature genetics* **20**, 31-36

51. Bittner, R. E., Anderson, L. V., Burkhardt, E., Bashir, R., Vafiadaki, E., Ivanova, S., Raffelsberger, T., Maerk, I., Hoger, H., Jung, M., Karbasiyan, M., Storch, M., Lassmann, H., Moss, J. A., Davison, K., Harrison, R., Bushby, K. M., and Reis, A. (1999) Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for limb girdle muscular dystrophy 2B. *Nature genetics* **23**, 141-142

52. Matsuda, C., Aoki, M., Hayashi, Y. K., Ho, M. F., Arahata, K., and Brown, R. H., Jr. (1999) Dysferlin is a surface membrane-associated protein that is absent in Miyoshi myopathy. *Neurology* **53**, 1119-1122

53. Bushby, K. M. (2000) Dysferlin and muscular dystrophy. *Acta neurologica Belgica* **100**, 142-145

54. Bansal, D., and Campbell, K. P. (2004) Dysferlin and the plasma membrane repair in muscular dystrophy. *Trends in cell biology* **14**, 206-213

55. Lek, A., Lek, M., North, K. N., and Cooper, S. T. (2010) Phylogenetic analysis of ferlin genes reveals ancient eukaryotic origins. *BMC Evol Biol* **10**, 231

56. Lek, A., Evesson, F. J., Sutton, R. B., North, K. N., and Cooper, S. T. (2012) Ferlins: regulators of vesicle fusion for auditory neurotransmission, receptor trafficking and membrane repair. *Traffic* **13**, 185-194

57. Humphrey, G. W., Mekhedov, E., Blank, P. S., de Morree, A., Pekkurnaz, G., Nagaraju, K., and Zimmerberg, J. (2012) GREG cells, a dysferlin-deficient myogenic mouse cell line. *Experimental cell research* **318**, 127-135

58. Han, W. Q., Xia, M., Xu, M., Boini, K. M., Ritter, J. K., Li, N. J., and Li, P. L. (2012) Lysosome fusion to the cell membrane is mediated by the dysferlin C2A domain in coronary arterial endothelial cells. *Journal of cell science* **125**, 1225-1234

59. McDade, J. R., and Michele, D. E. (2014) Membrane damage-induced vesicle-vesicle fusion of dysferlin-containing vesicles in muscle cells requires microtubules and kinesin. *Human molecular genetics* **23**, 1677-1686

60. Nagaraju, K., Rawat, R., Veszelovszky, E., Thapliyal, R., Kesari, A., Sparks, S., Raben, N., Plotz, P., and Hoffman, E. P. (2008) Dysferlin deficiency enhances monocyte phagocytosis: a model for the inflammatory onset of limb-girdle muscular dystrophy 2B. *The American journal of pathology* **172**, 774-785

61. Lennon, N. J., Kho, A., Bacskai, B. J., Perlmutter, S. L., Hyman, B. T., and Brown, R. H., Jr. (2003) Dysferlin interacts with annexins A1 and A2 and mediates sarcolemmal wound-healing. *The Journal of biological chemistry* **278**, 50466-50473

62. Waddell, L. B., Lemckert, F. A., Zheng, X. F., Tran, J., Evesson, F. J., Hawkes, J. M., Lek, A., Street, N. E., Lin, P., Clarke, N. F., Landstrom, A. P., Ackerman, M. J., Weisleder, N., Ma, J., North, K. N., and Cooper, S. T. (2011) Dysferlin, annexin A1, and mitsugumin 53 are upregulated in muscular dystrophy and localize to longitudinal tubules

of the T-system with stretch. *Journal of neuropathology and experimental neurology* **70**, 302-313

63. Roostalu, U., and Strahle, U. (2012) In vivo imaging of molecular interactions at damaged sarcolemma. *Developmental cell* **22**, 515-529

64. Ampong, B. N., Imamura, M., Matsumiya, T., Yoshida, M., and Takeda, S. (2005) Intracellular localization of dysferlin and its association with the dihydropyridine receptor. *Acta myologica : myopathies and cardiomyopathies : official journal of the Mediterranean Society of Myology / edited by the Gaetano Conte Academy for the study of striated muscle diseases* **24**, 134-144

