UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structural and Functional Studies of Three Human Hydrolases: DUF89, HtrA1, and Uba2

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Taylor N. Dennis

June 2019

Dissertation Committee:
Dr. Jeff Perry, Chairperson
Dr. Paul Larsen
Dr. Jikui Song

The Dissertation of Taylor N. Dennis is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my Ph.D. advisor Dr. John Jefferson Perry for his endless support. It has been an honor to be his first graduate student. The scientist I am today is due to his tutelage. I thank him for his guidance on these endeavors herein as well as critical editing of this thesis. I am grateful for his patience, insight, and funding, all of which were necessary to ensure my graduate success. It should go without saying, but I could not think of a better mentor for my graduate studies.

Besides my advisor, I must extend gratitude to my dissertation committee members, Dr. Paul Larsen and Dr. Jikui Song. These two illustrious professors have guided my studies over the last four years and have provided much needed advice regarding the broader impacts of my research. Also, thank you to members of my qualifying committee and collaborators on the SUMOylation project, Dr. Jiayu Liao and Dr. Dave Martin. And thank you to the graduate students in their designated labs, George Way and Bill Weigel, for your work in our FRET studies and natural product chemistry, respectively.

My time at UC Riverside was defined by the current and former graduate students who worked in, or rotated through our lab, Jen, Jenna, Tom, Vanessa, Logan, Kriti, Macon to name a few. The comradery we shared and the network of expertise you represent has been a boon to my research, and I am grateful. To all current and former undergraduate students to work in the lab, for which there are too many to list, thank you for your work, and especially thank you to Amrik and Andrew. None of this would have been possible

without the two of you. When miscellaneous responsibilities overburdened me, I knew I could always rely on your diligent and focused work.

I'd also like to thank the outstanding collaborators that assisted in these studies: Dr. Steven G. Clarke, Dr. Jay Kirkwood, and Dr. David Mobley. Your expertise and work assisted with our methyl transferase activity assays, LC-MS/MS metabolomics studies, and the CADD-identified compounds from the NCI diversity set. And to Dr. Eduardo Hilario, thank you for patiently teaching me the molecular biology techniques required for these studies. Gratitude must as well be extended to Dr. Seán O'Leary and, now, Dr. Caroline Sjogren for their ancillary support and personal guidance.

Acknowledgements are also due to the UC start-up, UCOP, UC CRCC, UCR Seed, and Metabolomics Seed grants awarded to the lab for financially supporting my training and this work.

Last, and certainly not least, I must extend the sincerest appreciation to my friends and family. This dissertation, nor these last five years, would not have been possible without their love and support.

DEDICATION

To those before and those to follow.

&

To Constance, my light.

ABSTRACT OF THE DISSERTATION

Structural and Functional Studies of Three Human Hydrolases: DUF89, HtrA1, and Uba2

by

Taylor N. Dennis

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, June 2019
Dr. Jeff Perry, Chairperson

Structural biology has offered deepened understanding of the biological world at the molecular level. An elucidated structure can be used to understand enzymatic mechanisms and design new, novel inhibitory molecules. Our studies have examined the structures of three proteins and paired them with *in vitro* assays to probe the activity and small-molecule inhibition of two therapeutically intriguing enzymes.

First, we have solved the three-dimensional structure of the gene product of *C6orf211* using X-ray crystallographic techniques. The architecture revealed a DUF89 fold with a coordinate magnesium in the deepest surface pocket. *In vitro* functional studies identified a metal-dependent phosphatase activity against a host of phosphometabolites, and strongest against fructose 1-phosphate. The enzymatic promiscuity, low Km, and ubiquitous low-level expression suggest the hDUF89 protein may play a biological role in metabolite repair.

Next, the extracellular protease HtrA1 was pursued as a possible therapeutic target for age-related macular degeneration (AMD). An AMD predisposing haplotype has been linked to *HTRA1* where elevated expression levels were previously noted. *In vitro* assays were developed to probe the efficacy of AI-CADD predicted binders. Our inhibition studies identified several low micromolar hits, and efforts to characterize the nature of the interaction between enzyme and ligand via macromolecular crystallographic approaches are ongoing.

Last, we pursued many avenues for inhibitor development against the Aos1/Uba2 heterodimer. The SUMOylation system has been widely identified as a suitable drug candidate due to its roles in many untreatable cancers. Of note, the most potent natural product inhibitor for Aos1/Uba2 is anacardic acid, a fat-soluble natural product with known off-target effects. Avenues for inhibitor identification therefore included natural product chemistry to produce anacardic acid derivatives, CADD, AI-CADD, and FBLD. Our *in vitro* binding studies have identified strong candidates for subsequent lead development via chemical modifications following ongoing crystallographic efforts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Observing the biological world at the molecular level is one of the most important tools in understanding the underlying chemistry that drives life. The field of structural biology owes itself to groundwork laid by early X-ray crystallographers. Since as early as 1840 [1], it was observed that proteins could form crystalline structures, but it wasn't until the early 1900s that X-rays were used to probe the underlying structure [2]. Interestingly, at that time X-rays were not a widely understood form of radiation and it was their scattering upon encountering crystals that gave evidence to the hypothesis that X-rays were in fact a form of light. Upon concentrated exposure to X-ray radiation, the sample diffracted the rays in an interpretable pattern that contains important information about the chemical structure. Sir William Henry Bragg and his son Lawrence Bragg received the Nobel prize in 1915 for their analysis of the X-ray diffraction measurements and determination of the chemical structure of sodium chloride, zinc sulfate, and diamond crystals [3]. The mathematical formula they derived to comprehend diffraction patterns would later be known as Bragg's Law, which now serves as one of the most important equations in macromolecular crystallography. In the 1950s X-ray diffraction data was collected on fibrous DNA by Rosalind Franklin and Maurice Wilkins, which aided in the characterization of "B" form DNA. These X-ray films provided independent confirmation of the proposed three-dimensional structure of DNA by Watson and Crick [4]. Unfortunately, at that time, interpreting the intricate nature of protein X-ray patterns remained elusive.

One of the key issues faced by these early crystallographers was that the diffraction pattern did not contain all the necessary information to obtain an image of the unit cell

contents. When observing a small object under a microscope, a lens is able to refocus the light reflected from an object and recreate the image at an appropriate scale. Thinking about X-rays in this manner, because X-rays cannot be refocused, essential information is lost in the data collection process [5]. In reality, the diffraction pattern is a Fourier transform of the electron-density distribution present in the crystal, and the issue arises when trying to reconstruct the structure by performing the inverse Fourier transformation [6]. Although the amplitude (intensity) and position (frequency) of each diffraction spot (diffracted ray) is known, this summation cannot be completed due to the relative position with respect to all other waves (phase) being systematically lost in any measurement [7]. Put simply, phase is a key variable in interpreting the diffraction pattern that cannot be directly measured. This "phase problem" was eventually overcome when Perutz and Kendrew introduced heavy atoms such as mercury into protein crystals. By comparing the data sets from both the native and heavy-atom soaked isomorphous crystals, the Patterson difference map between the two crystals could reveal the position of the heavy atom within the unit cell. Both the amplitude and phase of the heavy-atom contribution could be calculated and ultimately used to solve the structure. This breakthrough led to a calculation of an electron-density map for myoglobin and a structural model that was built into the three-dimensional map in 1960, making it the first protein structure ever solved [8], followed shortly by the structural determination of hemoglobin [9]. This work established that proteins form incredibly intricate architectures that are necessary for their catalytic function. Many key structures were solved in prevailing years that adorn today's textbooks [10]–[22], and in total 28 Nobel prizes have been awarded to crystallographers. But the

modern field of structural biology owes itself to advances in recombinant DNA and synchrotron radiation technologies.

The first protein crystals were grown from abundant proteins that could be easily extracted from large sources: pig blood, egg protein, whale sperm etc. However, to use X-ray crystallography to investigate more biologically-interesting targets, landmark discoveries in molecular biology proved integral. The bacterium *Escherichia coli*, first discovered in 1884 [23], was widely used in microbiology due to its hardy and non-pathogenic nature [24]. It therefore became the model organism of choice for work that cracked the genetic code and our understanding of transcription, translation, and replication [25], [26]. This deepened understanding of DNA biology allowed for manipulation of the organism to suit biotechnological needs using, then, newly discovered methodologies. Integrating recombinant DNA technology with macromolecular crystallography subsequently produced a revolution in structural biology [27]. The power of recombinant DNA technologies that accompanied the use of heterologous expression systems provided the field with exquisite control over the exact gene sequence being investigated. Such a breakthrough permitted the inclusion of a wide range of cleavable, high-affinity tags that facilitate purification and non-native amino acids such as selenomethionine that facilitate routine phasing [7], or the exclusion of select DNA via truncations or mutations that facilitate stability [28], [29]. The availability of nearly any protein-of-interest, in bulk quantities and with exogenous modifications, provides the field a unique opportunity to study human enzymes outside of native systems. There leaves no doubt that the production of recombinant proteins in microbial systems has revolutionized biochemistry [30].

Advancements in X-ray sources, detection technologies, and crystal handling have all defined the modern field of structural biology. The original sources for X-rays were vacuum tubes that converted electrical power into X-rays, but the generation of this form of radiation has evolved since the times of Röntgen [31]. Synchrotron facilities have become the mainstay of X-ray sources rapidly since their first usage in the 1970s, due to their strong beam intensity and high degree of collimation with powerful tuning capabilities [32], [33]. Synchrotrons not only offer unparalleled X-ray power, but also unique access to research labs without a "home" source of X-rays, as well as beamline support from dedicated staff. Such facilities are well-funded by their respective governments and the technology they house is state-of-the-art. Thankfully, gone are the days of X-ray films. Most synchrotron beamlines dedicated to protein crystallography have transitioned from charge-coupled device-based detector systems to the significantly more advanced hybrid photon counting (HPC) detectors [34]. HPC detectors are capable of measuring X-ray intensities with single-photon sensitivity over an extremely high dynamic range. Not just the X-ray and detector have been improved, but ways of handling the crystal itself are also implemented in modern macromolecular crystallography. In order to collect diffraction data, the protein crystal is exposed to significant amounts of radiation. Incoherent scattering of the X-rays will cause damage to the crystalline lattice and reduction in data quality. To preserve the crystal from such damage, it is conventionally flash frozen in liquid nitrogen, and data collection is conducted under a stream of nitrogen gas which keeps the crystal frozen throughout the experiment [35]. These protective measures greatly improve the crystal lifetime and the quality of the data [36]. As the crystals are frozen, they can be

kept harvested and submerged in liquid nitrogen, and an automated crystal handling system can be used to mount each crystal in the X-ray beam [37]. These advances define the modern macromolecular crystallography field.

Investigations of disease states are increasingly necessary and are correspondingly robust. In many instances, individual mutant or upregulated protein targets are known and their *apo* or *holo* three-dimensional (3D) structures have been elucidated. Developing novel therapeutic tools for these enzymes can therefore be conducted in a rational, evidence-based manner. Structure-based drug design (SBDD) has become an indispensable tool for the development of target-based therapies as it utilizes the 3D structure of the biological target (protein) in order to identify putative modulators of the enzymatic activity. Using the crystal structure of the enzyme, ligands can be designed to bind within a pocket or protuberance of the protein target ideally with potential hydrogen bond donors and acceptors and/or hydrophobic characteristics. Examples of the success of this strategy have been identified in as early as 1990 [38], and the process has only gotten more streamlined with advances in computation [39]. Virtual screening successes have been completed using a variety of programs for docking, QSAR, and pharmacophore modeling coupled with *in vitro* assays [40]–[45]. If substrate information is already known for the active site, modifications can be made to the molecule such as augmenting the electrostatics or lowering the energetic potential by making more negative van der Waals terms to improve complementarity with the receptor. One compelling example of the viability of this approach is the discovery of the reversible ACE inhibitor and hypertension drug captopril [46], [47]. Knowledge of the active-site similarities and substrate differences between ACE

and carboxypeptidase lead researchers to generate possible inhibitors [48]. Alanine-proline dipeptides were found to inhibit ACE and combining this with knowledge of the structure revealed that the dipeptide carboxy-terminus coordinates with the zinc ion within the protease [49]. This information was critical in guiding the structure-activity relationship (SAR) that lead to the replacement of the terminal amine with an ethanethiol to produce captopril. Captopril became the first ACE inhibitor to enter clinical use following its FDA approval in 1981 and soon it was established as a frontline therapeutic agent for the treatment of hypertension and heart failure [50].

Herein, we will use the power of modern macromolecular crystallography to investigate biologically-interesting targets only producible through recombinant techniques with a focus on structure determination for hDUF89 and SBDD for HtrA1 and Aos1/Uba2.

Chapter 1
DUF89

Chapter 1: DUF89

1.1 Abstract

Metabolite damage control is a critical but poorly defined aspect of cellular biochemistry, which likely involves many of the so far functionally uncharacterized protein domains that are referred to as domains of unknown function (DUF). We have determined the crystal structure of the human DUF89 protein product of the *C6ORF211* gene to 1.85 Å. The crystal structure contains of a core α-β-α fold, active site bound metal ion and α-helical bundle N-terminal cap, which are all conserved features of subfamily III DUF89 domains. The human DUF89 structure shares certain structural similarities to SAM-MT methyltransferases, but the recombinant protein did not have any readily detectable methyltransferase activity. Instead, *in vitro* phosphatase activity is supported by divalent cations that include Co, Ni, Mn, or Mg, and the enzyme targets a number of phosphate substrates. Highest activity was observed with fructose 1-phosphate, a potent glycating agent, and thus, the novel DUF89 phosphatase activity in human cells may potentially function to help limit the build-up of phospho-glycan species and their related damaged metabolites.

1.2 Introduction

Continuing genomic sequencing efforts have resulted in an ever-expanding prevalence of the annotation "domain of unknown function" (DUF) and in the 31.0 release of the Pfam database they totaled more than twenty percent of the 16,712 known protein domain families [51]. Over 800 of the DUFs are conserved between bacteria and eukaryotes, and approximately 300 of these are can also occur in archaeal species [52]. This widespread taxonomic distribution is suggestive of important roles in the cell, and therefore defining the functions of DUF proteins could provide critical new insights into cellular metabolism. Not all of the DUF functions will likely prove to be essential, and instead some functions may only be required under certain conditions, such as starvation [53]. This makes defining the cellular roles of DUFs through classical phenotypic screen all the more challenging, and thus other approaches that include in-depth structural and biochemical characterizations are likely to be needed.

The lesser-explored research area of metabolite damage control has been suggested to account for the functions of many of the DUF families [54]. Similar to DNA, metabolites are prone to damage that may occur through spontaneous reactions or by more promiscuous enzymes acting on compounds distinct from their primary substrates in side-reactions [55]. However, the sheer number, variety, and instability of damage-prone metabolites than can be produced may contribute to the limited understanding of the molecular mechanisms of metabolite damage control [56]. In addition, metabolite repair enzymes often have comparatively low activities [55], and there has been a lack of approaches sensitive enough to detect and define such molecules until recently, further limiting the characterization of

damage control enzymes [57]. Initial studies have demonstrated that metabolite damage control has roles in pre-empting [58], repairing [59], or removing side products or damaged metabolites [60], or in utilizing directed overflow mechanisms that remove disproportionally high levels of normal metabolites [61], [62]. The accumulation of side products, damaged metabolites, or abnormally high levels of normal metabolites may exert toxic effects in the cell. One such example is L-2-hydroxyglutaric aciduria, a rare neurological condition [63]. This disease occurs due to a mutation in the human L-2-hydroxyglutarate dehydrogenase metabolite repair enzyme (L2HGDH), which functions to oxidize the detrimental metabolic side-product L-2-hydroxyglutarate back to 2-oxoglutarate [60], [64]. Without this metabolite repair pathway intact, L-2-hydroxyglutarate builds up and inhibits key α-ketoglutarate dependent enzymes that unfortunately results in major skeletal and neurological growth defects, and eventually, premature death [65]. Therefore, defining the activities of more metabolite repair enzymes, notably DUF proteins, which have potentially evolved to mitigate metabolite-mediated toxicities under certain stresses, may reveal impacts on human health, as well as improve our understanding of cellular metabolism, and support on-going efforts in systems biology and metabolic engineering [54], [66], [67].

One DUF family likely to have key cellular functions, due to its conservation from certain bacterial species through to higher eukaryotes, is the DUF89 family (Pfam01937). A recent landmark study on recombinant DUF89 proteins that included specific bacterial, archaeal, fungi, and eukaryotic sequences revealed a conserved, novel, metal-dependent phosphatase activity [68]. The DUF89 family has been further divided into three

subfamilies, namely I, II and III, with subfamily I members occurring only in specific anaerobic or microaerophilic bacteria and archaea, and having activities directed towards nucleotides. Subfamily II family proteins belong to two subgroups, the first of which is observed in plant species with activity against sugar phosphates. The second subset consists of proteins that have a DUF89 domain fused onto the C-terminus of a panthothenate kinase (PanK) domain [69], and their respective genes are present in certain fungi, and in plants and animals. The human and Arabidopsis PanK-DUF89 protein homologs were demonstrated to target oxidized forms of phosphopantetheine, and the plant protein was also active against 4'-phosphopantothenate [68]. A recent study identified a mutation within the PanK-DUF89 gene (PanK4) of mice as a novel pathogenic variant for congenital cataract [70], and thereby providing the first evidence of a DUF89 domain impacting human health. Subfamily III is the most prevalent of the subfamilies, being present in some bacterial species through to higher eukaryotes, and the few members of this family characterized so far have activity against sugar phosphate substrates [68].

The human genome contains two DUF89 containing sequences, the subfamily II PanK-DUF89 fusion (PanK4) and a poorly characterized product of the *C6ORF211* gene, which shares strong sequence similarities to subfamily III members. Initial studies on *C6ORF211* had noted the gene to be co-expressed with the estrogen receptor, and siRNA knockdown of *C6orf211* expression was observed to reduce breast cancer cell proliferation independent of estrogen levels [71]. One study that found higher *C6orf211* gene expression levels in the 344 breast cancer tissue samples exhibited lower nuclear grade, positive ER status, positive PgR status, negative HER2 status, and no correlation with Ki67; gene

expression was also high in the $ER^{+}HER2^{-}$ subtype [72]. Additionally, the yeast DUF89 subfamily III homolog, YMR027W, in a gene deletion mutant, showed consistently elevated levels of spontaneous Rad52 foci [73], and in an untargeted metabolomics study, fructose 1-phosphate (F-1-P) was the only substrate that accumulated in a statistically significant way when cells were grown on either glucose or fructose media [68]. This evidence indicates a role for hDUF89 within cellular maintenance, and the proposed F-1-P activity may provide some insight.

The enzymatic attachment of a particular glycan to a protein or lipid is known as glycosylation. Much has been defined in terms of glycobiology, and generally speaking glycans can facilitate many varied functions when properly attached to their substrate [74]. However, non-enzymatic glycosylation, referred to as glycation, can have severe consequences due to the large charge-destabilization from the addition of a glucose or fructose moiety, and further due to the irreversible nature of the bond formation. Early glycation products, first initiated by the generation of acid-labile Schiff base adducts [75], [76], are considered to be intermediates of the glycation reaction that thus undergo slow and complex transformations to form irreversible advanced glycation end products (AGEs) [77]–[80]. AGEs, which include heterogeneous structures of complex modifications, tend to accumulate on long-lived biological molecules [81]. It has been reported that while glucose is the primary substrate for glycation events, fructose and its phosphate derivatives will predicate irreversible glycosylation—fructation—at a higher rate [79]. While the glycation of short-lived biomolecules can have deleterious effects, the modification of long-lived biomolecules such as collagen, DNA, lens crystalline, and myelin have been

implicated in the pathophysiology of aging [81]–[84], diabetes [85]–[88], and Alzheimer's disease [89], [90]. Thus, if such high-energy metabolites are left unchecked, they can exert severe cellular damage.

Here, we present the first detailed structural and biochemical characterizations of the human C6orf211 protein product. The crystal structure at 1.85 Å reveals a conserved DUF89 subfamily III fold, complete with an α-β-α sandwich that includes a bound metal in the active site and an N-terminal α-helical bundle, as well as key conserved structural motifs belonging to subfamily III members. The α-β-α core shares similarity to SAM-dependent methyltransferases, but methyltransferase activity was not readily observed in our assays with the recombinant human DUF89 protein. Instead, a number of different divalent metal cations were observed to support *in vitro* phosphatase activity against several distinct substrates, predominantly against phospho-metabolites involved in fructose metabolism, PPP, and glycolysis. The highest activity was observed against fructose 1-phosphate, and interestingly, fructose 1-phosphate is known to be a potent glycating agent. This suggests that human DUF89 could play a role in limiting the potentially damaging build-up of phospho-glycan species in human tissues.

1.3 Results

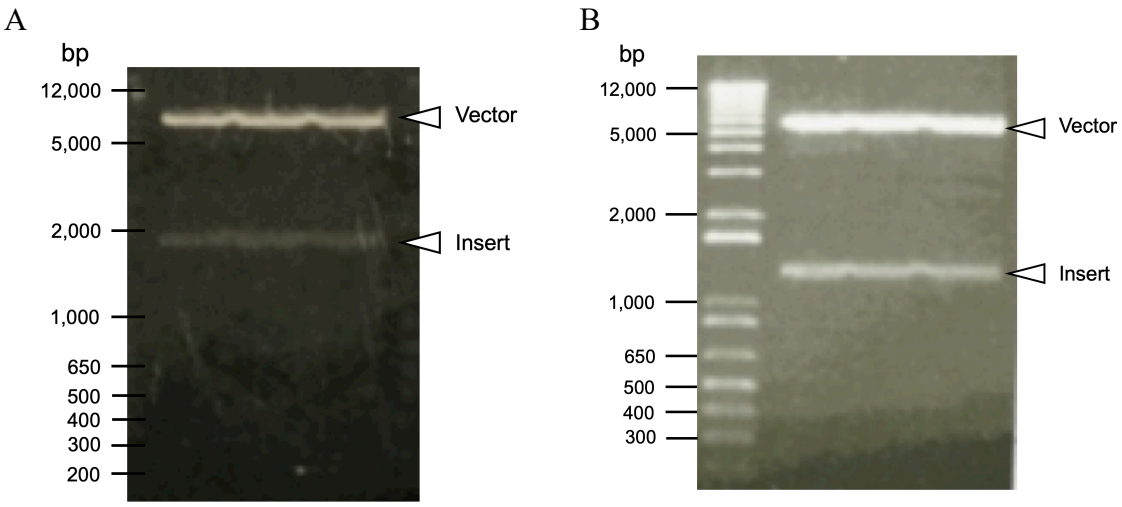


Figure 1.1 | YMR027W and C6orf211 cloning (A) Samples from the YMR027W restriction digestion was run on a 1% agarose gel and visualized under UV light (B) Samples from the C6orf211 restriction digestion were run on a 1% agarose gel and visualized under UV light. Marker used: 1 Kb Plus DNA Ladder (Invitrogen™).

1.3.1 Cloning and expression of DUF89 proteins

To begin work on these DUF89 proteins in *E. coli*, their genes needed to be isolated as recombinant DNA and re-cloned into bacterial expression systems. Successful cloning of the YMR027W gene into pET 22b can be determined via a restriction enzyme digestion reaction using the enzymes *Nde*I and *Xho*I, where the insert (1,436 bases) migrates faster that the vector (5.5 kb) (**Fig. 1.1A**). Likewise, successful cloning of the *C6orf211* gene into pET SUMO can be observed by the two bands (1,344 and 5,600 bases) following the four-hour incubation with restriction enzymes *Bam*HI-HF and *Sal*I-HF (**Fig. 1.1B**). These results, as well as the site-directed mutagenesis to generate the wild-type human enzyme were further validated by Sanger sequencing conducted by the UCR Core facility.

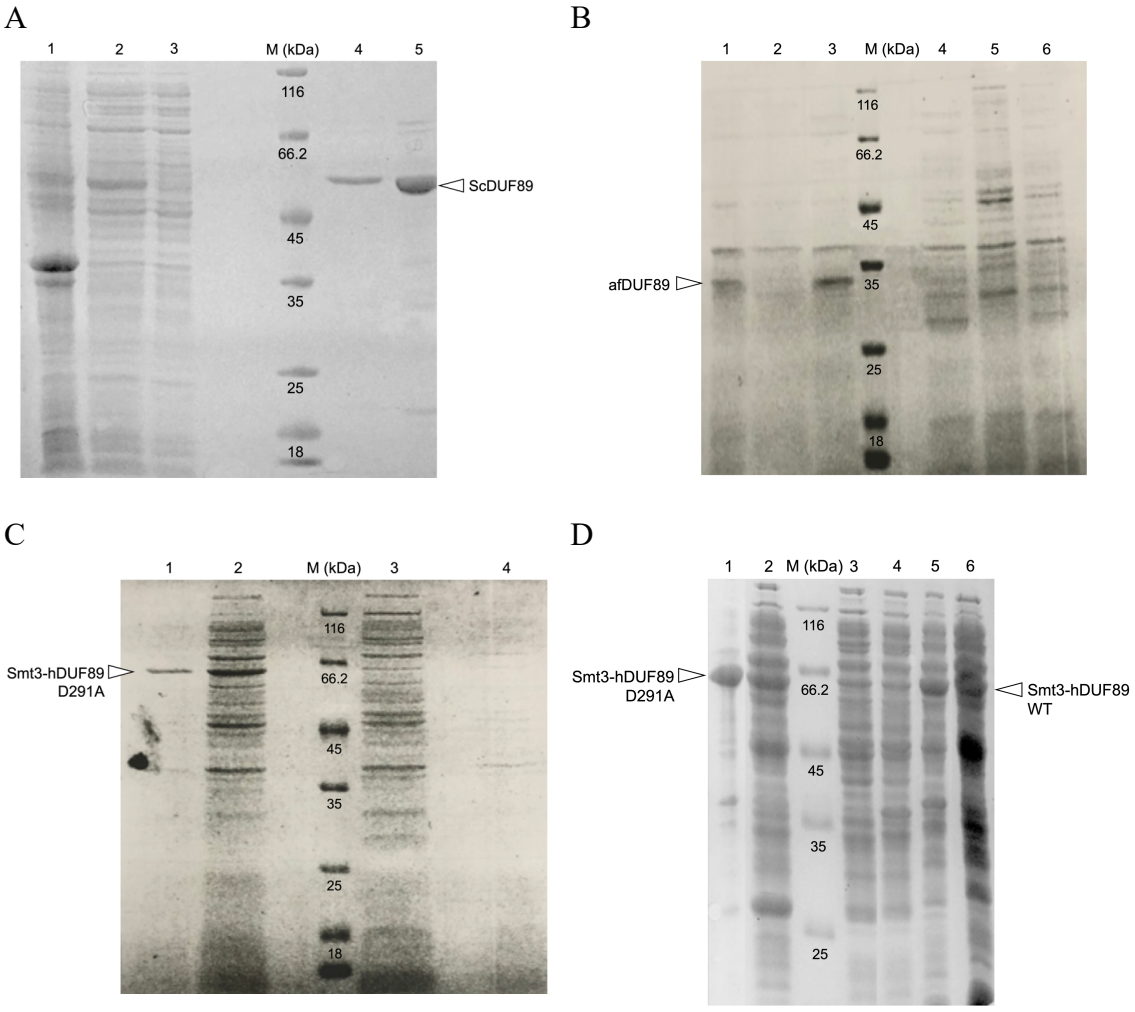


Figure 1.2 | Expression tests for recombinant DUF89 proteins (A) Expression of ScDUF89 (54 kDa) in BL21(DE3) as judged by 12% SDS-PAGE. (B) Expression of DUF89-fold containing protein AF1107 from *Archeaoglobus fulgidus* in BL21(DE3) as judged by 12% SDS-PAGE (C) Expression of D291A hDUF89 protein (51.2 kDa) in BL21(DE3) as judged by 12% SDS-PAGE (D) Expression of WT hDUF89 in BL21(DE3) RIL as judged by 12% SDS-PAGE. Marker used: EZ-run Protein Marker (Thermo)

Small-scale expression tests for the human, yeast, and archaeal DUF89 proteins were analyzed by SDS-PAGE. The ScDUF89 protein can be observed in lanes 4 and 5 following Ni-NTA elution (**Fig. 1.2A**), above the 45 kDa marker and below the 66 kDa marker, representative of the proposed MW of 54 kDa. Production of the AF1107 protein was

validated (**Fig. 1.2B**) by the predominant band present in lane 3 at 33 kDa. The result for the successful expression of the 6xHis-Smt3-DUF89 D291A human protein in BL21 (DE3) cells can be recognized in lanes 1 and 2 (**Fig. 1.2C**) presenting a prominent band at approximately 65 kDa, the predicted molecular weight of this fusion protein. Oddly, when expression tests were conducted for the WT hDUF89 protein in BL21(DE3), they failed. Following transformation in several cell-lines, expression was observed in BL21 (DE3) RIL, as evidenced by the predominant band in lanes 5 and 6 of **Fig. 1.2D**. For all subsequent purifications BL21 (DE3) was used for hDUF89 D291A, yDUF89, afDUF89, and BL21 (DE3) RIL was used for the WT hDUF89.

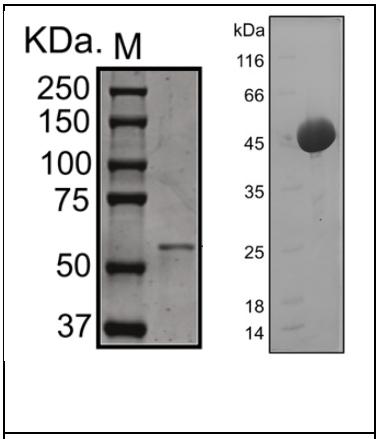


Figure 1.3 | Protein yield comparison At left is a 10% SDS-PAGE gel depicting the gene product of C6orf211 expressed in insect cells [91]. At right is our 12% SDS-PAGE gel showing hDUF89 purified after 6L *E. coli* expression.

Following the column chromatography final purification step for recombinant hDUF89 protein produced in *E. coli*, the protein sample was run and visualized on a 12% SDS-PAGE gel (**Fig. 1.3**). The left gel image depicted is from Perry *et al.,* 2015 [91], and indicates the expression level of the hDUF89 protein produced from the baculovirus system (1 μg/L). With our improved protocol, we were able to increase our yield by 5,000-fold to 5 mg/L, making crystallographic and activity studies more viable.

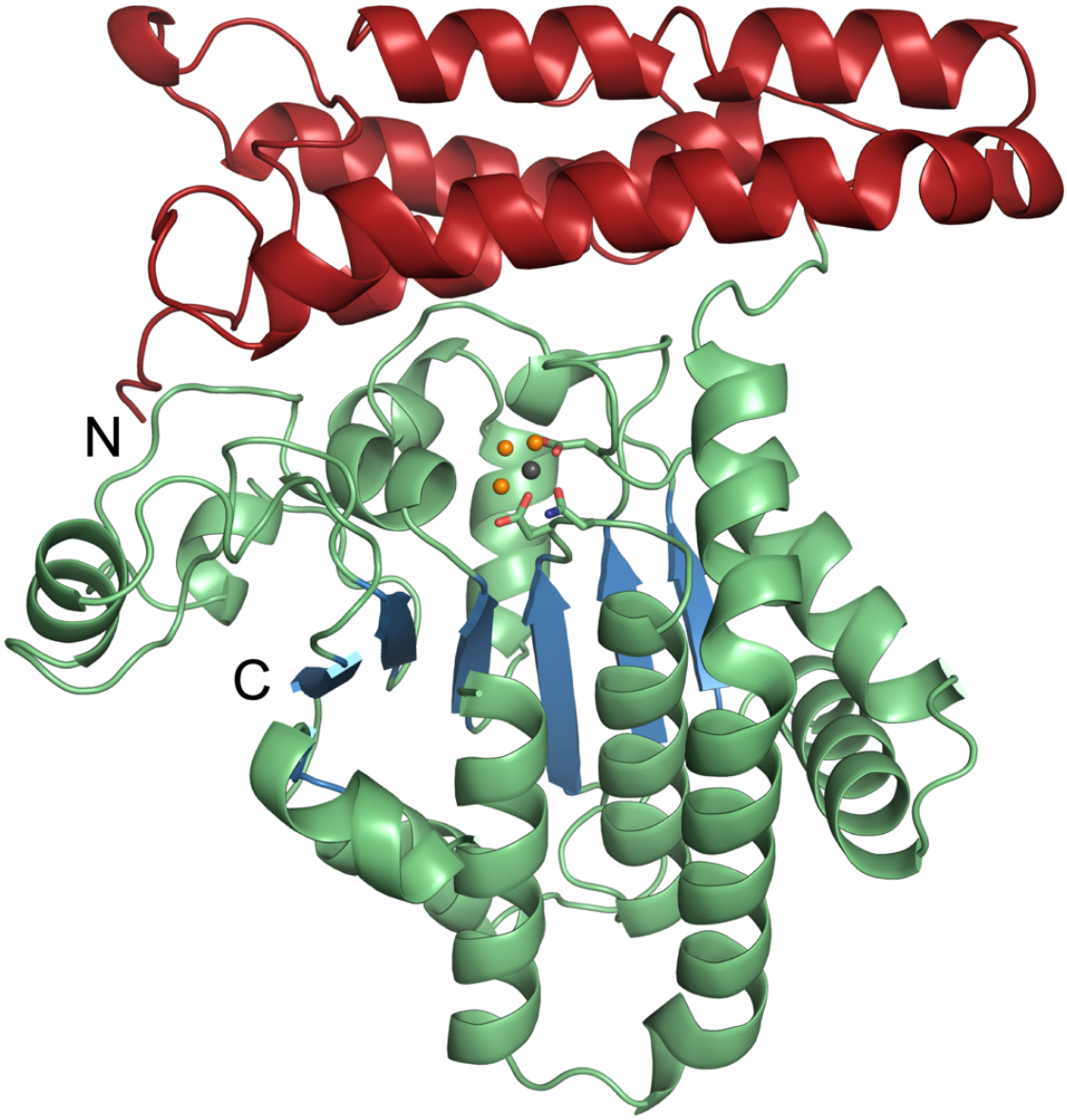


Figure 1.4 | Crystal structure of hDUF89 Cartoon representation of the monomeric protein with the α-helical N-lobe in red, and the central fold loops and α-helices in green, and the β-strands in blue. The N and C termini are labeled. The active-site magnesium is shown in grey, the coordinated waters are shown in orange, and the coordinating amino acids D253, N254, D291 are represented as sticks. Structural images were produced using PyMOL 2.0. [92]

Table 1.1 | Table of crystallographic data

	DUF89 WT	DUF89 D291A
Data collection		
Space group	$C222_1$	$C222_1$
a, b, c (Å)	90.14, 194.47, 114.20	91.26, 194.69, 114.94
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	66.49-1.85 (1.916-1.85)	99.99-2.21 (2.289-2.21)
R_{merge} (%)	11.8 (21.5)	13.4 (48.4)
I/σ (I)	8.4 (5.2)	8.1 (3.4)
Completeness (%)	98.54 (99.27)	98.66 (97.48)
Redundancy	6.1 (6.2)	6.0 (6.9)
Refinement		
No. reflections	84,394 (8,418)	50,879 (4,984)
R_{work}/R_{free} (%)	19.04/21.59 (20.98/25.28)	20.82/24.56 (28.62/33.63)
No. atoms		
Protein	7,739	6,877
Ligand/ion	4	2
Water	818	799
B-factors		
Protein	15.42	21.68
Ligand/ion	10.35	26.19
Water	26.3	29.7
r.m.s deviations		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	0.855	0.880

Highest resolution shell is shown in parenthesis
†These values are high due to presence of pseudosymmetry and tNCS
$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $I(hkl)$ is the average over all observations of reflection hkl.
$R_{factor} = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$ for all data excluding the 10% that comprised the R_{free} used for cross-validation.

