Serial Crystallographic Studies for Therapeutic Drug Advancement

by

Rebecca Jernigan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2022 by the
Graduate Supervisory Committee:

Petra Fromme, Co-Chair
Debra Hansen, Co-Chair
Po-Lin Chiu
Brenda Hogue

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

Macromolecular structural biology advances the understanding of protein function through the structure-function relationship for applications to scientific challenges like energy and medicine. The proteins described in these studies have applications to medicine as targets for therapeutic drug design. By understanding the mechanisms and dynamics of these proteins, therapeutics can be designed and optimized based on their unique structural characteristics. This can create new, focused therapeutics for the treatment of diseases with increased specificity — which translates to greater efficacy and fewer off-target effects. Many of the structures generated for this purpose are "static" in nature, meaning the protein is observed like a still-frame photograph; however, the use of time-resolved techniques is allowing for greater understanding of the dynamic and flexible nature of proteins. This work advances understanding the dynamics of the medically relevant proteins NendoU and Taspase1 using serial crystallography to establish conditions for time-resolved, mix-and-inject crystallographic studies.

DEDICATION

This dissertation is dedicated to my family, who have supported me always and helped

make all of this possible. I would like to honor my mother, Susan Jernigan, and my

father, Terry Jernigan, for their support. I would like to honor my girls, Claire, Kya, and

Rose who bring me great joy. I would like to also honor Quin-Anne Hinrichs whose

support and love has been invaluable.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Macromolecular crystallography (MX) examines fundamental biological

principles by characterizing the biomechanics of proteins to elucidate their roles within

larger pathways. This understanding applies towards fields such as medical research and

new energy technology development. Altering a pathway in disease through structure-

based drug discovery and design has been applied to cancer, infectious disease, addiction,

pain sensing, and more (Congreve et al., 2011; Engelman & Cherepanov, 2012; Neidle,

2011; H. Yang & Rao, 2021). Applications into energy technology development use

insights from mechanisms engineered by nature, such as photosynthesis (Henry N

Chapman et al., 2011; Kupitz, Basu, et al., 2014). The findings can then be applied into

today's emerging technology, such as working towards solving the climate crisis (Lewis

et al., 2022). Structural biology plays an important and growing role in biochemistry to

understand the fundamental mechanics of how proteins work to effect change in larger

biological systems – and use these findings to correct disfunctions and apply the elegance

of nature to our growing technologies.

To accomplish these goals, target proteins can be expressed in systems such as

bacteria (commonly *Escherichia coli*), insect cells, mammalian cells, and cell-free

models (McKenzie & Abbott, 2018; Rosano et al., 2019; Silverman et al., 2020). The

protein is commonly isolated through a combination of techniques such as: detergent

extraction, centrifugation, affinity and ion exchange chromatography, and size exclusion

chromatography (Youngchang Kim et al., 2011). Purity and homogeneity can be

examined through methods such as polyacrylamide gel electrophoresis (PAGE) and

dynamic light scattering (DLS). Protein characterization can be performed in solution with techniques like circular dichroism (CD) and small-angle X-ray scattering (SAXS) (Brosey & Tainer, 2019; Miles & Wallace, 2016). To determine high resolution structures, techniques such as cryogenic electron microscopy (CryoEM), nuclear magnetic resonance (NMR) and crystallography, or X-ray diffraction, can be employed (Palmer III, 2001; Saibil, 2022; Su et al., 2015).

X-ray crystallography was the first technique in macromolecular structural biology, that allowed for atomic resolution structure of biomolecules to be determined (Hodgkin et al., 1956). X-ray crystallography includes the crystallization of a target protein, subsequent diffraction by an X-ray light source, and solution of the phase problem and structure determination where the two-dimensional diffraction patterns are used to finally determine a three-dimensional structure. The technique can be challenging, but highly rewarding with near-atomic to atomic resolution structures. Advances in crystallography have led to exponential growth over the past century in understanding the structure-function relationship of biomolecules at the molecular level.

1.1 Fundamentals of Crystallography

The first critical development needed to bring X-ray crystallography to fruition was the advancement of X-ray light sources. X-rays were first discovered by Wilhelm Conrad Röentgen in 1895 (Röntgen, 1895). It wasn't until 1912 that this discovery was applied to crystals. Using the shorter wavelength of X-rays, crystals were hypothesized to interfere and diffract the X-ray beam by Max von Laue (Friedrich et al., 1913). Laue demonstrated the wave nature of X-rays with the diffraction of copper sulfate crystals and received the 1914 Nobel Prize in Physics (Eckert, 2012). William L. and William H. Bragg used this

information to analyze crystal structures and by calculating the positions of atoms within the crystal lattice, completing the foundation for crystallography(W. H. Bragg & Bragg, 1913; W. L. Bragg, 1913).

The structure solutions of biomolecules such as penicillin (1945), insulin (1969), and vitamin B_{12} (1955) by Dorothy Hodgkin paved the way to larger macromolecules (Abraham, 1987; Adams et al., 1969; Hodgkin et al., 1956). The structure of vitamin B_{12} made Hodgkin the third woman to receive a Nobel Prize in Chemistry. The first macromolecular structures in crystallography, myoglobin and hemoglobin, were solved in the late 1950s by John Kendrew and Max Perutz, who shared the 1962 Nobel Prize in Chemistry for their findings (Kendrew & Parrish, 1957; Perutz et al., 1960). Since then, X-ray diffraction is one of the most prolific techniques used in structural biology with the majority of deposited structures in the Protein Data Bank (PDB).