65. Kerr, J. P., Ziman, A. P., Mueller, A. L., Muriel, J. M., Kleinhans-Welte, E., Gumerson, J. D., Vogel, S. S., Ward, C. W., Roche, J. A., and Bloch, R. J. (2013) Dysferlin stabilizes stress-induced Ca2+ signaling in the transverse tubule membrane. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20831-20836

66. Kerr, J. P., Ward, C. W., and Bloch, R. J. (2014) Dysferlin at transverse tubules regulates Ca homeostasis in skeletal muscle. *Frontiers in physiology* **5**, 89

67. Evesson, F. J., Peat, R. A., Lek, A., Brilot, F., Lo, H. P., Dale, R. C., Parton, R. G., North, K. N., and Cooper, S. T. (2010) Reduced plasma membrane expression of dysferlin mutants is attributed to accelerated endocytosis via a syntaxin-4-associated pathway. *The Journal of biological chemistry* **285**, 28529-28539

68. Marty, N. J., Holman, C. L., Abdullah, N., and Johnson, C. P. (2013) The C2 domains of otoferlin, dysferlin, and myoferlin alter the packing of lipid bilayers. *Biochemistry* **52**, 5585-5592

69. Therrien, C., Di Fulvio, S., Pickles, S., and Sinnreich, M. (2009) Characterization of lipid binding specificities of dysferlin C2 domains reveals novel interactions with phosphoinositides. *Biochemistry* **48**, 2377-2384

70. Tucker, W. C., Weber, T., and Chapman, E. R. (2004) Reconstitution of Ca2+-regulated membrane fusion by synaptotagmin and SNAREs. *Science* **304**, 435-438

71. Hui, E., Johnson, C. P., Yao, J., Dunning, F. M., and Chapman, E. R. (2009) Synaptotagmin-mediated bending of the target membrane is a critical step in Ca(2+)-regulated fusion. *Cell* **138**, 709-721

72. Hui, E., Gaffaney, J. D., Wang, Z., Johnson, C. P., Evans, C. S., and Chapman, E. R. (2011) Mechanism and function of synaptotagmin-mediated membrane apposition. *Nature structural & molecular biology* **18**, 813-821

73. Towbin, H., Staehelin, T., and Gordon, J. (1992) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24**, 145-149

74. Weber, T., Zemelman, B. V., McNew, J. A., Westermann, B., Gmachl, M., Parlati, F., Sollner, T. H., and Rothman, J. E. (1998) SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759-772

75. Soderberg, O., Leuchowius, K. J., Gullberg, M., Jarvius, M., Weibrecht, I., Larsson, L. G., and Landegren, U. (2008) Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay. *Methods* **45**, 227-232

76. Fasshauer, D., Otto, H., Eliason, W. K., Jahn, R., and Brunger, A. T. (1997) Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. *The Journal of biological chemistry* **272**, 28036-28041

77. Fasshauer, D., Sutton, R. B., Brunger, A. T., and Jahn, R. (1998) Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 15781-15786

78. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395**, 347-353

79. Roux, I., Safieddine, S., Nouvian, R., Grati, M., Simmler, M. C., Bahloul, A., Perfettini, I., Le Gall, M., Rostaing, P., Hamard, G., Triller, A., Avan, P., Moser, T., and Petit, C. (2006) Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell* **127**, 277-289

80. Bhalla, A., Chicka, M. C., Tucker, W. C., and Chapman, E. R. (2006) Ca(2+)-synaptotagmin directly regulates t-SNARE function during reconstituted membrane fusion. *Nature structural & molecular biology* **13**, 323-330

81. Ma, C., Su, L., Seven, A. B., Xu, Y., and Rizo, J. (2013) Reconstitution of the vital functions of Munc18 and Munc13 in neurotransmitter release. *Science* **339**, 421-425

82. Chen, X., Lu, J., Dulubova, I., and Rizo, J. (2008) NMR analysis of the closed conformation of syntaxin-1. *Journal of biomolecular NMR* **41**, 43-54