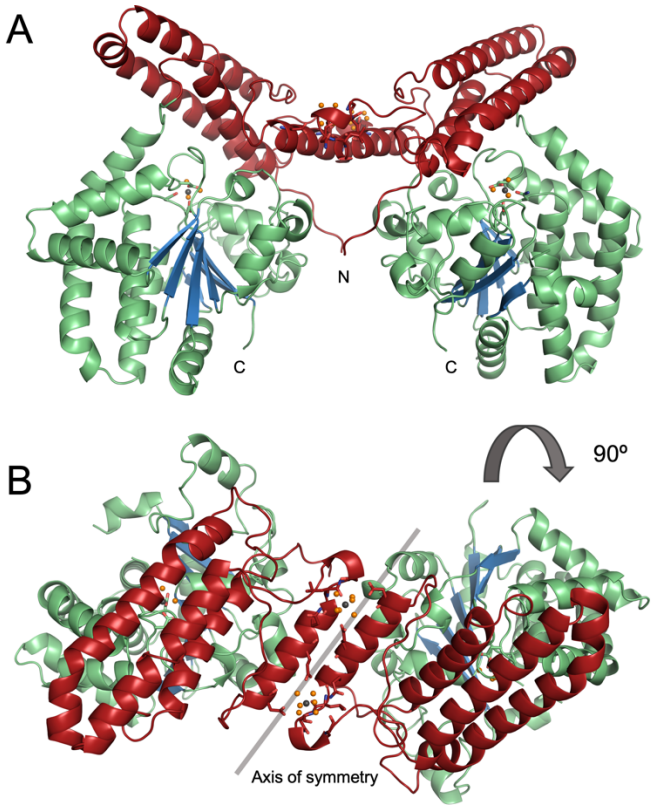


Figure 1.5 | Unit cell view of hDUF89 crystal structure (A) Cartoon representation of the unit cell forming a dimer with N and C termini labeled. (B) A 90º rotated, top-down view of a cartoon representation of the dimeric unit cell, displaying in more detail the crystal packing interface. Side chains for D86, E93, Q96, L99, N100 are depicted as sticks to show amino acids critical for magnesium coordination in the interface as well as all relevant coordinated water molecules. Structural images were produced using PyMOL 2.0

1.3.2 Structural studies of the C6orf211 protein

Human DUF89 crystal structure- The human DUF89 crystal structure was determined to a resolution of 1.85 Å, using the single-wavelength anomalous diffraction (SAD) method for phasing on the selenomethionine-modified protein crystals. The final model was built with the aid of the electron density maps and refined to produce a final model with R_{work} of 19.04 % and R_{free} of 21.59 % (**Fig. 1.4**) (**Table 1.1**). Two molecules of hDUF89 are present in the asymmetric unit, having 180º rotational symmetry along the

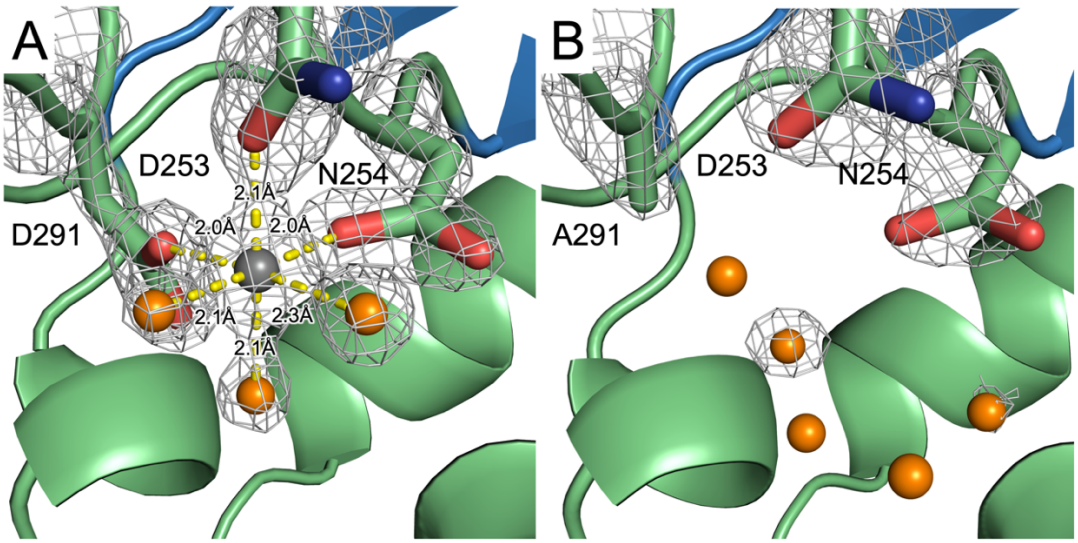


Figure 1.6 | Active site of hDUF89 WT and D291A mutant protein (A) The crystal structure of the DUF89 protein revealed a magnesium-binding site. The magnesium (gray sphere) is coordinated by D254, N254, D291, (labeled) and three water molecules (orange). The coordinate bonds are shown as yellow dashed lines. An omit map contoured at 3.0 σ is represented as a wire mesh. (B) The active site of the D291A DUF89 mutant protein shows distinct lack of metal coordination. An omit map contoured at 3.0 σ is represented as a wire mesh. Structural images were produced using PyMOL 2.0

axis of the interaction between the N-lobe of both moieties (**Fig. 1.5**). Residues 1-190 and 213-441 are readily observable in the electron density maps of each molecule in the crystallographic asymmetric unit, while residues 191-212 are disordered in both molecules. hDUF89 protein consists of 18 α-helices and 7 β-strands that form two regions of the protein structure: an N-lobe that is composed of an α-helical bundle followed by an α-β-α three-layered sandwich (**Fig. 1.4**). The two molecules in the asymmetric unit interact via their N- terminal α-helical lobes, and this interaction includes the coordination of two metal ions, which are likely to be magnesium, as judged from the electron density maps and the high concentration of this metal in the experimental precipitant solution. The first magnesium is bound by residues D253, N254, D291 (**Fig. 1.6A**) and the second is

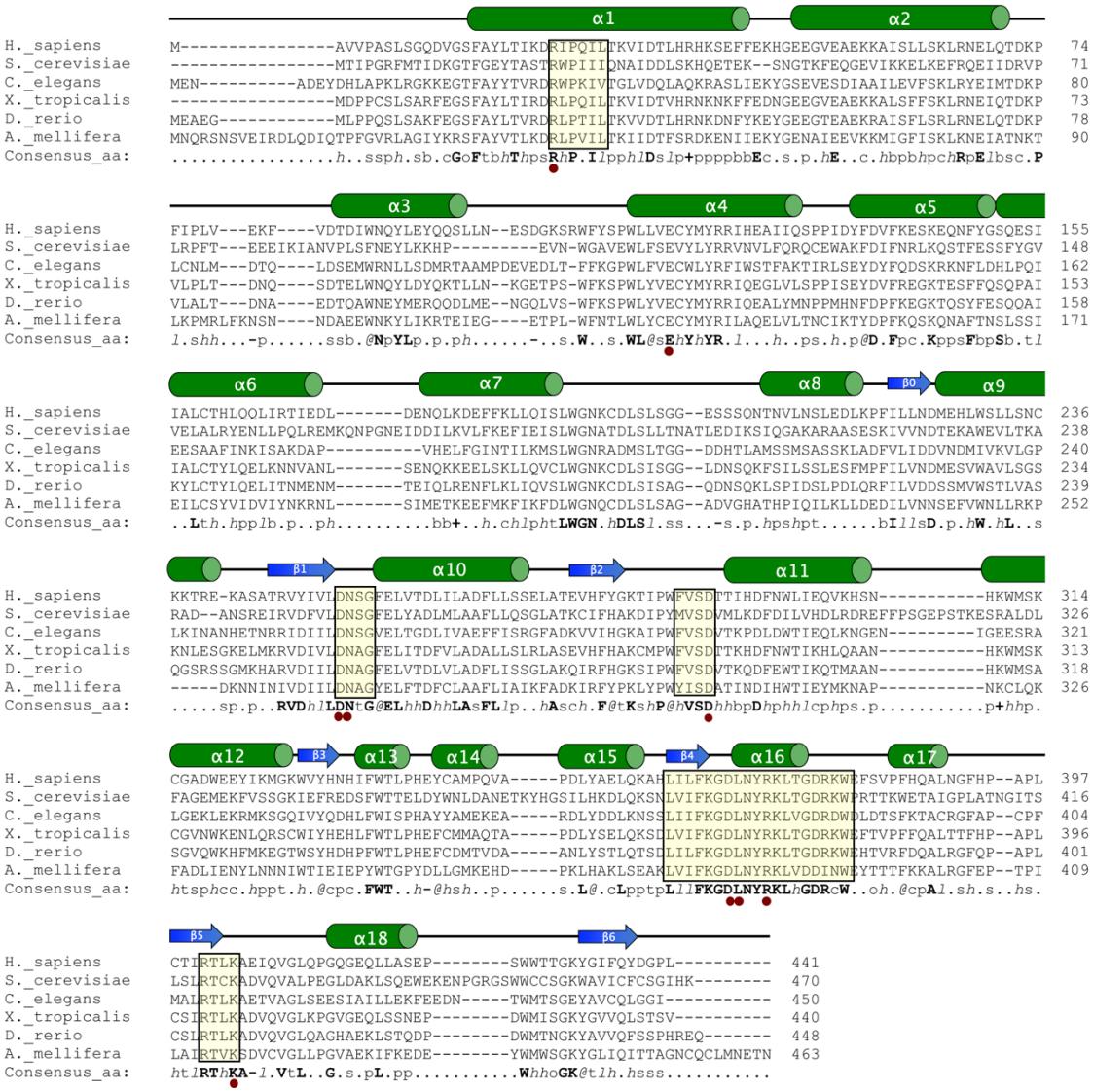


Figure 1.7 | Sequence alignment of DUF89 structural homologs Gene sequences were obtained from GENBANK or SEED using the following ascension codes: *H. sapiens* (fig|9606.3.peg.27930), *S. cerevisiae* (fig|559292.3.peg.4170), *C. elegans* (fig|6239.3.peg.22181), *X. tropicalis* (NP_001004883.1), *D. rerio* (NP_001014353.1), *A. mellifera* (XP_623407.3). Above the aligned sequence is the secondary-structural features consistent with the human structure where α-helices are shown as green cylinders and blue arrows represent β-strands. In yellow boxes are highly conserved motifs and maroon circles are conserved active-site amino acids.

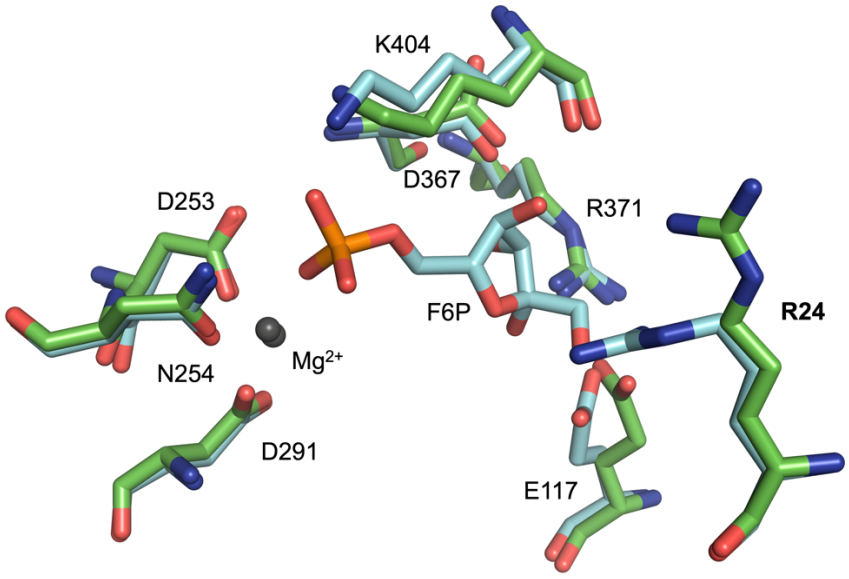


Figure 1.8 | A proposed substrate binding site Structural superimposition of *S. cerevisiae* F6P-bound protein YMR027W (PDB: 3PT1) in cyan and *H. sapiens* C6orf211 (Uniprot code: Q9H993) in green. The depicted amino acids are key to substrate hydrogen bonding or metal-coordination and are as follows with the yeast homolog in parenthesis: R24 (R23), E117 (E110), D253 (D254), N254 (N255), D291 (D292), D367 (D384), R371 (R388), K404 (K423). Structural images were produced using PyMOL 2.0

coordinated at all six points by H_2O molecules which are interacting with the following five amino acids: the side chains of D86 and E93 on one molecule, and Q96 side chain and the carboxyl backbones L99 and N100 on the second molecule (**Fig. 1.5**). Results from Xtriage [93] suggest the presence of translational non-crystallographic symmetry within all the data collected on these crystals.

Although hDUF89 packs as a dimer in the asymmetric unit cell, analysis by gel filtration chromatography indicates an in solution molecular weight of 55.3 kDa suggesting that it exists as a biological monomer (**Fig. 1.9**). The largest surface pocket in the hDUF89

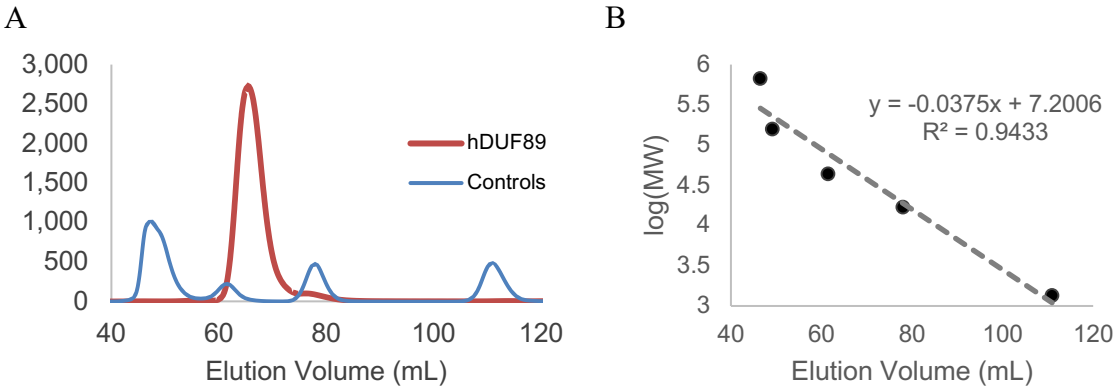


Figure 1.9 | Gel filtration analysis of hDUF89 indicates *in vitro* monomer (A) Absorbance at 280 nm against column volume of Bio-Rad Calibration standard (#1511901) is colored in blue and purified recombinant hDUF89 protein loaded on a HiLoad 16/600 Superdex 75 pg (GE Healthcare) gel filtration column is colored in red. (B) Standard line generated from log(MW) against elution volume (ml). Equation of line was used to calculate the in solution molecular weight of the hDUF89 protein.

structure contains a metal-ion that is coordinated by the side chains of D253 and N254 of the β1-α10 loop, and the D291 side chain present in the β2-α11 loop, in addition to three water molecules (**Fig. 1.6A**). The crystal structure of the hDUF89 D291A mutant protein was determined to 2.21 Å, and the alpha carbon backbone overlays with WT hDUF89 at 0.2 RMSD and shows no significant changes within secondary structure elements. The electron density maps revealed that the D291A mutation does not chelate a metal ion, in addition to a slight rotation of the N254 side chain and altered water coordination in this region (**Fig. 1.6B**).

Structural near-neighbors- The core α-β-α sandwich domain of hDUF89 starts at α6 and contains 13 α-helices, all 7 β-strands, and includes a cis-proline at position 132. This central seven stranded β-sheet has a topology similar to the Rossmann fold, and the

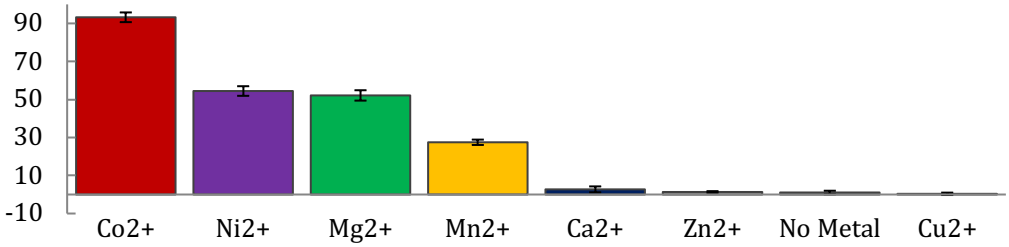


Figure 1.10 | Malachite green assay using hDUF89 with varied divalent cations Mean background (substrate alone) subtracted data are presented as means ± s.d. of three replicates in units of nmol min^{-1} mg^{-1}.

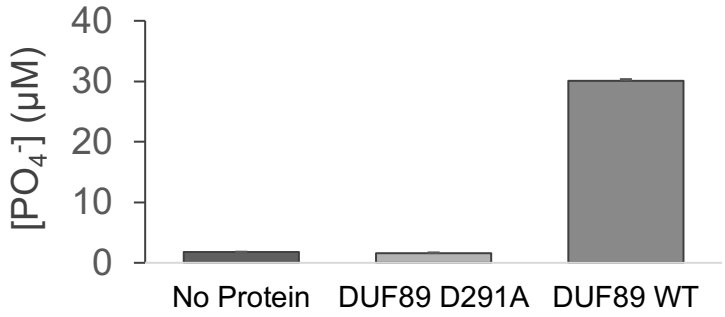


Figure 1.11 | Validation of D291A mutant inactivity Comparing free phosphate generated by WT and D291A hDUF89 proteins. Mean background (substrate alone) subtracted data are presented as means ± s.d. of three replicates

adopted numbering convention reflects that. It is this Rossmann-like fold pattern that draws the powerful resemblance between the DUF89 fold and the classical SAM-MT fold. [94], [95], but it is distinct in that strand 0 & 6 are in an anti-parallel conformation. A search for structural homologues using the DALI server [96] revealed that DUF89 from *Saccharomyces cerevisiae*, ScDUF89 (PDB: 3PT1) is the closest homolog to hDUF89 with the Cα atoms superimposing within 1.31 RMSD and the two sequences share 32.1% sequence identity (**Fig. 1.8**). ScDUF89 belongs to subfamily III of the larger DUF89

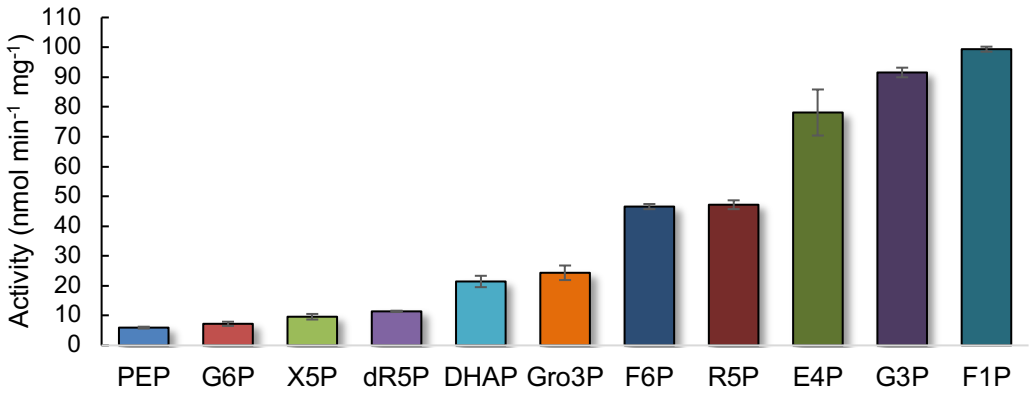


Figure 1.12 | Top 11 *in vitro* hDUF89 phosphometabolite substrates Mean background (substrate alone) subtracted data are presented as means ± s.d. of three replicates in units of nanomole free phosphate per second per milligram protein. Abbreviations can be found in Table S1.1.

family, and members of this subset were previously noted to contain conserved DNxG and RTxK motifs which are also present in hDUF89 (**Fig. 1.7**).

The hDUF89 metal binding residues D253 and N254 belong to the DNxG motif and the D291 residue belongs to the PΦxVSD motif, where Φ represents bulky hydrophobic residues, (F/W/Y) [97], [98] that we define. These two motifs involved in metal binding are conserved in all DUF89 proteins that have been structurally determined to date. The RTxK motif contains Lys404 in hDUF89, and this residue in ScDUF89 (K423) is involved binding with the phosphate group on the previously identified and crystallized substrate fructose 6-phosphate (F-6-P) in the ScDUF89 structure (3PT1.pdb). In addition to PΦxVSD we observed two further highly conserved regions shared by the human and yeast proteins, one that is composed of residues 361-378, L(I/L/V)(I/L/V)FKGDLNYRKL(V/T)GDR(K/D/N)W, which form β-strand 4 and α-helix 16 in hDUF89 and these residues are likely important to substrate interactions, as observed in the ScDUF89 structure. The second newly observed motif is a well-conserved

R(I/W/L)PxI(I/L/V) present in the N-terminal cap region, and R24 of this motif in hDUF89 belongs to α-helix 1. The Arg24 side chain is observed in distinct orientations, where, in the absence of substrate, Arg24 is pointing away from the active site in hDUF89 and it is not observed in the electron density maps of the apo ScDUF89 structure (residue R23 PDB: 5F13). In the presence of the substrate F-6-P, the *S. cerevisiae* Arg23 side chain rotates 45.8 º to move a full 4.5 Å towards the substrate and forms a hydrogen bond to the hydroxyl group on carbon 1 of the sugar-phosphate substrate (**Fig. 1.8**).

1.3.3 Functional studies of the C6orf211 protein

Metal-dependent phosphatase activity- To demonstrate metal-dependence of the enzyme, an assay was developed to define activity against a single substrate with varied metallo-cations. Buffers with each cation at 5 mM, and EDTA at 0.5 mM to chelate any contaminant metals, were used in the assay. We observed the increased activity of the enzyme in the presence of cobalt when compared to the likely more biologically relevant magnesium (**Fig. 1.10**). The "No Metal" data point represents a sample that used a buffer absent of a metal cation with 0.5 mM EDTA to chelate any divalent metal ion cofactor remaining in the active site of the protein during expression or purification. Metal-ion dependent activity was reinforced by an assay that incubated the D291A mutant protein, which was not observed to bind metal ions in our crystal structures with a known phosphometabolite substrate and that resulted in no substrate degradation (**Fig. 1.11**).

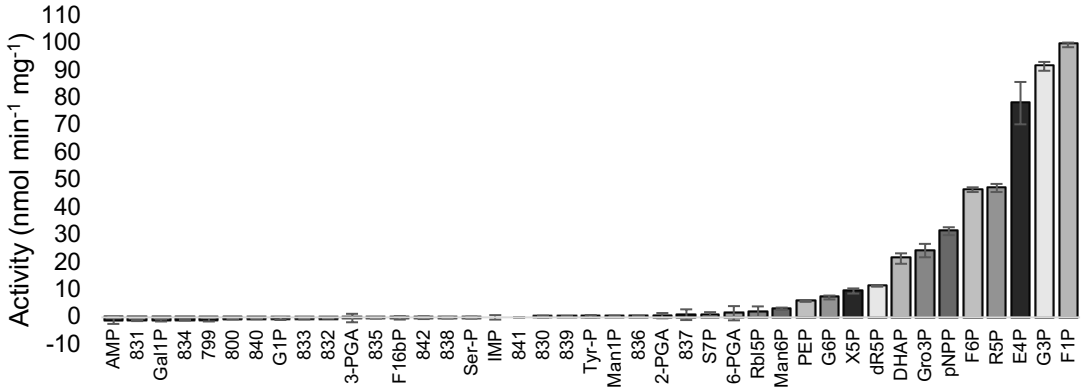


Figure 1.13 | hDUF89 activity against diverse phospho-metabolites Mean background (substrate alone) subtracted data are presented as means ± s.d. of three replicates in units of nanomole free phosphate per second per milligram protein. Abbreviations can be found in **Table S1.1**.

Ultimately, the human DUF89 protein was screened against a diverse set of 41 phospho-metabolites (**Fig. 1.13**) in the presence of 5 mM $CoCl_2$. Activity of the enzyme was observed for many phospho-metabolites related to fructose metabolism and glycolysis. The substrates for which hDUF89 has strongest activity against are shown in **Figure 1.12** where fructose 1-phosphate is hydrolyzed at a rate of 99.4 nanomoles per minute per milligram enzyme. Data obtained from a malachite green-based assay was compared against background auto-hydrolysis of the substrates and in triplicate. Although the hDUF89 is shown to have *in vitro* activity against fructose 1-phosphate, (F-1-P) it also shows pronounced activity for glyceraldehyde 3-phosphate and erythrose 4-phosphate. Fructose 6-phosphate and ribose 5-phosphate share ~50% activity, likely due to their structural similarities to F-1-P. The next pairing of substrates are dihydroxyacetone phosphate and glycerol 3-phosphate, which share ~25% activity and are structurally identical aside from the internal ketone of DHAP is hydrolyzed to a hydroxyl group on

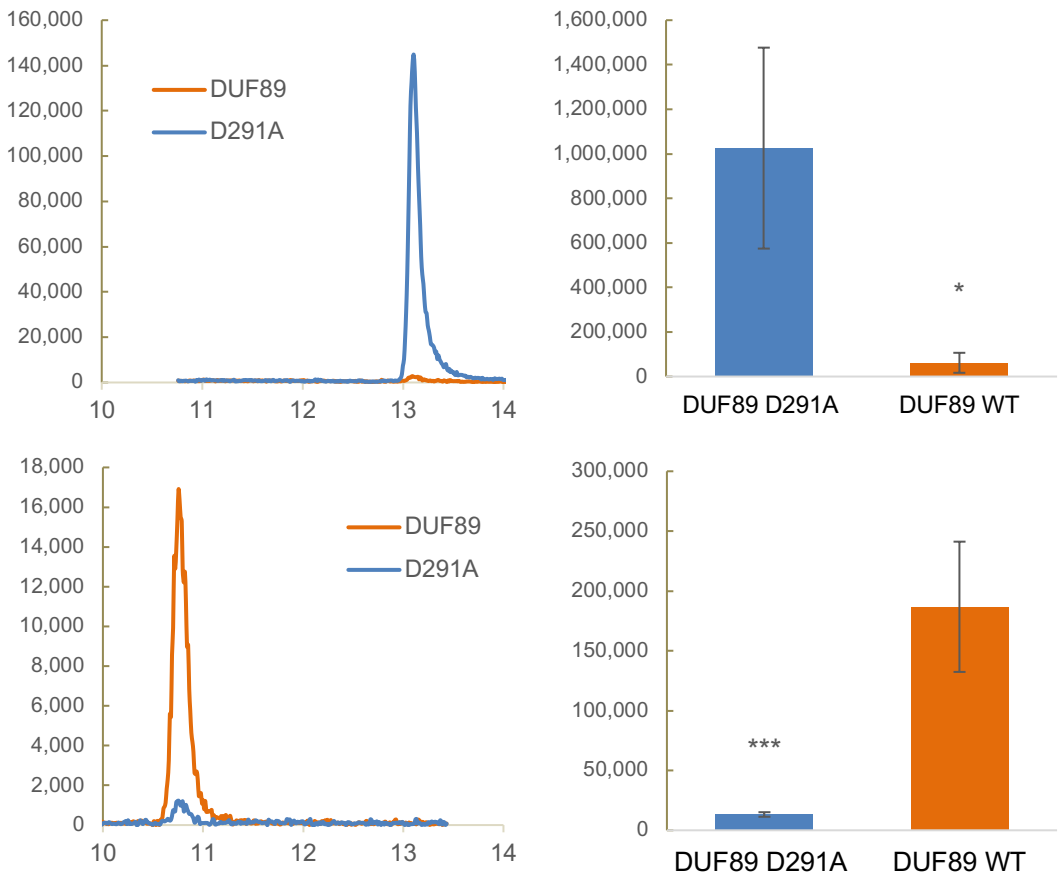


Figure 1.14 | hDUF89 F-1-P phosphatase LC-MS/MS study The top panel shows LC-MS/MS peaks in the m/z region for fructose 1-phosphate, and the bottom panel shows such peaks in the m/z region for fructose, with bar graphs, at right, representing the integration for the fructose 1-phosphate peak, and the fructose peak, respectively, from the LC-MS/MS analysis after 4 hour incubation with the wild-type and mutant DUF89 proteins presented in arbitrary units. Data are means ± s.d. of four replicates and comparisons made using Student's t test. * $p < 0.05$, *** $p < 0.001$

Gro3P. The final grouping of compounds, at ~10% activity, are structurally dissimilar: deoxyribose 5-phosphate, xylulose 5-phosphate, glucose 6-phosphate, and phosphoenolpyruvate; a 5-carbon ring, a 5-carbon chain, a 6-carbon ring, and a 2-carbon chain. Substrates below the threshold of 5% activity were excluded from **Figure 1.12** and can be found in **Figure 1.13**. These results indicate a level of specificity, as well as promiscuity, for the enzyme in terms of metal cofactor and biological substrate. Michaelis–

Menten kinetics were not determined for hDUF89, due to saturation conditions for the enzyme not being met: rate did not diminish when enzyme was at 400 nM and substrate was 1 mM (2,500-fold). It would therefore not be surprising for evolution to conserve such enzymatic features for a damage repair enzyme of being slow, promiscuous, and unsaturable.

LC-MS/MS study- To validate both the phosphatase activity of hDUF89 towards fructose 1-phosphate and the requirement of an active site metal for catalysis, hDUF89 WT and its mutant, DUF89 D291A, were incubated with F-1-P. Following incubation, extraction, and LC-MS/MS analysis, the WT sample contained only trace amounts of F-1-P when compared to the near background levels of F-1-P in the D291A mutant sample (**Fig. 1.14**). This suggests DUF89 depleted F-1-P and metal-coordination was required for this depletion. In addition, fructose, the product of this proposed reaction, was only observed in the DUF89 WT sample, suggesting that DUF89 depleted F-1-P by phosphate hydrolysis. Moreover, in an untargeted metabolomics study conducted using polar metabolites extracted from *E. coli,* revealed that most substrates for which hDUF89 has activity toward (p-value < 0.01) are uncharacterized metabolites that are not in online small molecule databases (**Table S1.4**).

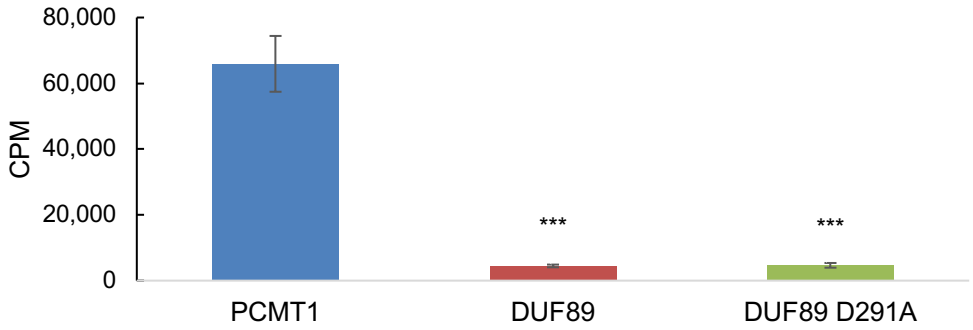


Figure 1.15 | Methyltransferase scintillation study ^{3}H-methanol scintillation counts using BT-PCNA as substrate presented in counts per million. Data are means ± s.d. of four replicates and comparisons made using Student's t test.
(Data collected in collaboration with Dr. Steven G. Clarke of UCLA)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Methyltransferase study- It was previously reported that *C6orf211* had methyltransferase activity against PCNA [91]. To probe this, an assay was developed to look for protein-o-methyl esterification adducts through the rapid degradation of the linkage and formation of methanol. This assay was originally developed by Professor Steven G. Clarke at UCLA [99] and the following enzymatic study was conducted by his lab through our ongoing collaboration. Using ^{3}H-methyl-labeled *S*-adenosyl methionine as the reagent and base treated PCNA as the target, the WT and D291A mutant hDUF89 proteins were incubated and the reaction was measured via scintillation (**Fig. 1.15**). The success of the protocol is demonstrated as evidenced by the PCMT1 control, a known protein-o-methyltransferase. A lack of signal for DUF89 WT enzyme and the D291A mutant was also observed, which would suggest the hDUF89 protein does not have methyltransferase activity.

1.4 Discussion

Posttranslational modifications (PTM) of PCNA have been studied for the last 16 years and have been determined to regulate the many key roles of PCNA in the cell [100]. One such PTM is labile methyl-esterification first reported 10+ years ago [101] and later characterized [102], yet the enzyme responsible remained elusive. A decade later in early 2015, a publication proposed that the gene product of *C6orf211* possessed such an activity. *In vitro* methyltransferase assays were performed, and structural homology studies strongly suggested that the core of this protein had a classical SAM-MT fold based on the closest observable homologs at the time [91]. This publication also suggested that the PTM imparted on PCNA operates in the DNA damage response. Our work to determine the role of this enzyme, and its potential therapeutic viability, began with cloning efforts that allowed for viable expression of the protein in *E. coli*. Shortly after cloning studies were confirmed and completed, a noteworthy study categorized DUF89 proteins across all domains of life into three subgroups, and identified activities for certain enzymes within each group [68]. This study indicated that the *C6orf211* gene encoded for a DUF89-fold protein with metal-dependent phosphatase activity, which we tested through our biochemical and structural analysis.