Macromolecular crystallography advances the understanding of protein science through structural discovery. While there have been significant challenges, numerous techniques have been developed that address many of these aspects. Crystallization of the protein can often be a limiting step in the crystallography pipeline as crystals of sufficient size and quality can be difficult to obtain. Vapor diffusion is one of the most common methods for growing large crystals, where a drop of concentrated protein is mixed with a precipitating agent (various salts, buffers, and additives) and placed into a sealed compartment. As water evaporates from the drop, the concentration of the components is increased until an equilibrium is reached. This moves the solution through the phase diagram and, if successful, nucleation occurs in the nucleation zone and the protein crystals grow in the metastable zone as protein is added to the crystal (**Figure 1.1**)
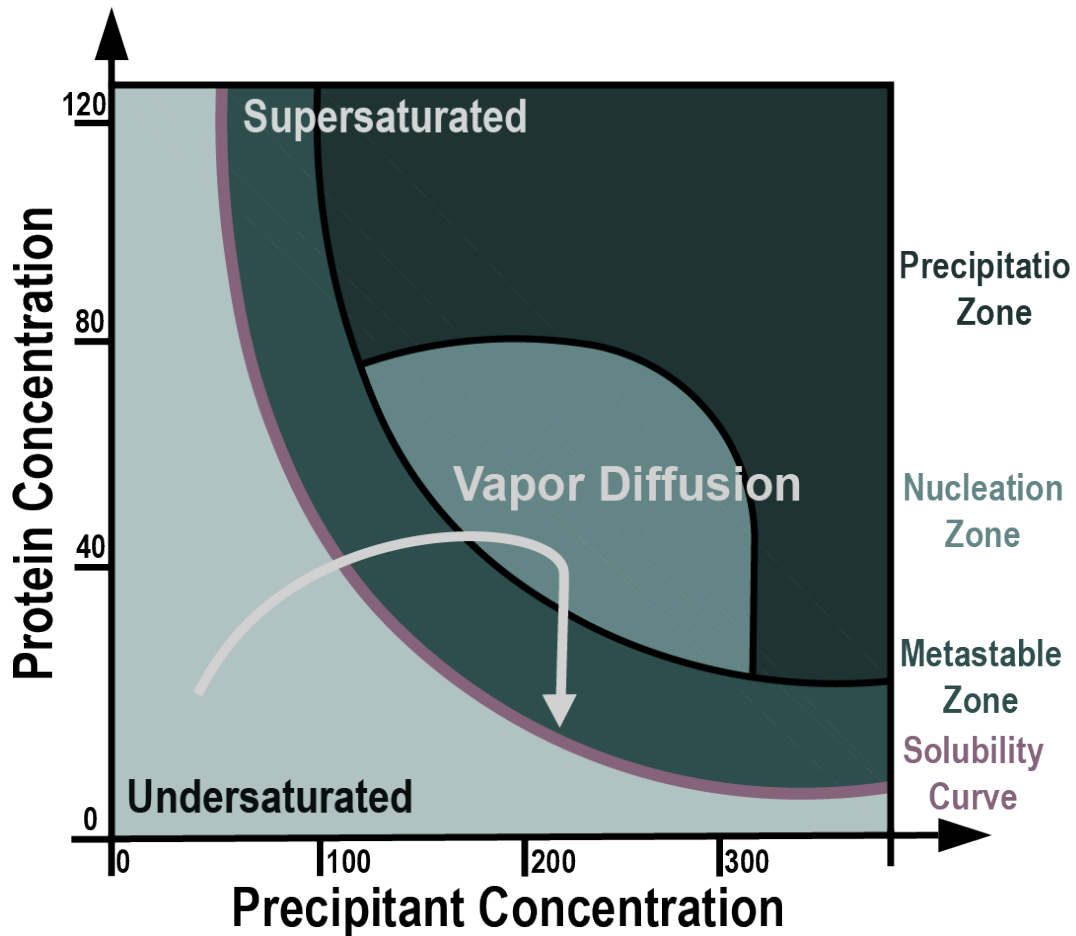
Figure 1.1. Phase diagram for protein crystallization by vapor diffusion. Protein concentration and precipitant concentration increase as water evaporates from the crystallization drop. However, once nucleation occurs, the concentration of free protein decreases as it is binding to the growing protein crystal.

(Dessau & Modis, 2011). Many other techniques have been established for crystallization including free interface diffusion, dialysis, and batch, which approach the phase diagram in different ways, but ultimately need to bring the protein into the nucleation and metastable regions to grow protein crystals (Kupitz, Grotjohann, et al., 2014).

Despite the difficulty of protein crystallization, advances in high throughput screening for crystallography have improved the ease, time, and sample constraints for crystallization. Automation of the process through liquid handling robots has allowed

samples to be set up using smaller quantities of protein and under more conditions, allowing sample to be screened in shorter periods of time (Abola et al., 2000; Stevens, 2000). Crystallization screening kits use common crystallization conditions and reagents to identify possible combinations that had in the past been successfully used for protein crystallization. Robots such as the Phoenix® and the Mosquito® can be used to set up 96-well crystallization plates with two different drops per condition (i.e., a 1:1 and a 1:2 ratio) for a total of 192 conditions in under ten minutes. Subsequently, plates can then be stored in crystallization "hotels" that will store the plates in environmentally controlled conditions and periodically photograph each drop with options such as polarized and UV light. Overall, it is possible to set up thousands of crystallization conditions by high-throughput robotic screening in a day with only 2 – 4 milligrams of protein.

Automation extends beyond the standard crystallization for soluble proteins. Membrane proteins are commonly known for being a difficult structural biology target in general; however, they also gained a reputation for challenges in crystallization (Bill et al., 2011). Advances for membrane protein crystallography include screens designed specifically for membrane proteins and lipid cubic phase (LCP) crystallization – which can all be performed with robotics in addition to by-hand crystallization (Caffrey, 2015). These systems can also be applied to drug discovery, one of the largest applications of crystallography. Libraries of compounds can be screened using high throughput methods and directly applied to crystallization for protein-ligand complex structures (Blundell et al., 2002). The combination of these methods has been especially important for the study of G-protein coupled receptors (GPCRs), which fall into both categories as membrane proteins and promising drug targets (Johansson et al., 2009).

The final experimental stage is the X-ray diffraction of a protein crystal. For the modern crystallization pipeline, this commonly means X-ray diffraction at a synchrotron source. Synchrotron light sources provide a powerful source of X-rays (compared to most laboratory home sources) but remain relatively accessible to the scientific community. The first large-scale synchrotron, Bevatron, was built in the 1950s at Lawrence-Berkley Laboratories. Today, around 70 synchrotrons are operating, or under construction ranging from first generation to third generation, with increase of photon flux with each new generation. Notably among the third-generation light sources are the Advanced Photon Source (**APS**) in Chicago, Illinois; Stanford Synchrotron Radiation Lightsource (**SSRL**) in Menlo Park, California; and the European Synchrotron Radiation Facility (**ESRF**) in Grenoble, France. Of the three facilities, SSRL can accelerate electrons to 3 GeV whereas the larger synchrotrons, ESRF and APS, can operate at 6 and 7 GeV, respectively.

Synchrotrons are comprised of four parts: an electron gun, a linear accelerator (linac), a booster ring, and a storage ring. Electrons produced by the electron gun are accelerated to nearly the speed of light in the linac by using radiofrequency cavities. The electrons are transferred to the booster ring where they increase in energy to >3 GeV and then placed in the storage ring. A cone of electromagnetic radiation is emitted from the electrons as they are deflected through the magnetic field. Tangential experimental end stations (called hutches) are placed around the synchrotron that allow for a variety of X-ray based experimentation, including X-ray crystallography (Ishikawa, 2019). This gives the synchrotron accessibility as experiments can be multiplexed and allow broad use of the facility by scientists. X-rays between 3.5 and 20 keV are used for X-ray crystallography to obtain X-ray diffraction datasets from diffraction of large protein

crystals. While microfocus beamlines have advanced to allow the diffraction of crystals of 10 µm dimensions, X-ray damage is severe and thereby crystal size can still be a limiting factor for X-ray structure analysis of many proteins.