83. Yang, B., Gonzalez, L., Jr., Prekeris, R., Steegmaier, M., Advani, R. J., and Scheller, R. H. (1999) SNARE interactions are not selective. Implications for membrane fusion specificity. *The Journal of biological chemistry* **274**, 5649-5653

84. Kubista, H., Edelbauer, H., and Boehm, S. (2004) Evidence for structural and functional diversity among SDS-resistant SNARE complexes in neuroendocrine cells. *Journal of cell science* **117**, 955-966

85. Chiu, Y. H., Hornsey, M. A., Klinge, L., Jorgensen, L. H., Laval, S. H., Charlton, R., Barresi, R., Straub, V., Lochmuller, H., and Bushby, K. (2009) Attenuated muscle regeneration is a key factor in dysferlin-deficient muscular dystrophy. *Human molecular*

genetics **18**, 1976-1989

86. Redpath, G. M., Woolger, N., Piper, A. K., Lemckert, F. A., Lek, A., Greer, P. A., North, K. N., and Cooper, S. T. (2014) Calpain cleavage within dysferlin exon 40a releases a synaptotagmin-like module for membrane repair. *Molecular biology of the cell* **25**, 3037-3048

87. Azakir, B. A., Di Fulvio, S., Salomon, S., Brockhoff, M., Therrien, C., and Sinnreich, M. (2012) Modular dispensability of dysferlin C2 domains reveals rational design for mini-dysferlin molecules. *The Journal of biological chemistry* **287**, 27629-27636

88. Padmanarayana, M., Hams, N., Speight, L. C., Petersson, E. J., Mehl, R. A., and Johnson, C. P. (2014) Characterization of the Lipid Binding Properties of Otoferlin Reveal Specific Interactions Between PI(4,5)P2 and the C2C and C2F Domains. *Biochemistry* **53**, 5023-5033

89. Ramakrishnan, N. A., Drescher, M. J., and Drescher, D. G. (2009) Direct interaction of otoferlin with syntaxin 1A, SNAP-25, and the L-type voltage-gated calcium channel Cav1.3. *The Journal of biological chemistry* **284**, 1364-1372

90. Achanzar, W. E., and Ward, S. (1997) A nematode gene required for sperm vesicle fusion. *Journal of cell science* **110 (Pt 9)**, 1073-1081

91. Lostal, W., Bartoli, M., Roudaut, C., Bourg, N., Krahn, M., Pryadkina, M., Borel, P., Suel, L., Roche, J. A., Stockholm, D., Bloch, R. J., Levy, N., Bashir, R., and Richard, I. (2012) Lack of correlation between outcomes of membrane repair assay and correction of dystrophic changes in experimental therapeutic strategy in dysferlinopathy. *PloS one* **7**, e38036

92. Klenchin, V. A., and Martin, T. F. (2000) Priming in exocytosis: attaining fusion-competence after vesicle docking. *Biochimie* **82**, 399-407

93. Xie, J., and Schultz, P. G. (2005) Adding amino acids to the genetic repertoire. *Current opinion in chemical biology* **9**, 548-554

94. Speight, L. C., Muthusamy, A. K., Goldberg, J. M., Warner, J. B., Wissner, R. F., Willi, T. S., Woodman, B. F., Mehl, R. A., and Petersson, E. J. (2013) Efficient synthesis and in vivo incorporation of acridon-2-ylalanine, a fluorescent amino acid for lifetime and Forster resonance energy transfer/luminescence resonance energy transfer studies. *Journal of the American Chemical Society* **135**, 18806-18814

95. Goldberg, J. M., Speight, L. C., Fegley, M. W., and Petersson, E. J. (2012) Minimalist probes for studying protein dynamics: thioamide quenching of selectively excitable fluorescent amino acids. *Journal of the American Chemical Society* **134**, 6088-6091

96. Hernandez-Deviez, D. J., Howes, M. T., Laval, S. H., Bushby, K., Hancock, J. F., and Parton, R. G. (2008) Caveolin regulates endocytosis of the muscle repair protein, dysferlin. *The Journal of biological chemistry* **283**, 6476-6488