Our crystallographic studies did indeed reveal that the *C6orf211* gene product contained this newly characterized DUF89 fold. The active site of the enzyme resides in the central concave surface pocket on one face of the core of the Rossmann-like region. We have defined a remarkably conserved region of the protein, residues L361-W378 L(I/L/V)(I/L/V)FKGDLNYRKL(V/T)GDR(K/D/N)W, that forms β-strand 4 and α-helix

16, which serves as a substrate binding scaffold. This region also contains two substrate-binding residues, Asp367 which hydrogen bonds to the carbon-3 hydroxyl of F-6-P and Arg371 which hydrogen bonds to the carbon-2 hydroxyl of F-6-P (**Fig. 1.8**). Support for our scaffolding conclusion is evidenced by the complete lack of translation for these key amino acids when comparing the apo and substrate-bound structures. One notable difference between these two models is the 45º rotation of Arg24 to form a hydrogen bond with F-6-P carbon-1 hydroxyl. This substrate-induced motion is significant as it may be critical to substrate recognition and it does not belong to the Rossmann-like core of the protein, rather it is located in the alpha-helical N-lobe. We performed a structural homology search with our DUF89 structure using the DALI server, and the first result following other known DUF89-fold containing protein structures, a carboxy methyltransferase was highlighted (PDB: 1RJG). Although the structure does share Rossmann-like similarities to the classical SAM-MT fold, the coordinated metal in the active site is a feature not common to methyl transferases that were structurally more similar to our protein.

Investigations into the activity of the hDUF89 protein revealed a metal-dependent phosphatase activity. This was evidenced by malachite green assay (**Fig. 1.12, 1.13**), as well as our LC-MS/MS study (**Fig 1.14**). The metal-dependence of the enzyme was demonstrated by using our active site alanine mutation in both the studies and that produced values near background for all substrates tested. This point was further confirmed by conducting a malachite green assay in the presence of no metal cofactor and high levels of EDTA, a well-known chelating agent, where this reaction sample demonstrated no phosphatase activity. The renowned methyltransferase expert Steven Clarke at UCLA to

assist our studies on potential methyltransferase activity. The previous investigation of C6orf211 that noted the methyl transferase activity [91] used professor Clarke's 1984 methodology [99]. Wild-type hDUF89 protein was provided to the Clarke lab as was PCNA, the proposed substrate. Following the ^{3}H-methanol scintillation counting, methyltransferase activity was not observed (**Fig. 1.15**); in the control (**Fig. S1.1**), the hDUF89 distinctly lacked activity when ovalbumin was used as substrate. hDUF89 also failed to show the auto-methylation activity as reported previously, as in the absence of substrate, scintillation counts did not rise above background measurements. Thus, our studies reveal that hDUF89 has metal dependent phosphatase activity, but not a readily observable protein-o-methyltransferase activity.

In searching for the preferred substrate for hDUF89, a wide variety of phosphometabolites were tested, determining that fructose 1-phsophate was the most hydrolyzed substrate. This result agreed with a previous study that discovered yeast ScDuf89 likewise has the strongest activity against fructose 1-phosphate. This previous study on ScDUF89 also determined through an untargeted metabolomics study on a yeast knockout mutant of YMR027W, that fructose 1-phosphate accumulated at a statistically significant margin when grown on both glucose and fructose medium. This result was somewhat surprising, as F-1-P is a non-classical yeast metabolite, and if grown on glucose media, it should therefore not be produced. Additionally, the active site of the hDUF89 protein showed striking promiscuity for a range of distinct phosphometabolite, both in chemical structure and metabolic pathway (**Diagram 1.1**). As the enzyme has activity against a variety of phosphometabolites, it may play a role in regulation of these distinct

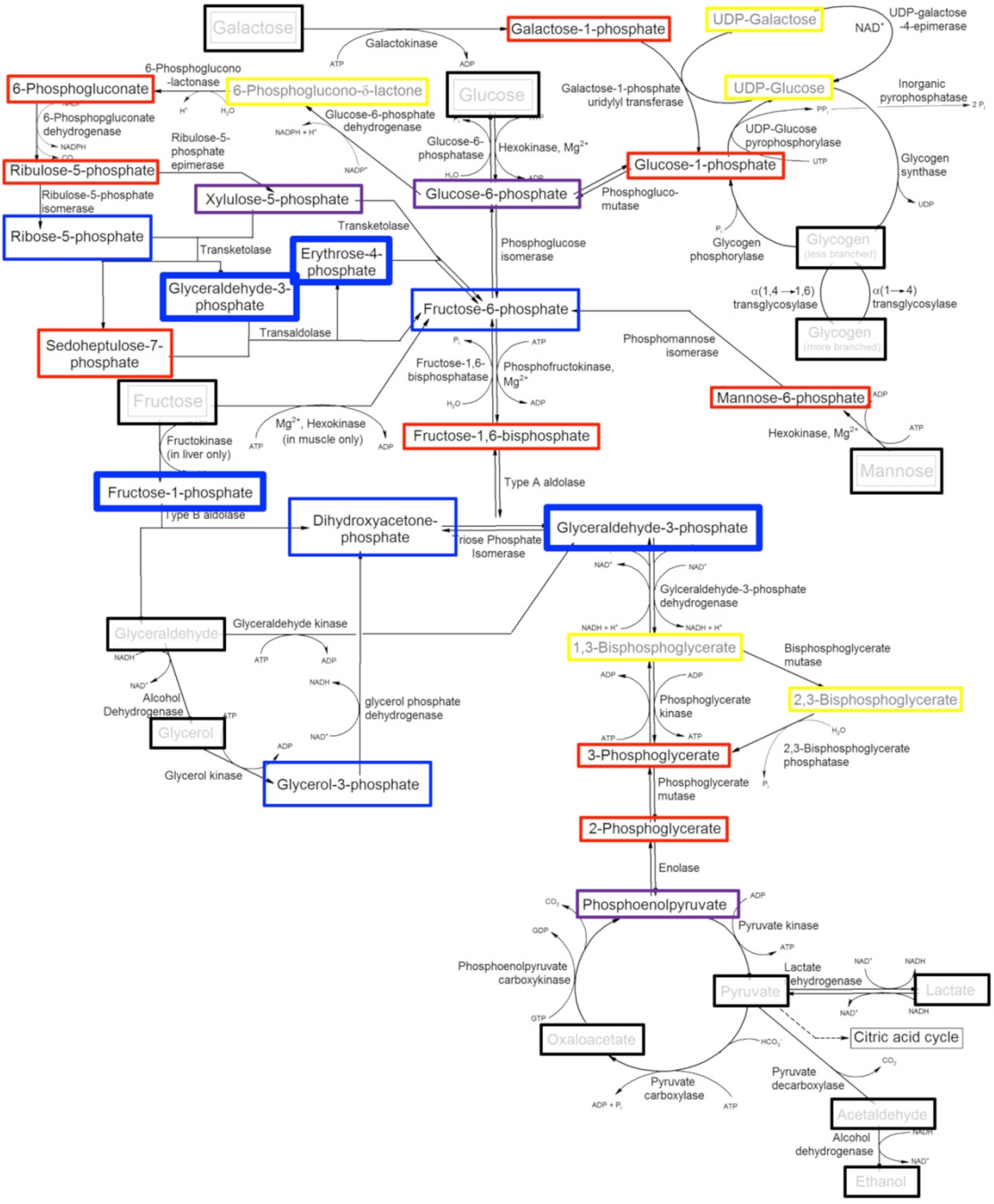


Diagram 1.1 | Proposed DUF89 activity within fructose metabolism Diagram modified from [103] wherein blue boxes represent substrates for which hDUF89 has greater than 10% activity, purple boxes are 1-5% activity, red boxes have no activity, yellow boxes were untested in our studies, and black boxes are unphosphorylated substrates.

pathways. As a means of controlling the possible overflow of metabolites from each of these pathways, the cell would likely conserve an enzyme that can dephosphorylate many high energy metabolic substrates.

For enzymes involved in pathways of secondary metabolism, lower selective pressure is exerted on their evolution. One study found that such enzymes are, on average, ~30-fold slower than those of central metabolism [104]. This may provide some insight into the relatively low activity of the hDUF89: the cell does not need it to be fast. The enzyme was tested under conditions of increased substrate concentration, and, even to the point of 10,000,000-to-1 ratio, the enzymatic rate was not diminished, indicating that the protein was not saturated with ligand of any kind. The hDUF89 enzyme therefore does not clearly satisfy Michaelis-Menten kinetics and the kcat, and Km could not be determined as such. This would further support our hypothesis that hDUF89 is a metabolite repair enzyme, as the cell would not maintain an enzyme that compete for primary metabolites, but the cell would likely evolutionarily select for promiscuous dephosphorylation activity to protect the cell from aberrant phosphorolytic damage.

A search through online bioinformatic databases found agreement with our hypothesis that hDUF89 is a metabolite repair enzyme. Expression patterns of other known regulatory proteins are observed to have low-level expression values conserved across most tissue types. Using RanGAP as an example, expression was observed in all tissues demonstrated in the UCSC Genome Browser [105] with a transcript per million (TPM) value never falling below 15. TPM is a normalization method that defines the number of RNA-seq reads for every 1,000,000 RNA molecules in the sample associated with the

gene/transcript being investigated. Similar qualities to the expression pattern of RanGAP are observed for hDUF89, which stayed within the range of confident mRNA production and cellular maintenance levels in all tissues (**Fig. S1.2**). hDUF89 expression was also be assessed using a proteomics database [106]. The *C6orf211* gene product has been associated with 27 unique peptides from 50 deposited proteomics projects totaling 94% sequence coverage. As with the microarray data, LC-MS/MS proteomic data demonstrates expression levels (log 10 normalized iBAQ intensities) consistent across most tissue types for the hDUF89 protein (**Fig. S1.3**). These data are an interesting point in the explanation of this strongly conserved gene's cellular role.

This enzyme having an evolutionarily conserved fold and being found in all tissues with activity for high energy metabolites serves as an important point in the new understanding of cellular biology as a system. For fear of overstating the important role of this gene, it remains non-essential. While the *in vitro* activity of this enzyme may show significant levels of promiscuity, the *in vivo* activity has yet to be studied. Untargeted metabolomic studies of human cell knockdowns would prove vital to this understanding, especially if activity were varied under stress conditions, notably starvation or uncontrollable (cancerous) growth.

The *C6orf211* gene product was originally proposed to be a SAM-dependent protein-o-methyltransferase, or the acidic residue methyltransferase (ArmT1), targeted to PCNA and linked to DNA damage response. Our data, however, indicates that *C6orf211* encodes for a DUF89 fold-containing protein that possessed metal-dependent phosphatase

activity against a diverse set of phosphometabolites and may play a role in metabolite damage repair through directed overflow.

1.5 Methods

1.5.1 Cloning

C6orf211 cloning - Earlier studies of the standalone human DUF89-fold containing protein (not PanK4) were conducted by one of our collaborators. We received plasmid with hDUF89 that contained an active site mutation (D291A) as a gift from Dr. Derek Hoelz. Primers (IDT) were designed for the insertion of these gene into the pET SUMO expression system vector, which had been modified in our lab by adding a multiple cloning site (*Bam*HI, *Eco*RI, *Sac*I, *Sal*I, *Hind*III, *Not*I), and can be found in **Table S1.2**. These primers were used in the following 300 µL PCR experiment, adapted from NEB: 4.32 µL of 139.3 pmol/µL forward primer, 4.24 µL of 141.8 pmol/µL reverse primer, 60 µL of 5X OneTaq Standard Reaction Buffer, 6 µL of 10mM dNTPs (NEB), 7.6 µL of 79 ng/µL hDUF89 D291A containing plasmid, and 1.2 µL at 5 units/µL OneTaq DNA Polymerase were added to 216.65 µL autoclaved ddH_2O, pipetted up and down, and aliquoted into six 50 µL PCR reaction tubes. These tubes were loaded into our Eppendorf Mastercycler with the lid set to 105 °C and the protocol set to 30 cycles of denaturation at 94 °C, annealing at 56 °C, and 90 second extension at 68 °C.

Amplification products were pooled and purified from remaining reagents, via ethanol precipitation (**Appendix A**). The pellet that remained after the ethanol precipitation was resuspended in water to which 10X CutSmart buffer and the appropriate restriction enzymes were added: 1 µL of *Bam*HI-HF and *Sal*I-HF. To prepare the vector for ligation, it was restriction enzyme digested where 50 µL of miniprepped pET SUMO MCS was incubated overnight with *Bam*HI/*Sal*I, 10X CutSmart buffer, and RNase I. After overnight

digestion, these samples run on a 1% agarose gel, and the digested plasmid was excised from the gel for gel extraction (**Appendix B**).

To perform the ligation reaction, the following reagents were mixed sequentially and incubated at 16 °C overnight: 10.5 µL autoclaved ddH_2O, 2 µL of 10X T4 Ligase Reaction Buffer (NEB), 5 µL gel-extracted and double-digested pET SUMO MCS, 2 µL of gel-extracted and double-digested hDUF89 D291A PCR product, and 0.5 µL of T4 DNA Ligase at 400 U/µL (NEB). XL-10 Gold® ultracompotent cells (Stratagene) were used for the transformation (**Appendix C**), wherein 4 µl β-mercaptoethanol was added to 100 µl of thawed XL-10 Gold® cells 2 µL of ligation reaction was then added to the cells, which were then stored on ice for 30 min, heat shocked at 42 °C for 30 sec, iced for >2 min, diluted in 300 µL SOC media, incubated at 37 °C shaking at 225 rpm for an hour, and plated onto LB-agar plates with kanamycin. The following day, 10 colonies were picked and grown overnight in a 5 ml LB culture tubes, to be miniprepped (**Appendix D**).

These ten purified plasmids first underwent a double-digest reaction. To begin, a master-mix (5X) of the restriction enzymes *Bam*HI-HF and *Sal*I-HF, autoclaved water, and 10X CutSmart buffer were added together with 1 µL of each enzyme, which was aliquoted and added to each plasmid so that each reaction uses 0.2 µL of each enzyme. These reactions were incubated at 37 °C for 4 hours, then run on a 1% agarose gel (**Fig. 1.1**). Colonies 3, 6, and 9 showed PCR product sized bands on the double-digest agarose gel, so these plasmids were sent off for sequencing that confirmed the successful insertion of the gene into the pET SUMO MCS vector; hDUF89-6 D291A from the sixth colony picked will be used in all future experiments.

YMR027W cDNA cloning- To clone the YMR027W gene into an expression vector, similar protocols to the human gene were run, with different required template DNA, primers, and restriction enzymes. The ScDUF89 protein was cloned from a *Saccharomyces cerevisiae* cDNA library (gift from Dr. Nick Boddy) and the primers (**Table S1.2**.; IDT) used were designed for insertion into the pET 22b vector using the restriction enzymes *Nde*I and *Xho*I. The annealing temperature for the PCR reaction was changed to 45 ºC. Gel extraction and ligation were performed in much the same manner with changes for the corresponding restriction enzymes. Likewise, the final restriction digest analysis (**Fig. 1.1A**) was conducted as above, but with *Nde*I and *Xho*I. Successful cloning of the insert was confirmed by Sanger sequencing.

C6orf211 SDM D291A to WT- We generated a WT *C6orf211* construct by using a single nucleotide substitution (GCU = A → GAU = D), using the Q5® Site-Directed Mutagenesis (SDM) Kit protocol from NEB to complete this reaction. Primers were designed ten nucleotides upstream of the intended change, to provide adequate time for annealing, and where the forward and reverse primers 5' ends were back-to-back to allow for linear exponential amplification. Primer (IDT) sequences are in **Table S1.2**. The SDM reaction was conducted by adding 1.25 µL of each primer at 10 µM to 12.5 µL of Q5 Hot Start High-Fidelity 2X Master Mix, 1 µL of hDUF89-6 D291A pET SUMO template plasmid, and 9 µL of autoclaved water into a PCR tube, mixing up and down, and placing the reaction tube in a thermocycler set to 25 cycles of 10 second denaturation step at 98 °C, 30 second annealing step at 56 °C, and 5 minutes of extension at 72 °C. The PCR reaction was

used immediately in the ligation reaction with 1 µL of the PCR product, 5 µL of 2X KLD Reaction Buffer, 10X KLD Enzyme Mix, and 3 µL of autoclaved water mixed and incubated at room temperature for one hour. Completed SDM reactions were transformed into DH5α by adding 5 µL of reaction mixture into 50 µL of chemically competent cells, stored on ice for 30 minutes, heat shocked in a water bath set to 42 °C for 30 seconds, and plated onto LB-agar plates with added kanamycin. 5 colonies were inoculated, mini-prepped, and sequenced, which found that colonies 2-5 had been successfully mutated into the wild-type, which we refer to as hDUF89.

Expression plasmid acquisition- The AF1104 gene, encoding for the *Archaeoglobus fulgidus* DUF89-fold containing protein, was purchased from DNASU (Clone ID: AsCD00530645). The gene was provided in the bacterial expression vector pMH4 which has ampicillin resistance and the gene was inserted following an N-terminal 6xHis-tag. The plasmid was provided as a bacterial stock in DH5α which was grown on an ampicillin supplemented selection plate, individual colonies were secondarily inoculated in LB, and the next day a miniprep (**Appendix D**) was conducted and this purified plasmid was used in subsequent expression studies.

1.5.2 Expression

Recombinant protein expression- To begin, these four plasmids were transformed into the following *E. coli* cell strains designed for bacterial expression of recombinant proteins: hDUF89 D291A into BL21 (DE3) and BL21 pLysS, ScDUF89 into BL21 (DE3), AF1104

into BL21 (DE3) and BL21 pLysS, hDUF89 WT into BL21 (DE3), BL21 (DE3) RIL, BL21 (DE3) pLysS, and Rosetta™ 2(DE3). Individual colonies from the corresponding cell strains, were inoculated in 5 mL of LB with antibiotic O/N at 37 ºC. 50 mL of LB with antibiotic was secondarily inoculated with the corresponding O/N culture and grown at 37 ºC. Once the 50 mL culture reached A_{600} = 0.6, protein expression was induced with 0.4 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG). After 4 hours, the cells were harvested by 4 ºC centrifugation at 5,000 rpm for 20 minutes. Harvested cells were resuspended in Resuspension Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 1 mM β-ME), according to the ratio of 3 mL buffer to 1 gram of cell pellet. The resuspended cells were sonicated (Q-Sonica Q125, ⅛ inch probe) at 30% amplitude for 3 minutes, with pulsing for 10 seconds on and 10 seconds off. The mixture was then centrifuged at 12,000 rpm for 20 minutes. Samples of both the soluble and insoluble fractions were run on 12 % SDS-PAGE gels (**Fig. 1.2**).

After determining cell strains that expressed our proteins of interest, production efforts were scaled up. Overnight (O/N) 7 mL cultures were inoculated using glycerol stocks generated from previously uninduced samples of plasmids in recombinant expression cell strains. From these O/N cultures 1 mL was secondarily inoculated into 1 L of LB with the appropriate antibiotic, and incubated at 37 ºC and 225 rpm for approximately 4 hours until the OD_{600nm} = 0.6 before induction with 0.4 mM IPTG. To optimally express the hDUF89 the temperature was lowered to 18 ºC, and for ScDUF89 lowered to 25 ºC, and for arcDUF89 lowered to 21 ºC and set to incubate and shake O/N

for 15 hours. The following morning the cells were harvested by 4ºC centrifugation at 5,000 rpm for 20 minutes, and the cell pellets were collected and stored at -80 ºC.

Selenomet expression- To produce our selenomethionine-containing hDUF89, the inactive mutant D291A plasmid was transformed into the B834(DE3) methionine auxotroph competent cells (Novagen) and plated onto LB-agar plates supplemented with kanamycin. For the expression of the SeMet protein, the EMBL Seleno-methionine Labeling of Proteins in *E. coli* was followed as such: one liter of medium A [107] with 50 mg of methionine was inoculated and grown at 37 °C 225 rpm until the A_{600} reached 1.04, at which point the cells were pelleted at 5,000 rpm for 20 minutes at 4 °C, and subsequently resuspended in one liter of methionine-deficient medium A and shaken for another five hours. The media was then cooled to 18 °C before being supplemented with selenomethionine for 30 minutes. Finally, protein expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside for 15 hours, followed by harvesting via centrifugation.

1.5.3 Purification

C6orf211 gene product- To purify the hDUF89 protein, WT, D291A mutant, and SeMet, the collected cells were resuspended in Resuspension Buffer at a ratio of 3 mL of buffer per 1 gram of pellet. These resuspended pellets were then disrupted by sonication (Q-Sonica Q125, ¼ inch probe) on ice at 70% amplitude for 5 minutes with pulsing for 10 seconds on and 10 seconds off. The insoluble fraction of this suspension was separated via

4 ºC centrifugation for 30 minutes at 30,000 g. The supernatant was carefully decanted and loaded onto a Ni-NTA column using an NGC Scout Liquid Chromatography System (Bio-Rad). After binding the column, the 6xHis-tagged hDUF89-Smt3 fusion protein was washed with buffer containing 30 mM imidazole and eluted against a gradient to 500 mM imidazole. Following elution, the 6xHis-Smt3 tag was removed after a fifteen-hour incubation at 8º C with recombinant ubiquitin-like protease 1 (Ulp1) at a ratio of 1 µg Ulp1:1 mg 6xHis-Smt3-hDUF89. The digest reaction was diluted to 50 mM NaCl and loaded onto HiTrap Q FF (GE Health Sciences) and eluted against a gradient to 500 mM NaCl. Fractions containing hDUF89 were concentrated to 5 mL and loaded onto and run down an S-75 column (GE Health Sciences), fractionating every 2 mL. Samples containing purified hDUF89 protein were determined by SDS-PAGE gel and were concentrated to 13 mg/ml ($\varepsilon = 1.982$) in 20 mM Tris-HCl pH 7.5, 120 mM NaCl, 1 mM β-mercaptoethanol, and 0.1 % sodium azide storage buffer, flash-frozen in liquid nitrogen, and stored at -80 ºC. Similar purification protocols were conducted on the yeast and archaeal proteins.

1.5.4 Crystallography

Vapor diffusion- To find conditions suitable for crystallization we used both the Phenix (Art-Robbins) and the Mosquito (TTP Labtech) automated liquid handling robots to set up the crystal plates with purified protein at 10 mg/mL against the following sparse matrix kits: Hampton Research Index, PEG-Ion Screen, Salt Rx, Grid Screen, Crystal Screen 1-2, and Anatrace kits Top 96, and MCSG 1-4. Small crystals were first observed in both the Crystal Screen 1 condition A6 and MCSG Top96 condition A1: 30% PEG 4K, 200 mM

$MgCl_2$, 100 mM Tris pH 8.5 at 4 ºC. Following a variation of protein concentration from 8-20 mg/mL and PEG concentration from 10-40% the final condition which grew full sized (250 x 100 x 80 μm) crystals in 5 days was refined to 18% PEG 4K, 200 mM $MgCl_2$, 100 mM Tris pH 8.5 in a 2:2 μL sitting drop vapor diffusion tray (HR1-002) with 400 μL reservoir solution under 70 μL Al's oil (HR3-413). These conditions were used for the three protein samples being investigated (SeMet, D291A, and WT). The SeMet and D291A crystals were flash-frozen directly, whereas for the WT crystals, 30% (w/v) xylitol was used as cryoprotection before shipment to two synchrotron facilities. SAD-based X-ray diffraction data for the SeMet hDUF89 D291A mutant crystal was collected at BL9.0.3, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford, CA. To obtain single-wavelength anomalous signal, a fluorescence scan was conducted and determined the peak wavelength to be 0.979 Å, 12,662 eV, with 15º wedges at +/- 180º. Crystals belonged to the $C222_1$ space group. Data was obtained for both WT and D291A crystals at the SIBYLS synchrotron beamline BL12.3.1, Advanced Light Source, Lawrence Berkeley National Laboratories, Berkeley, CA. For the wild-type, a 6M Pilatus detector was used with an 11,110-eV energy beam collecting 0.75 second exposures with 0.2º oscillations for 180º for a total of 900 images. For the D291A crystal, a CCD detector was still installed so the crystal was exposed to an 11,999 eV energy beam for 2 second exposures, 1.0º oscillations for the full 180º rotation.

Structural determination- Diffraction data were indexed, integrated, and scaled using iMOSflm [108], pointless, and aimless scripts in the CCP4 program suite [109]. Phasing

information from the SeMet-based data was obtained using the AutoSol extension of PHENIX [110], [111], with the single-wavelength anomalous data and an I-TASSER [112]–[114] generated homology model based on the yeast homolog (3PT1.pdb). The resultant structure was used as the molecular replacement model to define phasing information for the next two structures. Molecular replacement revealed two molecules in the asymmetric unit, no systematic absences, and was successful only when the "Account for translational NCS if present" box was unchecked using the Phenix.phaser GUI [115]. Model building and refinement of the three crystal structures was completed using phenix.refine [116], and placement of waters, the 4 magnesium atoms, and amino acids as well as mutations were all accomplished using COOT [117]. The WT and D291A hDUF89 crystal structures will be deposited into the Protein Data Bank (PDB) at the time of publication submission. Structural representations and models were generated using PyMOL 2.0 [92].

1.5.5 Enzymatic activity

Malachite green- To determine the enzymatic preference for divalent cation, phosphatase activity was measured with varied metal cofactors using the Malachite Green Phosphate Assay Kit (MAK307 Sigma-Aldrich). To conduct the enzymatic assay, a reactions mixture was generated with 10 µL 800 nM hDUF89, 10 µL HEPES-NaOH buffer, 10 µL 10:0.1 mM metal co-factor:EDTA, 10 µL 1 mM fructose 1-phosphate at 30 °C and at incremental time points the reaction was quenched. Absorbance was measured in a 96-well format at 620 nm using a Multiskan™ FC Microplate Photometer (Thermo Scientific). Free

phosphate concentration was estimated by producing a standard curve from a stock solution serial dilution of ammonium phosphate (**Fig. 1.10 and 1.11**).

A more sensitive assay was developed that utilized a spectrophotometer, as opposed to the 96-well format plate reader used for the kit. To achieve the results seen in **Figure 1.12** and **1.13** hDUF89 protein at 100 nM was incubated at 37 ºC for one hour with phosphometabolite substrates at 100 µM concentration in a reaction buffer of 20 mM HEPES pH 7.4, 120 mM NaCl, 5 mM $CoCl_2$, 0.5 mM EDTA. The reaction mixture was then diluted to 1000µL with H_2O, and 250 µL of malachite green reagent was added to quench the reaction. The malachite green reagent solution was composed of 14.95 mM malachite green oxalate, 1.19 µM ammonium molybdate, >2.3 M H_2SO_4, and 0.17% Tween-20. This quenched reaction was then incubated 30 minutes at room temperature for color development, and absorbance was measured at 620 nm using a UV-2600 Shimadzu spectrophotometer. Studies on the activity of WT hDUF89 and hDUF89 D291A mutant enzyme were conducted in the presence and absence of divalent metal ions, using the same reaction and buffer conditions. Controls were run in the absence of enzyme and/or substrate, and all measurements were conducted in triplicate. Free phosphate concentration was estimated by producing a standard curve from a serial dilution of ammonium phosphate.

LC-MS/MS studies- To conduct our untargeted metabolomics study, *E. coli* strain BL21(DE3) was grown normally, harvested, and the polar and non-polar solutes were extracted. Both sets of solutes were incubated with recombinant wild-type and inactive

human DUF89 protein. Following background analysis and a statistical comparison between the active and inactive protein samples (**Table S1.4**). Of note, several of the features with lowest p-values are uncharacterized or unknown metabolites, meaning their m/z and retention times do not correspond to any known metabolite.

Next, we wanted to verify our results from the malachite green study which found activity against fructose-1-phosphate. To conduct our targeted LC-MS/MS study, the reaction mixture was a dilution of both active and inactive protein at a final concentration of 400 nM in a buffer composed of 5 mM HEPES pH 7.5, 50 mM NaCl, 2.5 mM $CoCl_2$, 0.25 mM EDTA. Reactions were conducted against a final concentration of fructose 1-phosphate at 10 μM for 4 hours at 37 ºC. Each reaction mixture was conducted in quintuplicate (n = 4) and provided to the Metabolomics core at UCR (**Fig. 1.14**).

Separation and detection of fructose 1-phosphate and fructose were performed on a TQ-XS triple quadrupole mass spectrometer (Waters) coupled to an I-class UPLC system (Waters). Separations were carried out on a ZIC-pHILIC column (2.1 x 150 mm, 5 μM) (EMD Millipore). The mobile phases were (A) water with 15 mM ammonium bicarbonate adjusted to pH 9.6 with ammonium hydroxide and (B) acetonitrile. The flow rate was 200 μL/min and the column temperature was held at 50 °C. The injection volume was 1 μL. The gradient was as follows: 0 min, 90% B; 1.5 min, 90% B; 16 min, 20% B; 18 min, 20% B; 20 min, 90% B; 28 min, 90% B. The MS was operated in selected reaction monitoring mode (**Table S1.3**).

Methyltransferase activity- Conducted by Dr. Steven G. Clarke's laboratory at UCLA. To conduct the scintillation count experiment, 20 µL ^{3}H-methyl-labeled *S*-adenosyl methionine was mixed together with 5 µL of enzyme (DUF89, DUF89 D291A, and PCMT1 as control) at 0.4 µg/µL and 5 µL of substrate (base-treated PCNA and ovalbumin as control) at 2 µg/µL in 20 µL HEPES buffer at pH 7.5 with 100 mM NaCl and incubated for 30 minutes at 37 ºC in a water bath. The tubes were then very briefly spun in a microcentrifuge and frozen at -20 °C. To quantitate the number of methyl-esters generated by the DUF89 and isoaspartyl methyltransferase (PCMT), the reactions were thawed on ice and 100 µL of 0.2 M NaOH was added to each, capping each immediately and vortexing. Next, 100 µL of the mixture was added to filter paper and squeezed into the neck of a 20 mL scintillation vial containing 5 mL of scintillation cocktail and securing the cap. Vials for all reactions were then incubated at room temperature for 2 hours before being counted on a scintillation counter (Beckman). ^{3}H-Methanol counts were measured and used to compute the values for methyl transferase activity (**Fig. 1.15**).

1.6 Supplemental Tables and Figures

Table S1.1 | Phosphometabolite abbreviations

Abbr.	Compound
F1P	Fructose 1-phosphate
G3P	Glyceraldehyde 3-phosphate
E4P	Erythrose 4-phosphate
R5P	Ribose 5-phosphate
F6P	Fructose 6-phosphate
pNPP	p-Nitrophenyl Phosphate
Gro3P	Glycerol 3-phosphate
DHAP	Dihydroxyacetone phosphate
dR5P	Deoxyribose 5-phosphate
X5P	Xylulose 5-phosphate
G6P	Glucose 6-phosphate
PEP	Phosphoenolpyruvic acid
Man6P	Mannose 6-phosphate
Rbl5P	Ribulose 5-phosphate
6-PGA	6-phosphogluconic acid
S7P	Sedoheptulose 7-phosphate
837	D-myo-Inositol-1,2,3,4-tetraphosphate (sodium salt)
2-PGA	2-Phosphoglyceric acid
836	D-myo-Inositol-1,3-diphosphate (sodium salt)
Man1P	Mannose 1-phosphate
Tyr-P	o-Tyrosine-L-phosphate
839	D-myo-Inositol-1,2,3,6-tetraphosphate (sodium salt)
830	D-myo-Inositol-4,5-diphosphate (sodium salt)
841	D-myo-Inositol-1,2,4,5,6-pentaphosphate (sodium salt)
IMP	Inositol monophosphate
Ser-P	o-Phospho-L-serine
838	D-myo-Inositol-1,2,3,5-tetraphosphate (sodium salt)
842	D-myo-Inositol-2,3,4,5-tetraphosphate (ammonium salt)
F16bP	Fructose 1,6-bisphosphate
835	D-myo-Inositol-1,5-diphosphate (sodium salt)
3-PGA	3-phosphoglyceric acid
832	D-myo-Inositol-3,4,5-triphosphate (sodium salt)
833	D-myo-Inositol-4-phosphate (ammonium salt)
G1P	Glucose 1-phosphate
840	D-myo-Inositol-1,2,4,5-tetraphosphate (sodium salt)
800	D-myo-Inositol-1,3,4,5-tetraphosphate (sodium salt)
799	D-myo-Inositol-1,4,5-triphosphate (potassium salt)
834	D-myo-Inositol-1,4-diphosphate (sodium salt)
Gal1P	Galactose 1-phosphate
831	D-myo-Inositol-2,4-diphosphate (sodium salt)
AMP	Adenosine monophosphate

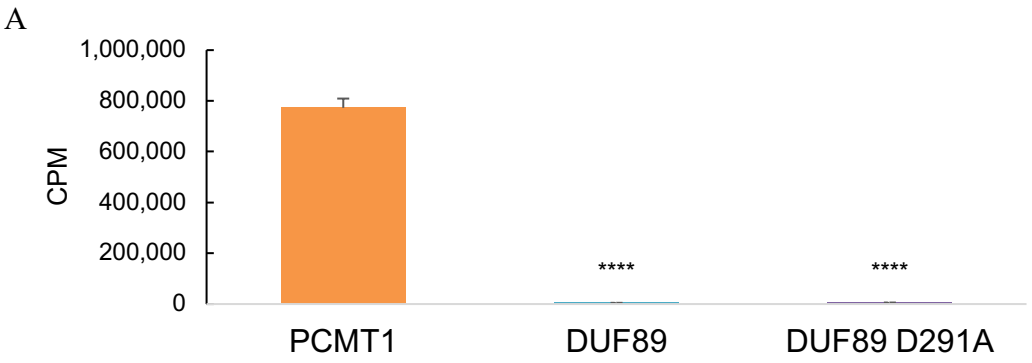


Figure S1.1 | Methyltransferase scintillation study against ovalbumin ^{3}H-methanol scintillation counts using ovalbumin as substrate.
(* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

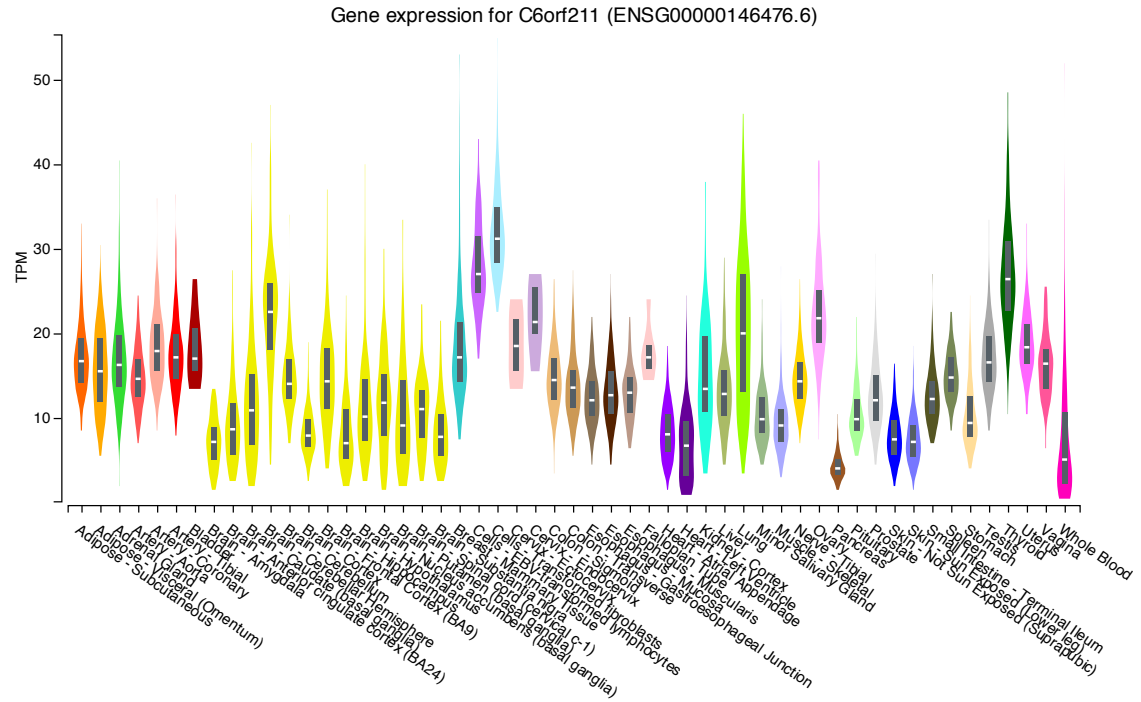


Figure S1.2 | Gene expression profile of C6orf211 Data was obtained from the UCSC Genome Browser Human Assembly Dec. 2013 (GRCh38/hg38) and is sorted alphabetically and colored according to tissue [105]. Units are in TPM.