The majority of crystallographic structures are determined from X-ray diffraction data collected at synchrotron sources, and most studies are performed under cryogenic conditions to diminish X-ray damage. To diffract crystals at a synchrotron source, crystals generally need to be mounted in a loop with a technique often called "fishing". The technique uses a small synthetic fiber "loop" (0.05 – 0.4+ µm based on crystal size) to catch the crystals from the crystallization drop. These crystals can be soaked or coated in a cryoprotectant (e.g., 30% glycerol, high molecular weight polyethylene glycol, sugar additives). This requires a second fishing step or addition of the cryoprotectant to the crystallization drop prior to the "fishing" and subsequent freezing steps. The crystals,
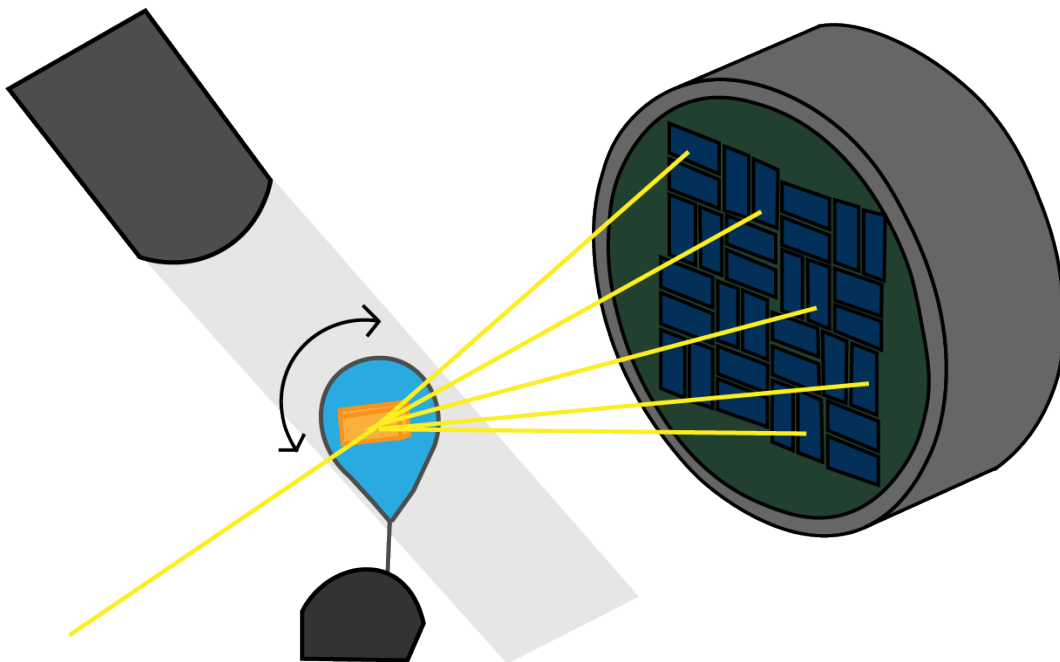
Figure 1.2. Schematic for cryo-crystallography experiment for large crystal X-ray diffraction (Cryo-MX) commonly used for X-ray data collection at a synchrotron or home light sources.

harvested in loops mounted on a magnetic base, are placed into cryo-pucks (shaped similar to a hockey puck, with small holes to hold the loops) in liquid nitrogen and hold the crystals during shipping to a synchrotron facility. Many synchrotrons have remote user operation options that will automate the process of placing the loop onto a remote-controlled goniometer, which allows the crystal to be oriented and rotated in the X-ray beam for user operation on site and remotely from their home institutions. A cryo-transfer holder and cryostream maintains the crystals during transfer to the goniometer and data collection under cryogenic conditions (Figure 1.2).

Diffraction under cryogenic conditions has the primary advantage of diminishing radiation damage. Definitions of primary, secondary, and tertiary radiation damage of protein crystals have been noted as inconsistent in the literature (Garman, 2010). Primary X-ray damage describes the initial ionization event where electrons are ejected from the inner orbitals of the atoms by the X-ray interaction. In secondary radiation damage, the free radicals emitted diffuse through the crystal and cause X-ray damage by reduction of atoms, thereby altering oxidation states of atoms and can lead to changes of chemical structures and chemical bonds. The tertiary X-ray damage definition is less well defined but generally describes the decrease of the order of the crystals by the chemical modification of the molecules. Exposure of the protein to the X-ray beam causes ionization and excitation events, resulting in subsequent heating of the crystals. Freezing the crystal and keeping it under cryogenic conditions does not mitigate the primary ionization event but limits the diffusion of the radicals in the crystals and the alteration of chemical bonds. This also reduces the heating effect thereby allowing for data sets of diffraction images to be collected without significant alteration of the protein and the

crystal contact sites (Garman, 2010). Some proteins, such as metalloproteins and membrane proteins, can be especially sensitive to radiation damage. This radiation damage to the crystal can cause the alteration of the proteins, causing the diffraction quality to worsen (going from high resolution patterns to low resolution patterns), breakage of disulfide bonds, alteration amino acids, and increase of the B-factors of the protein (Burmeister, 2000; Garman & Owen, 2006; Ravelli & McSweeney, 2000).

1.2 Rise of the XFELs

One of the largest recent advancements in crystallography has been the development and use of X-ray Free Electron Lasers (XFELs). The development of the first Free Electron Lasers (FELs) began in the 1970s and produced light in the infrared range (Madey, 1971; Pellegrini, 2012). The first X-ray XFEL in the soft X-ray regime was FLASH in Hamburg, Germany, and LCLS was the first XFEL in the world to generate hard X-rays. To generate X-rays, all current large XFELs combine a linac with a magnetic undulator and generate ultrashort X-rays by the principle of Self-Amplified Spontaneous Emission (**SASE**). As the electrons are forced on an oscillating path, they emit X-rays and then interact with their emitted radiation and form ultrashort coherent bunches. The bunching of electrons results in the emittance of coherent, femtosecond X-ray pulses of up to 10^{13} photons per pulse (Pellegrini & Stöhr, 2003).

One advantage to ultrashort femtosecond X-ray pulses with high flux is outrunning secondary radiation damage in a concept known as "diffraction before destruction," which was proposed nine years before the first hard X-ray XFEL, LCLS at Stanford started operation (Neutze et al., 2000). The molecular dynamic simulations of T4 lysozyme exposure to an XFEL beam by Neutze et al. showed the process of the

Coulomb explosion at time points between 2 and 50 fs under XFEL conditions. Their findings suggested that if the pulses were short enough to avoid the predicted Coulomb explosion that occurs at $5 - 10$ fs, with sufficient flux, diffraction of the sample could be observed before the sample is destroyed (Neutze et al., 2000). This was supported by the first XFEL diffraction experiments in 2009 where the structure of Photosystem I (PSI) was determined with data collected on protein nanocrystals at the Linac Coherent Light Source (Henry N Chapman et al., 2011). XFEL X-ray data collecting significantly differs from standard cryo-X-ray crystallography. As the XFEL pulse destroys solid material, the sample has to be replaced between each shot, but the pulses are so short that X-ray diffraction is observed before secondary X-ray damage destroys the crystals. Data is collected in a serial fashion where a stream of crystals is delivered to the XFEL beam at room temperature where they crystals in random orientation interact with the X-rays leading to tens of thousands of still diffraction images from the crystals (**Figure 1.3**). This allows for the solution of protein structures at room temperature without the severe radiation damage of their synchrotron counterparts. The first XFEL structure of Photosystem I (Henry N Chapman et al., 2011) was based on diffraction from over 30,000 nanocrystals.