97. Sula, A., Cole, A. R., Yeats, C., Orengo, C., and Keep, N. H. (2014) Crystal structures of the human Dysferlin inner DysF domain. *BMC structural biology* **14**, 3

98. Carpenter, E. P., Hawkins, A. R., Frost, J. W., and Brown, K. A. (1998) Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. *Nature* **394**, 299-302

99. Nango, E., Kumasaka, T., Hirayama, T., Tanaka, N., and Eguchi, T. (2008) Structure of 2-deoxy-scyllo-inososе synthase, a key enzyme in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics, in complex with a mechanism-based inhibitor and NAD+. *Proteins* **70**, 517-527

100. Bender, S. L., Mehdi, S., and Knowles, J. R. (1989) Dehydroquinate synthase: the role of divalent metal cations and of nicotinamide adenine dinucleotide in catalysis. *Biochemistry* **28**, 7555-7560

101. Mahmud, T. (2009) Progress in aminocyclitol biosynthesis. *Curr Opin Chem Biol* **13**, 161-170

102. Soonthornpun, S., Rattarasarn, C., Thamprasit, A., and Leetanaporn, K. (1998) Effect of acarbose in treatment of type II diabetes mellitus: a double-blind, crossover, placebo-controlled trial. *J Med Assoc Thai* **81**, 195-200

103. Stratmann, A., Mahmud, T., Lee, S., Distler, J., Floss, H. G., and Piepersberg, W. (1999) The AcbC protein from Actinoplanes species is a C7-cyclitol synthase related to 3-dehydroquinate synthases and is involved in the biosynthesis of the alpha-glucosidase inhibitor acarbose. *J Biol Chem* **274**, 10889-10896

104. Dong, H., Mahmud, T., Tornus, I., Lee, S., and Floss, H. G. (2001) Biosynthesis of the validamycins: identification of intermediates in the biosynthesis of validamycin A by Streptomyces hygroscopicus var. limoneus. *J Am Chem Soc* **123**, 2733-2742

105. Balskus, E. P., and Walsh, C. T. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* **329**, 1653-1656

106. Wu, X., Flatt, P. M., Xu, H., and Mahmud, T. (2009) Biosynthetic gene cluster of cetoniacytone A, an unusual aminocyclitol from the endosymbiotic Bacterium Actinomyces sp. Lu 9419. *Chembiochem* **10**, 304-314

107. Schlorke, O., Krastel, P., Muller, I., Uson, I., Dettner, K., and Zeeck, A. (2002) Structure and biosynthesis of cetoniacytone A, a cytotoxic aminocarba sugar produced by an endosymbiontic Actinomyces. *J Antibiot (Tokyo)* **55**, 635-642

108. Asamizu, S., Yang, J., Almabruk, K. H., and Mahmud, T. Pseudoglycosyltransferase catalyzes nonglycosidic C-N coupling in validamycin a biosynthesis. *J Am Chem Soc* **133**, 12124-12135

109. Yu, Y., Bai, L., Minagawa, K., Jian, X., Li, L., Li, J., Chen, S., Cao, E., Mahmud, T.,

Floss, H. G., Zhou, X., and Deng, Z. (2005) Gene cluster responsible for validamycin biosynthesis in Streptomyces hygroscopicus subsp. jinggangensis 5008. *Appl Environ Microbiol* **71**, 5066-5076

110. Srinivasan, P. R., Rothschild, J., and Sprinson, D. B. (1963) The Enzymic Conversion of 3-Deoxy-D-Arabino-Heptulosonic Acid 7-Phosphate to 5-Dehydroquinate. *J Biol Chem* **238**, 3176-3182

111. Flatt, P. M., and Mahmud, T. (2007) Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat Prod Rep* **24**, 358-392

112. Bender, S. L., Widlanski, T., and Knowles, J. R. (1989) Dehydroquinate synthase: the use of substrate analogues to probe the early steps of the catalyzed reaction. *Biochemistry* **28**, 7560-7572