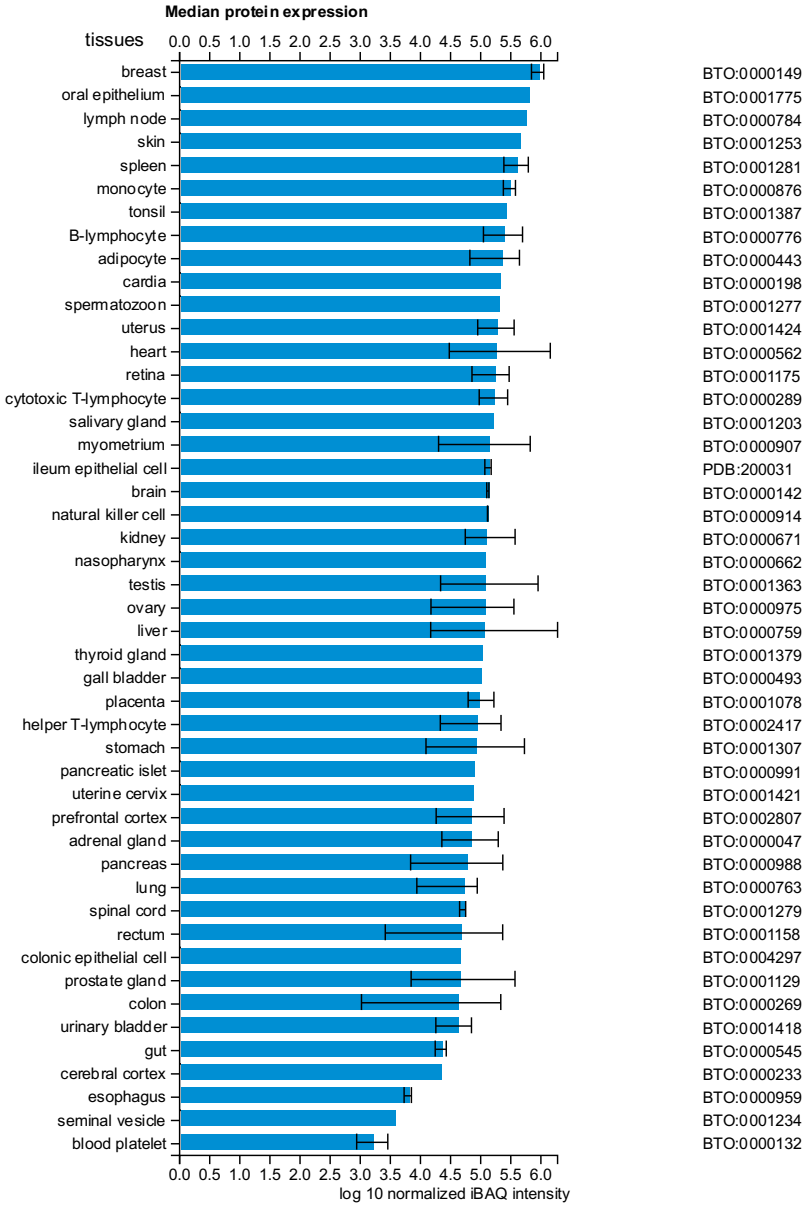


Figure S1.3 | Proteomic database entry for C6orf211 Data was obtained from Proteomicsdb.com and tissues are sorted in descending order [118], [119]. Units in normalized log 10 normalized iBAQ intensities.

Table S1.2 | All DUF89 primers	
Primer	Sequence
C6orf211 Forward	5'-CGC GGA TCC ATG GCT GTC GTC CCG GCG-3'
C6orf211 Reverse	5'-ACG CGT CGA CTC AAA GGG GAC CAT CGT ACT-3'
YMR027W Forward	5'-GGG AAT TCC ATA TGA TGA CTA T C CTG GAA GA-3'
YMR027W Reverse	5'-CCG CTC GAG TAT ACC AGA GCA GAA ACA-3'
C6orf211 SDM F	5'-TTT GTT TCT GAT ACT ACT ATA CAT G-3'
C6orf211 SDM R	5'-CCA TGG AAT TGT TTT TCC-3'

Table S1.3 | LC-MS/MS specifications					
Molecule	**Parent (m/z)**	**Collision (eV)**	**Product (m/z)**	**Dwell (ms)**	**Cone (V)**
F-1-P	259	33	79	200	35
F-1-P	259	15	97	200	35
Fructose	179	14	71	200	35

Table S1.4 | Untargeted metabolomics

Feature ID	Retention Time	m/z	neutral mass	Annotation	p-value
11.06_231.0267	11.06	231.0267		C5 sugar alcohol	5.269E-06
9.28_285.0447	9.28	285.0447		no confident matches	3.729E-05
11.20_261.0372	11.2	261.0372		C6 sugar alcohol	0.0003505
9.29_249.0235	9.29	249.0235		no confident matches	0.001085
8.35_209.1040	8.35		209.104	no confident matches	0.003006
8.73_224.0917	8.73	224.0917			0.0032086
14.01_145.0969	14.01	145.0969			0.0040055
8.34_438.1702	8.34	438.1702			0.0083842
8.42_313.1281	8.42	313.1281			0.0084177
8.84_292.1152	8.84	292.1152			0.0091779
8.31_386.1809	8.31	386.1809			0.0098649
4.07_238.1718	4.07	238.1718		no confident matches	0.0099619

Features with p values > 0.01 were excluded.

Chapter 2
HtrA1

Chapter 2: HtrA1

2.1 Abstract

HtrA1 is a secreted protease that degrades misfolded or unfolded proteins in the extracellular matrix. A predominant disease haplotype in the region of *HTRA1* has been associated with major genetic susceptibility to AMD, where a gene expression study discovered the mutation increased expression of *HTRA1* in patients. Herein, we outline our early-stage drug discovery efforts to target the HtrA1 serine protease that has clear links to wet age-related macular degeneration (AMD) pathology, and potential links to dry AMD. *In vitro* assays were developed to apply a structure-based drug discovery (SBDD) approach to this therapeutic target. Testing 81 compounds identified through artificial intelligence-based computational aided drug discovery (CADD) resulted in 4 compounds (4.9% hit rate) having observed low to mid micromolar IC_{50} values. In our real-time fluorometric assay. Co-crystallographic studies were conducted with our strongest hits, where data collection and subsequent analyses are currently on-going.

2.2 Introduction

The HtrA1 protein belongs to a family of evolutionarily conserved serine proteases that play important roles as modulators of physiological processes such as mitochondrial homeostasis, cell signaling, and apoptosis [120]–[123]. HtrA proteases function by recognizing aberrant hydrophobic regions of proteins and degrading them. The characteristic feature of the HtrA family members is the evolutionarily conserved protease domain (PD) which adopts a trypsin-like fold with His, Asp, and Ser residues comprising the active site catalytic triad [124]. HtrA enzymes also maintain at least one C-terminal PDZ domain. The PDZ abbreviation is an initialism combining the first letters of the three proteins first identified that shares the domain: post-synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1) [125]. Subtle differences exist between the mature forms of HtrA1, 3, and 4, but they share an N-terminal signal peptide (SP) followed by a Mac25-like region (tandem IGFBD and Kazal-type motif), which acts in a regulatory role (**Fig. 2.1**).

In contrast, HtrA2 contains an N-terminal mitochondrial targeting sequence (MTS) followed by a transmembrane domain (TM) and a tetrapeptide AVPS motif that mediates interaction with inhibitor of apoptosis proteins (IAPs). As evidenced by their targeting sequences, HtrA1, 3, and 4 are extracellular proteases where HtrA2 is localized to the mitochondrion, more specifically the PD and PDZ domains are exposed into the mitochondrial intermembrane space. There, HtrA2 plays a dual role in cellular physiology; it participates in the maintenance of mitochondrial homeostasis and functions as an important inductor of cell death triggering apoptosis of irreversibly

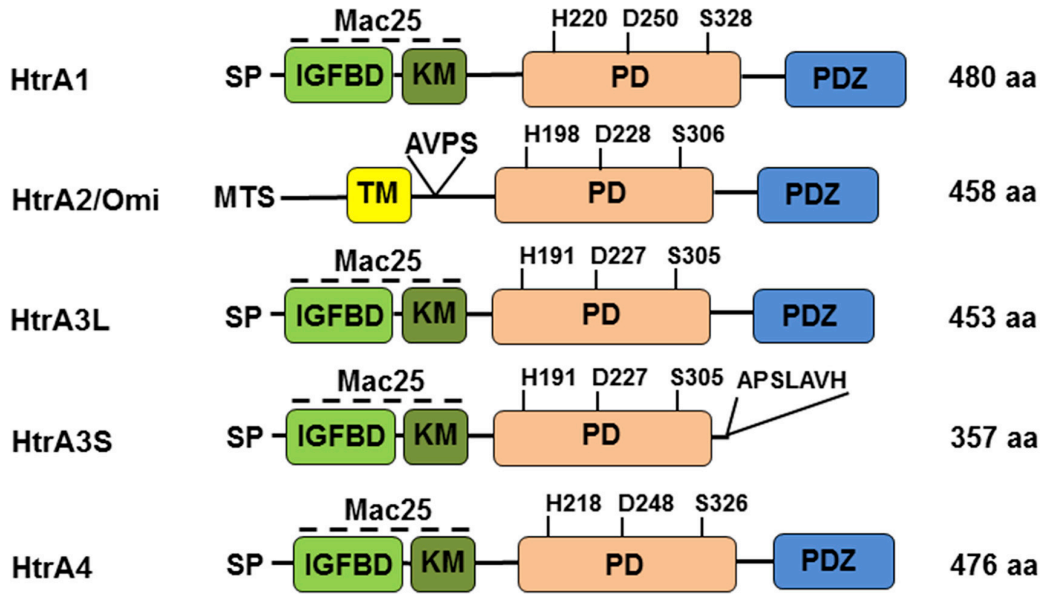


Figure 2.1 | Domain architecture of HtrA proteins Adapted from [121].

damaged cells [126], [127]. Notably, in missense and knockout mutation of HtrA2, both mouse strains exhibited typical phenotypic features for parkinsonian syndrome [128]–[130].

In the extracellular matrix, HtrA1 was observed to cleave a large number of substrates, many of which are extracellular matrix components, such as amyloid precursor protein, biglycan, decorin, aggrecan, fibromodulin, and elastin [131]. The HtrA1-mediated proteolytic cleavage of some of these substrates may be directly linked to HtrA1's biological functions and the pathological phenotypes, including apoptosis, tumor suppression, arthritis, bone mineralization, Alzheimer's disease, and age-related macular degeneration (AMD) [131]. Therefore, identifying inhibitors targeted to HtrA1 could prove therapeutically interesting. However, due to high levels of structural homology with HtrA2

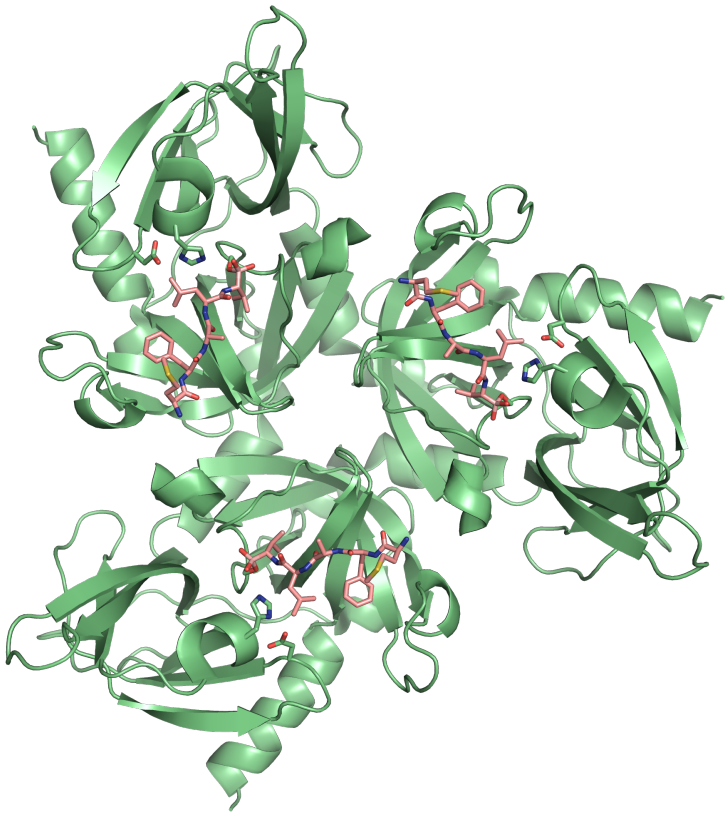


Figure 2.2 | Crystal structure of HtrA1 with substrate bound Structural representation where three HtrA1 protein molecules are in green and active site bound peptide substrate is in pink as generated from 3NZI.pdb using PyMOL 2.0 [92].

and its cellular homeostatic role, inhibitors for HtrA1 must be remarkably selective, and thus counter-screening against HtrA2 will be a key component of drug discovery efforts.

For the HtrA1 protein, structural information has already been elucidated. The structure revealed PD domain mediated homo-trimerization with all three active sites on one face of the superstructure resembling a flat disk [132] (**Fig. 2.2**). Although included in the crystallization construct (158-480), the PDZ domain was not present in the electron density maps. In the peptide-mimic substrate-bound form of the structure, the amino acids of the catalytic triad are positioned properly in distances appropriate for electron transfer

and catalysis. Loops surrounding the active site are also displaced, and notably L3 interacts with the peptide and contributes to the formation of the hydrophobic core of the substrate binding site. This crystal structure therefore represents the most accurate snapshot of a catalytically active HtrA1 enzyme, and this structure will be used in SBDD efforts.

Through our active collaboration with Dr. Kang Zhang at UCSD, a renowned ophthalmologist, we are pursuing HtrA1 as a target for age related macular degeneration (AMD). AMD is the leading cause of visual impairment among the elderly of developed countries [133]. The key characteristic of this disease is an irregularity in the retinal pigment epithelium (RPE) which results in loss of central vision [134]. Advanced AMD can be classified in two forms: geographic atrophy (GA) or choroidal neovascularization (CNV). GA, or dry AMD, is characterized by a regional loss of the RPE and eventual degeneration of the overlaid photoreceptors. The second type of AMD is CNV, or wet AMD, which is characterized by abnormal growth of blood vessels that leads to bleeding and fluid leakage the eye. Wet AMD accounts for more than 80% of severe blindness caused by AMD.

Early reporting has suggested the pathogenesis of AMD is related to abnormal levels of vascular endothelial growth factor (VEGF) [135]–[137]. Following these reports, several anti-VEGF therapies have become the mainstay of treating wet AMD [138], [139], all of which employ monoclonal antibodies (mAbs) that bind to VEGF and attenuate its angiogenic function in the retina [140]–[142]. These therapies have been shown to be effective in slowing wet AMD progression and are the primary treatment for the disease. However, studies have found that less than 40% of patients can benefit from maximal

treatment response [143], and intravitreal injection of anti-VEGF agents may be associated with devastating complications [144]. Further, the developed mAbs only target VEGF, the terminal effector of the neovascularization pathway. Therefore, designing an inhibitor for an earlier contributor to the disease-state may prove valuable.

Previous research has associated the *HTRA1* gene with AMD [145]–[147], specifically with mutations in a region on chromosome 10q26 [148]. One clinical study found that diminished protease activity due to mutations in the promotor region of *HTRA1* resulted in dysregulation of vascular growth of small blood vessels in the brain which caused a non-hypertensive cerebral small-vessel arteriopathy (CARASIL) [149]. This finding indicated that HtrA1 is involved in mediating vascular activity. In a recent study using *HTRA1* knockout (*htra1-/-*) mice, significant down-regulation of VEGF gene levels and decreased vascular development in the retinal pigment epithelium layer in comparison to the wild-type mice [148]. Inversely, it has also been reported that increased VEGF expression and polypoidal choroidopathy (a type of wet AMD) in transgenic mice over-expressing *HTRA1* [147]**.** More recently, an independent study has found that *HTRA1* inactivation led to inhibition of tumor growth through deregulation of angiogenesis [150]. These results establish HtrA1 as a promising therapeutic target for AMD, and possibly arthritis through its activity against fibronectin in the extracellular matrix, where fibronectin breakdown fragments can stimulate the matrix metalloproteinase MMP-14 that has known roles in driving arthritis [151].

New small molecules designed to inhibit the activity of HtrA1 could overcome the limitations of anti-VEGF antibodies. The biotechnology company Atomwise has

developed a deep learning-based computational drug discovery platform (AtomNet^{TM}) that can virtual screen millions of commercially available chemical compounds for those predicted to most likely to bind the chosen target protein [152]. As the crystal structure of HtrA1 is known both in its active [132] and inactive forms [153], the neural net could be trained robustly. In a partnership with this company, we identified 81 compounds for potentially binding the active site of HtrA1. These compounds were acquired through Atomwise and delivered to our lab so that we may determine their efficacy and eventually elucidate the atomic nature of the proposed interactions through macromolecular crystallography.

We therefore hypothesize that finding *in vitro* inhibitors for HtrA1 could serve as new therapeutic avenue for treating AMD. Herein, we will discuss our initial efforts in pursuing this potential therapeutic target.

2.3 Results

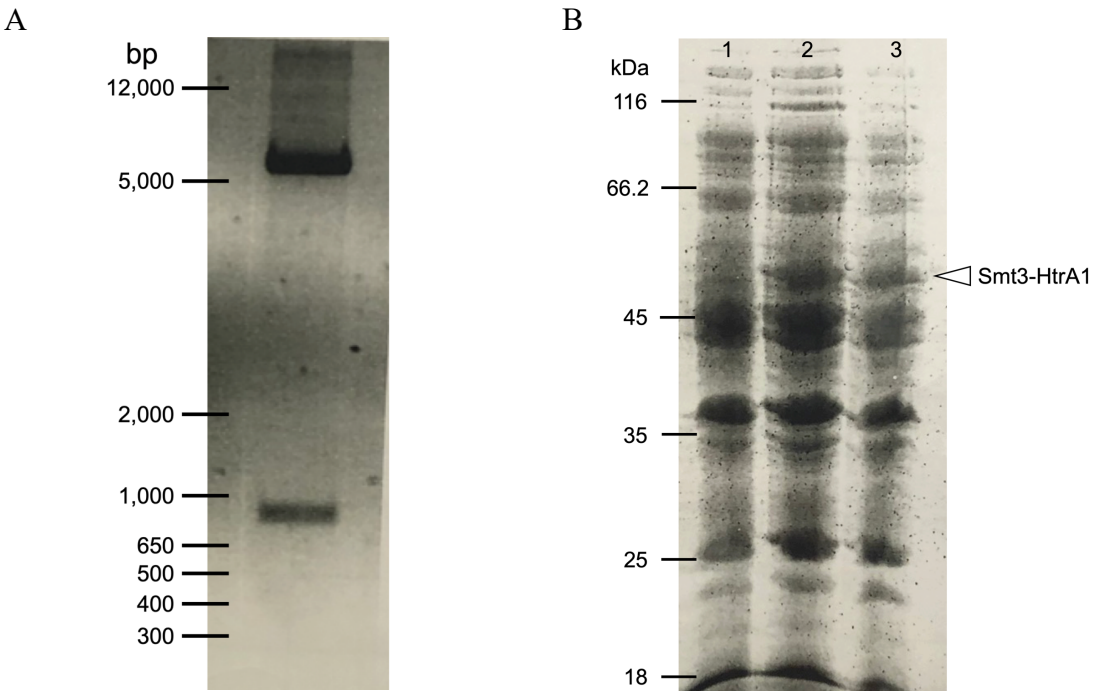


Figure 2.3 | Restriction enzyme digest on potential clones and HtrA1 expression test (A) Sample from restriction enzyme digest with *Bam*HI and *Sal*I following ligation were run on a 1% agarose gel and visualized under UV light (B) Expression of the HtrA1-Smt3 fusion protein (51 kDa) as judged by Coomassie stained 12% SDS-PAGE. DNA marker used was 1 Kb Plus DNA Ladder (Invitrogen™). Protein marker used was Unstained Protein MW Marker (Pierce™).

2.3.1 Cloning and expression of HtrA1

Following a digestion reaction` with the restriction enzymes *Bam*HI and *Sal*I, the successful ligation of the insert (900 bp) into the pET SUMO vector (~5 kb) as evidenced as two bands on a 1% agarose gel (**Fig. 2.3A**). This successfully cloned plasmid was then transformed and expressed in BL21 (DE3) cells. The Smt3-HtrA1 fusion protein can be readily observed around the 51 kDa range in both lanes 2 and 3 of the 12% SDS-PAGE gel (**Fig. 2.3B**). This plasmid and these expression conditions were therefore used in the large-scale recombinant production of this enzyme.

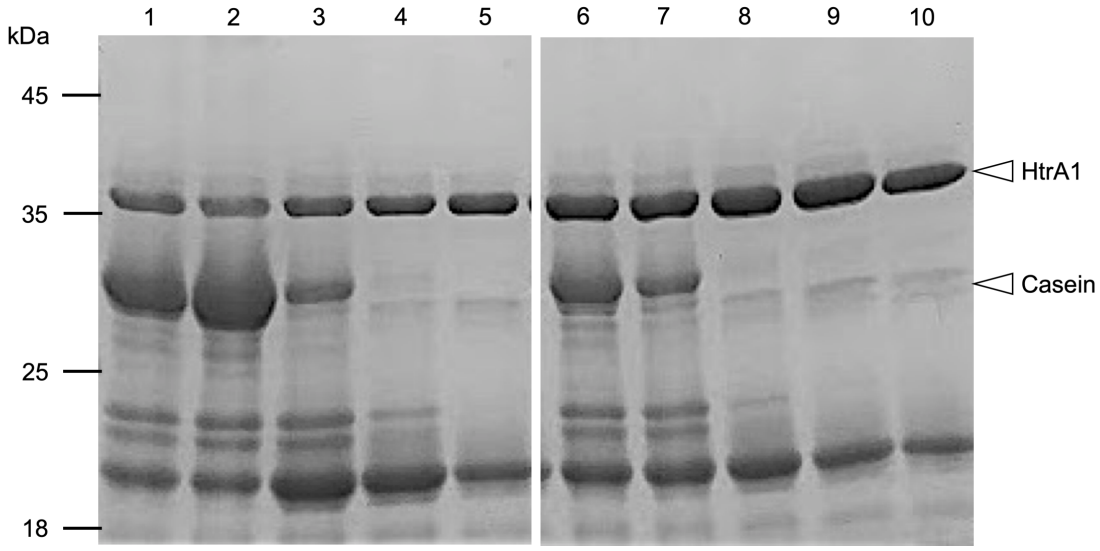


Figure 2.4 | Gel-based proteolysis assay Enzymatic activity of HtrA1 as judged by Coomassie stained 12% SDS-PAGE. The protease HtrA1 is 37.5 kDa, the substrate β-casein is 24 kDa, and the products of proteolysis are <24 kDa. Protein marker used was Unstained Protein MW Marker (Pierce™).

2.3.2 Assay development

Gel-based assay- Initial pursuits to develop an enzymatic assay suitable for early stage drug discovery studies of the HtrA1 serine protease enzyme began with a gel-based proteolysis assay (**Fig. 2.4**). The consistent band around 35 kDa is the HtrA1 enzyme being tested, and below it is β-casein, the substrate. Bands appearing below are likely proteolytic products or protein contaminants. Depletion of the substrate can be observed for lanes 1-5 where HtrA1 was preincubated with 5 mM leupeptin, a known trypsin inhibitor. However, when compared to the lanes 6-10 control, the remaining substrate, an indication of inhibition, was not easily analyzed, nor particularly strong even at 5 mM (2,000-fold excess; trypsin K_i = 3.5 nM). Further, lane-to-lane variability led to inconsistent data

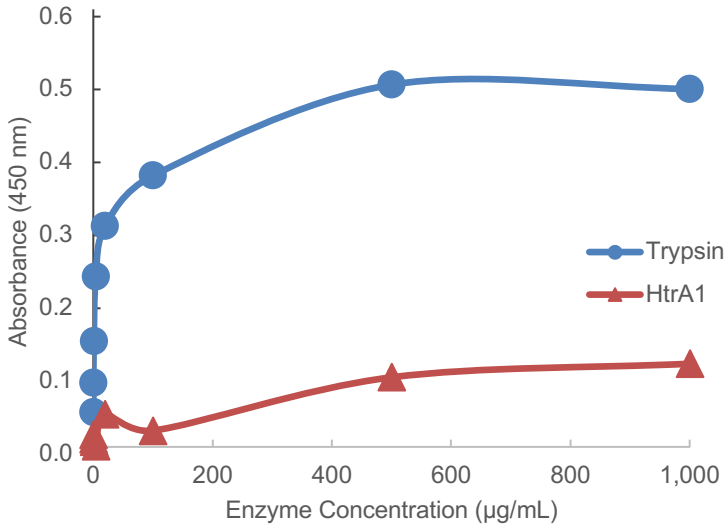


Figure 2.5 | Trypsin and HtrA1 digestion of succinylated casein Data are presented as background (enzyme only) subtracted absorbance values at 450 nm against protein concentration in units of µg/mL where trypsin is colored in blue and HtrA1 is colored in red.

integration. For these reasons, and scalability issues, this assay was not chosen for subsequent inhibition studies.

Colorimetric assay- Next, a colorimetric assay was developed using a succinylated trypsin substrate. It has been previously noted that the catalytic domain of HtrA1 is trypsin-like. This reaction was therefore considered for assay development. Absorbance values, thus product formation, for the trypsin reaction increases to a plateau as the enzyme concentration increases (**Fig. 2.5**), as expected. In this assay, studies on the HtrA1 enzyme revealed that both the product formation was much lower than trypsin, and the background was much higher. This is likely resultant of either low catalytic activity or specificity for this modified casein, or a high level of exposed primary amines in the HtrA1 protein. To

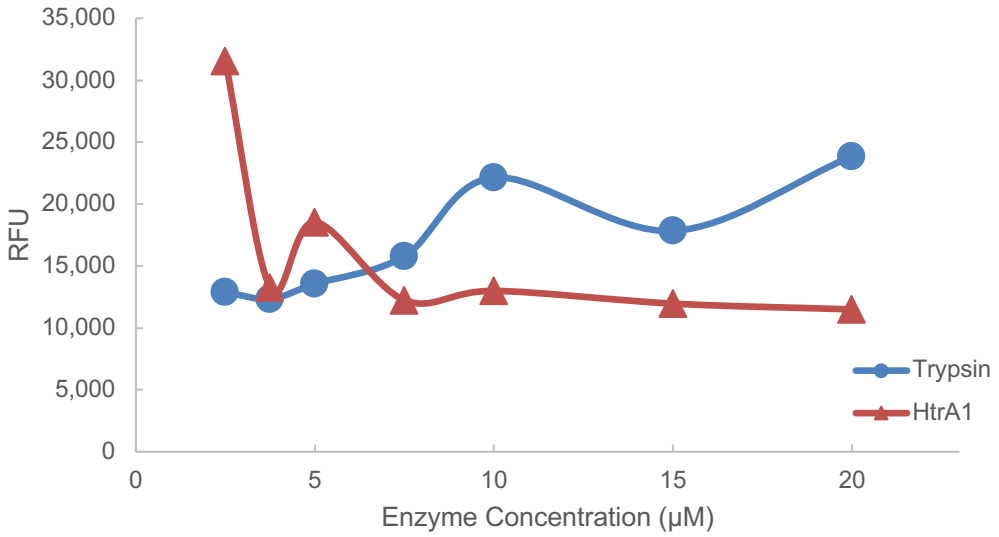


Figure 2.6 | End-point fluorescence kinetics with HtrA1 and trypsin control Raw fluorescence values from 485/535 nm are plotted against increasing concentrations of protease in units of µM following 1-hour incubation. The trypsin reaction is colored in blue and the HtrA1 plot is colored in red.

overcome these shortcomings, this reaction would require very high levels of HtrA1, therefore, this assay was also not taken further for our inhibition studies.

End-point fluorometric assay- To develop an assay sensitive enough to detect HtrA1 activity and the inhibitory values of the compounds being tested, a fluorometric assay was also conducted. We used a similar reaction to before in our colorimetric assay with HtrA1 incubation with a modified casein substrate, using trypsin for control at increasing enzyme concentrations. For fluorescence detection, a fluorescein isothiocyanate (FITC)-casein derivative (ThermoFisher; 23267) was used. FITC is a derivative of fluorescein and can be covalently linked to casein, and in this form its fluorescence is quenched. After acid precipitating the reaction products, the supernatant comprised of FITC-peptides was

measured for fluorescence. For the reaction with trypsin, increasing fluorescence was correlated with increasing enzyme concentration. We noted that the opposite was true for HtrA1, where higher enzyme concentrations resulted in lower activity levels (**Fig. 2.6)**. Again, the product of the enzymatic reaction is monitored by fluorescence and should therefore increase with increasing protein concentration. Trypsin goes up, HtrA1 goes down. Due to inconsistencies between datasets and the inverse correspondence of HtrA1 protease concentration and the production of FITC-casein peptides, we decided to pursue other assay avenues.

Real-time fluorometric assay- A real-time florescence-based assay was developed that probed the activity of the enzyme was also conducted for our inhibition studies. BODIPY® is a low molecular weight fluorophore than can be intramolecularly quenched by casein protein. If this casein-BODIPY® derivative is used as the proteolytic substrate, the fluorophores would be unquenched, and this release could be actively monitored to continuously measure enzyme kinetics. In our assay we therefore conducted inhibition studies with titrating amounts of CADD identified compounds. Relative fluorescence (505/513 nm) was measured in regular intervals at 37 ºC (**Fig. 2.7A**). The initial slopes of these curves were used to determine the rate of catalysis and to calculate the percent inhibition (**Fig. 2.7B**). For hit compounds that showed inhibition initially at 100 μM such titrations were completed (**Table S2.1**) and their IC_{50} values were calculated (**Fig. S2.1**). Using this assay, we were able to determine low to mid micromolar inhibition for 4 SBDD compounds from the 81 molecules tested (**Table 2.1**).

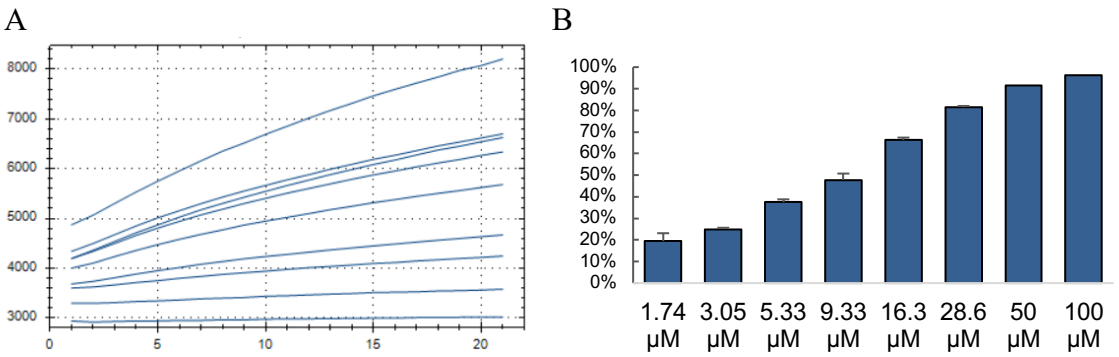


Figure 2.7 | Fluorometric assay and demonstration of percent inhibition (A) Raw fluorescence reads from titration series over time. (B) Percent inhibition for compound G7 across titration series. Data are background-subtracted (substrate alone) means presented ± s.d. of three replicates.

Table 2.1 | AW IC_{50} values	
Compound	IC_{50} (µM)
B4	17.0 ± 1.01
B8	41.5 ± 0.75
G4	7.02 ± 0.55
G7	10.3 ± 0.65
Data are means of three replicates ± standard error of curve (**Fig. S2.1**).	