As of 2022, there are currently five operational hard X-ray XFELs worldwide: the Linac Coherent Light Source (**LCLS**) at the SLAC National Accelerator Laboratory in Menlo Park, California, USA; the European XFEL (**EuXFEL**) in Schenefeldt, Germany; Pohang Accelerator Laboratory XFEL (**PAL-XFEL**) in Pohang, South Korea; RIKEN's Spring-8 Angstrom Compact Free Electron Laser (**SACLA**) in Japan; and Paul Scherrer

Institute **SwissFEL** in Switzerland (Hasegawa et al., 2016; Ko et al., 2017; Milne et al., 2017; Weise & Decking, 2017).

With the advancement of XFELs came new challenges; however, the solutions to these challenges have also driven the field of crystallography forward to create new opportunities. Three of the biggest challenges presented by XFELs include: sample preparation, sample delivery, and data analysis. Ultimately the solutions of these challenges open new possibilities and advantages for X-ray crystallography beyond XFEL research to include new options for developing crystallization conditions, methods of obtaining well diffracting micro and nanocrystals, and abilities to process copious amounts of complex data sets consisting of tens of thousands of single-shot still diffraction patterns.

The technique of serial crystallography was originally designed for the advancement of the X-ray Free Electron Laser, where sample, destroyed by the power of the X-ray beam, had to be continuously replaced. Where large, single crystals had been traditionally
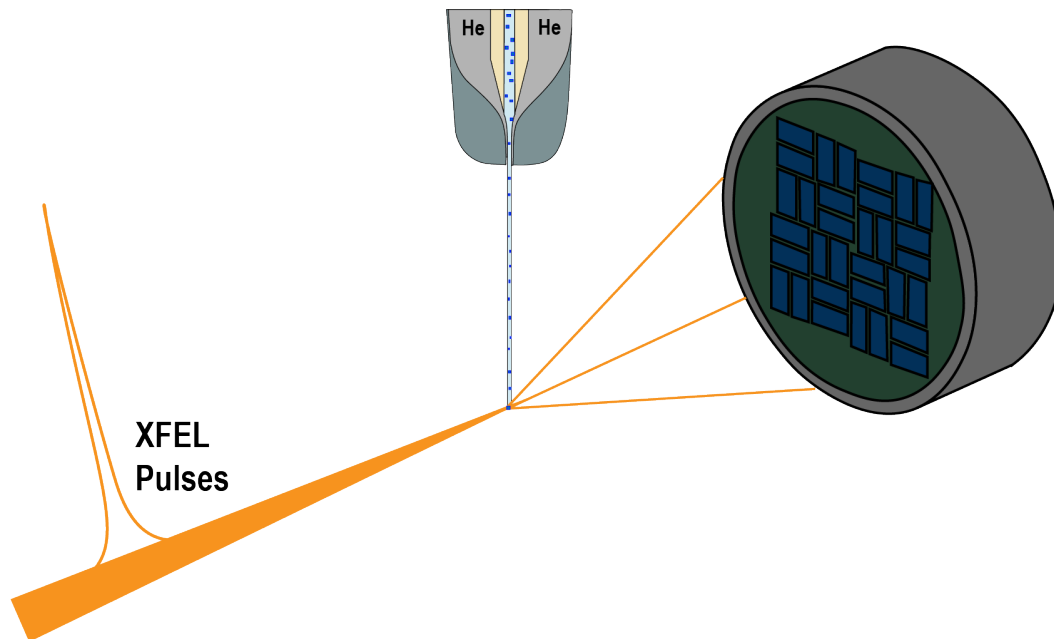
Figure 1.3. Schematic picture for serial femtosecond crystallography at XFELs. X-ray pulses (orange) interact with a liquid jet containing protein crystals (light blue). When an X-ray pulse hits a crystal, the crystal diffraction is observed, and the pattern is collected on a detector.

used in crystallography at synchrotrons, XFELs would require many very small crystals to diffract. To accomplish this, nano/micrometer sized crystals would be grown in supersaturated solutions to create crystal slurries, which are then "injected" by forming a liquid jet that intersects with the X-ray beam (**Figure 1.3**). This is the basis for serial femtosecond crystallography (SFX), where many crystals are diffracted by femtosecond pulses (Henry N Chapman et al., 2011).

Prior to this, nano and microcrystalline material were viewed as a way to obtain larger crystals by seeding, but had no practical use as they were too small to show high resolution diffraction and suffer strongly from radiation damage – though the idea of

collecting data sets from multiple small crystals had been suggested (Cusack et al., 1998). Since this advancement, many new techniques have been developed to obtain nano/microcrystals. Microcrystals can be advantageous as that they can sometimes be easier to obtain than large crystals and have less long-range disorder; however, for XFEL experiments, large sample volumes are often required. Methods such as batch and free interface diffusion are popular for obtaining microcrystals and can be assisted by techniques such as seeding. Small microcrystals can also be grown in lipid cubic phase (LCP), which is of special interest for membrane proteins and especially GPCRs as crystals are grown in a membrane mimicking environment (Gusach et al., 2020). Growth of crystals in LCP also has the advantage that they can be delivered by viscous injection which delivers samples at 10 to 100 times lower flow rates compared to liquid injection and thereby saving precious sample (Andersson, 2020; Kupitz, Grotjohann, et al., 2014; Weierstall, 2014).

The batch method of crystallization is performed by the rapid mixing of protein and precipitant. Upon mixing, the nucleation zone is directly reached. As crystals begin to grow, the concentration of free protein decreases, slowly moving into the metastable zone. Crystals settle at the bottom of the experiment (typically performed in and Eppendorf tube), forming a crystal pellet with time. Batch with agitation works along the same principles; however, by performing it in crystallization vials with a tiny stir bar, crystals grow in a smaller, more homogenous fashion as they are unable to settle and nucleation is favored over crystal growth. Free interface diffusion (FID) techniques utilize a layering of the protein over the solution with the precipitant with higher density. The layer between the two is the "interface" where diffusion occurs as the layers

13

interact. Nucleating crystals then sink into the precipitant layer to stop growing (Kupitz, Grotjohann, et al., 2014). While other techniques exist, such as ultrafiltration and dialysis, all methods utilize the principle of a fast run through the phase diagram to reach the nucleation zone (**Figure 1.1**), except for conditions where seeding is used to grow microcrystals in the metastable zone.

Crystal samples for diffraction at XFEL sources need to have adequate density and size distribution. The density of the crystals being diffracted is important for collecting full data sets in a timely fashion. As the crystals flow in the liquid jet, they can be "hit" by the X-ray beam; however, the majority of crystals flow past the X-ray interaction region without interacting with the beam. The frequency of this interaction where a crystal is diffracted in relationship to the total number of pulses is known as a hit rate (value given in a percentage). Because complete data sets require high multiplicity, thousands of indexable diffraction patterns are required to solve the structure. If the hit rate is low (ex. 0.1 – 1%) data collection could take half to multiple shifts of 12 hours to complete. Comparatively, with a high hit rate (ex. 40%) it may only take a few hours or less to complete a data set, depending also on how many patterns can be indexed. Crystal homogeneity is especially important for sample injection as large crystals can lead to clogging of the injector nozzle – replacement of which can often take 15 – 30 minutes of valuable time.