113. Widlanski, T., Bender, S. L., and Knowles, J. R. (1989) Dehydroquinate synthase: the use of substrate analogues to probe the late steps of the catalyzed reaction. *Biochemistry* **28**, 7572-7582

114. Nango, E., Eguchi, T., and Kakinuma, K. (2004) Active site mapping of 2-deoxy-scyllo-inosose synthase, the key starter enzyme for the biosynthesis of 2-deoxystreptamine. Mechanism-based inhibition and identification of lysine-141 as the entrapped nucleophile. *J Org Chem* **69**, 593-600

115. Winter, J. M., and Tang, Y. (2012) Synthetic biological approaches to natural product biosynthesis. *Curr Opin Biotechnol* **23**, 736-743

116. Kean, K. M., Codding, S. J., Asamizu, S., Mahmud, T., and Karplus, P. A. (2014) Structure of a sedoheptulose 7-phosphate cyclase: ValA from Streptomyces hygroscopicus. *Biochemistry* **53**, 4250-4260

117. Asamizu, S., Xie, P. F., Brumsted, C. J., Platt, P. M., and Mahmud, T. (2012) Evolutionary Divergence of Sedoheptulose 7-Phosphate Cyclases Leads to Several Distinct Cyclic Products. *J Am Chem Soc* **134**, 12219-12229

118. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica, Section D: Biological Crystallography* **67**, 271-281

119. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica, Section D: Biological Crystallography* **67**, 235-242

120. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica, Section D: Biological Crystallography* **62**, 72-82

121. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674

122. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica, Section D: Biological Crystallography* **66**, 213-221

123. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica, Section D: Biological Crystallography* **66**, 486-501

124. Terwilliger, T. C., Dimaio, F., Read, R. J., Baker, D., Bunkoczi, G., Adams, P. D., Grosse-Kunstleve, R. W., Afonine, P. V., and Echols, N. (2012) phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *Journal of Structural and Functional Genomics* **13**, 81-90

125. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637

126. Joosten, R. P., te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Research* **39**, D411-419

127. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**, W545-549

128. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948

129. Golovin, A., Dimitropoulos, D., Oldfield, T., Rachedi, A., and Henrick, K. (2005) MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. *Proteins* **58**, 190-199

130. Park, A., Lamb, H. K., Nichols, C., Moore, J. D., Brown, K. A., Cooper, A., Charles, I. G., Stammers, D. K., and Hawkins, A. R. (2004) Biophysical and kinetic analysis of wild-type and site-directed mutants of the isolated and native dehydroquinate synthase domain of the AROM protein. *Protein science : a publication of the Protein Society* **13**, 2108-2119

131. D'Arcy, A., Villard, F., and Marsh, M. (2007) An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* **63**, 550-554

132. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58

133. Terwilliger, T. C., Dimaio, F., Read, R. J., Baker, D., Bunkoczi, G., Adams, P. D., Grosse-Kunstleve, R. W., Afonine, P. V., and Echols, N. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics* **13**, 81-90

134. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and Andre, I. (2011) Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6**, e20450

135. D'Alonzo, D., Guaragna, A., Napolitano, C., and Palumbo, G. (2008) Rapid access to 1,6-anhydro-beta-L-hexopyranose derivatives via domino reaction: synthesis of L-allose and L-glucose. *J Org Chem* **73**, 5636-5639

136. Guaragna, A., Napolitano, C., D'Alonzo, D., Pedatella, S., and Palumbo, G. (2006) A versatile route to L-hexoses: synthesis of L-mannose and L-altrose. *Organic letters* **8**, 4863-4866

137. Ahronowitz, I., Xin, W., Kiely, R., Sims, K., MacCollin, M., and Nunes, F. P. (2007) Mutational spectrum of the NF2 gene: a meta-analysis of 12 years of research and diagnostic laboratory findings. *Hum Mutat* **28**, 1-12

138. Fehon, R. G., LaJeunesse, D., Lamb, R., McCartney, B. M., Schweizer, L., and Ward, R. E. (1997) Functional studies of the protein 4.1 family of junctional proteins in Drosophila. *Soc Gen Physiol Ser* **52**, 149-159