2.3.3 Crystallization

Understanding the atomic positioning of an inhibitor in complex with its substrate is the holy grail of structure-based drug discovery (SBDD). To accomplish this, we first needed to crystallize the HtrA1 protein and collect data at resolutions suitable for the observation of a ligand to be bound. Using our purified protein and modified conditions from the literature [153], we were able to obtain a few large 250 x 250 x 250 µm HtrA1 crystals (**Fig. 2.8A**). This crystallization approach was further refined by conducting micro-seeding trials, which significantly improved the growth rate (3-fold) and also the reproducibility of large crystal growth (6-fold). These crystals were then soaked overnight in xylitol-supplemented cryoprotectant solution, inoculated with 500 µM of SBDD-compound G7. The next day the crystals appeared yellow in hue (**Fig. 2.8B**) under unpolarized light, matching the color of the G7 compound in solution. These yellow-

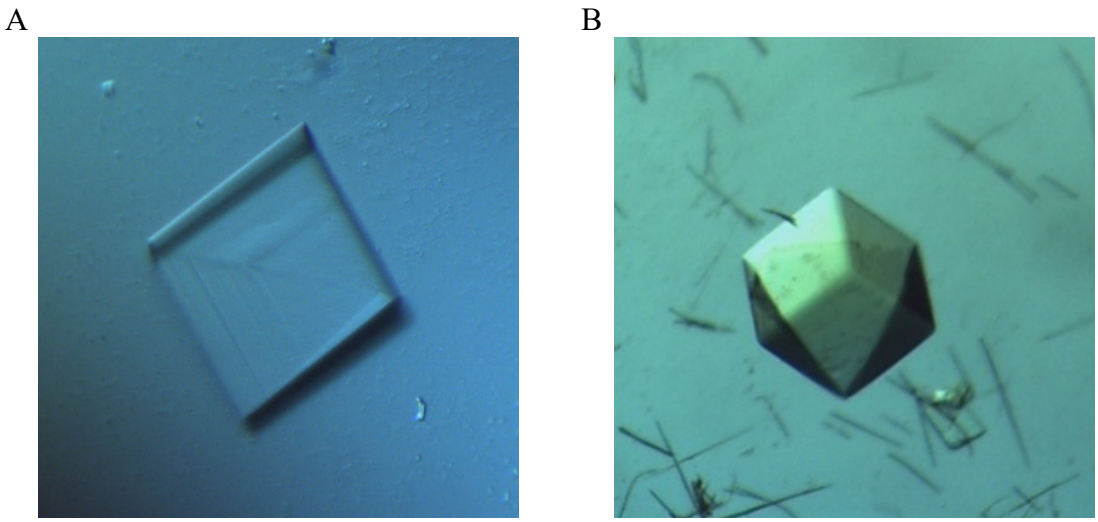


Figure 2.8 | Crystalline HtrA1 enzyme (A) HtrA1 protein crystal visualized under polarized light. (B) HtrA1 protein crystal following 15-hour soak with G7 visualized unpolarized light.

colored HtrA1 crystals were harvested and X-ray diffraction data was collected. After structural determination was completed using the homotrimer (3TJN.pdb) as the molecular replacement model to 2.44 Å (R_{work}/R_{free} 0.2486/0.2801), G7 was not readily detectable in the electron density.

Previous observations of the diffraction pattern for HtrA1 crystals revealed strong X-ray diffuse scattering. When analysis was conducted, a clear anisotropic component was observable from our cryo-crystallography procedure. In an attempt to drive the crystalline form towards greater disorder, we conducted room temperature (22 ºC) crystallography. Preliminary data has been collected, and data processing is underway, as part of a separate on-going project in the Perry laboratory on developing X-ray diffuse scatter technologies.

2.4 Discussion

HtrA1 is an attractive therapeutic target for AMD and arthritic disease states. In our studies, we have used *in vitro* enzymatic assays to characterize the inhibitory actions of several AI-CADD identified compounds that could potentially inhibit HtrA1 enzymatic activity. We used a number of assays to see define which was optimal for our studies. Our gel-based assay clearly demonstrated proteolysis for casein (**Fig. 2.4**) but there were challenges with quantification due to variability in sample-loading and reaction-timing. Scalability would also become an issue when the inhibition studies are undergone for the full suite of 81-compounds. Although other techniques were pursued, the gel-based sensitivity issue could be resolved in future efforts by conducting a fluorescent peptide zymography [154] experiments. In a proposed future assay, a standard SDS polyacrylamide gel is copolymerized with a peptide-MCA (4-methyl-coumaryl-7-amide) substrate, and electrophoretic zymographic analysis is conducted. Inhibition could therefore me monitored with more sensitive fluorescence values, thus overcoming one limitation of our Coomassie staining analysis. While this assay would have suited our needs, scalability is still a concern and lack of detection instrumentation remains.

When compared to the trypsin control, HtrA1 severely lacked activity for both the succinylated casein and FITC-casein, thus detection of each proved inconsistent. We observed that detection by colorimetric assay lacked sensitivity, as the difference between the proteolytic product, when comparing 1 μg/mL to 1,000 μg/mL of enzyme, was merely 0.07 absorbance units (**Fig 2.5**). Our studies using a fluorescein-derivative fluorometric reagent observed that increased HtrA1 concentration in the assay lowered proteolytic yield

(**Fig. 2.6**). Thus, the fluorescein-labeling of the casein substrate in this assay, as compared to native casein in our SDS-PAGE gel assay, perturbed HtrA1 activity. One possible explanation of this phenomenon could be that this fluorescein modification blocks the site of proteolytic cleavage by HtrA1, but not its substrate recognition. This would satisfy the observation, as HtrA1 would still bind to casein, and perhaps at the high concentration used in our assay quench the fluorescence but have minimal or depleted proteolytic activity against the substrate.

A BODIPY-casein substrate was found to be readily proteolyzed by HtrA1, hence we developed a real-time fluorometric assay to define enzymatic inhibition by compounds of interest. Here, four compounds with low to mid micromolar IC_{50} values, Atomwise B4 (17 μM), B8 (41.5 μM), G4 (7 μM), and G7 (10 μM), were identified from the set of AI-CADD designed molecules (**Fig. 2.7**). This result is an important evaluation of the viability of the AtomNet™ AI-CADD platform as a virtual screening methodology. Our findings represent successful first round of structure-based drug discovery, where these 4 compounds, and likewise not the other 77 compounds, can be used to inform the next set of chemicals to be evaluated.

To better understand the binding characteristics between these compounds and HtrA1, we are now pursuing co-crystallographic efforts. Using seeding methods, we can consistently produce HtrA1 crystals, and this has allowed for soaking methods, which we have been conducted on our top three hits. The in-solution hue of concentrated G7 is yellow and soaking with this hit turned the HtrA1 crystals yellow, which is suggestive of binding (**Fig. 2.8)**.

Initial studies on one well-diffracting HtrA1 crystal (2.44 Å) did not reveal any observable difference between our apo model and the obtained structure as evidenced by the Fo-Fc map. Subsequent efforts could perform co-crystallization with this compound at the lowest possible DMSO concentration, and highest compound concentration while preserving its solubility, in order to maintain the strong diffraction nature of these crystals. X-ray crystallographic structure determination provides the electron density map from the most populated state within the unit cell, or the average structure. This higher compound concentration might result in more binding, and thus improving the contribution of the compound to the X-ray diffraction data. Likewise, the enzyme-inhibitor complex could be screened for new crystallization conditions, as there may be conformational changes associated with the ligand binding. A crystal structure of HtrA1 that diffracts to a higher resolution (2.3 Å) has been previously reported, however the construct is a smaller (160-370) inactive mutant (S328A) that lacks the PDZ domain. Cloning efforts for this construct have begun, but if the active site serine is required for inhibitor binding, this crystal may not suit our needs. Lastly, initial analyses of the X-ray diffuse scatter data for the room temperature HtrA1 crystals are ongoing and are showing promising results.

To limit off-target effects, counter-screens against HtrA2 that has roles in Parkinson's disease must be conducted. Thus, SBDD approaches are ideal for designing compounds that are selective to substrate binding pocket for HtrA1, especially due to the high degree of conservation in the active site region. Cloning and expression procedures have been conducted on HtrA2, in preparation of counter screens, during the hits-to-leads development steps of this project, so as to limit off-target effects.

Thus, overall, our work has recombinantly produced the HtrA1 protein, purified and crystallized it, and developed powerful tools for understanding its activity. These results and identification of initial HtrA1 inhibitors can be leveraged for early stage drug discovery and development for the treatment of AMD.

2.5 Methods

2.5.1 Recombinant cloning and expression

Cloning- To begin studying the HtrA1 protein, primers (**Table 2.2**; IDT) were designed around amino acids 158-480 of the *HTRA1* gene from donor vector obtained from DNASU (Clone Id = 353964) for insertion into pET SUMO MCS via restriction enzymes *Bam*HI and *Sal*I. A similar PCR reaction as in 1.5.1 was used with a annealing temperature lowered to 55 ºC. Ethanol precipitation (**Appendix A**), agarose gel extraction (**Appendix B**), ligation reaction, transformation (**Appendix C**), and miniprep (**Appendix D**) were followed as outlined in 1.5.1. Likewise, the final restriction enzyme digest analysis (**Fig. 2.3A**) was conducted as in 1.5.1 with *Bam*HI and *Sal*I. Successfully ligated constructs were further confirmed by Sanger sequencing (UCR core facility).

Table 2.2 | Primers used in HtrA1 cloning

Forward primer	5'-CGC GGA TCC ATG GGG CAG GAA GAT CCC AAC AG-3'
Reverse primer	5'-ACG CGT CGA CTT TGG GTC AAT TTC TTC GGG AAT -3'

Expression- To determine suitable protein expression conditions for the HtrA1-Smt3 fusion protein, tests were conducted on a small scale. The successfully cloned vector was transformed into a variety of bacterial expression strains of *E. coli*: BL21 (DE3), BL21 (DE3) RIL, BL21 (DE3) pLysS, and Rosetta™ 2(DE3). Individual colonies from the corresponding cell strains, were inoculated in 5 mL of LB with kanamycin O/N at 37 ºC. 50 mL of LB supplemented with 35 µg/mL kanamycin was secondarily inoculated with the corresponding O/N culture and grown at 37 ºC. Once the 50 mL culture reached OD_{600nm}

= 0.6, protein expression was induced with 0.4 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG). After 4 hours, the cells were harvested by 4 ºC centrifugation at 5,000 rpm for 20 minutes. Harvested cells were resuspended in Resuspension Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 1 mM β-ME), according to the ratio of 3 mL buffer to 1 gram of cell pellet. The resuspended cells were sonicated (Q-Sonica Q125, ⅛ inch probe) at 30% amplitude for 3 minutes, with pulsing for 10 seconds on and 10 seconds off. The mixture was then centrifuged at 12,000 rpm for 20 minutes. Samples of both the soluble and insoluble fractions were run on 12 % SDS-PAGE gels (**Fig. 2.3B**).

After determining that BL21 (DE3) was suitable for HtrA1-Smt3 expression, production efforts were scaled up. An overnight (O/N) 7 mL culture was inoculated using a glycerol stock generated from a previously uninduced sample. The O/N culture was secondarily inoculated into 1 L of LB with kanamycin and incubated at 37 ºC and 225 rpm (New Brunswick Scientific Innova©43R) for approximately 4 hours until the OD_{600nm} = 0.6. To optimally express the HtrA1 the temperature was lowered to 16 ºC, induced with 0.4 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) and set to incubate and shake O/N for 15 hours. The following morning the cells were harvested by 4ºC centrifugation at 5,000 rpm for 20 minutes (Sorvall LYNX 4000 Superspeed Centrifuge with a Fiberlite™ F9-6 x 1000 LEX Fixed Angle Rotor), and the cell pellets were collected and stored at -80 ºC.

To purify the HtrA1 protein the collected cells were resuspended in Resuspension Buffer at a ratio of 3 mL of buffer per 1 gram of pellet. These resuspended pellets were then disrupted by sonication (Q-Sonica Q125, ¼ inch probe) on ice at 70% amplitude for

5 minutes with pulsing for 10 seconds on and 10 seconds off. The insoluble fraction of this suspension was separated via 4 ºC centrifugation for 30 minutes at 30,000 g. The supernatant was carefully decanted and loaded onto a Ni-NTA column (HisTrap™ FF; GE Life Sciences) using an NGC Scout Liquid Chromatography System (Bio-Rad). After binding the column, the 6xHis-tagged HtrA1-Smt3 fusion protein was washed with buffer containing 30 mM imidazole and eluted against a gradient to 500 mM imidazole. Following elution, the 6xHis-Smt3 tag was removed after a fifteen-hour incubation at 8º C with recombinant ubiquitin-like protease 1 (Ulp1) at a ratio of 1 µg Ulp1:1 mg 6xHis-Smt3-HtrA1. The digest reaction was diluted to 50 mM NaCl and loaded onto HiTrap Q FF (GE Life Sciences) and eluted against a gradient to 500 mM NaCl. Fractions containing HtrA1 were concentrated to 5 mL and loaded onto and run down a HiLoad® 16/600 Superdex® 200 pg gel filtration column (GE Life Sciences), fractionating every 2 mL. Samples containing purified HtrA1 protein were determined by SDS-PAGE gel and were concentrated to 10 mg/ml ($\varepsilon = 0.2044$) in 10 mM HEPES pH 7.5, and 50 mM ammonium sulfate storage buffer, flash-frozen in liquid nitrogen, and stored at -80 ºC.

2.5.2 *In vitro* assay development

SDS-PAGE proteolytic assay- To conduct the gel-based assay, a reaction using 2.5 µM HtrA1 enzyme was preincubated in a 10 mM HEPES-NaOH pH 7.5, 50 mM ammonium sulfate buffer with 100 µM compound or DMSO as control for 30 minutes. After which, the reaction volume was doubled with 170 µM of β-casein substrate, initiating proteolysis. At specific time intervals (0.5, 1, 3, 5, 10 minutes) 10 µL samples were removed from the

reaction tube and heated to 95 ºC to denature the protein and stop the reaction. These time-trial samples were then run on a 12% SDS-PAGE gel and visualized under white light after Coomassie staining (**Fig. 2.4**). Quantification of the gels was conducted using ImageJ [155].

Colorimetric assay- To conduct the HtrA1 digestion reaction of succinylated-casein (Pierce Protease Kit), 50 µL of titrated concentrations (0.0064, 0.16, 0.8, 4, 20, 100, 500, 1,000 µg/mL) of HtrA1 and trypsin in 10 mM HEPES-NaOH pH 7.5, 50 mM ammonium sulfate buffer was mixed with 100 µL of resuspended succinylated casein. The reaction was incubated for 1 hour, after which TNBSA (2,4,6-trinitrobenzene sulfonic acid), 5% (w/v) in methanol was added to each well where it reacts with the exposed primary amines from the peptidase reaction to produce an orange-yellow product. Absorbance at 450 nm was measured 20 minutes later to allow for color development in a Multiskan™ FC Microplate Photometer (Thermo Scientific). Absorbance values were output in Microsoft Excel and results were background subtracted using a no substrate control (**Fig. 2.5**).

*End-point fluorometric assay***-** Following the Protease Fluorescent Detection Kit (Sigma-Aldrich), titrating concentrations (2.5, 3.75, 5, 7.5, 10, 15, and 20 µM) of trypsin and HtrA1 were incubated with fluorescein isothiocyanate (FITC) labeled casein for 1 hour at 37 ºC. These reaction mixtures were then acid precipitated with trichloroacetic acid (TCA) leaving the FITC-casein peptide fragments remaining. Debris was centrifuged and the reaction product was decanted and distributed into a black 96-well plate and measured for

fluorescence at 485/535 nm (Typhoon 9410; GE Healthcare) (**Figure 2.6**). Data are presented as raw fluorescence reads with no background subtraction of a single replicate.

Real-time fluorometric kinetic assay- Finally, a fluorometric assay was developed to monitor enzymatic activity in real-time for our inhibition studies. To prepare the reaction, 500 nM HtrA1 enzyme was preincubated in 10 mM HEPES-NaOH pH 7.5 and 50 mM ammonium sulfate with titrating amounts of test compounds at 37 ºC for one hour. To begin the reaction, 5 μg/mL of casein derivative heavily labeled with BODIPY® was added, sealed with an optically clear seal, and delivered into a RT-PCR apparatus. There the plate was incubated at 37 ºC to facilitate the reaction and fluorescence was monitored every 30 seconds at excitation and emission wavelengths 503/513 nm. Over the time course protease-catalyzed hydrolysis releases the highly-fluorescent BODIPY® dye-labeled peptides, allowing for quantitative detection of protease activity in solution (**Fig. 2.7A**). The initial slope of these lines (3-10 minutes) were used to determine the rate of catalysis, and to calculate the percent inhibition by comparing slopes of increasing concentrations of compound against the DMSO control slope (**Fig. 2.7B**). Data are means of triplicate. To interpret the IC_{50} values from the percent inhibition calculation, these values were input into an online four parameter logistic curve calculator [156] (**Fig. S2.1**). The inflection point of these curves was used to ascribe the IC_{50} value for the compound tested (**Table 2.1**).

2.5.3 Crystallization

Methylation- To conduct the *in vitro* methylation of HtrA1 required for crystallization, for every 1 mL of protein at 10 mg/mL, 20 µL of 1 M borane-dimethylamine complex (ABC) and 40 µL of 1 M formaldehyde was added and the solution was gently mixed and incubated on ice for 2 hours. This step was repeated and then incubated for another 2 hours. Finally, 10 µL of 1 M ABC per 1 mL of protein was added and the solution was incubated on ice overnight. The reaction was quenched by addition of 1 mL Tris-HCl pH 8.8, incubated for 1 hour on ice and then the methylated protein was exchanged into the storage buffer and concentrated as mentioned previously.

Vapor diffusion- Methylated HtrA1 was crystallized using the sitting drop vapor diffusion method at 16 ºC. Initially, 2 µL of the methylated HtrA1 (7 mg mL^{-1} in 10 mM HEPES-NaOH pH 7.5, 50 mM ammonium sulfate was mixed with 2 µL reservoir solution (1.0 M $LiSO_4$, 0.1 M sodium citrate pH 6.0, and 0.5 M ammonium sulfate in the pedestal above 400 µL reservoir solution and sealed. Cubic crystals grew to their final size of 150 µm x 150 µm x 150 µm after 21 days. To improve the size and reproducibility of these crystals, micro-seeding was conducted by mixing 6 crystal-containing wells in a microcentrifuge tube with a PTFE bead (HR2-320; Hampton Research). The tube was then vortexed for three minutes, stopping every 30 seconds to cool the tube on ice. The vortexing of the crystal and bead mixture pulverized the crystals to generate a microcrystal seed stock [157]. A serial dilution of this seed stock revealed that the most consistent crystals were grown using a 1:100 dilution in reservoir solution. Seeded crystals were grown by adding 2 µL of

protein at 7 mg mL^{-1} to 2 µL of the 1:100 seed bead dilution via sitting drop vapor diffusion. Cubic crystals grew to their final size of ~ 250 µm x 250 µm x 250 µm after 7 days. Before harvesting the crystal by flash-freezing in liquid nitrogen, crystals were briefly submerged in a cryoprotectant solution (1:1 (v/v) reservoir solution and 60% xylitol). To perform crystal soaks, 2 µL of a soaking solution (5:95 (v/v) ratio of test compound dissolved in DMSO and the reservoir solution) where the compound was 500 µM was added atop the well and allowed 16 hours to diffuse before harvesting.

X-ray diffuse scatter- Diffraction data were collected at 297 K at beamline 8.3.1 at the Advance Light Source Synchrotron Radiation Facility in Berkeley, California which is equipped with a PILATUS 6M detector. A WaterShed™ (Mitegen) humidifying system is installed in the hutch and when positioned one half inch from the goniometer, ambient temperature crystals stay hydrated. HtrA1 apo crystals were mounted on a 500 µm Mitegen loop and exposed to 11,111 Ev x-ray for 0.3 second exposures oscillating 1 degree for 45 frames total. Radiation damage was evident after 15 frames, so subsequent datasets collected from the same crystal were conducted 15 degrees advanced from the previous collection. All diffraction data were processed using DIALS. Diffuse scatter was calculated using LUNUS.

2.6 Supplemental Information

Table S2.1 | Percent inhibition values for AI-CADD compounds

Compound	Inhibition at 100 μM (%)	Compound	Inhibition at 100 μM (%)
A01	25.97 ± 3.29	D08	-7.44 ± 7.14
A02	-23.06 ± 0.20	D09	-2.43 ± 5.63
A03	69.51 ± 2.42	D10	-19.75 ± 4.33
A04	12.58 ± 5.07	D11	10.64 ± 5.42
A05	63.50 ± 4.26	D12	73.15 ± 2.37
A06	-8.90 ± 1.13	E01	9.01 ± 8.72
A07	58.55 ± 0.35	E02	-6.49 ± 1.43
A08	33.35 ± 0.64	E03	-13.70 ± 3.25
A09	-0.20 ± 2.97	E04	-7.70 ± 1.13
A10	37.20 ± 5.52	E05	6.16 ± 3.04
A11	24.41 ± 3.87	E06	-10.84 ± 4.25
A12	2.45 ± 1.06	E07	39.05 ± 3.75
B01	22.00 ± 0.71	E08	-13.40 ± 0.69
B02	-5.28 ± 3.67	E09	53.55 ± 0.64
B03	15.90 ± 0.71	E10	-8.15 ± 3.37
B04*	90.95 ± 8.49	E11	-17.81 ± 2.95
B05	38.45 ± 0.92	E12	-9.25 ± 1.91
B06	0.13 ± 2.98	F01	6.35 ± 0.78
B07	-2.69 ± 5.20	F02	-7.35 ± 2.64
B08*	85.45 ± 1.34	F03	-1.60 ± 0.28
B09	75.00 ± 0.99	F04	10.12 ± 1.47
B10	32.90 ± 2.97	F05	41.09 ± 6.52
B11	-4.20 ± 1.98	F06	2.66 ± 6.75
B12	58.20 ± 1.98	F07	93.75 ± 1.06
C01	-4.40 ± 0.99	F08	71.35 ± 1.06
C02	63.10 ± 0.42	G01	-11.55 ± 1.64
C03	16.70 ± 6.58	G02	12.73 ± 4.01
C04	-6.38 ± 8.37	G03	21.35 ± 2.90
C05	37.45 ± 9.97	G04*	93.45 ± 0.64
C06	7.50 ± 5.66	G05	14.65 ± 7.42
C07	10.13 ± 2.95	G06	10.30 ± 0.85
C08	-6.40 ± 13.44	G07*	95.83 ± 3.43
C09	5.48 ± 3.44	G08	10.69 ± 4.43
C10	20.95 ± 1.91	G09	-2.38 ± 7.12
C11	36.45 ± 16.33	G10	7.05 ± 6.29
C12	-1.23 ± 8.26	G11	4.00 ± 2.55
D01	54.50 ± 10.18	G12	23.66 ± 7.59
D02	73.75 ± 4.74	H01	7.90 ± 10.32
D03	24.60 ± 6.36	H02	26.00 ± 3.54
D04	9.00 ± 6.36	H03	58.23 ± 2.03
D05	8.92 ± 14.25	H04	-1.60 ± 2.40
D06	-15.65 ± 5.87	H05	61.92 ± 9.48
D07	-5.45 ± 1.91		

*Titrations were conducted (**Table 2.1**)

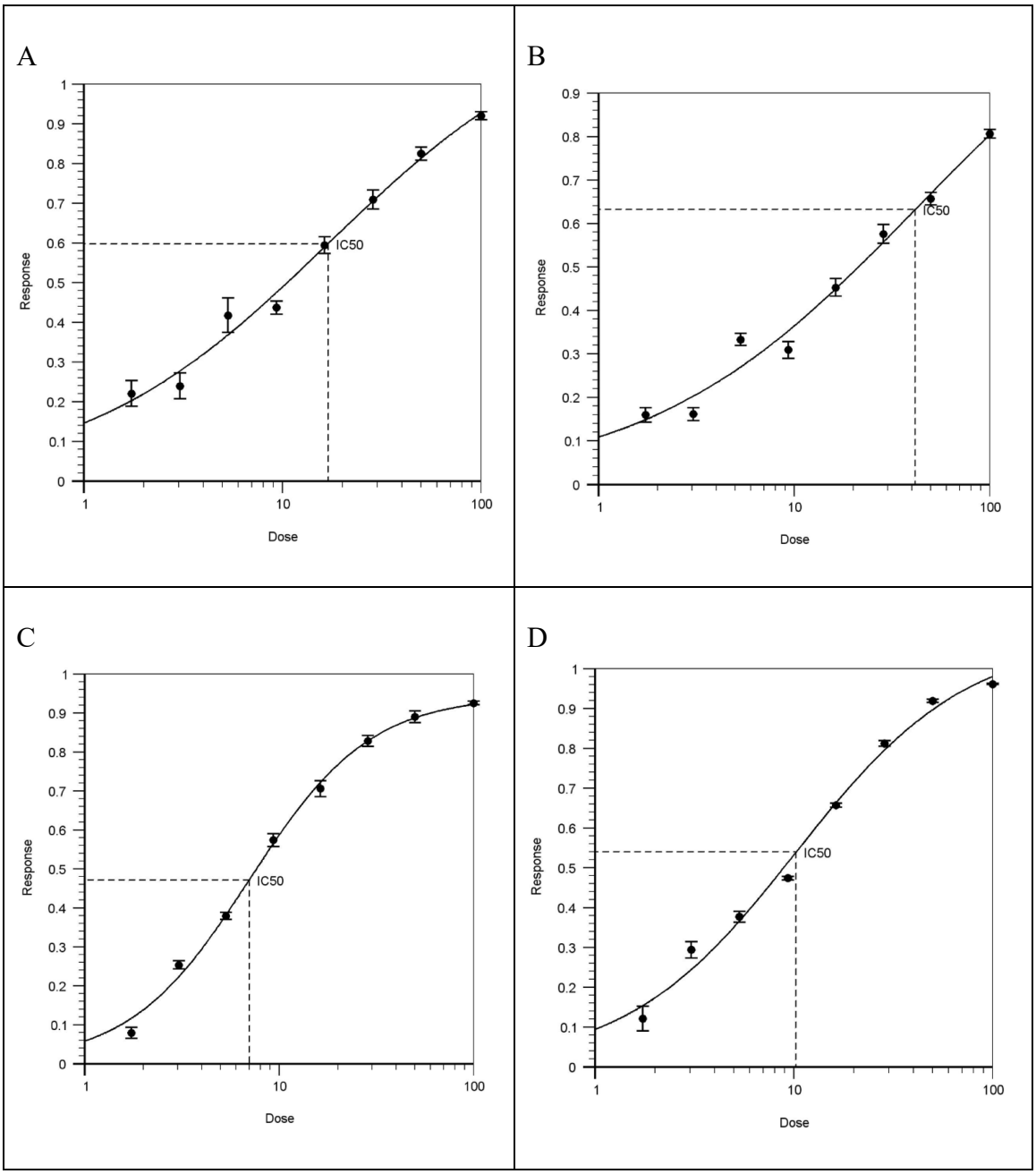


Figure S2.1 | Four parameter logistic plots for CADD compound hits Percent inhibition values for four CADD hit compounds (A) B4 (B) B8 (C) G4 (D) G7 were input into an online curve generator [156] and plotted as percent inhibition against compounds concentration.

Chapter 3
Aos1/Uba2

Chapter 3: Aos1/Uba2

3.1 Abstract

The small ubiquitin-like modifier (SUMO) is a 12 kDa post-translationally conjugated protein for which the reaction scheme is analogous to ubiquitination. SUMO conjugation occurs in most cells and is involved in various cellular processes. Interestingly, the SUMOylation pathway is one such system that has been upregulated and critical in KRAS mutant tumors, while it has moreover been deemed less critical to normal cell viability. KRAS oncogenic mutations represent 60% of pancreatic carcinomas, 50% of colorectal cancers, >25% of lung cancers, and are presently a challenge to treat. Other known untreatable cancers with upregulated SUMOylation are those involving c-Myc overexpression (>50% of human cancers have deregulated MYC oncogenes) or Notch-1 deregulation. Synthetic lethality approaches are being increasingly used to treat many disease states deemed intractable when targeted directly. Thus, our efforts are to develop small molecule inhibitors for the SUMO system, with an aim towards treating these hard to selectively target and aggressive tumors. We have pursued initial studies in natural product chemistry, virtual screening, SAR by catalog, and fragment-based drug discovery avenues for developing such a small molecule, which we explore herein.

3.2 Introduction

Pursuing oncogene addiction mechanisms has allowed for the development of a first generation of cancer-targeted therapies [158], [159]. However, in spite of recent successes there is a key subset of oncogenes that are prevalent in a large spectrum of cancers and that have so far proved intractable when targeted directly [160], [161]. These genes could instead be targeted instead through their 'non-oncogene addiction' pathways [162], [163]. In these cancer cells, non-oncogene addiction is where viability has become reliant on systems not immediately responsible for tumorigenesis, but are essentially for survival. Furthermore, these genes/pathways are not required to the same degree for the viability of normal levels [164]. One such noted non-oncogene addiction pathway is the SUMOylation cascade, shown to be essential to the viability of KRAS [161], [165]–[168], c-Myc [169], [170], and Notch-1 cancers [171]–[173], and appearing critical to drug resistance to acute myeloid leukemia (AML) therapy [164], [174], [175], all difficult to treat tumors.

The small ubiquitin-like modifier (SUMO) proteins are a group of short polypeptides with four distinct isoforms being encoded by the human genome that are added to specific target proteins as a post-translational modification (PTM). SUMOylation occurs both in the nucleus and the cytoplasm and it controls a number of important cellular processes that includes transcription, DNA repair, nuclear transport, response to stress, proliferation/progression through the cell cycle and apoptosis [176]–[180]. A sequential cascade of conjugation enzymes is used to dynamically control the addition of a SUMO moiety to the target protein. The SUMO E1 enzyme is a heterodimer made up of Aos1 and Uba2 subunits that activate this mature SUMO protein as the C-terminal adenylate in an

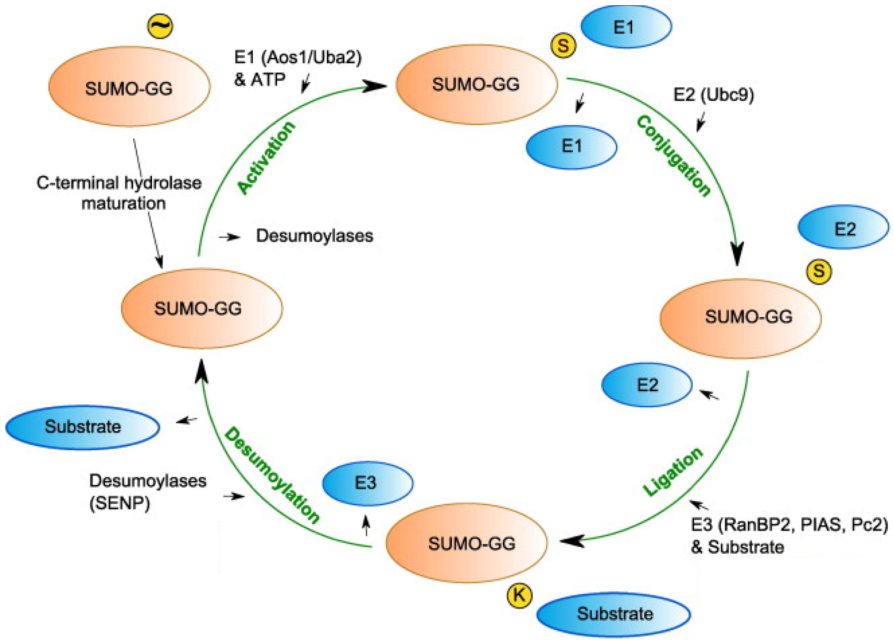


Figure 3.1 | SUMOylation cycle Maturation, activation, conjugation, and recycling of SUMO. Adapted from [181].

ATP-dependent manner. SUMO is then transferred to a cysteine residue on the E1 enzyme to form a transient thioester to the Uba2 subunit following release of AMP [182]. The activated SUMO group is further transferred to a cysteine on the E2 enzyme (Ubc9) that now serves as the SUMO-conjugating species to target proteins [182], often mediated by an E3 enzyme such as a PIAS protein family member, to impart substrate specificity [183]. SUMOylation occurs on target lysine side chains, and similar to ubiquitination, target proteins can be either mono-SUMOylated or have chains of SUMO added [184]–[186].

In recent years, dysregulation of SUMOylation has been increasingly linked to carcinogenesis and chemotherapeutic resistance for a variety of cancers [181]. Thus, there has been a growing interest in a more complete understanding of the role of SUMO in cancer and the mechanisms by which dysregulation of this system takes place. Our efforts

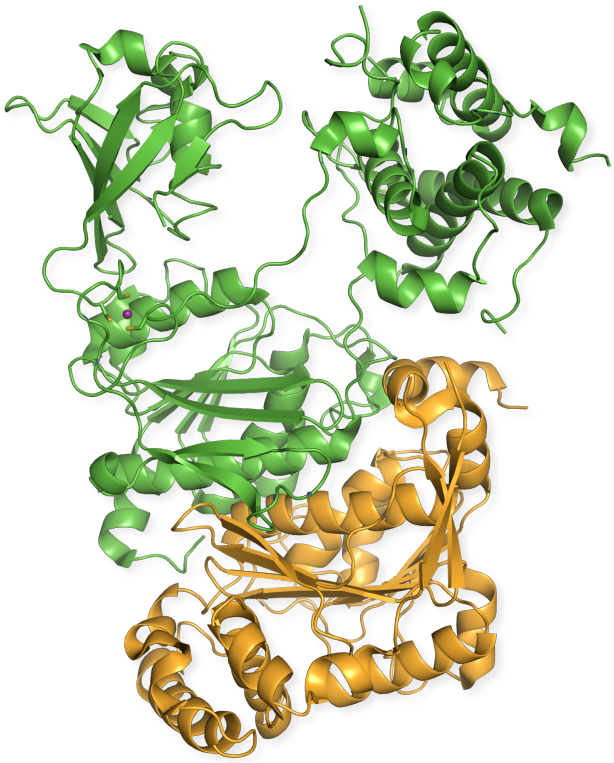


Figure 3.2 | Crystal structure of apo Aos1/Uba2 Structural representation from model (6CWZ.pdb) generated in PyMOL [92].

are therefore to design and develop small molecules targeted for the SUMO E1 enzyme, through a variety of means. One such way of developing drugs will be to model the compounds after a known inhibitor. To date, the most potent natural product inhibitor of Aos1/Uba2 is anacardic acid, a cashew nut shell extract natural product with an *in vitro* IC_{50} value of 2.2 μM [187]. Thus, we aim to be the first to develop anacardic acid analogues into suitable leads for future drug development [188]. Additionally, several *ab initio* approaches to drug discovery are being pursued. The first of which is through a partnership with Atomwise. As with the relationship established in 2.2, Atomwise will be training their neural net with the published crystal structure of Aos1/Uba2 (PDB code: 6CWZ) and

conducting virtual screening against a library of all commercially available compounds and providing a subset of them to our lab for *in vitro* studies. A second virtual screening approach is being conducted in collaboration with Dr. David Mobley's group at UC Irvine. His team will be scoring their possible hits by conducting alchemical free energy calculations by turning off interactions between the protein and the proposed ligand in a set of simulations, then allowing transfer of the ligand from the binding site to solution [189]. This methodology entitled Binding Modes of Ligands Using Enhanced Sampling (BLUES) yields binding free energies for possible ligands. The compounds that the Mobley group will be testing against the active site of Aos1/Uba2 are from the NCI Diversity set, and hits will likewise be provided for *in vitro* studies. [1]

The final approach we will be pursuing is fragment-based lead discovery (FBLD). In an FBLD screen, small fragments of 150-220 Daltons in size are screened for binding to the target protein. The aim of FBLD is to find initial hits that are then easy to optimize by using a carefully selected fragment library [190]. Within fragment libraries, compounds consist of distinct functional groups, they can be used to very efficiently sample chemical space, and while most fragments will not bind, those that in fact bind, do so with extremely high ligand efficiency (LE). LE is the number of atoms within the fragment involved in binding the protein, and while a typical fragment can have an LE of 70% [191], a typical HTS hit will have an LE of only 20-30% [192], [193]. This means that a fragment can be more readily developed into a final hit/lead compound that is much smaller than those

[1] **Nomenclature Note:** AAD: anacardic acid derivative; AW##: numbered compounds received from Atomwise virtual screen; NCI##: numbered compounds received from Mobley Lab virtual screen; SD##: STE-derivatives found through SAR by catalog

produced by other methods, such as by chemical modification of an HTS hit. Thus, FBLD yields compounds with greater specificity and selectivity, which can reduce off-target effects and hence toxicities, all of which are key requirements for passing pre-clinical and clinical trials. This is evidenced by the recent development of currently over 30 lead compounds generated from fragment methods that are in either late-stage clinical trials or in the clinic.