Sample injection is the process of delivering the sample, a slurry of crystals, to the X-ray beam. This has traditionally been accomplished with injectors that focus the crystal slurry delivered by a thin glass nozzle of 50-100 μm diameter into a liquid jet of 4 – 10 μm with a surrounding jet of helium gas. This system is a called Gas Dynamic Virtual

Nozzle (GVDN) and delivers the sample at flow rates of 10-20 µl/min to the interaction region with the beam (Weierstall, 2014). The challenge of sample injection often comes down to maintaining a stable jet and sample waste, which can be strenuous based on high flow rate and large jets. In addition to the jets, the jetting media and the environment are equally important. Not all buffer solutions behave equally in different types of injectors and environments. The viscosity of the buffer can play a major role in the observation if a sample jets or not and PEG, a common component in crystallization, can strongly affect this. Additionally, if the experiment is performed in vacuum versus in atmosphere or helium can play a major role.

Important injectors in serial crystallography include the double-flow focused nozzle (DDFN) liquid injector, the gas dynamic virtual nozzle (GDVN), and high viscosity injectors or LCP injectors (Martin-Garcia et al., 2019; Nelson et al., 2016; Wang et al., 2014; Weierstall et al., 2012). Fixed target options are also available for serial crystallography (Gilbile et al., 2021; Meents et al., 2017). Common issues for liquid injection such as difficulties of jet formation, clogging and low hit rate are largely avoided; however, challenges like increased background, sample environment (humidity, drying out of chips), and ability to move the chip do exist. Other methods under development include various types of droplet injection, separating the sample into small drops that are synchronized with the x-ray pulses, co-flow with oil, or a conveyer belt style sample delivery method. The goal of emerging sample delivery mechanisms is relatively universal: deliver the least amount of sample for the maximal hit rate and best quality of data.

1.3 Time-Resolved Crystallography

Time-resolved studies for the creation of "molecular movies" that show the dynamics of biomolecules in action has been dubbed as the "holy grail" of structural biology (Martin-Garcia, 2021). Structurally biology has grown exponentially in the deposition of PDB structures that show static images of macromolecules. Many strategies have been used to solve structures at certain time points of a biochemical reaction. For example the solution of the structure of an apoprotein, the protein bound to a substrate, the protein bound to a molecule mimicking a hypothesized intermediate structure, and the structure with the remaining product (post-cleavage) bound to the protein represent four different images at different stages of the reaction (Meredith N Frazier et al., 2021; Y. Kim et al., 2020; Y. Kim et al., 2021). Together these static structures create a more dynamic picture of the conformational changes that the protein undergoes during a reaction, as proposed in the concept of induced fit by David Koshland in 1958 (Koshland Jr, 1958). As crystallography grows as an expanding, interdisciplinary field that includes microbiologists, biochemists, biophysicists, accelerator physicists, computer scientists and data analysts, engineers, and more, more "challenging approaches" become accessible and practical to the scientific community. Time-resolved crystallography is still a challenging task for protein science; however, the growing support from novel methods development, technology, and new facilities that become available to the scientific community (the user) are enabling more scientists to take on these challenges to reap the benefits of time-resolved data.

Early on, the first reports were published where protein crystals have been shown to retain dynamics and activities within the crystalline lattice, although with some restraints

in terms of rate constants and crystal contacts (Quiocho & Richards, 1966). Early time resolved methods used large, single crystal diffraction and monitored very slow or irreversible reactions, making use of the technology that was available at the time. Clever strategies like flow cell systems and the placement of a large crystal into a glass capillary for substrate to diffuse in from either end allowed data collection at different timepoints (Gouaux et al., 1990; Wyckoff et al., 1967).

Time-resolved crystallography advanced again with Laue diffraction techniques, so named for the polychromatic spectrum (frequently called "pink beam") X-rays, which had been used by Max von Laue to diffract copper sulfate crystals. With the use of Laue diffraction, each diffraction pattern covers a larger "cup" of the Ewald sphere thereby a full data set can be achieved with diffraction from 10-20 different crystal orientations. As data can be collected faster, this significantly improved time resolution from minutes to seconds and millisecond time resolution for light-driven systems (Hajdu & Johnson, 1990). The applications of Laue diffraction have been used for light sensitive reactions (photoreceptors, ligand photolysis), enzymatic reactions, and virology advances (Bourgeois & Royant, 2005; Neutze et al., 2000; Schotte et al., 2003; Šrajer & Schmidt, 2017; Stoddard, 2001). Time scales and quality of Laue diffraction have further improved with the third generation synchrotron light sources and XFELs (Martin-Garcia, 2021). Laue diffraction is a useful technique that has enabled many scientific discoveries and continues to be used today (Moffat, 2019; Ren et al., 1999; Ren & Moffat, 1994).

The use of serial crystallography at the first femtosecond X-ray protein nanocrystallography (FPN) experiments, or SFX experiments, at the XFELs opened additional pathways for time-resolved studies (Henry N Chapman et al., 2011; Fromme,

2015; Fromme et al., 2020). Pump-probe lasers can be used "on the fly" to illuminate the stream of crystals with specific timing intervals between illumination and diffraction, resulting in change of state for photosensitive proteins (Kupitz, Basu, et al., 2014; Nogly et al., 2018; Weinert et al., 2019). Mix-and-inject systems allow for streams of substrate/ligand to be mixed with streams of protein crystals, allowing for diffusion to be observed at specific time points (Christopher Kupitz et al., 2017; J. L. Olmos et al., 2018; Pandey et al., 2020; Pandey et al., 2021). These new technology developments have now also be successfully implemented at synchrotron beamlines leading to the rise of serial millisecond crystallography (SMX) and time resolved SMX.

Synchrotrons, with their greater ability to multiplex experiments, are more accessible to the scientific community than XFELs. As a result, serial synchrotron crystallography (SSX) is becoming more accessible to the scientific community. Previously, large interdisciplinary teams were needed to manage such beamtimes using dedicated groups for sample preparation, injection, and data analysis. While these are beneficial, smaller groups can now conduct experiments with more established and automated systems employed at synchrotrons, which are notably more user friendly. Viscous media (LCP) injectors and fixed target systems help sample delivery by reducing sample consumption. The viscous injectors run at low flowrates that dramatically reduces sample consumption. More complicated time-resolved experiments can also be performed at synchrotrons, as shown by the pump-probe experiments of bacteriorhodopsin (Weinert et al., 2019), but they are still in their exploratory phase. With growing interest, facilities such as ESRF are even commissioning a new beamline (ID-29) that will specifically support SMX experiments (de Sanctis, 2021).