139. Giovannini, M., Robanus-Maandag, E., van der Valk, M., Niwa-Kawakita, M., Abramowski, V., Goutebroze, L., Woodruff, J. M., Berns, A., and Thomas, G. (2000) Conditional biallelic Nf2 mutation in the mouse promotes manifestations of human neurofibromatosis type 2. *Genes Dev* **14**, 1617-1630

140. McClatchey, A. I., and Giovannini, M. (2005) Membrane organization and tumorigenesis--the NF2 tumor suppressor, Merlin. *Genes Dev* **19**, 2265-2277

141. Nguyen, R., Reczek, D., and Bretscher, A. (2001) Hierarchy of merlin and ezrin N- and C-terminal domain interactions in homo- and heterotypic associations and their relationship to binding of scaffolding proteins EBP50 and E3KARP. *J Biol Chem* **276**, 7621-7629

142. Shaw, R. J., Paez, J. G., Curto, M., Yaktine, A., Pruitt, W. M., Saotome, I., O'Bryan, J. P., Gupta, V., Ratner, N., Der, C. J., Jacks, T., and McClatchey, A. I. (2001) The Nf2 tumor suppressor, merlin, functions in Rac-dependent signaling. *Dev Cell* **1**, 63-72

143. Bretscher, A., Edwards, K., and Fehon, R. G. (2002) ERM proteins and merlin: integrators at the cell cortex. *Nat Rev Mol Cell Biol* **3**, 586-599

144. Fehon, R. G., McClatchey, A. I., and Bretscher, A. (2010) Organizing the cell cortex: the role of ERM proteins. *Nat Rev Mol Cell Biol* **11**, 276-287

145. Reczek, D., Berryman, M., and Bretscher, A. (1997) Identification of EBP50: A PDZ-

containing phosphoprotein that associates with members of the ezrin-radixin-moesin family. *J Cell Biol* **139**, 169-179

146. Garbett, D., LaLonde, D. P., and Bretscher, A. (2010) The scaffolding protein EBP50 regulates microvillar assembly in a phosphorylation-dependent manner. *The Journal of cell biology* **191**, 397-413

147. Reczek, D., and Bretscher, A. (1998) The carboxyl-terminal region of EBP50 binds to a site in the amino-terminal domain of ezrin that is masked in the dormant molecule. *J Biol Chem* **273**, 18452-18458

148. Finnerty, C. M., Chambers, D., Ingraffea, J., Faber, H. R., Karplus, P. A., and Bretscher, A. (2004) The EBP50-moesin interaction involves a binding site regulated by direct masking on the FERM domain. *J Cell Sci* **117**, 1547-1552

149. Terawaki, S., Maesaki, R., and Hakoshima, T. (2006) Structural basis for NHERF recognition by ERM proteins. *Structure* **14**, 777-789

150. Smith, W. J., Nassar, N., Bretscher, A., Cerione, R. A., and Karplus, P. A. (2003) Structure of the active N-terminal domain of Ezrin. Conformational and mobility changes identify keystone interactions. *The Journal of biological chemistry* **278**, 4949-4956

151. Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chembiochem* **5**, 170-176

152. Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U., and Derewenda, Z. S. (2002) The structure of the FERM domain of merlin, the neurofibromatosis type 2 gene product. *Acta Crystallogr D Biol Crystallogr* **58**, 381-391

153. Shimizu, T., Seto, A., Maita, N., Hamada, K., Tsukita, S., Tsukita, S., and Hakoshima, T. (2002) Structural basis for neurofibromatosis type 2. Crystal structure of the merlin FERM domain. *J Biol Chem* **277**, 10332-10336

154. Yogesha, S. D., Sharff, A. J., Giovannini, M., Bricogne, G., and Izard, T. (2011) Unfurling of the band 4.1, ezrin, radixin, moesin (FERM) domain of the merlin tumor suppressor. *Protein Sci* **20**, 2113-2120