The compounds provided from each of these drug discovery platforms will be analyzed through a differential scanning fluorimetry (thermal shift) assay. Proteins exist in thermodynamic equilibrium between multiple conformational states and the binding of any molecule to the protein will alter the populations of these states [194]. Typically, specific binding of a small-molecule to a structurally defined site of a protein will increasingly populate the folded state more than any nonspecific interaction with the unfolded state. This population change would increase the free energy difference between the two states, $\Delta G_{U\text{-}F}$. The effect of this will cause a shift of the melting temperature of the protein (T_m), the temperature at which there is 50% denaturation, to a higher value (ΔT_m). Consequently, by measuring T_m in the presence and absence of a potential ligand, it is possible to detect any protein-ligand binding.

To perform such an HTS format assay into ligand-dependent changes to the T_m, a thermal shift assay (TSA) can be utilized. Although the idea that fluorescence increases when dyes bind to a protein's hydrophobic regions was first described over sixty years ago [195], it has only recently gained popularity as a method for screening protein-ligand interactions [196]–[198]. As an assay, monitoring thermal stability is nothing new [199],

but the modern version of this approach measures the temperature dependence of the fluorescence signal of a dye that binds preferentially to the unfolded population state of a protein [200]. SYPRO® Orange is a commercially available dye whose fluorescence is quenched in aqueous solution, and upon binding to a hydrophobic surface, emits a fluorescent signal.

Using this dye in the presence of protein, it is possible to determine the extent of denaturation when fluorescence is monitored in temperature increasing conditions. To reiterate, upon denaturation, the hydrophobic core of the protein becomes solvent exposed and thus the fluorescent dye now has a larger hydrophobic surface area to bind relative to a folded protein. The resultant plot will be a melt curve for each individual protein-ligand complex. Calculating the differences in the temperature dependent fluorescence profiles of protein plus dye in the presence and absence of a potential ligand may reveal a change in T_m, which could be indicative of binding. For well-behaved systems, a plot of fluorescence signal against temperature should produce a sigmoidal curve. The melting temperature is observed by determining the point of inflection of this curve. This can most easily be assessed by plotting the derivative of the fluorescent signal against temperature (dRFU/dT vs. T). This, so-called, thermal shift assay has also been referred to as differential scanning fluorimetry (DSF), an allusion to the related technique differential scanning calorimetry. DSF is being extensively used in high-throughput assays for small-molecule hit identification, as it can be easily implemented in microplate formats using a real-time thermal cycler (QPCR) instrument.

Together, these studies will likely provide a novel set of chemicals that can be used to further probe the roles of the SUMO pathways in cancer, and importantly such compounds can be developed into leads for preclinical/clinical development. Thus, this innovative research could provide a new therapeutic avenue that is urgently needed to treat tumors with some of the most-poor treatment prognoses.

3.3 Results

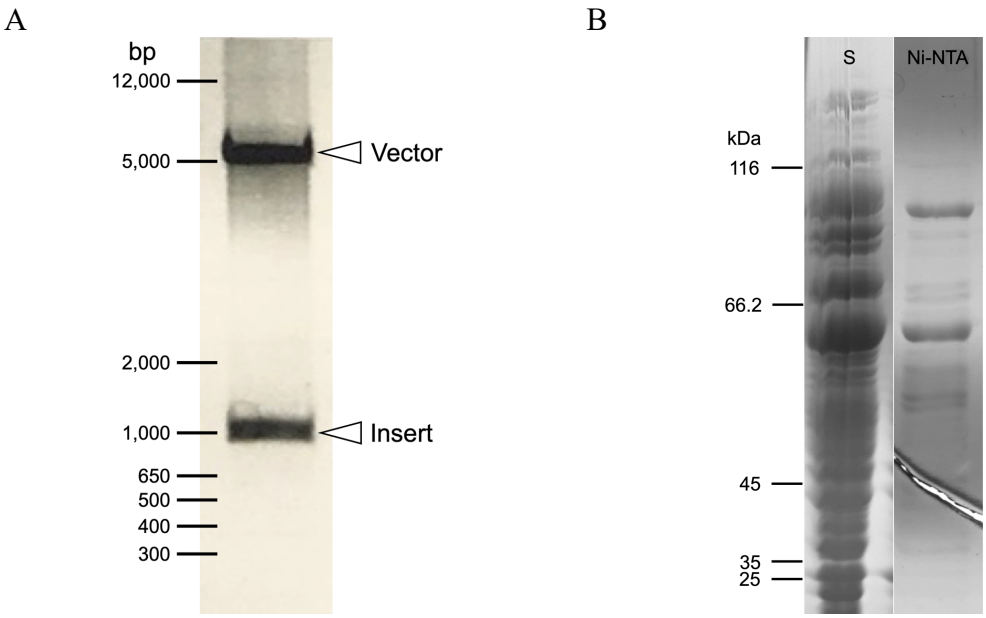


Figure 3.3 | Cloning and expression for Aos1 and Uba2 (A) Restriction digest analysis of pGEX-6p1 Aos1 with *Bam*HI and *Sal*I as judged by UV illumination of 1% agarose gel. DNA marker used was 1 Kb Plus DNA Ladder (Invitrogen™) (B) Co-expression test for Aos1/Uba2 and nickel bead elution sample. Protein marker used was Unstained Protein MW Marker (Pierce™).

3.3.1 Recombinant DNA and expression studies

To conduct our intended assays with the SUMO system, we first needed produce the proteins recombinantly. Standard molecular biology techniques were utilized to re-clone the Aos1 gene into a pGEX-6P1 GST vector to pair with our pre-existing construct of 6xHis-Uba2 in pET 28b. Success of this cloning was validated by restriction digestion (**Fig. 3.3A**) and subsequently by Sanger sequencing. Transforming both plasmids into BL21 (DE3) yielded successful co-expression conditions for the recombinant Aos1-GST/6x-His-Uba2 heterodimer complex as supernatant and Ni-NTA elution (**Fig. 3.3B**). The final step of the purification was a Superdex 200 16/60 size-exclusion column, and when analyzed,

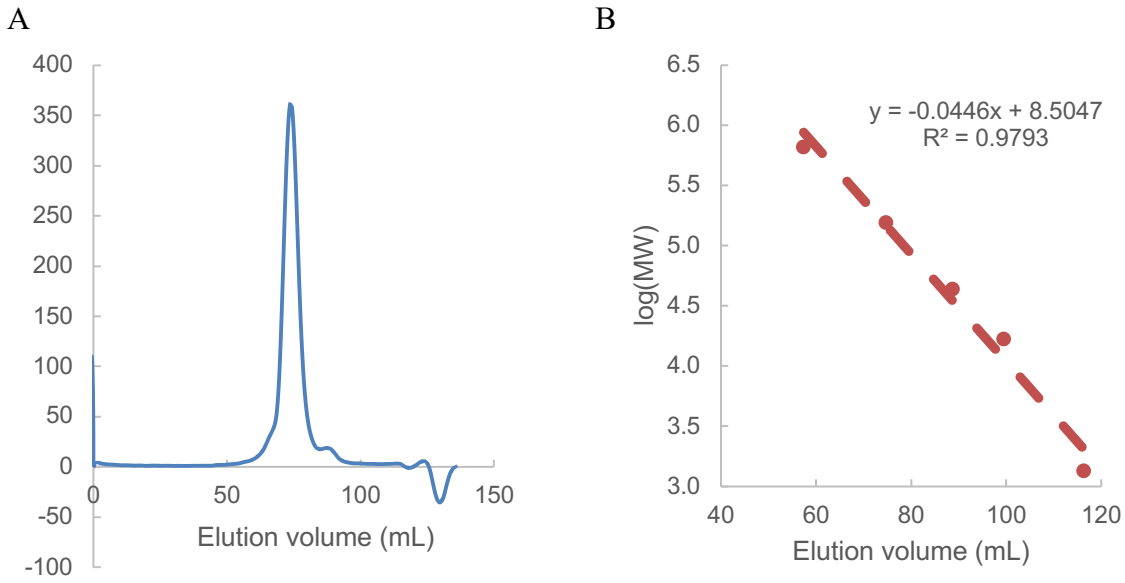


Figure 3.4 | Gel filtration analysis of Aos1/Uba2 (A) Elution of Aos1/Uba2 heterodimeric complex plotted as absorbance at 280 nm versus elution volume in units of mL. (B) Standard curve from calibration run. Equation of the line was used to determine in solution molecular weight of Aos1/Uba2. Data are known log(MW) values plotted against empirically determined elution values for calibration proteins (Bio-Rad #1511901).

the protein complex elutes at a volume associated with an in-solution molecular weight of 161 kDa (**Fig. 3.4**), 50 kDa above the predicted heterodimeric mass.

3.3.2 SUMOylation inhibition by anacardic acid derivatives

Western blot- The first series of possible drug candidates we tested against the SUMO activating enzyme complex were anacardic acid and derivatives thereof. Through our collaboration with Dr. Dave Martin in Chemistry at UCR, several derivatives of anacardic acid (AADs) were produced. To test the efficacy of these compounds, an *in vitro* SUMOylation reaction of RanGAP was conducted. The first two lanes are controls with or without ATP and the following four lanes are the results of the SUMOylation reaction in the presence of 1 mM anacardic acid (Sigma), AAD #1, #4, and #6 (**Fig. 3.5**). Inhibition

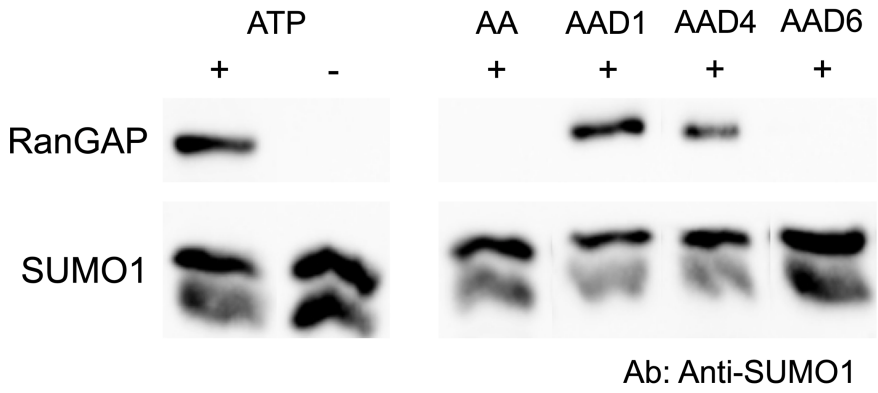
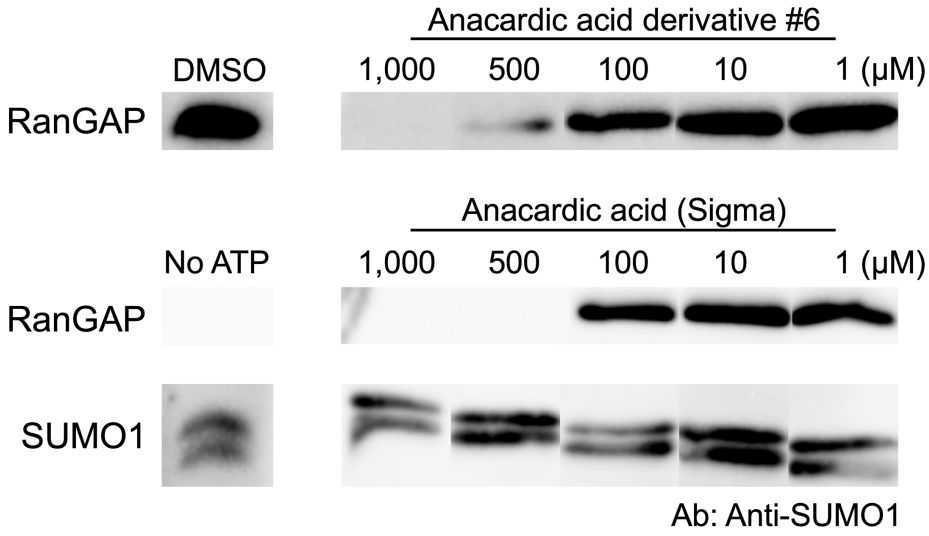


Figure 3.5 | Western blot inhibition analysis of anacardic acid derivatives SUMO-transfer to RanGAP substrate evidenced by SUMO1 antibody blotting. Image captured following chemiluminescence and detection on a Gel Doc™ (Bio-Rad).

of the SUMOylation reaction was observed for the commercial anacardic acid and AAD #6. Each lane was run with excess SUMO1 to demonstrate the successful western transfer and SUMO1-Ab blotting.

To further understand the nature of this inhibition, titration series were conducted for both anacardic acid and AAD #6. Each reaction was conducted with final compound concentrations of 1,000, 500, 100, 10, and 1 μM (**Fig. 3.6A**). The intensity of each resultant RanGAP-SUMO band observed was analyzed (ImageLab; Bio-Rad) and compared to a DMSO control to calculate percent inhibition (**Figure 3.6B**). Calculating the IC_{50} value from our Western blot titration analysis, we find for each of the compounds ~90 μM. This result (**Table 3.1**) does not reflect that of the literature where anacardic acid has been found to have an IC_{50} against SUMOylation of RanGAP of 2.2 μM as determined by Western blot titration [187]. Due to the insensitivity and difficulties in reproducibility, other assays were pursued.

A



B

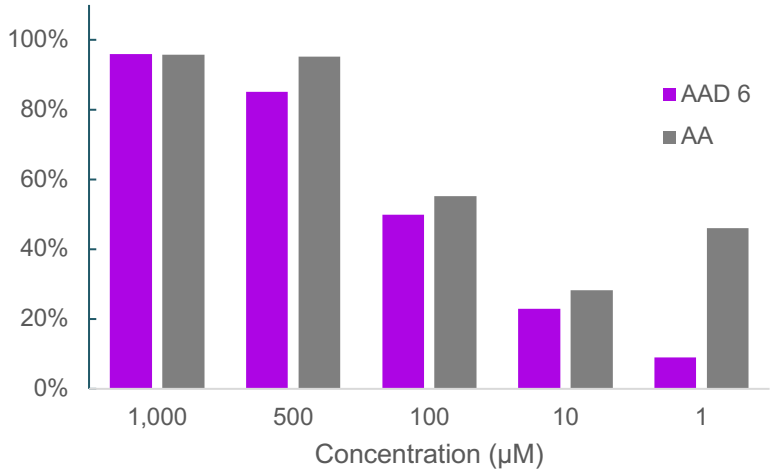


Table 3.1 | AAD ΔT_m

Compound	IC_{50} (μM)
AAD #6	91.17
Anacardic acid	89.07

Figure 3.6 | Western blot inhibition titration analysis (A) SUMO-transfer to RanGAP substrate evidenced by SUMO1 antibody blotting. Image captured following chemiluminescence and detection on a Gel Doc™ (Bio-Rad). (B) Percent inhibition plotted against compound concentration in units of μM where the anacardic acid derivative is depicted in purple and anacardic acid (Sigma) is depicted in grey. Band intensity determined in ImageLab (Bio-Rad) and percent inhibition calculated by comparing to DMSO control. IC_{50} values calculated using AAT Bioquest® 4PL curve calculator [156].

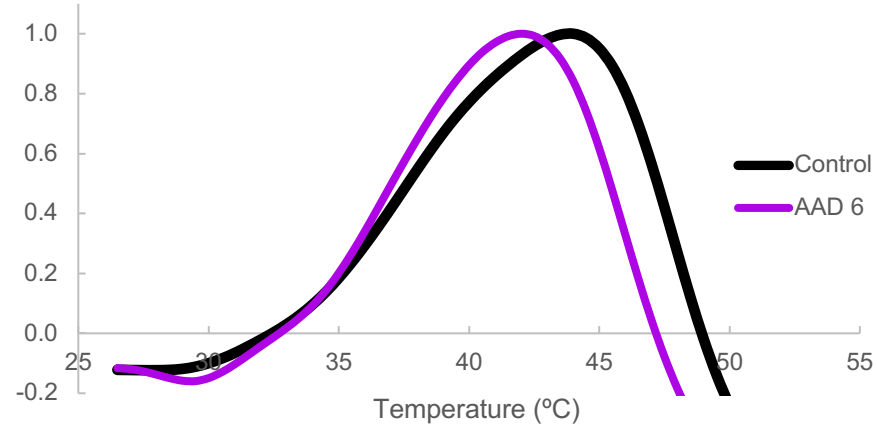
Differential scanning fluorimetry- The thermal stability of a macromolecule is altered upon interaction with a ligand and can be analyzed using differential scanning fluorimetry (DSF). In fact, the DSF assay we've developed proved quite insightful for understanding protein-ligand interactions. Here we show that the nature of the interaction between the compound and target is destabilizing (**Fig. 3.7A**). The inhibition is not due to full denaturation, but rather minor destabilization of the complex, lowering the melting temperature by 2 ºC.

FRET- Additionally, these anacardic acid derivatives were tested in collaboration with Dr. Jiayu Liao's laboratory in Bioengineering at UCR. They have developed a highly sensitive FRET-based assay to determine full SUMOylation activity *in vitro* [201]. In their assay, CyPet-SUMO1 and YPet-RanGAP are incubated with SUMOylation machinery of Aos1/Uba2, Ubc9, together with Mg-ATP, and the fluorescence is measured and analyzed. Therefore, fluorescence detected above background levels would correspond to successful transfer of the SUMO moiety to the target substrate RanGAP [202]. When inhibition studies were conducted using anacardic acid and our derivatives, full catalytic depletion was observed for AAD #6 at 300 μM (**Fig. 3.7B**). This indicates that the IC_{50} is lower than previously published, and more in keeping with our initial western blotting protocol.

3.3.3 SUMOylation inhibition studies with SBDD compounds

FRET- To begin determining the potential interactions between the virtual screening compounds from our collaboration with the Mobley Lab (UCI) and our partnership with

A



	ΔT_m (ºC)
AAD6	-2

B

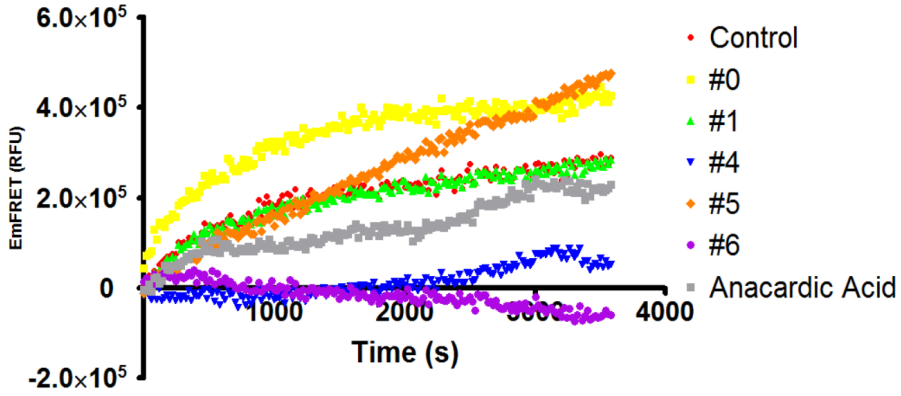


Figure 3.7 | AAD interactions with Aos1/Uba2 analyzed by DSF and FRET (A) Thermal stability of Aos1/Uba2 with and without AAD #6 presented as normalized dRFU/dT values against temperature in units of degrees Celsius. Data are means of three replicates. DMSO-only control is plotted in black for contrast. (B) FRET analysis conducted by Zhehao (Michael) Xiong from Jiayu Liao's group.

Atomwise and the SUMO E1 enzyme, we utilized the FRET-based assay. The SUMOylation reaction was conducted in the presence of 1, 5, and 10 μM of these compounds and measured for fluorescence (**Fig. 3.8**). Depletion of the FRET signal was

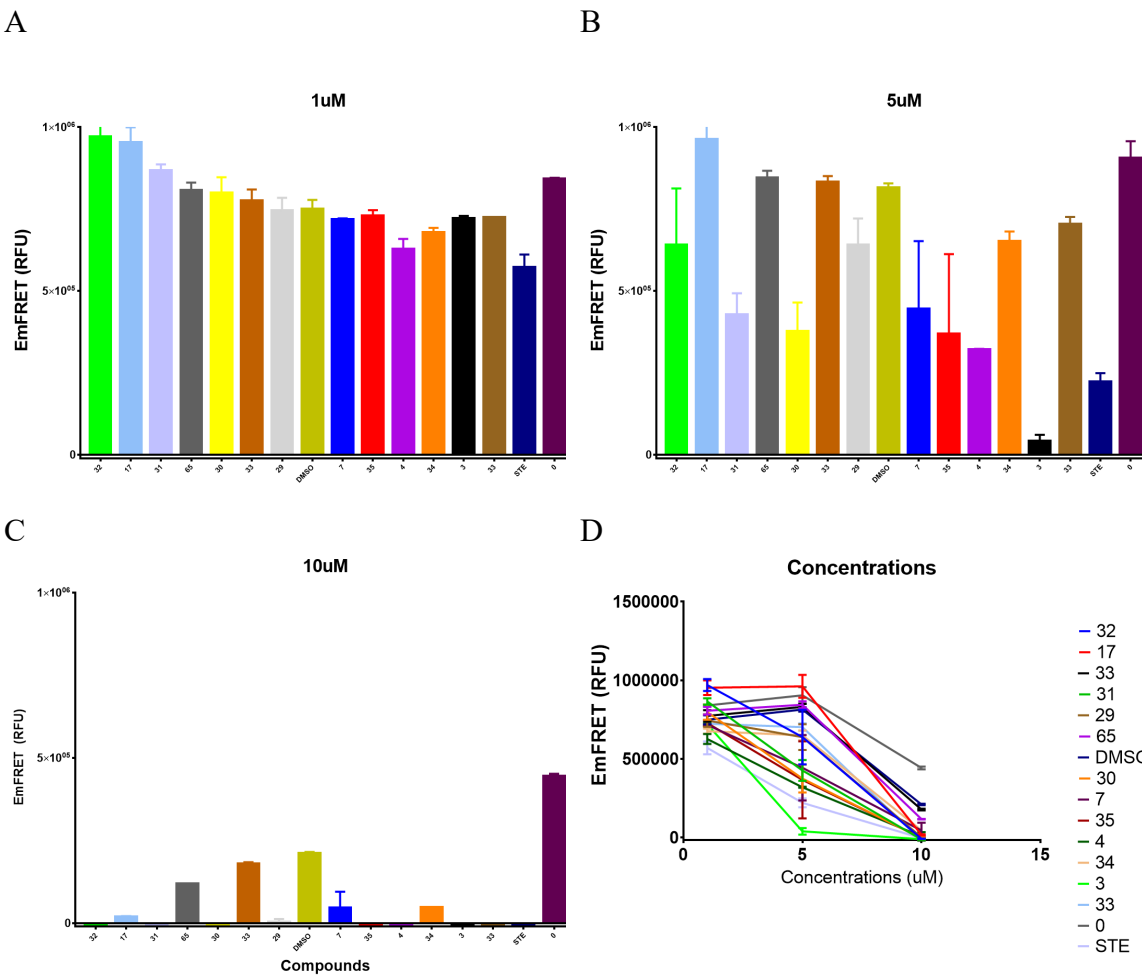


Figure 3.8 | SUMOylation inhibition by SBDD compounds Titrations for the compounds were tested at (A) 1 μM, (B) 5 μM, and (C) 10 μM using a FRET-based assay. FRET analysis conducted by Zhehao (Michael) Xiong from Jiayu Liao's group.

observed for many of these compounds at 10 μM, but further investigation was needed to understand the nature of the interactions.

NCI DSF- After our own wider test, we selected our hit compounds to conduct a DSF assay in triplicate (**Fig. 3.9**). We were able to observe significant shifts in thermal stability for many of these NCI compounds (**Table 3.2**), but they also heavily diminished the amplitude of the RFU value. ΔT_m values were calculated as before (3.2). DSF data herein has been

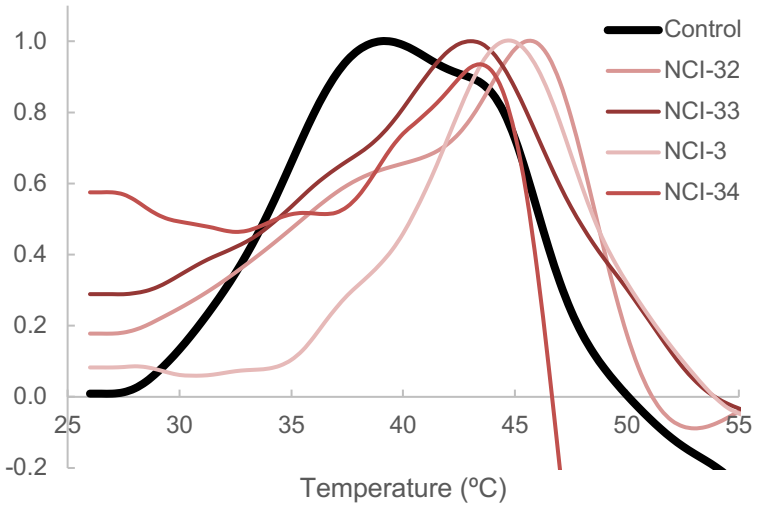


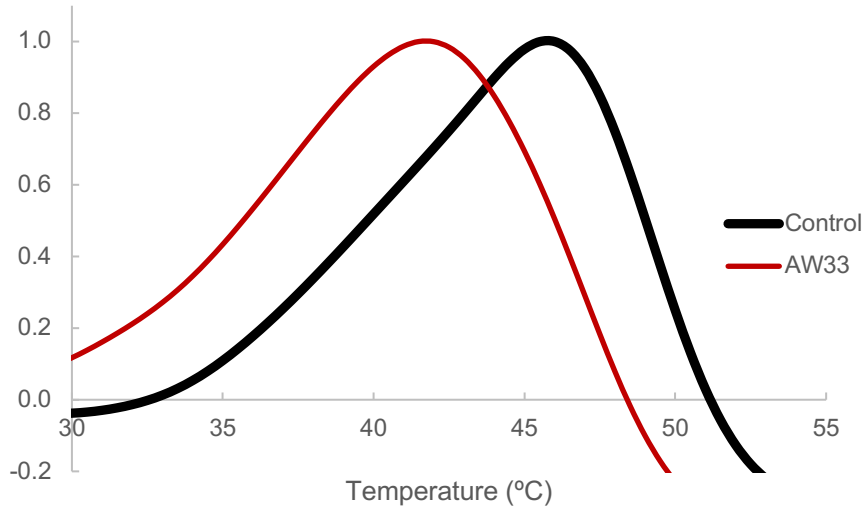
Table 3.2 | NCI ΔT_m	
Compound	ΔT_m (ºC)
NCI-3	5*
NCI-32	6*
NCI-33	3.5*
NCI-34	4*
*Lowers fluorescence	

Figure 3.9 | DSF analysis of NCI virtual screening compounds (A) Thermal stability of Aos1/Uba2 with NCI-3, 32, 33, and 34 presented as normalized dRFU/dT values against temperature in units of degrees Celsius. Data are means of three replicates. DMSO-only control is plotted in black for contrast.

normalized for presentational reasons, but the overall signal is lowered by 10,000-fold. Nevertheless, any compound that improves the thermal stability by greater than 4 ºC is of strong interest.

AW DSF and FRET- From the Atomwise SBDD compounds, only one gave a confident result in the DSF assay. AW33 negatively affected the thermal stability of the Aos1/Uba2 heterodimer complex by 3 ºC (**Fig. 3.10A**). As DSF only provides binding information, the same FRET-based assay as before was used to further investigate the interaction. Minor inhibition was observed (**Fig. 3.10B**), but 50% activity was not reached even with 100 µM AW33. Roughly speaking, that would mean the IC_{50} for AW33 is likely well above 100 µM. While this compound obviously is not suitable now, there is evidence of an

A



	ΔT_m (ºC)
AW33	-3

B

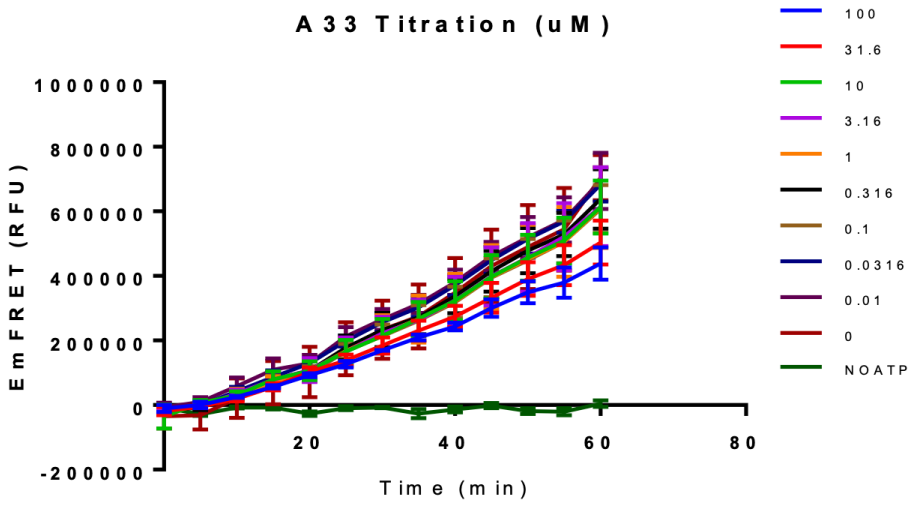


Figure 3.10 | Atomwise compound 33 analyzed by DSF and titrating FRET (A) Thermal stability of Aos1/Uba2 with and without AW33 presented as normalized dRFU/dT values against temperature in units of degrees Celsius. Data are means of three replicates. DMSO-only control is plotted in black for contrast. (B) FRET analysis conducted by George Way from Jiayu Liao's group.

interaction, and this compound could therefore serve as a scaffold for subsequent SAR efforts.

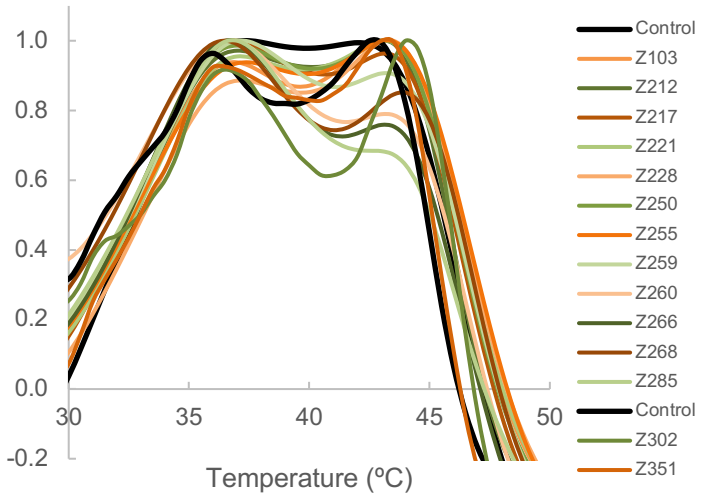


Table 3.3 | Fragment ΔT_m	
Name	ΔT_m (ºC)
Z103	1.5
Z212	1.5
Z217	1.5
Z221	1.5
Z228	1.5
Z250	1.5
Z255	1.5
Z259	1.5
Z260	1.5
Z266	1.5
Z268	2
Z285	1.5
Z302	1
Z351	0.5
Z425	1

Figure 3.11 | DSF analysis of Aos1/Uba2 and fragment library Thermal stability of Aos1/Uba2 with 10 mM of fragment molecules presented as normalized dRFU/dT values against temperature in units of degrees Celsius. Data are means of three replicates. DMSO-only control is plotted in black for contrast.

3.3.4 Aos1/Uba2 binding studies with fragment library

In addition to our SBDD approaches, we also pursued fragment-based studies. To perform FBLD, purified Aos1/Uba2 protein (2.5 μM) was mixed with 10 mM of each fragment in our library, and the potential binding interactions between the fragment and the enzyme were probed using the DSF assay (**Fig. 3.11**). From the 428 fragments tested, we found 15 fragments (3.5%) that significantly affected the thermal stability of the E1 protein (**Table 3.3**). ΔT_m values were calculated as before (3.2). The nature of the interactions between these fragments and Aos1/Uba2 require further investigation and replicates.

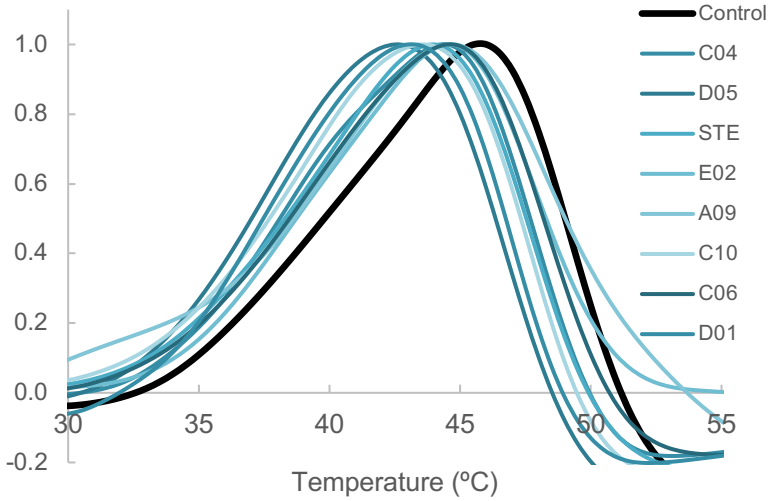


Table 3.4 | SD ΔT_m	
Compound	ΔT_m
STE	-2
SD-A09	-1.5
SD-C04	-1.5*
SD-C06	-1.5*
SD-C10	-2*
SD-D01	-3*
SD-D05	-3.5*
SD-E02	-1.5
*Lowers fluorescence	

Figure 3.12 | DSF analysis of Aos1/Uba2 with STE-derivatives Thermal stability of Aos1/Uba2 with 1 mM of each STE-derivative presented as normalized dRFU/dT values against temperature in units of degrees Celsius. Data are means of three replicates. DMSO-only control is plotted in black for contrast.