The experiments presented in this dissertation represent a culmination of these techniques in crystallography for structure-based rational drug design. The proteins NendoU and Taspase1 are medically relevant proteins to the diseases of COVID-19 and cancer, respectively. If their biological function is inhibited, the downstream effects of their pathways in promoting disease would be downregulated. Standard and serial crystallographic techniques are used to establish the foundation for time-resolved studies. Additional components of this thesis include extensive crystallization and co-crystallization method developments, activity assays to study function and propose inhibitory compounds, protein mutagenesis, microscopy, and more. The comprehensive goal is to better elucidate the mechanism and dynamics of these proteins for rational drug design while continually improving the methodology used to achieve these goals.

CHAPTER 2

ROOM TEMPERATURE STRUCTURAL STUDIES OF SARS COV-2 PROTEIN

NENDOU WITH AN X-RAY FREE ELECTRON LASER

Text and Figures in this Chapter were reprinted with permission from **Rebecca Jernigan**, Dhenu Logeswaran, Diandra Doppler, Mukul Sonker, Nirupa Nagaratnam, Jose Martin-Garcia, Sabine Botha, Jay-How Yang Thomas Grant, Gihan Ketawala, Valerio Mariani, Alexandra Tolstikova, Michelle Sheikh, Megan Shelby, Mimi Cho Yung, Matt Coleman, Sahba Zahre, Marc Messerschmidt, Emily Kaschner, Michele Zacks, Meng Liang, Raymond Sierra, Mark Hunter, Stella Lisova, Alexander Batyuk, Christopher Kupitz, Sebastien Boutet, Debra T. Hansen, Richard Kirian, Marius Schmidt, Raimund Fromme, Matthias Frank, Alexandra Ros, Julian Chen and Petra Fromme.

2.1 Abstract

The first room temperature structure of the SARS-CoV-2 endoribonuclease NendoU is solved by serial femtosecond crystallography at an X-ray Free Electron Laser (XFEL) to 2.5 Å resolution. NendoU is a key protein in the high virulence of SARS-CoV-2. It cleaves the poly-uridine (poly-U) leader sequence of the viral negative strand RNA, which is recognized and activates host sensors and rapid degradation of the virus in macrophages. By cleaving the poly-U leader, it hides the presence of the CoV-2 RNA from the immune system. We describe micro-crystallization of NendoU as well as novel functional studies on RNA cleavage specificity in solution and in crystals, showing that the enzyme is fully active in crystal form and providing new insights into the substrate specificity of NendoU. This work paves the way for future time-resolved studies of NendoU, towards a molecular movie of its catalytic mechanism.

2.2 Introduction

Understanding the fundamental biology and mechanisms of the SARS-CoV-2 virus, the causative agent of the COVID-19 pandemic, is not only critical for scientific understanding, but also a rapid drug discovery pipeline. One leading target for drug discovery, conserved in the *Coronaviridae* family, is the NendoU protein (nonstructural protein 15 (Nsp15)), which has been implicated in the virus's evasion from the host immune system (Hackbart et al., 2020). As the high rate of transmission among asymptomatic individuals is a differentiating factor from SARS and MERS, early detection of SARS-CoV-2 and response by the immune system is hypothesized to make an impactful difference in reducing the spread and severity (R. J. Khan et al., 2022; Saramago et al., 2022) .

Figure 2.1. Infection of SARS-CoV-2 and role of NendoU. The translated antisense strand of RNA contains a polyuridine (Poly-U) leader. NendoU cleaves the Poly-U tail, interfering with immune system's pattern recognition receptors.
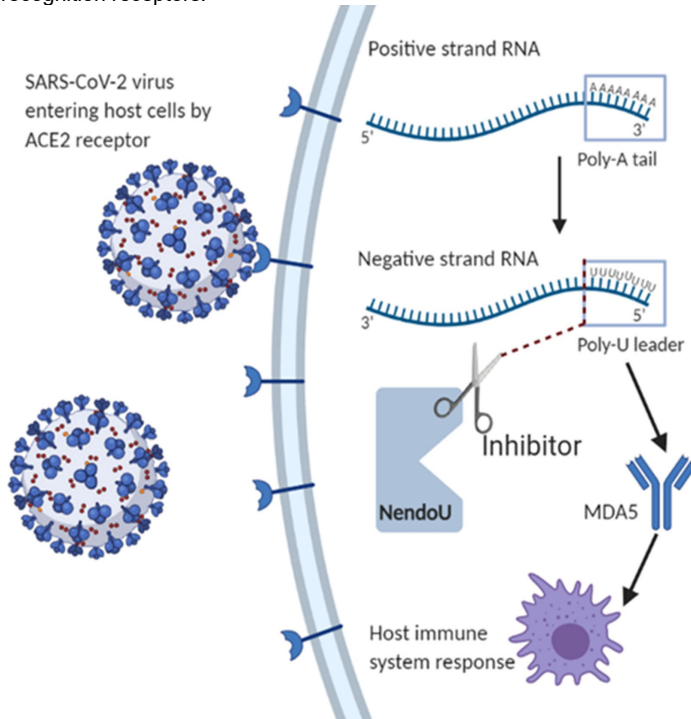
The **U**ridine specific **n**idoviral **endo**ribonuclease (NendoU) cleaves the poly-uridine (Poly-U) sequence of viral anti-sense RNA (**Figure 1**) (Hackbart et al., 2020). Synthesis of anti-sense genomic RNA is a vital step in the replication of SARS-CoV-2 (V'kovski et al., 2021). The poly-A tail of the positive strand RNA

genome's 3′ untranslated region (3′ UTR) corresponds to the 5′ poly-U leader sequence in the anti-sense RNA of SARS-CoV-2. The intermediate structure, double stranded RNA (dsRNA) is recognized as a pathogen-associated molecular pattern (PAMP). The dsRNA stimulates the innate immune system via type 1 interferon expression by recognition through melanoma differentiation-associated protein 5 (MDA5). Loss of functional NendoU by mutation during *in vitro* studies with the mouse hepatitis virus strain A59 (MHV-A59) resulted in a dramatic increase in double stranded RNA dependent immune response (Kindler et al., 2017). The loss of NendoU activity resulted in greatly attenuated disease in mice and stimulated a protective immune response (Kindler et al., 2017). These results were supported by mouse-model studies using mouse coronavirus showing that NendoU is essential for allowing virus replication with minimal response by macrophages and natural killer cells (X. Deng et al., 2017). It was recently discovered that SARS-CoV-2 NendoU cleaves and removes the poly-U leader sequence of the negative strand RNA genome of SARS-CoV-2, which indirectly enables the virus to evade host immune responses (Hackbart et al., 2020).

As NendoU plays a critical role in hiding the SARS-CoV-2 virus from the host immune system, it is an important enzyme to understand the catalytic mechanism as a target for drug discovery and rational drug design. NendoU is a 39 kDa enzyme (41 kDa construct, including a Histidine tag) that forms a hexamer made up of a dimer of trimers(Y. Kim et al., 2020). It has been shown that the hexameric assembly is critical for activity (Guarino et al., 2005). A two-step catalytic mechanism has been proposed in the past based on RNaseA which shares some active site similarities with NendoU (Y. Kim et al., 2021), however several critical amino acids in the catalytic center are not conserved and there is

also currently no mechanism proposed that could explain the manganese-dependency of the catalytic mechanism.