155. Sher, I., Hanemann, C. O., Karplus, P. A., and Bretscher, A. (2012) The tumor suppressor merlin controls growth in its open state, and phosphorylation converts it to a less-active more-closed state. *Dev Cell* **22**, 703-705

156. Okada, T., You, L., and Giancotti, F. G. (2007) Shedding light on Merlin's wizardry. *Trends Cell Biol* **17**, 222-229

157. Rose, J. C., Huang, P. S., Camp, N. D., Ye, J., Leidal, A. M., Goreshnik, I., Trevillian, B. M., Dickinson, M. S., Cunningham-Bryant, D., Debnath, J., Baker, D., Wolf-Yadlin, A., and Maly, D. J. (2017) A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. *Nature chemical biology* **13**, 119-126

158. Hou, J. H., Kralj, J. M., Douglass, A. D., Engert, F., and Cohen, A. E. (2014) Simultaneous mapping of membrane voltage and calcium in zebrafish heart in vivo reveals chamber-specific developmental transitions in ionic currents. *Frontiers in physiology* **5**, 344

159. Illarioshkin, S. N., Ivanova-Smolenskaya, I. A., Greenberg, C. R., Nylen, E., Sukhorukov, V. S., Poleshchuk, V. V., Markova, E. D., and Wrogemann, K. (2000) Identical dysferlin mutation in limb-girdle muscular dystrophy type 2B and distal myopathy. *Neurology* **55**, 1931-1933

160. de Morree, A., Flix, B., Bagaric, I., Wang, J., van den Boogaard, M., Grand Moursel, L., Frants, R. R., Illa, I., Gallardo, E., Toes, R., and van der Maarel, S. M. (2013) Dysferlin regulates cell adhesion in human monocytes. *The Journal of biological chemistry* **288**, 14147-14157

161. Codding, S. J., Marty, N., Abdullah, N., and Johnson, C. P. (2016) Dysferlin Binds SNAREs (Soluble N-Ethylmaleimide-sensitive Factor (NSF) Attachment Protein Receptors) and Stimulates Membrane Fusion in a Calcium-sensitive Manner. *The Journal of biological chemistry* **291**, 14575-14584

162. Matsuda, C., Kameyama, K., Tagawa, K., Ogawa, M., Suzuki, A., Yamaji, S., Okamoto, H., Nishino, I., and Hayashi, Y. K. (2005) Dysferlin interacts with affixin (beta-parvin) at the sarcolemma. *Journal of neuropathology and experimental neurology* **64**, 334-340

163. Lek, A., Evesson, F. J., Lemckert, F. A., Redpath, G. M., Lueders, A. K., Turnbull, L., Whitchurch, C. B., North, K. N., and Cooper, S. T. (2013) Calpains, cleaved mini-dysferlinC72, and L-type channels underpin calcium-dependent muscle membrane repair. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 5085-5094

164. de Luna, N., Gallardo, E., Soriano, M., Dominguez-Perles, R., de la Torre, C., Rojas-Garcia, R., Garcia-Verdugo, J. M., and Illa, I. (2006) Absence of dysferlin alters myogenin expression and delays human muscle differentiation "in vitro". *The Journal of biological chemistry* **281**, 17092-17098

165. Demonbreun, A. R., Fahrenbach, J. P., Deveaux, K., Earley, J. U., Pytel, P., and McNally, E. M. (2011) Impaired muscle growth and response to insulin-like growth factor 1 in dysferlin-mediated muscular dystrophy. *Human molecular genetics* **20**, 779-789

166. Sharma, A., Yu, C., Leung, C., Trane, A., Lau, M., Utokaparch, S., Shaheen, F., Sheibani, N., and Bernatchez, P. (2010) A new role for the muscle repair protein dysferlin in endothelial cell adhesion and angiogenesis. *Arteriosclerosis, thrombosis, and vascular biology* **30**, 2196-2204

167. Oulhen, N., Onorato, T. M., Ramos, I., and Wessel, G. M. (2014) Dysferlin is essential for endocytosis in the sea star oocyte. *Developmental biology* **388**, 94-102