3.3.5 Aos1/Uba2 binding studies with STE-derivatives

The final arm of our drug discovery effort was using derivatives generated from a FRET-based high-throughput screen (HTS) identified by the Liao group. This compound's chemical structure will not be presented herein. For intellectual property reasons the chemical will be referred to as STE. We have conducted our own homology searches and found our own STE-derivatives (SD), which we have tested to probe the effects of certain modifications to the molecules through DSF (**Fig. 3.12**). When STE was analyzed through DSF, a -2 ºC shift was observed (**Table 3.4**). Our derivatives show similar values, and in some cases, an increased shift. Analysis of the derivations made to STE are ongoing.

A B

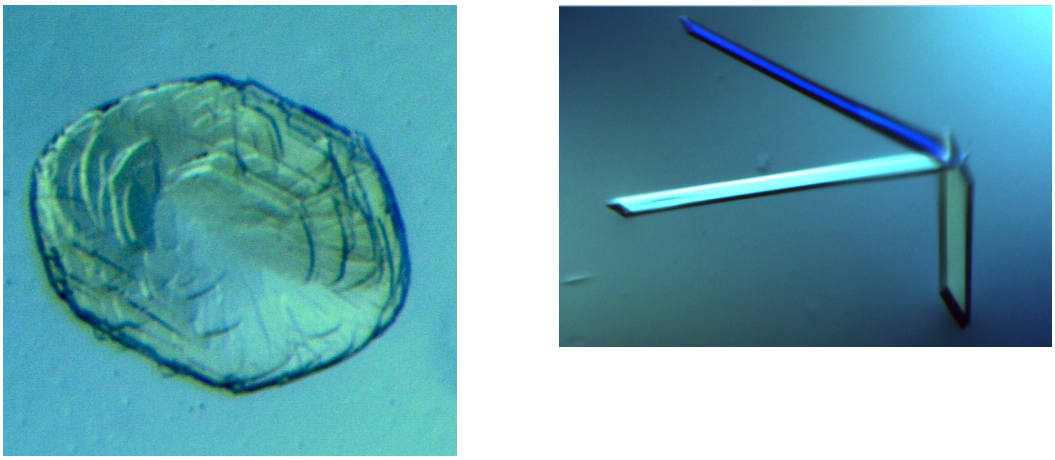


Figure 3.13 | SUMOylation system crystals (A) Pseudo-crystals formed of the E1 Aos1/Uba2 heterodimer visualized under polarized light (B) Long (300 μm) crystals formed by Ubc9 protein visualized under polarized light.

3.3.6 Crystallization attempts for the SUMOylation system

We are now initiating protein crystallography studies to define interactions of our small molecule or fragment hits at atomic resolutions, to thereby aid the hit-to-lead development state. Crystals have been obtained for the wild-type Ubc9 protein that diffracted natively to 2.0 Å (**Fig. 3.13B**). Fragment hits are being investigated for Ubc9 binding, but the crystals are highly DMSO-sensitive so new crystallization conditions will be needed before further binding studies are conducted. For the Aos1/Uba2 crystals, further refinement of the apo conditions will be needed as these pseudo-crystals (**Fig. 3.13A**) grew inconsistently and diffracted poorly (> 10 Å). Efforts to adjust the plasmid construct, cleave the 6x-His with thrombin, and reduction in contaminants during purification are ongoing to assist in crystal formation.

3.4 Discussion

Discovering novel therapeutics for intractable diseases such as KRAS, c-Myc, and Notch-1 cancers is of strong interest to the scientific community. The notion that these diseases, and many other disease states including viral infections including HIV-1 [203] and certain neurodegenerative diseases such as Huntington's disease [204], are reliant on non-oncogene pathways provides a unique avenue for their treatment. SUMOylation is one such pathway in which inhibition would impart synthetic lethality. Our efforts herein lay the groundwork for our multi-pronged endeavors in developing small-molecule inhibitors for the SUMO E1 activating enzyme, Aos1/Uba2, and now also FBDD approaches with the E2, Ubc9.

Anacardic acid is the best-known natural product inhibitor of SUMOylation to date. Our lab has experience working with anacardic acids [205], [206], including chemically derived analogs [207]. Our initial anacardic acid-based derivatives did show inhibition at high concentrations, yet they failed to inhibit at therapeutically feasible concentrations, and thus further analogs/rational design is required. It should be noted that the results from our DSF assay and AAD #6 corresponded to a negative shift, likely due to a destabilization of the heterodimeric state, but even at high concentrations it did not lead to nonspecific denaturation. Anacardic acid has a long hydrocarbon chain making it poorly soluble. The modifications made in generating AAD #6 improved solubility and generally improve the ADME values, the drug-like character of the molecule. More remains to be defined about the nature of the interaction between anacardic acid and Aos1/Uba2, and this information would greatly aid the lead-development strategies.

In vitro assays were conducted on the *ab initio* virtual screening compounds. Our powerful DSF assay demonstrated strong alteration to protein stability in the presence of many of these molecules. Combined with the FRET studies on SUMOylation inhibition, these compounds can be fed back into the pharmacophore scoring to reinitiate another iteration of *in silico* studies. Likewise, the STE-derivatives we tested can be used to better understand what functional groups on the drug-like molecule increase or decrease the binding to the substrate, so as to develop a further improved analogue with optimized binding features.

Possibly the technique most reliant on crystallographic information is FBDD. Although several fragment hits were observed in our DSF assay, elucidating the nature of these interactions is a critical step in the lead development. In the previous SBDD techniques, larger molecules with multiple rings and functional groups were used. Fragments however are small molecules, in the 150-220 dalton range. As they are in fact so small, the DSF technique is one of the only techniques that has been demonstrated to be sensitive enough for initial FBLD analysis, albeit at remarkably high fragment concentrations. Enzymatic inhibition assay could not be used as it is unlikely that fragments would show strong, or any inhibition, as it has been previously noted that fragments have dissociation constants in the micro to low millimolar range for known targets [208]. A crystal structure of the fragment-protein complex would greatly aid our efforts, and thus new purification schemes will be implemented in that include a thrombin cleavage step to remove the N-terminal 6x-His tag from Uba2 and new cloning efforts will be undergone to relocate the 6x-His tag to the carboxy terminus of the protein.

Our hits from natural product chemistry, *in silico*, and SAR by catalog endeavors serve as an exciting starting point in our pursuits in developing novel tools to probe the effect of inhibition of the SUMOylation system in intractable disease states.

3.5 Methods

3.5.1 Recombinant cloning and expression

Cloning- To begin studying the Aos1/Uba2 heterodimeric protein complex, first the gene would need to be re-cloned. Both genes were available in pET28b plasmids, however if co-expression was going to be conducted, each plasmid would need to house different antibiotic resistances in order to select for colonies transfected with both vectors. For this reason, the Aos1 protein was re-cloned into the pGEX6P1 GST-fusion vector. Primers (**Table 3.5**; IDT) were designed for the full length Aos1 gene from the available pET28b construct and the insertion of the PCR product into pGEX6P1 following digestion with *Bam*HI and *Sal*I. The PCR reaction was conducted much the same as before (1.5.1, 2.5.1), only the annealing temperature here was lowered to 52 ºC. Ethanol precipitation (**Appendix A**), agarose gel extraction (**Appendix B**), ligation reaction, transformation (**Appendix C**), and miniprep (**Appendix D**) were followed as outlined in 1.5.1. Likewise, the final restriction enzyme digest analysis (**Fig. 3.3A**) was conducted as in 1.5.1 with *Bam*HI and *Sal*I. Successfully ligated constructs were further confirmed by Sanger sequencing (UCR core facility).

Table 3.5 | Primers used in Aos1 cloning	
Forward primer	5'- CGC GGA TCC ATG GTG GAG AAG GAG -3'
Reverse primer	5'- ACG CGT CGA CTC ACT TGG GGC CAA -3'

Expression- To determine suitable protein expression conditions for the Uba2 and the Aos1-GST fusion protein, small-scale tests were conducted. Here, miniprepped plasmid for Aos1-GST and Uba2 were co-transformed into BL21 (DE3) and expression was

determined much the same as demonstrated in 1.5.1 and 2.5.1. After running a sample of the supernatant and pellet on an SDS-PAGE gel, the supernatant was run down Ni-NTA beads and eluted in high imidazole. The elution was run on the SDS-PAGE gel as well and clearly demonstrates two prominent bands eluted (**Fig. 3.3B**).

After determining that BL21 (DE3) was suitable for Aos1/Uba2 expression, production efforts were scaled up. An overnight (O/N) 7 mL culture was inoculated using a glycerol stock generated from a previously uninduced sample. The O/N culture was secondarily inoculated into 1 L of 2X-YT supplemented with ampicillin, chloramphenicol, and kanamycin and incubated at 37 ºC and 225 rpm (New Brunswick Scientific Innova©43R) for approximately 4 hours until the OD_{600nm} = 1.0. To optimally express the Aos1/Uba2 heterodimer, the temperature was lowered to 25 ºC, induced with 0.4 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG), and set to shake O/N for 15 hours. The cells were harvested by 4ºC centrifugation at 5,000 rpm for 20 minutes (Sorvall LYNX 4000 Superspeed Centrifuge with a Fiberlite™ F9-6 x 1000 LEX Fixed Angle Rotor), and the cell pellets were collected and stored at -80 ºC.

To purify the Aos1-GST/Uba2 complex the cells were resuspended, sonicated, and centrifuged as before (1.5.1, 2.5.1). The supernatant was carefully decanted and loaded onto a Ni-NTA column (HisTrap™ FF; GE Life Sciences) using an NGC Scout Liquid Chromatography System (Bio-Rad). The eluted fractions were then loaded onto a GSTPrep FF 16/10 (GE Life Sciences) column and eluted against buffer supplemented with 10 mM reduced glutathione. The fractions containing the protein complex were pooled and a ratio of 1 µg PreScission Protease to 1 mg Aos1/Uba2 was added and incubated for 15 hours at

8 ºC. The digest reaction was diluted to 50 mM NaCl and loaded onto HiTrap Q FF (GE Life Sciences) and eluted against a gradient to 500 mM NaCl. Fractions containing Aos1/Uba2 were concentrated to 5 mL and run down a HiLoad® 16/600 Superdex® 200 pg gel filtration column (GE Life Sciences), fractionating every 2 mL. Samples containing purified Aos1/Uba2 complex were determined by SDS-PAGE gel and concentrated to 10 mg/ml (ε = 0.6136) in 10 mM Tris-PH 7.5, 87 mM NaCl, 1mM DTT storage buffer, flash-frozen in liquid nitrogen, and stored at -80 ºC.

3.5.2 SUMOylation reaction

To study the effects of anacardic acid on the SUMOylation reaction, a kit was purchased, and the enclosed protocol was followed. The kit provides all components required for an *in vitro* SUMOylation reaction and the antibodies required for subsequent Western blot. To conduct our inhibition assay, 19 µL of a reaction master-mix containing 13 µL buffer and 1 µL of each E1, E2, RanGAP, SUMO1, and Mg-ATP solution was added to 1 µL of each anacardic acid derivative at a 1 mM final concentration. As a positive control the reaction was conducted using 1µL of DMSO (5% final), and as a negative control the reaction was conducted in the absence of Mg-ATP. Each reaction was pipetted, thoroughly mixed, and incubated at 37 ºC for one hour. Following this hour, the reaction was stopped with 2X SDS-PAGE loading buffer containing Coomassie, glycerol, and β-mercaptoethanol. These reactions were run on a 12% SDS-PAGE gel for 70 minutes. The gel was then assembled into a Western transfer apparatus neighboring a PVDF membrane. After the 70-minute 250mA transfer, the PVDF membrane was washed in PBS, followed

by a 2-hour room temperature blocking using 1X PBS with 1% casein (Bio-Rad). After the PVDF was blocked, the membrane was washed in 1X PBS again, then in an overnight 4 ºC step the membrane was blotted with the 1:1000 primary rabbit SUMO1 polyclonal antibody. The next day, the membrane was washed, then blotted with a 1:5000 HRP conjugated, anti-rabbit, goat monoclonal antibody (Bio-Rad) for an hour. After a final wash, the membrane was exposed using the SuperSignal West Pico Chemiluminescent Substrate (Fisher) enhancer and conjugator duo, then imaged in a Gel Doc™ EZ Gel Documentation System (Bio-Rad). For the follow-up titration experiment, the reaction, blot, and analysis were all conducted in the same manner, only with anacardic acid and AAD #6 at 1,000, 500, 100, 10, and 1 μM final concentrations. Bands were analyzed using ImageLab (Bio-Rad) and plotted using Microsoft Excel.

3.5.3 Differential scanning fluorimetry

To conduct the differential scanning fluorimetry (DSF) assay for Aos1/Uba2, 19 μL from a master mix of 2.5 μM Aos1/Uba2, 20X SYPRO® Orange Dye, in a 10 mM HEPES pH 8.0, 50 mM NaCl buffer was added to 1 μL of each compound to be tested pre-aliquoted into a white, 0.2 mL semi-skirted 96-well tray (AB-0900/W; Thermo Scientific). The anacardic acid derivatives, Atomwise, NCI diversity set, and STE-derivative, and fragment compounds tested in this way and were therefore diluted to a final concentration of 1 mM. The 96-well plate was then sealed with optically clear Microseal® 'B' seal Seals (MSB1001; Bio-Rad) and loaded into a CFX Connect™ Real-Time PCR Detection System (Bio-Rad) set to the following protocol: thermostatic increase from 25 ºC to 95 ºC for 70

minutes taking fluorescence scans at excitation and emission wavelengths of 470/570 nm every 30 seconds, or every 0.5 ºC interval. The melting temperature (T_m) of the protein can be inferred from the inflection point of this generated fluorescence curve **(Fig. 3.7, 3.9, 3.10, 3.11, 3.12)**. Therefore, the thermal shift (ΔT_m) is calculated by subtracting the control T_m (DMSO only) from that of tested compounds melting temperatures **(Table 3.2, 3.3, 3.4)**.

3.5.4 FRET studies

The FRET-based experiments were entirely conducted by the Liao lab in UCR Bioengineering. Methods are modified from [201]. For the anacardic acid derivatives, NCI Diversity compounds, and Atomwise molecules 1 μL (1.67% DMSO final) was mixed in the presence of 1 μM CyPet–SUMO1, 2 μM YPet–RanGAP, 0.1 μM Aos1/Uba2, 0.5 μM Ubc9, and 2 mM ATP in a buffered solution containing 50 mM Tris-HCl (pH 7.4), 1 mM DTT, and 4 mM $MgCl_2$ to a total volume of 60 μL. Samples were incubated at 22 °C and fluorescence emission was monitored. CyPet–SUMO1 and YPet–RanGAP were mixed with or without 10 mM ATP as a control. Their fluorescence signals were determined by the fluorescent plate reader Flexstation II 384 (Molecular Devices) in a 384-well plate over a period of time. After the emission intensities were corrected by subtracting the background fluorescence from the plate, the FRET indexes [209] were calculated and compared to reactions with varied concentration of compound.

Chapter 4
Conclusions

Chapter 4: Conclusions

4.1 Introduction

Nearly every fundamental biological process necessary for life is conducted by proteins. Structural biology is a field of research that studies the intricate molecular architecture of these macromolecules. Elucidating the three-dimensional structure of a protein provides unparalleled insights into the mechanism of its catalysis. Here, we determine the crystal structure of one enzyme, hDUF89, and utilize the crystal structures of two other therapeutically intriguing enzymes, HtrA1 and Aos1/Uba2, to initiate structure-based drug discovery studies.

4.2 DUF89 Conclusion

Our initial investigation of the hDUF89 protein began in earnest due to the implications of its roles in the DNA damage response where the reported phenotype upon hDUF89 knockdown generated DNA damage sensitivity in select breast cancer cell lines. This would have made hDUF89 an intriguing target for therapeutic endeavors, however, our characterization of the structure and *in vitro* activity, and a comprehensive re-evaluation of *C6orf211* knockouts in breast cancer cell lines, strongly refuted many of those claims. Nevertheless, the hDUF89 structure and function proved independently intriguing.

Our elucidated crystal structure contains of a core α-β-α fold, an active-site coordinated magnesium ion, and an N-terminal α-helical bundle, which are all conserved features of subfamily III DUF89 enzymes. While the human DUF89 structure shares certain structural similarities to SAM-MT methyltransferases, the recombinant protein did

not demonstrate any readily detectable methyltransferase activity. In fact, enzymatic incubation with a host of biological phospho-metabolites revealed substrate depletion in our LC/MS-MS studies and dephosphorylated product formation in our malachite green-based assays. Therefore, we conclude that the *C6orf211* gene product encodes for a subfamily III DUF89-fold containing metal-dependent phosphatase, which displays the highest dephosphorylation activity against fructose 1-phosphate *in vitro*.

4.3 DUF89 Future Perspectives

The determinations we present herein beg the inquiry of deeper biological questions. Several new motifs were discussed here, yet investigations of mutated disease states or alanine-scanning type mutations could provide insights into the explicit roles of these motifs. Relatedly, it has been noted in proteomic databases that post-translation modifications (acetylation) are observed on specific residues in the N-lobe region of the protein. Cell-based studies could attempt to understand what the roles of these PTMs play *in vivo*. Other than Arg24, we conducted no analysis of other aspects of the N-lobe region of the protein, which could also prove insightful upon further, *in vivo*, investigations.

More comprehensive studies of the enzymatic activity could also be conducted. We do not present any evidence of the *in vivo* function of this enzyme, which requires a distinct skill set, and such analysis would be required to reinforce many of our suggestions about the cellular role of this enzyme. If the biological function matches that of the activity of our recombinant enzyme, as we predict it does, *in vivo* and *in vitro* investigations conducted on the ScDUF89 protein found the same activities, analysis of the metabolome of these

cells could yield rich details about the enzyme's role. Further, analysis of cellular metabolomics data under stress conditions, such as starvation, high glucose/fructose, or tumorigenic conditions could prove of interest with *C6orf211* knockouts.

4.4 HtrA1 Conclusions

HtrA1 is an extracellular protease which recognizes hydrophobic peptide regions, a characteristic of misfolded proteins, and degrades them. Mutations in the promotor region of *HTRA1* resultant in overexpression have been linked to disease states, such as AMD and arthritis. Efforts to discover small molecule inhibitors against HtrA1 were therefore pursued, as they could serve as important therapeutic tools. Our real-time fluorescence detection *in vitro* assays identified four inhibitors with low to mid micromolar IC_{50} values, from a suite of 83 AI-CADD compounds predicted to bind the active site of HtrA1.

4.5 HtrA1 Future Perspectives

These compounds represent the first round of an iterative drug discovery pipeline in which the neural net at Atomwise will be fed these inhibition values to improve its training set for virtual screening. Nevertheless, elucidation of direct atomic interactions between these compounds and HtrA1 would significantly improve the AtomNet™ platform. Therefore, co-crystallographic efforts are underway with more soaked crystals being produced with even higher compound concentrations. We have also begun pursuing efforts to crystallize an HtrA1 construct containing just the trypsin-like protease domain, as this has been shown

to improve resolution in an inactive S328A mutant, and thus it may be a suitable path forward, assuming that this residue is not critical for inhibitor compound binding.

Due to the high levels of conservation between the active sites of HtrA1 and HtrA2, and the potential inhibition of HtrA2 resulting in Parkinson's disease, counter-screens for potential therapeutic compounds must be conducted against HtrA2. Thus, cloning and expression procedures have begun on HtrA2, in preparation of counter-screens, and preliminary tests are awaiting better performing inhibitors. One note, the cellular compartmentalization might favor our efforts due to potential HtrA1 inhibitors not needing to be membrane-permeable, as HtrA1 in extracellular while Htra2 is intracellular.

Previous observations of the diffraction pattern for HtrA1 crystals revealed strong X-ray diffuse scattering. We therefore collected X-ray scattering data on seed-grown HtrA1 crystals at ambient (22 ºC) temperature. Data processing is underway, as part of a separate on-going project in the Perry laboratory on developing X-ray diffuse scatter technologies.

4.6 Aos1/Uba2 Conclusions

The SUMOylation cascade is initiated by a SUMO-activating enzyme (E1), for which there is only one in humans: the heterodimeric Aos1/Uba2 (SAE1/2). KRAS cancers constitute 60% of pancreatic cancers, among others, and cannot be targeted directly, However, such tumors are reliant on the SUMOylation cascade for viability. The KRAS mutant tumor is therefore 'addicted' to the activity of this non-oncogenic system, which is heavily upregulated in this cancer. Thus, developing inhibitors for the SUMOylation system would

prove therapeutically interesting. Our efforts to identify such compounds occupy multiple chemical spaces through natural product chemistry, virtual screening, SAR by catalog, and fragment-based drug discovery avenues.

4.7 Aos1/Uba2 Future Perspectives

Many preliminary results have been discovered for each of the arms of our inhibitor search, all of which will require follow-up studies. Although AAD #6 did not show strong inhibition, the derivation process significantly improved the ADME characteristics of the molecule, and further efforts following those modifications could prove interesting. For our virtual screening angle, subsequent investigations must be conducted on the exact nature of the fluorescence quenching from the compounds, as it may be significantly skewing our results, which indicate significant binding (> 3 ºC ΔT_m). Our derivatives of the HTS hit from the Liao research group supplied from commercial chemical sources are of great interest, and SAR studies must be conducted to analyze which modifications to STE drove varied binding character. Last, because the fragments are so small, they are unlikely to inhibit SUMOylation on their own, so the FBDD efforts cannot be validated with *in vitro* activity assays and must be coupled with a more powerful interaction elucidating endeavor, such as X-ray crystallography.

To generate protein crystals for Aos1/Uba2, new purification schemes will be implemented in future studies that include a thrombin cleavage step to remove the N-terminal 6x-His tag from Uba2. Analysis of the crystal packing interface (PDB: 3KYC) suggests that such extraneous amino acids may disrupt crystallization (in that specific unit

cell, with those specific conditions). Additionally, we are pursuing FBDD from an angle not reliant on co-crystallization. We've recently purchased a covalent fragment library of over 1,000 compounds. DSF analysis from the covalent library would be persuasive, especially if the linkage disrupted the active site as this enzyme-fragment complex would be catalytically deficient. Fragments are also being investigated for Ubc9, and although crystals form readily and diffract to 2.0 Å, the crystals are highly DMSO sensitive, making them presently unsuitable for FBDD efforts until we further optimize our soaking or co-crystallization conditions.

APPENDICES

Appendix A Ethanol Precipitation

1. Consolidate PCR reactions.
2. Add ½ volume of 5M KOAc.
3. Add 2½ volume of 100% isopropanol.
4. Pipette up and down several times.
5. Keep at -20ºC for ≥ 30 minutes.
6. Spin at 12,000 rpm at 4 ºC for 20 minutes.
7. Carefully decant the supernatant taking extra precaution to not lose the very small pellet of DNA.
8. Wash and dry twice with 500 µL 70% ethanol. Vortex, spin down, decant.
9. Leave tube open to dry.
10. Dissolve in 50 µL of ddH_2O.
11. Store in -20 ºC freezer.

Appendix B Agarose Gel Extraction

1. Excise bands with a razor blade corresponding to the molecular weight.
2. Solubilize gel slice in 3X volumes of GEX Buffer (likely 6M NaI).
3. Incubate the gel at 55 °C for 10 minutes, inverting regularly.
4. Once dissolved, load onto the provided column and centrifuge 5,000 rpm for 30 seconds.
5. Wash the column with 500 µL of WN buffer and spin 5,000 rpm for 30 seconds.
6. Wash the column with 500 µL of WS buffer and spin 5,000 rpm for 60 seconds.
7. Spin down at 12,000 rpm for 3 minutes to remove residual ethanol from the column.
8. Transfer column to microcentrifuge tube.
9. Add 30 µL of elution buffer to center of column.
10. Centrifuge for 1 minute at 12,000 rpm.
11. Store in -20 ºC freezer.

Adapted from GenCatch™ Advanced Gel Extraction Kit

Appendix C XL-10 Gold Transformation

1. Thaw provided β-mercaptoethanol solution and XL-10 Gold® cells on ice.
2. Once thaw, add 4 µL of β-ME mix to 100 µL of thawed competent cells.
3. Swirl to ensure mixture and keep on ice for 10 minutes.
4. Add 2 µL of ligation reaction.
5. Flick tube to mix and chill on ice for 30 minutes.
6. Heat shock cells in a water bath at 42 °C for 30 seconds.
7. Store on ice for 2 minutes.
8. Add 300 µL of prewarmed SOC medium.
9. Incubate at 37 °C shaking at 225 rpm for one hour.
10. Plate 200 µL and 20 µL onto two LB-agar plates with correct antibiotic.
11. Incubate 37 °C overnight.

Adapted from XL10-Gold® Ultracompetent Cells Transformation protocol (Catalog #: 200314)

Appendix D Miniprep Plasmid Isolation

1. Harvest 2 mL of fresh culture in 2 mL tube by centrifugation 1 min at 10,000 rpm.
2. Discard supernatant and blot inverted tube on paper towel.
3. Add 300 µL Buffer F1 containing RNase1 and completely dissolve pellet by vortex.
4. Add 200 µL Buffer F2 and mix by inverting 10 times at room temp until clear.
5. Add 200 µL Buffer F3 and mix by inverting 10 minutes until cloudy.
6. Centrifuge at 12,000 rpm for 30 seconds.
7. Transfer supernatant to 'Lysate Clearance Column.'
8. Centrifuge at 12,000 rpm for 30 seconds.
9. Add 200 µL of 100% isopropanol to the flow-through. Store in -20 °C for 30 min.
10. Pipette up and down, then transfer sample to a 'DNA Mini Column.'
11. Centrifuge at 12,000 rpm for 30 seconds.
12. Discard flow-through and wash column with 600 µL of DNA Wash Buffer.
13. Centrifuge at 12,000 rpm for 30 seconds.
14. Discard flow-through and reinsert the column adding no solution and spinning again for 1 minute to remove any residual ethanol.
15. Transfer column to labeled microcentrifuge tube.
16. Add 100 µL of elution buffer to column and centrifuge for 30 seconds at 12,000 rpm.
17. Store DNA at -20 °C.

Adapted from BioMiga Mini-Prep Protocol (Catalog Number: PD1218-02)

REFERENCES

[1] F. L. Hünefeld, *Chemistry in animal organization: Physiological-chemical investigations of the material changes, or of the educational life in the animal organism, in particular of the process of blood formation, of the nature of the corpuscles and their nuclei.* Leipzig: Brockhaus, 1840.

[2] W. H. Bragg and W. L. Bragg, "The Reflection of X-rays by Crystals," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 88, no. 605, pp. 428–438, Jul. 1913.

[3] W. henry Bragg, "The Nobel Prize in Physics 1915," *Nobelprize.org*, 1915. [Online]. Available: https://www.nobelprize.org/prizes/physics/1915/summary/. [Accessed: 11-May-2019].

[4] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[5] G. Rhodes, *Crystallography made crystal clear : a guide for users of macromolecular models*. Elsevier/Academic Press, 2006.

[6] H. Rondeau, Jean-Michel; Schreuder, "Protein Crystallography and Drug Design," in *The Practice of Medicinal Chemistry*, 4th ed., G. W. Camille and A. David, Eds. 2015, pp. 511–537.

[7] G. L. Taylor, "Introduction to phasing," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 325–338, 2010.

[8] J. Kendrew *et al.*, "Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution," *Nature*, vol. 185, no. 185, pp. 422–427, Feb. 1960.

[9] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, and A. C. T. North, "Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis," *Nature*, vol. 185, no. 4711, pp. 416–422, Feb. 1960.

[10] M. G. Rossmann *et al.*, "Structure of a human common cold virus and functional relationship to other picornaviruses.," *Nature*, vol. 317, no. 6033, pp. 145–53, 1985.

[11] J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel, "Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3 resolution," *Nature*, vol. 318, no. 6047, pp. 618–624, Dec. 1985.

[12] D. A. Doyle *et al.*, "The structure of the potassium channel: molecular basis of K+ conduction and selectivity.," *Science*, vol. 280, no. 5360, pp. 69–77, Apr. 1998.

[13] Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, and R. MacKinnon, "Crystal structure and mechanism of a calcium-gated potassium channel," *Nature*, vol. 417, no. 6888, pp. 515–522, May 2002.

[14] S. G. F. Rasmussen *et al.*, "Crystal structure of the human β2 adrenergic G-protein-coupled receptor," *Nature*, vol. 450, no. 7168, pp. 383–387, Nov. 2007.

[15] J. P. Abrahams, A. G. W. Leslie, R. Lutter, and J. E. Walker, "Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria," *Nature*, vol. 370, no. 6491, pp. 621–628, Aug. 1994.

[16] J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister, and R. Huber, "Crystal structure of the 20S

proteasome from the archaeon T. acidophilum at 3.4 Å resolution.," *Science*, vol. 268, no. 5210, pp. 533–9, Apr. 1995.

[17] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, Sep. 1997.

[18] D. E. Brodersen, W. M. Clemons, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan, "The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit.," *Cell*, vol. 103, no. 7, pp. 1143–54, Dec. 2000.

[19] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.," *Science*, vol. 289, no. 5481, pp. 905–20, Aug. 2000.

[20] P. Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz, "The structural basis of ribosome activity in peptide bond synthesis.," *Science*, vol. 289, no. 5481, pp. 920–30, Aug. 2000.

[21] M. Pioletti *et al.*, "Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3," *EMBO J.*, vol. 20, no. 8, pp. 1829–1839, Apr. 2001.

[22] P. Cramer, D. A. Bushnell, and R. D. Kornberg, "Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution," *Science (80-.).*, vol. 292, no. 5523, pp. 1863–1876, Jun. 2001.

[23] T. H. Escherich, "The intestinal bacteria of the neonate and breast-fed infant," *Rev. Infect. Dis.*, 1989.

[24] Z. D. Blount, "The unexhausted potential of E. coli," *Elife*, vol. 4, Mar. 2015.

[25] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, "General nature of the genetic code for proteins," *Nature*, vol. 192, no. 4809, pp. 1227–1232, 1961.

[26] M. Nirenberg *et al.*, "RNA codewords and protein synthesis, VII. On the general nature of the RNA code.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 53, no. 5, pp. 1161–8, May 1965.

[27] A. McPherson and J. A. Gavira, "Introduction to protein crystallization.," *Acta Crystallogr. Sect. F, Struct. Biol. Commun.*, vol. 70, no. Pt 1, pp. 2–20, Jan. 2014.

[28] D. S. Waugh, "Making the most of affinity tags," *Trends Biotechnol.*, vol. 23, no. 6, pp. 316–320, 2005.

[29] D. L. Minor, "The Neurobiologist's Guide to Structural Biology: A Primer on Why Macromolecular Structure Matters and How to Evaluate Structural Data," *Neuron*, vol. 54, no. 4, pp. 511–533, 2007.

[30] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in Escherichia coli: Advances and challenges," *Front. Microbiol.*, vol. 5, no. APR, pp. 1–17, 2014.

[31] W. C. Röntgen, "On a new kind of rays," *Science (80-.).*, vol. 3, no. 59, pp. 227–231, Feb. 1896.

[32] G. Rosenbaum, K. C. Holmes, and J. Witz, "Synchrotron radiation as a source for x-ray diffraction," *Nature*, vol. 230, no. 5294, pp. 434–437, Apr. 1971.

[33] J. C. Phillips, A. Wlodawer, M. M. Yevitz, and K. O. Hodgson, “Applications of synchrotron radiation to protein crystallography: preliminary results.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 73, no. 1, pp. 128–32, Jan. 1976.

[34] A. Förster, S. Brandstetter, and C. Schulze-Briese, “Transforming X-ray detection with hybrid photon counting detectors,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 377, no. 2147, p. 20180241, Jun. 2019.

[35] E. Garman, “Cool data: quantity AND quality.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 55, no. Pt 10, pp. 1641–53, Oct. 1999.

[36] D. W. Rodgers, “Cryocrystallography,” *Structure*, vol. 2, no. 12, pp. 1135–40, Dec. 1994.

[37] A. E. Cohen, P. J. Ellis, M. D. Miller, A. M. Deacon, and R. P. Phizackerley, “An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot,” *J. Appl. Crystallogr.*, vol. 35, no. 6, pp. 720–726, Dec. 2002.

[38] N. Roberts *et al.*, “Rational design of peptide-based HIV proteinase inhibitors,” *Science (80-.).*, vol. 248, no. 4953, pp. 358–361, Apr. 1990.

[39] R. V Agafonov, C. Wilson, and D. Kern, “Evolution and intelligent design in drug development,” *Front. Mol. Biosci.*, vol. 2, p. 27, 2015.

[40] M. Danishuddin and A. U. Khan, “Structure based virtual screening to discover putative drug candidates: Necessary considerations and successful case studies,” *Methods*, vol. 71, pp. 135–145, Jan. 2015.

[41] A. Kumar, A. Ito, M. Takemoto, M. Yoshida, and K. Y. J. Zhang, “Identification of 1,2,5-Oxadiazoles as a New Class of SENP2 Inhibitors Using Structure Based Virtual Screening,” *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 870–880, Mar. 2014.

[42] M. Danishuddin *et al.*, “Structure-based screening of inhibitors against KPC-2: designing potential drug candidates against multidrug-resistant bacteria,” *J. Biomol. Struct. Dyn.*, vol. 32, no. 5, pp. 741–750, May 2014.

[43] L. Xu *et al.*, “Discovery of Novel Inhibitors Targeting the Macrophage Migration Inhibitory Factor via Structure-Based Virtual Screening and Bioassays,” *J. Med. Chem.*, vol. 57, no. 9, pp. 3737–3745, May 2014.

[44] H.-J. Zhong *et al.*, “Structure-based repurposing of FDA-approved drugs as inhibitors of NEDD8-activating enzyme,” *Biochimie*, vol. 102, pp. 211–215, Jul. 2014.

[45] Q. Wang, J. Park, A. K. Devkota, E. J. Cho, K. N. Dalby, and P. Ren, “Identification and Validation of Novel PERK Inhibitors,” *J. Chem. Inf. Model.*, vol. 54, no. 5, pp. 1467–1475, May 2014.

[46] M. A. Ondetti, B. Rubin, and D. W. Cushman, “Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents.,” *Science*, vol. 196, no. 4288, pp. 441–4, Apr. 1977.

[47] D. W. Cushman, H. S. Cheung, E. F. Sabo, and M. A. Ondetti, “Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids,” *Biochemistry*, vol. 16, no. 25, pp. 5484–5491, Dec. 1977.

[48] L. D. Byers and R. Wolfenden, “Binding of the by-product analog benzylsuccinic acid by carboxypeptidase A,” *Biochemistry*, vol. 12, no. 11, pp. 2070–2078, May 1973.