Structures of NendoU from SARS-CoV-2 have been determined under cryogenic conditions in the last 2 years both by X-ray crystallography (Y. Kim et al., 2020; Y. Kim et al., 2021) and cryo-EM (Pillon et al., 2020, 2021) (M. N. Frazier et al., 2021). The first structure of NendoU from SARS-CoV-2 was determined by the group of Jankowiak and colleagues in 2020 (Y. Kim et al., 2020), which included the apo protein and the complex with citrate. The same group, in 2021, solved crystal structures of the enzyme in complex with single nucleotides, a proposed intermediate analog, and proposed inhibitor Tripiracil (Y. Kim et al., 2021). Cryo-EM investigations of NendoU included studies of the apo structure of NendoU and the complex with uridine triphosphate (UTP) (Pillon et al., 2021). Then, in 2021 Frazier et al published two cryo-EM structures of NendoU in the journal Nucl Acid Research, which are proposed to represent the pre- and post-cleavage structure of NendoU (M. N. Frazier et al., 2021). For their studies, NendoU was incubated with a 3 nucleotide RNA and the post cleavage structure identifies 2 nucleotides in the binding site. All structures of NendoU so far have been solved either by standard crystallography under cryogenic conditions or by cryo-EM. Currently there are no structures published that have investigated the NendoU structure at room temperature.

This study investigates the function and substrate specificity of NendoU and reports on the first room temperature structure of NendoU, laying the groundwork for future time-resolve studies by serial femtosecond X-ray crystallography (TR-SFX) at X-ray Free Electron Lasers (XFELs). X-ray free electron lasers produce X-ray pulses that are so short that the resulting diffraction patterns can be collected before X-ray radiation damage can

manifest throughout the crystal. This allows for structure determination at ambient temperature (A. Barty et al., 2012; H. N. Chapman et al., 2011), with radiation doses well beyond the generally allowable Henderson limit. Reactions can be triggered "on the fly" to study the mechanism of NendoU through use of mix-and-inject technology (Ishigami et al., 2019; J. L. Olmos, Jr. et al., 2018; Pandey et al., 2020). These techniques have been used in previous studies to study the enzymatic function of BlaC, which cleaves antibiotics (C. Kupitz et al., 2017; J. L. Olmos, Jr. et al., 2018; Pandey et al., 2020; Pandey et al., 2021), and enabled dynamic studies of a riboswitch that regulates gene translation (Stagno et al., 2017). For SFX studies, thousands of single crystal diffraction snapshots are collected from a liquid jet of nano or microcrystals at room temperature which interact with ultrashort, ultrabright X-ray pulses from the XFEL with 10 – 40 femtosecond pulse duration. The intensity of the pulses so strong, that they destroy any solid material, but they are short enough to collect X-ray diffraction images prior to the onset of secondary X-ray damage (A. Barty et al., 2012). As data collection occurs at room temperature and "outruns" secondary radiation damage, structures are solved at ambient "near native" conditions.

Time-resolved studies for SFX pose many challenges, including the primary challenge of serial fs crystallography itself, which is obtaining quality nano- or micro-crystallization conditions. Micro-crystallization conditions for SFX studies require millions of small, uniform crystals. These crystals are delivered by a liquid jet to continuously replenish the sample interacting with the X-rays, leading to large sample consumption which can require hundreds of milligrams to a gram of protein for a single experiment. Even more challenging is establishing crystallization conditions where the

protein remains active, allowing for substrate to diffuse into the crystals and conformational changes without dissolving the crystals or ruining the sensitive RNA substrate.

Here we report on SFX experiments of NendoU at an XFEL leading to the first structure of NendoU at room temperature, which show differences in the flexibility of the two trimers in the hexamer, with potential large implications for the catalytic function of NendoU. In preparation for future time-resolved SFX studies, RNA substrates of varying lengths were designed based on analysis of the NendoU protein structure and genomic analysis of the SARS-CoV-2 RNA. This work paves the way for future time-resolved analysis, where mix-and-inject studies could show substrate binding, the catalytic reaction, and product release. We could thereby discover the mechanism of RNA cleavage by NendoU, where the unravelling of the transition state could provide the basis for development of novel drugs for severe coronavirus infections and to investigate the role of this protein in suppression of the immune response.

2.3 Methods

Vector design and Cloning

The expression plasmid pHis-TEV-Nsp15 was designed based on the available information in (Kim et al 2020) for Protein Data Bank accession no. 6VWW. pHis-TEV-Nsp15 expresses a protein sequence identical to 6VWW and was purchased from GenScript (4 µg). The parent vector is pET-11b, which, like the vector for 6VWW, confers ampicillin-resistance, uses the T7*lac* promoter and T7 terminator, contains the *lacI* gene and has a ColE1 replication origin. Like 6VWW, the DNA sequence of pHis-

25

TEV-Nsp15 was optimized for expression in *Escherichia coli* using GenScript's

OptimumGene algorithm. For use in *in vitro* expression, an additional 100 µg was

purchased from GenScript. Supplementary Figures S1 A-C show the vector map,

complete DNA sequence and expressed protein sequence, respectively.

In vivo Protein Expression

Expression of wildtype NendoU was optimized from prior methodology

(Makowska-Grzyska et al., 2014 & Kim et al., 2020). The plasmid for Nsp15-TEV-His$_6$

was obtained from genscript and transformed into commercial competent *E. coli* cells

using the BL21-Gold (DE3) strain from Agilent Technologies. Transformation was

performed by adding 20 ng of plasmid to 100 µL of competent cells and incubating on

ice for 30 minutes. Cells were then heat shocked by 42 °C water bath for 45 seconds and

then rested on ice for two minutes. Super Optimal broth with Catabolite repression (SOC)

was added to the cells for a total volume of 1 mL and grown at 37 °C for 1 hour with 250

RPM shaking. Cells were then plated to luria broth plates with 50µg/mL carbenicillin and

grown overnight at 37 °C. Starter cultures were prepared by inoculation of a single

colony into a 50 mL LB Lennox media (+50 µg/mL carbenicillin) and grown at 37 °C

overnight with 180 RPM shaking. Cell cultures of 1 L were inoculated with 1 % of

overnight starter culture and grown at 37 °C with 180 RPM shaking to an OD$_{600}$ of ~1.0.

Cultures were transferred to incubator at 18 °C with and induced with 0.2 mM IPTG,

0.1 % w/v glucose, and 40 mM of potassium phosphate dibasic. Cells were grown for

18 hours with 180 RPM shaking at 18 °C. Whole cell contents were analyzed by SDS-

PAGE gel with InstantBlue Coomassie stain and anti-His western. Cells were harvested

by centrifugation at 7500 xg, 15 minutes, 4 °C and stored at -80 °C.