[49] S. H. Ferreira, D. C. Bartelt, and L. J. Greene, “Isolation of bradykinin-potentiating peptides from Bothrops jararaca venom,” *Biochemistry*, vol. 9, no. 13, pp. 2583–2593, Jun. 1970.

[50] T. T. Talele, S. A. Khedkar, and A. C. Rigby, “Successful applications of computer aided drug discovery: moving drugs from concept to the clinic.,” *Curr. Top. Med. Chem.*, vol. 10, no. 1, pp. 127–41, 2010.

[51] A. Bateman, P. Coggill, and R. D. Finn, “DUFs: Families in search of function,” *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 66, no. 10, pp. 1148–1152, 2010.

[52] N. F. Goodacre, D. L. Gerloff, and P. Uetz, “Protein Domains of Unknown Function Are Essential in Bacteria,” *MBio*, vol. 5, no. 1, pp. e00744-13, 2014.

[53] R. Häuser *et al.*, “RsFA (YbeB) proteins are conserved ribosomal silencing factors,” *PLoS Genet.*, vol. 8, no. 7, pp. 1–12, 2012.

[54] V. de Crécy-Lagard, D. Haas, and A. D. Hanson, “Newly-discovered enzymes that function in metabolite damage-control,” *Curr. Opin. Chem. Biol.*, vol. 47, pp. 101–108, 2018.

[55] C. L. Linster, E. Van Schaftingen, and A. D. Hanson, “Metabolite damage and its repair or pre-emption,” *Nat. Chem. Biol.*, vol. 9, no. 2, pp. 72–80, Jan. 2013.

[56] B. P. Bowen and T. R. Northen, “Dealing with the unknown: Metabolomics and metabolite atlases,” *J. Am. Soc. Mass Spectrom.*, vol. 21, no. 9, pp. 1471–1476, 2010.

[57] S. Naz *et al.*, “Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition,” *Anal. Chem.*, vol. 89, no. 15, pp. 7933–7942, 2017.

[58] M. Y. Galperin, O. V. Moroz, K. S. Wilson, and A. G. Murzin, “House cleaning, a part of good housekeeping,” *Molecular Microbiology*, vol. 59, no. 1. pp. 5–19, 2006.

[59] M. L. Reaves, B. D. Young, A. M. Hosios, Y. Xu, and D. Rabinowitz, “Pyrimidine homeostasis is accomplished by directed overflow metabolism,” *Nature*, vol. 500, no. 7461, pp. 237–241, 2015.

[60] E. Van Schaftingen, R. Rzem, A. Marbaix, F. Collard, M. Veiga-Da-Cunha, and C. L. Linster, “Metabolite proofreading, a neglected aspect of intermediary metabolism,” *J. Inherit. Metab. Dis.*, vol. 36, no. 3, pp. 427–434, 2013.

[61] A. Sharma, F. U. Hartl, M. Hayer-Hartl, A. Starling-Windhof, and A. Bracher, “Degradation of potent Rubisco inhibitor by selective sugar phosphatase,” *Nat. Plants*, vol. 1, no. 1, p. 14002, 2015.

[62] L. B. Johnson, D. J. Mulla, U. G. Munderloh, and P. T. Redig, “A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis,” *J. Biol. Chem.*, vol. 12, no. 1, pp. 152–163, 2015.

[63] M. Duran, J. P. Kamerling, H. D. Bakker, A. H. van Gennip, and S. K. Wadman, “L-2-Hydroxyglutaric aciduria: an inborn error of metabolism?,” *J. Inherit. Metab. Dis.*, vol. 3, no. 4, pp. 109–112, 1980.

[64] R. Rzem *et al.*, “A Mouse Model of L-2-Hydroxyglutaric Aciduria, a Disorder of Metabolite

Repair," *PLoS One*, vol. 10, no. 3, p. e0119540, 2015.

[65] B. Tabarki *et al.*, "A gene encoding a putative FAD-dependent L-2-hydroxyglutarate dehydrogenase is mutated in L-2-hydroxyglutaric aciduria," *Proc. Natl. Acad. Sci.*, vol. 101, no. 48, pp. 16849–16854, 2004.

[66] C. S. Henry, A. D. Hanson, S. D. Bruner, J. G. Jeffryes, and J. Sun, "Metabolite damage and repair in metabolic engineering design," *Metab. Eng.*, vol. 44, no. September, pp. 150–159, 2017.

[67] A. D. Hanson, C. S. Henry, O. Fiehn, and V. de Crécy-Lagard, "Metabolite Damage and Metabolite Damage Control in Plants," *Annu. Rev. Plant Biol.*, vol. 67, no. 1, pp. 131–152, 2015.

[68] L. Huang *et al.*, "A family of metal-dependent phosphatases implicated in metabolite damage-control," *Nat. Chem. Biol.*, vol. 12, no. 8, pp. 621–627, 2016.

[69] G. B. Tilton, W. J. Wedemeyer, J. Browse, and J. Ohlrogge, "Plant coenzyme A biosynthesis: Characterization of two pantothenate kinases from Arabidopsis," *Plant Mol. Biol.*, vol. 61, no. 4–5, pp. 629–642, 2006.

[70] M. Sun *et al.*, "A novel mutation of PANK4 causes autosomal dominant congenital posterior cataract," *Hum. Mutat.*, vol. 40, no. 4, pp. 380–391, 2019.

[71] A. K. Dunbier *et al.*, "ESR1 is co-expressed with closely adjacent uncharacterised genes spanning a breast cancer susceptibility locus at 6q25.1," *PLoS Genet.*, vol. 7, no. 4, 2011.

[72] M. Yamamoto-Ibusuki *et al.*, "C6ORF97-ESR1 breast cancer susceptibility locus: Influence on progression and survival in breast cancer patients," *Eur. J. Hum. Genet.*, vol. 23, no. 7, pp. 949–956, 2015.

[73] D. Alvaro, M. Lisby, and R. Rothstein, "Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination," *PLoS Genet.*, vol. 3, no. 12, pp. 2439–2449, 2007.

[74] R. A. Dwek, "Glycobiology: Toward Understanding the Function of Sugars," *Chem. Rev.*, vol. 96, no. 2, pp. 683–720, 1996.

[75] S. P. Wolff and R. T. Dean, "Glucose autoxidation and protein modification The potential role of 'autoxidative glycosylation' in diabetes," 1987.

[76] J. J. Harding, "Nonenzymatic Covalent Posttranslational Modification of Proteins In Vivo," *Adv. Protein Chem.*, vol. 37, pp. 247–334, Jan. 1985.

[77] M. X. Fu, K. J. Wells-Knecht, J. A. Blackledge, T. J. Lyons, S. R. Thorpe, and J. W. Baynes, "Glycation, glycoxidation, and cross-linking of collagen by glucose. Kinetics, mechanisms, and inhibition of late stages of the Maillard reaction.," *Diabetes*, vol. 43, no. 5, pp. 676–83, May 1994.

[78] R. Bucala and A. Cerami, "Advanced Glycosylation: Chemistry, Biology, and Implications for Diabetes and Aging," *Adv. Pharmacol.*, vol. 23, no. C, pp. 1–34, 1992.

[79] B. Levi and M. J. Werman, "Fructose and related phosphate derivatives impose DNA damge and apoptosis in L5178Y mouse lymphoma cells," *J. Nutr. Biochem.*, vol. 14, no. 1, pp. 49–60, 2003.

[80] H. Vlassara, "Recent progress on the biologic and clinical significance of advanced glycosylation end products.," *J. Lab. Clin. Med.*, vol. 124, no. 1, pp. 19–30, Jul. 1994.

[81] D. R. Sell and V. M. Monnir, “Structure elucidation of a senescence cross-link from human extracellular matrix. Implication of pentoses in the aging process,” *J. Biol. Chem.*, vol. 264, no. 36, pp. 21597–21602, Dec. 1989.

[82] P. Gkogkolou and M. Böhm, “Advanced glycation end products,” *Dermatoendocrinol.*, vol. 4, no. 3, pp. 259–270, Jul. 2012.

[83] V. M. Monnier and a Cerami, “Nonenzymatic browning in vivo: possible process for aging of long-lived proteins.,” *Science*, vol. 211, no. 4481, pp. 491–493, Jan. 1981.

[84] E. J. Masoro, “Biology of aging: facts, thoughts, and experimental approaches,” *Lab Invest*, vol. 65, no. 5, pp. 500–510, Nov. 1991.

[85] H. Vlassara, R. Bucala, and L. Striker, “Pathogenic effects of advanced glycosylation: biochemical, biologic, and clinical implications for diabetes and aging.,” *Lab. Invest.*, vol. 70, no. 2, pp. 138–151, Feb. 1994.

[86] M. A. M. van Boekel, “The role of glycation in aging and diabetes mellitus,” *Mol. Biol. Rep.*, vol. 15, no. 2, pp. 57–64, 1991.

[87] G. Suarez, R. Rajaram, A. L. Oronsky, and M. A. Gawinowicz, “Nonenzymatic glycation of bovine serum albumin by fructose (fructation). Comparison with the Maillard reaction initiated by glucose,” *J. Biol. Chem.*, vol. 264, no. 7, pp. 3674–3679, 1989.

[88] M. Brownlee, “Nonenzymatic Glycosylation and the Pathogenesis of Diabetic Complications,” *Ann. Intern. Med.*, vol. 101, no. 4, p. 527, Oct. 1984.

[89] M. A. Smith *et al.*, “Advanced Maillard reaction end products are associated with Alzheimer disease pathology,” *Neurobiology*, vol. 91, no. 12, pp. 5710–5714, Jun. 1994.

[90] M. P. Vitek *et al.*, “Advanced glycation end products contribute to amyloidosis in Alzheimer disease.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 11, pp. 4766–70, May 1994.

[91] J. J. P. Perry, G. D. Ballard, A. E. Albert, L. E. Dobrolecki, L. H. Malkas, and D. J. Hoelz, “Human C6orf211 Encodes Armt1, a Protein Carboxyl Methyltransferase that Targets Perry, J. J. P., Ballard, G. D., Albert, A. E., Dobrolecki, L. E., Malkas, L. H., & Hoelz, D. J. (2015). Human C6orf211 Encodes Armt1, a Protein Carboxyl Methyltransferase tha,” *Cell Rep.*, vol. 10, no. 8, pp. 1288–1296, 2015.

[92] L. Schrödinger, “The PyMOL Molecular Graphics System, Version 2.0,” Nov. 2015.

[93] P. D. Zwart, P. H., Grosse-Kunstleve, R. W., Adams, “Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation,” *CCP4 Newsl.*, vol. 43, pp. 27–35, 2005.

[94] J. L. Martin, F. M. Mcmillan, and T. S-adenosylmethionine-dependent, “SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold,” *Curr. Opin. Struct. Biol.*, vol. 13, no. 1, p. 140, 2003.

[95] H. L. Schubert, R. M. Blumenthal, and X. Cheng, “Many paths to methyltransfer: A chronicle of convergence,” *Trends Biochem. Sci.*, vol. 28, no. 6, pp. 329–335, 2003.

[96] L. Holm and L. M. Laakso, “Dali server update,” *Nucleic Acids Res.*, vol. 44, no. W1, pp. W351–W355, 2016.

[97] P. Tompa, N. E. Davey, T. J. Gibson, and M. M. Babu, “A Million Peptide Motifs for the Molecular Biologist,” *Mol. Cell*, vol. 55, no. 2, pp. 161–169, Jul. 2014.

[98] R. Aasland *et al.*, “Normalization of nomenclature for peptide motifs as ligands of modular protein domains,” *FEBS Lett.*, vol. 513, no. 1, pp. 141–144, 2002.

[99] E. D. Murray and S. Clarke, “Synthetic Peptide Substrates for the Erythrocyte Protein Carboxyl Methyltransferase,” *J. Biol. Chem.*, vol. 259, no. 17, pp. 10722–10732, 1984.

[100] G. Maga and U. Hübscher, “Proliferating cell nuclear antigen (PCNA): a dancer with many partners,” *J. Cell Sci.*, vol. 116, no. 15, pp. 3051–3060, Aug. 2003.

[101] D. J. Hoelz *et al.*, “The discovery of labile methyl esters on proliferating cell nuclear antigen by MS/MS,” *Proteomics*, vol. 6, no. 17, pp. 4808–4816, Sep. 2006.

[102] L. H. Malkas *et al.*, “A cancer-associated PCNA expressed in breast cancer has implications as a potential biomarker.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 51, pp. 19472–7, Dec. 2006.

[103] R. D. Feinman and E. J. Fine, “Fructose in perspective,” *Nutr. Metab.*, vol. 10, no. 1, 2013.

[104] A. Bar-Even *et al.*, “The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters,” *Biochemistry*, vol. 50, no. 21, pp. 4402–4410, 2011.

[105] B. Berghuis *et al.*, “The Genotype-Tissue Expression (GTEx) project,” *Nat. Genet.*, vol. 45, no. 6, pp. 580–585, 2013.

[106] T. Schmidt *et al.*, “ProteomicsDB,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1271–D1281, Jan. 2018.

[107] EMBL, “Protein Expression - E.coli - Expression - Seleno-methionine labelling of proteins in E. coli - EMBL.” [Online]. Available: https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/seleno/. [Accessed: 22-Mar-2019].

[108] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, “iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 271–81, Apr. 2011.

[109] M. D. Winn *et al.*, “Overview of the CCP4 suite and current developments,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 235–242, Apr. 2011.

[110] P. D. Adams *et al.*, “PHENIX: a comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 213–221, Feb. 2010.

[111] T. C. Terwilliger *et al.*, “Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 65, no. 6, pp. 582–601, Jun. 2009.

[112] Y. Zhang, “I-TASSER server for protein 3D structure prediction,” *BMC Bioinformatics*, vol. 9, no. 1, p. 40, Jan. 2008.

[113] A. Roy, A. Kucukural, and Y. Zhang, “I-TASSER: a unified platform for automated protein structure and function prediction,” *Nat. Protoc.*, vol. 5, no. 4, pp. 725–738, Apr. 2010.

[114] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, “The I-TASSER Suite: protein structure and function prediction,” *Nat. Methods*, vol. 12, no. 1, pp. 7–8, 2014.

[115] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.

[116] P. V. Afonine *et al.*, “Towards automated crystallographic structure refinement with phenix.refine,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 68, no. 4, pp. 352–367, Mar. 2012.

[117] P. Emsley and K. Cowtan, “Coot: Model-building tools for molecular graphics,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2126–2132, Dec. 2004.

[118] J. Schlegl *et al.*, “ProteomicsDB,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1271–D1281, 2017.

[119] M. Wilhelm *et al.*, “Mass-spectrometry-based draft of the human proteome,” *Nature*, vol. 509, no. 7502, pp. 582–587, 2014.

[120] B. Lipinska, O. Fayet, L. Baird, and C. Georgopoulos, “Identification, characterization, and mapping of the Escherichia coli htrA gene, whose product is essential for bacterial growth only at elevated temperatures.,” *J. Bacteriol.*, vol. 171, no. 3, pp. 1574–84, Mar. 1989.

[121] D. Zurawa-Janicka *et al.*, “Structural insights into the activation mechanisms of human HtrA serine proteases,” *Arch. Biochem. Biophys.*, vol. 621, pp. 6–23, May 2017.

[122] M. L. Diaz-Torres and R. R. B. Russell, “HtrA protease and processing of extracellular proteins of Streptococcus mutans,” *FEMS Microbiol. Lett.*, vol. 204, no. 1, pp. 23–28, Oct. 2001.

[123] T. Clausen, C. Southan, and M. Ehrmann, “The HtrA Family of Proteases: Implications for Protein Composition and Cell Fate,” *Mol. Cell*, vol. 10, no. 3, pp. 443–455, Sep. 2002.

[124] P. Glaza *et al.*, “Structural and Functional Analysis of Human HtrA3 Protease and Its Subdomains,” *PLoS One*, vol. 10, no. 6, p. e0131142, Jun. 2015.

[125] M. B. Kennedy, “Origin of PDZ (DHR, GLGF) domains.,” *Trends Biochem. Sci.*, vol. 20, no. 9, p. 350, Sep. 1995.

[126] H.-G. Goo, H. Rhim, and S. Kang, “HtrA2/Omi influences the stability of LON protease 1 and prohibitin, proteins involved in mitochondrial homeostasis,” *Exp. Cell Res.*, vol. 328, no. 2, pp. 456–465, Nov. 2014.

[127] Y. Suzuki, Y. Imai, H. Nakayama, K. Takahashi, K. Takio, and R. Takahashi, “A Serine Protease, HtrA2, Is Released from the Mitochondria and Interacts with XIAP, Inducing Cell Death,” *Mol. Cell*, vol. 8, no. 3, pp. 613–621, Sep. 2001.

[128] K. M. Strauss *et al.*, “Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson’s disease,” *Hum. Mol. Genet.*, vol. 14, no. 15, pp. 2099–2111, Aug. 2005.

[129] L. M. Martins *et al.*, “Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice.,” *Mol. Cell. Biol.*, vol. 24, no. 22, pp. 9848–62, Nov. 2004.

[130] J. M. Jones *et al.*, “Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice,” *Nature*, vol. 425, no. 6959, pp. 721–727, Oct. 2003.

[131] C. Ciferri *et al.*, “The trimeric serine protease HtrA1 forms a cage-like inhibition complex with an anti-HtrA1 antibody,” *Biochem. J.*, vol. 472, no. 2, pp. 169–181, 2015.

[132] L. Truebestein *et al.*, “Substrate-induced remodeling of the active site regulates human HTRA1 activity,” *Nat. Struct. Mol. Biol.*, vol. 18, no. 3, pp. 386–388, Mar. 2011.

[133] D. S. Friedman *et al.*, “Prevalence of Age-Related Macular Degeneration in the United States,” *Arch. Ophthalmol.*, vol. 122, no. 4, p. 564, Apr. 2004.

[134] I. Bhutto and G. Lutty, “Understanding age-related macular degeneration (AMD): relationships between the photoreceptor/retinal pigment epithelium/Bruch’s membrane/choriocapillaris complex.,” *Mol. Aspects Med.*, vol. 33, no. 4, pp. 295–317, Aug. 2012.

[135] J. L. Haines *et al.*, “Functional candidate genes in age-related macular degeneration: Significant association with VEGF, VLDLR, and LRP6,” *Investig. Ophthalmol. Vis. Sci.*, vol. 47, no. 1, pp. 329–335, Jan. 2006.

[136] A. J. Churchill *et al.*, “VEGF polymorphisms are associated with neovascular age-related macular degeneration,” *Hum. Mol. Genet.*, vol. 15, no. 19, pp. 2955–2961, Aug. 2006.

[137] S. Grisanti and O. Tatar, “The role of vascular endothelial growth factor and other endogenous interplayers in age-related macular degeneration,” *Progress in Retinal and Eye Research*, vol. 27, no. 4. pp. 372–390, Jul-2008.

[138] C. Kent, “Treating Wet AMD with Anti-VEGF Drugs,” *Review of Ophthalmology*, 2016. [Online]. Available: https://www.reviewofophthalmology.com/article/treating-wet-amd-with-antivegf-drugs. [Accessed: 27-Apr-2019].

[139] J. L. Kovach, S. G. Schwartz, H. W. Flynn, I. U. Scott, and I. U. Scott, “Anti-VEGF Treatment Strategies for Wet AMD.,” *J. Ophthalmol.*, vol. 2012, p. 786870, 2012.

[140] J.-H. Chang, N. K. Garg, E. Lunde, K.-Y. Han, S. Jain, and D. T. Azar, “Corneal neovascularization: an anti-VEGF therapy review.,” *Surv. Ophthalmol.*, vol. 57, no. 5, pp. 415–29, Sep. 2012.

[141] G. Donati, “Emerging Therapies for Neovascular Age-Related Macular Degeneration: State of the Art,” *Ophthalmologica*, vol. 221, no. 6, pp. 366–377, 2007.

[142] S. Lien and H. B. Lowman, “Therapeutic Anti-VEGF Antibodies,” in *Handbook of experimental pharmacology*, no. 181, 2008, pp. 131–150.

[143] N. M. Bressler *et al.*, “Estimated Cases of Legal Blindness and Visual Impairment Avoided Using Ranibizumab for Choroidal Neovascularization,” *Arch. Ophthalmol.*, vol. 129, no. 6, pp. 709–717, Jun. 2011.

[144] K. G. Falavarjani and Q. D. Nguyen, “Adverse events and complications associated with intravitreal injection of anti-VEGF agents: a review of literature.,” *Eye (Lond).*, vol. 27, no. 7, pp. 787–94, Jul. 2013.

[145] Z. Yang *et al.*, “A Variant of the HTRA1 Gene Increases Susceptibility to Age-Related Macular Degeneration,” *Science (80-.).*, vol. 314, no. 5801, pp. 992–993, Nov. 2006.

[146] C.-C. Chan *et al.*, “Human HtrA1 in the archived eyes with age-related macular degeneration.,”

Trans. Am. Ophthalmol. Soc., vol. 105, pp. 92–7; discussion 97-8, 2007.

[147] A. Jones *et al.*, "Increased expression of multifunctional serine protease, HTRA1, in retinal pigment epithelium induces polypoidal choroidal vasculopathy in mice," *Proc. Natl. Acad. Sci.*, vol. 108, no. 35, pp. 14578–14583, Aug. 2011.

[148] L. Zhang *et al.*, "High temperature requirement factor A1 (HTRA1) gene regulates angiogenesis through transforming growth factor-β family member growth differentiation factor 6," *J. Biol. Chem.*, vol. 287, no. 2, pp. 1520–1526, Jan. 2012.

[149] K. Hara *et al.*, "Association of HTRA1 Mutations and Familial Ischemic Cerebral Small-Vessel Disease," *N. Engl. J. Med.*, vol. 360, no. 17, pp. 1729–1739, Apr. 2009.

[150] R. Klose *et al.*, "Inactivation of the serine protease HTRA1 inhibits tumor growth by deregulating angiogenesis," *Oncogene*, vol. 37, no. 31, pp. 4260–4272, Aug. 2018.

[151] Y. Hou *et al.*, "The Inhibitory Effect of IFN- on Protease HTRA1 Expression in Rheumatoid Arthritis," *J. Immunol.*, vol. 193, no. 1, pp. 130–138, 2014.

[152] I. Wallach, M. Dzamba, and A. Heifets, "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery," Oct. 2015.

[153] C. Eigenbrot *et al.*, "Structural and functional analysis of HtrA1 and its subdomains," *Structure*, vol. 20, no. 6, pp. 1040–1050, 2012.

[154] S. Yasothornsrikul and V. Y. H. Hook, "Detection of Proteolytic Activity by Fluorescent Zymogram In-Gel Assays," *Biotechniques*, vol. 28, no. 6, pp. 1166–1173, Jun. 2000.

[155] C. T. Rueden *et al.*, "ImageJ2: ImageJ for the next generation of scientific image data," *BMC Bioinformatics*, vol. 18, no. 1, p. 529, Dec. 2017.

[156] A. B. Inc, "Quest Graph™ Four Parameter Logistic (4PL) Curve Calculator," 2019. .

[157] J. R. Luft and G. T. DeTitta, "A method to produce microseed stock for use in the crystallization of biological macromolecules," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 55, no. 5, pp. 988–993, May 1999.

[158] R. Nahta, M.-C. Hung, and F. J. Esteva, "The HER-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells.," *Cancer Res.*, vol. 64, no. 7, pp. 2343–6, Apr. 2004.

[159] E. L. Kwak *et al.*, "Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer," *N. Engl. J. Med.*, vol. 363, no. 18, pp. 1693–1703, Oct. 2010.

[160] E. V Prochownik and P. K. Vogt, "Therapeutic Targeting of Myc," *Genes Cancer*, vol. 1, no. 6, pp. 650–659, Jun. 2010.

[161] A. G. Stephen, D. Esposito, R. G. Bagni, and F. McCormick, "Dragging ras back in the ring," *Cancer Cell*, vol. 25, no. 3, pp. 272–281, Mar. 2014.

[162] L. H. Hartwell, P. Szankasi, C. J. Roberts, A. W. Murray, and S. H. Friend, "Integrating genetic approaches into the discovery of anticancer drugs.," *Science*, vol. 278, no. 5340, pp. 1064–8, Nov. 1997.

[163] J. Luo, N. L. Solimini, and S. J. Elledge, “Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction,” *Cell*, vol. 136, no. 5, pp. 823–837, Mar. 2009.

[164] M. P. Licciardello and S. Kubicek, “Pharmacological treats for SUMO addicts,” *Pharmacol. Res.*, vol. 107, pp. 390–397, 2016.

[165] B. Yu *et al.*, “Oncogenesis driven by the Ras/Raf pathway requires the SUMO E2 ligase Ubc9.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 14, pp. E1724-33, 2015.

[166] J. Barretina *et al.*, “The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity,” *Nature*, vol. 483, no. 7391, pp. 603–607, Mar. 2012.

[167] A. V. Ivanov *et al.*, “PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing,” *Mol. Cell*, vol. 28, no. 5, pp. 823–837, Dec. 2007.

[168] Y.-K. Lee, S. N. Thomas, A. J. Yang, and D. K. Ann, “Doxorubicin Down-regulates Krüppel-associated Box Domain-associated Protein 1 Sumoylation That Relieves Its Transcription Repression on p21 $^{WAF1/CIP1}$ in Breast Cancer MCF-7 Cells,” *J. Biol. Chem.*, vol. 282, no. 3, pp. 1595–1606, Jan. 2007.

[169] J. D. Kessler *et al.*, “A SUMOylation-Dependent Transcriptional Subprogram Is Required for Myc-Driven Tumorigenesis,” *Science (80-.).*, vol. 335, no. 6066, pp. 348–353, Jan. 2012.

[170] A. Hoellein *et al.*, “Myc-induced SUMOylation is a therapeutic vulnerability for B-cell lymphoma,” *Blood*, vol. 124, no. 13, pp. 2081–2090, Sep. 2014.

[171] M. P. Licciardello *et al.*, “NOTCH1 activation in breast cancer confers sensitivity to inhibition of SUMOylation,” *Oncogene*, vol. 34, no. 29, pp. 3780–3790, Jul. 2015.

[172] E. R. Andersson and U. Lendahl, “Therapeutic modulation of Notch signalling — are we there yet?,” *Nat. Rev. Drug Discov.*, vol. 13, no. 5, pp. 357–378, May 2014.

[173] B. De Strooper and L. Chávez Gutiérrez, “Learning by Failing: Ideas and Concepts to Tackle γ-Secretases in Alzheimer’s Disease and Beyond,” *Annu. Rev. Pharmacol. Toxicol.*, vol. 55, no. 1, pp. 419–437, Jan. 2015.

[174] E. H. Estey, “Acute myeloid leukemia: 2012 update on diagnosis, risk stratification, and management,” *Am. J. Hematol.*, vol. 87, no. 1, pp. 89–99, Jan. 2012.

[175] G. Bossis *et al.*, “The ROS/SUMO Axis Contributes to the Response of Acute Myeloid Leukemia Cells to Chemotherapeutic Drugs,” *Cell Rep.*, vol. 7, no. 6, pp. 1815–1823, 2014.

[176] R. T. Hay, “SUMO: A history of modification,” *Molecular Cell*, vol. 18, no. 1. pp. 1–12, 01-Apr-2005.

[177] S. P. Jackson and D. Durocher, “Regulation of DNA Damage Responses by Ubiquitin and SUMO,” *Mol. Cell*, vol. 49, no. 5, pp. 795–807, Mar. 2013.

[178] X.-D. Zhang, J. Goeres, H. Zhang, T. J. Yen, A. C. G. Porter, and M. J. Matunis, “SUMO-2/3 Modification and Binding Regulate the Association of CENP-E with Kinetochores and Progression through Mitosis,” *Mol. Cell*, vol. 29, no. 6, pp. 729–741, Mar. 2008.

[179] U. Rothenbusch, M. Sawatzki, Y. Chang, S. Caesar, and G. Schlenstedt, “Sumoylation regulates Kap114-mediated nuclear transport,” *EMBO J.*, vol. 31, no. 11, pp. 2461–2472, May 2012.

[180] T. Makhnevych *et al.*, “Global Map of SUMO Function Revealed by Protein-Protein Interaction and Genetic Networks,” *Mol. Cell*, vol. 33, no. 1, pp. 124–135, Jan. 2009.

[181] K. Bettermann, M. Benesch, S. Weis, and J. Haybaeck, “SUMOylation in carcinogenesis,” *Cancer Lett.*, vol. 316, no. 2, pp. 113–125, 2012.

[182] R. Geiss-friedlander and F. Melchior, “Concepts in sumoylation: a decade on,” *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 12, pp. 947–956, 2007.

[183] M. M. Rytinki, S. Kaikkonen, P. Pehkonen, T. Jääskeläinen, and J. J. Palvimo, “PIAS proteins: pleiotropic interactors associated with SUMO,” *Cell. Mol. Life Sci.*, vol. 66, no. 18, pp. 3029–3041, Sep. 2009.

[184] V. Bernier-Villamor, D. A. Sampson, M. J. Matunis, and C. D. Lima, “Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1.,” *Cell*, vol. 108, no. 3, pp. 345–56, Feb. 2002.

[185] V. Hietakangas *et al.*, “PDSM, a motif for phosphorylation-dependent SUMO modification,” *Proc. Natl. Acad. Sci.*, vol. 103, no. 1, pp. 45–50, Jan. 2006.

[186] S.-H. Yang, A. Galanis, J. Witty, and A. D. Sharrocks, “An extended consensus motif enhances the specificity of substrate modification by SUMO,” *EMBO J.*, vol. 25, no. 21, pp. 5083–5093, Nov. 2006.

[187] I. Fukuda *et al.*, “Ginkgolic Acid Inhibits Protein SUMOylation by Blocking Formation of the E1-SUMO Intermediate,” *Chem. Biol.*, vol. 16, no. 2, pp. 133–140, 2009.

[188] W. K. Weigel, T. N. Dennis, A. S. Kang, J. J. P. Perry, and D. B. C. Martin, “A Heck-Based Strategy To Generate Anacardic Acids and Related Phenolic Lipids for Isoform-Specific Bioactivity Profiling,” *Org. Lett.*, vol. 20, no. 19, pp. 6234–6238, Oct. 2018.

[189] S. C. Gill *et al.*, “Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo,” *J. Phys. Chem. B*, vol. 122, no. 21, pp. 5579–5598, May 2018.

[190] M. Congreve, R. Carr, C. Murray, and H. Jhoti, “A ‘rule of three’ for fragment-based lead discovery?,” *Drug Discov. Today*, vol. 8, no. 19, pp. 876–7, Oct. 2003.

[191] M. M. Hann, A. R. Leach, and G. Harper, “Molecular complexity and its impact on the probability of finding leads for drug discovery.,” *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 3, pp. 856–64.

[192] C. Abad-Zapatero, “Ligand efficiency indices for effective drug discovery,” *Expert Opin. Drug Discov.*, vol. 2, no. 4, pp. 469–488, Apr. 2007.

[193] A. L. Hopkins, C. R. Groom, and A. Alex, “Ligand efficiency: a useful metric for lead selection,” *Drug Discov. Today*, vol. 9, no. 10, pp. 430–431, May 2004.

[194] A. Ciulli, “Biophysical screening for the discovery of small-molecule ligands,” *Methods Mol. Biol.*, vol. 1008, pp. 357–388, 2013.

[195] G. WEBER and D. J. LAURENCE, “Fluorescent indicators of adsorption in aqueous solution and on the solid phase.,” *Biochem. J.*, vol. 56, no. 325th Meeting, p. xxxi, Jan. 1954.

[196] J. K. Kranz and C. Schalk-Hihi, “Protein Thermal Shifts to Identify Low Molecular Weight Fragments,” *Methods Enzymol.*, vol. 493, pp. 277–298, 2011.

[197] M. D. Cummings, M. A. Farnum, and M. I. Nelen, “Universal Screening Methods and Applications of ThermoFluor®,” *J. Biomol. Screen.*, vol. 11, no. 7, pp. 854–863, Oct. 2006.

[198] F. J. Sorrell, G. K. Greenwood, K. Birchall, and B. Chen, “Development of a differential scanning fluorimetry based high throughput screening assay for the discovery of affinity binders against an anthrax protein,” *J. Pharm. Biomed. Anal.*, vol. 52, no. 5, pp. 802–808, Sep. 2010.

[199] C. N. Pace and T. McGrath, “Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation.,” *J. Biol. Chem.*, vol. 255, no. 9, pp. 3862–5, May 1980.

[200] M. N. Schulz, J. Landström, and R. E. Hubbard, “MTSA—A Matlab program to fit thermal shift data,” *Anal. Biochem.*, vol. 433, no. 1, pp. 43–47, Feb. 2013.

[201] L. Jiang *et al.*, “Specific substrate recognition and thioester intermediate determinations in ubiquitin and SUMO conjugation cascades revealed by a high-sensitive FRET assay,” *Mol. Biosyst.*, vol. 10, no. 4, p. 778, Mar. 2014.

[202] A. D. Capili and C. D. Lima, “Taking it Step by Step: Mechanistic insights from structural studies of Ubiquitin/Ubiquitin-like protein modification pathways,” *Curr. Opin. Struct. Biol.*, vol. 17, no. 6, p. 726, Dec. 2007.

[203] I. G. Madu *et al.*, “A Novel Class of HIV-1 Antiviral Agents Targeting HIV via a SUMOylation-Dependent Mechanism.,” *Sci. Rep.*, vol. 5, no. 3, p. 17808, 2015.

[204] V. Dorval and P. E. Fraser, “SUMO on the road to neurodegeneration,” *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1773, no. 6. pp. 694–706, 2007.

[205] A. Omanakuttan *et al.*, “Anacardic Acid Inhibits the Catalytic Activity of Matrix Metalloproteinase-2 and Matrix Metalloproteinase-9,” *Mol. Pharmacol.*, vol. 82, no. 4, pp. 614–622, Oct. 2012.

[206] A. Hollands *et al.*, “Natural Product Anacardic Acid from Cashew Nut Shells Stimulates Neutrophil Extracellular Trap Production and Bactericidal Activity.,” *J. Biol. Chem.*, vol. 291, no. 27, pp. 13964–73, Jul. 2016.

[207] W. K. Weigel, T. N. Dennis, A. S. Kang, J. J. P. Perry, and D. B. C. Martin, “A Heck-Based Strategy to Generate Anacardic Acids and Related Phenolic Lipids for Isoform-Specific Bioactivity Profiling,” *Org. Lett.*, vol. 20, no. 19, pp. 6234–6238, Oct. 2018.

[208] H. Krische *et al.*, *Fragment-Based Drug Discovery and X-Ray Crystallography*, vol. 317. 2012.

[209] Y. Song, V. Madahar, and J. Liao, “Development of FRET Assay into Quantitative and High-throughput Screening Technology Platforms for Protein–Protein Interactions,” *Ann. Biomed. Eng.*, vol. 39, no. 4, pp. 1224–1234, Apr. 2011.