Development of an in vitro *expression system for NendoU*

Expression screening: Initial screening was performed by running 50 µl cell-free reactions using an Invitrogen Expressway Kit for four hours with 0.5 – 2 µg pHis-TEV-Nsp15 plasmid and 2 µg Green Fluorescent Protein control incorporating 1 µL Fluorotect reagent into each reaction. Lysates were denatured and run in reducing conditions on 4-12% Bis-Tris SDS-Page gels and imaged at 600 nm and 700nm using a LI-COR Odyssey Fc Imaging system (Figure 2-1-2A).

Expression scale-up and purification: Preparative 1 ml reactions were carried out using the Rabbit Biotech RTS 500 ProteoMaster E. coli HY Kit for protein expression. In brief, lyophilized reaction components (E. coli lysate, reaction mixture, amino acid mixture, and methionine) were dissolved in reconstitution buffer and combined as specified by the manufacturer. A total of 20 µg of pHis-TEV-Nsp15 plasmid DNA was added to the lysate mixture and the reactions were incubated at 37 °C for 5 hours.

Gravity-flow immobilized nickel affinity chromatography was used to isolate the protein of interest using buffers containing 50 mM Sodium Phosphate pH 8, 150 mM NaCl. Expression products were bound to the Ni column for 1 hour in 5 mM Imidazole, washed with 6 column volumes of 10 mM imidazole, and eluted with 1.8 column volumes of 250 mM imidazole and 0.3 column volumes of 500 mM imidazole. Fractions were assayed with SDS-PAGE (Figure 2-1-2B) and for protein concentration using a Q-bit protein assay and Fluorimeter from Molecular Probes. For size exclusion chromatography (SEC) purification, fractions E2-E5 were pooled, dialyzed into 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP-HCl running buffer and concentrated to 5 mg/ml via 10 kDa molecular weight cut-off spin filtration. Protein was purified with a

Superdex 200 Semi-prep column at 1 ml/min flow rate (Figure 2-1-2C). (Levine et al., 2019; Lu, 2017)

Protein Purification

Purification was based on Kim et al., 2020; cell pellets were resuspended in 10% (w/v) lysis buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 20 mM imidazole) and lysed by sonication (2 sec on 2 sec off for 2-minute for five cycles with 2 minutes rest on ice). The cytosol was isolated by centrifugation at 45,000 xg for 30 minutes. Immobilized Metal Affinity Chromatography (IMAC) was performed with 5 mL Ni-NTA Chromatography Cartridges from GE on an ÄKTA Pure FPLC at 4 °C. After application to the column, the protein was washed with buffer (50mM HEPES, pH 7.5, 500 mM NaCl) for 20 column volumes (CV) with 20 mM imidazole and 20 CV of 50 mM imidazole. Eluted protein was concentrated with 30 kDa, 0.5 mL spin concentrators to 0.5 mL volume and applied to a Superdex 200 10-300 increase column. The NendoU hexamer was collected between 11-12 mL and analyzed by SDS-PAGE gel with InstantBlue Coomassie stain and anti-His western. The protein was concentrated to 10 mg/mL for storage in -80°C until future use.

Crystallization

Ultrafiltration

Purified protein stored at 10 mg/mL was concentrated to 60 – 80 mg/mL at 7000 xg, 4° C in 10 minute increments with pipette homogenization between steps. Once past the 60 mg/mL mark, the concentrated protein appears turbid after homogenization. At this point the protein/crystal solution is transferred to an Eppendorf tube and

homogenized until the solution is entirely turbid. Crystallization was verified under the microscope and the crystal suspension stored at 4° C to mature overnight for a denser pellet.

Batch with Agitation

NendoU protein was concentrated to 40 mg/mL using the same procedure as for ultrafiltration. A glass vial is placed on a magnetic stir plate at room temperature with a micro-stir bar gently spinning at ~200 rpm. The precipitant (300 µL: 0.1 M sodium citrate, pH 5.6, 10 % MPD, 10 % PEG 2000) is placed on the bottom of the vial. The protein (50 µL) is added dropwise to the vial with gentle stirring and turn turbid on addition and left for 24 hours. The sample is observed under the microscope by placing a drop between two coverslips to reduce the motion of small crystals within the drop. Same info without the faff:

NendoU protein solution was concentrated to 40 mg/mL using the same procedure as for ultrafiltration. To 300 uL of precipitant solution (0.1 M sodium citrate, pH 5.6, 10 % MPD, 10 % w/v PEG 2000) stirring at ~200 rpm at room temperature, 50 uL of protein solution was added drop by drop. The resulting turbid solution was stirred for an additional 24h.

RNA design and functional essays

Multiple sequence alignment

Multiple sequence alignment was performed manually using the BioEdit program (Hall et al., 2011).

NendoU in solution endoribonuclease assay

Purified NendoU protein solution was diluted to a final concentration of 6 µM with assay buffer (50 mM Tris-HCl pH 7.5, 50 mM KCl and 1 mM DTT). Lyophilized NIR-800 3′ labeled RNA oligos (Integrated DNA Technologies) maintained under dark conditions were dissolved to a final concentration of 100 µM in nuclease free water and was serially diluted to 5 mM with nuclease free water. Ten µl reactions were setup with 5 µl of 2X assay buffer, 2 µl water, 1 µl each of 6 µM NendoU solution (final concentration 0.6 µM) and 5 µM RNA solution (final concentration 0.5 µM). Reactions were incubated for 5 mins at 25 °C followed by addition of 1 µl water or 1 µl 50 mM $MnCl_2$ for a final concentration of 5 mM of what? and further incubation for 3 mins. Reactions were stopped by adding equal volume of 2X Urea loading dye (8 M Urea, 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.05% (w/v) Xylene cyanol and 0.05% (W/V) Bromophenol blue) followed by incubating at 95 °C for 2 mins and placing on ice. One-tenth of the reactions were size fractionated on a 10% Urea-PAGE or 15% Urea-PAGE gel for short substrates at 15 W power for 1.5 hr. Following electrophoresis, gels were scanned directly on a Typhoon NIR-IP scanner.

NendoU in-crystal endoribonuclease assay

1 µl of crystal suspension (equivalent to 15 mg NendoU protein) was incubated with 4.42 µl of crystallization buffer (0.1 M sodium citrate pH 5.5, 10% MPD, 20% w/v PEG 1000) and 3.58 µl of 100 µM RNA substrate. The reaction was set up in duplicate. A separate reaction mixture without the RNA but with 3.58 µl of crystallization buffer was set up as control. All three reaction mixtures were incubated at 22 °C for 30 mins. 1 µl of 50 mM $MnCl_2$ was added to one of the reaction mixtures containing RNA and 1 µl crystallization buffer each to the duplicate and control. T reaction mixtures were further

30

incubated for 1 min and 1 µl of reaction was added to 19 µl 2X Urea loading dye followed by incubation at 95 °C for 2 mins and placing on ice until loading. The rest of the reaction mixture was centrifuged at 500 g / 2 min at 4 °C and 5 µl of the supernatant was used for a Bradford assay according to manufacturer's instructions (BioRad protein assay – cuvette format). Crystals were further allowed to settle and checked under polarized microscope.

Data analysis

Images were recorded at the full 120Hz repetition rate of the X-ray FEL at the LCLS using an epix10k detector. The online hitrate and background correction for the live display were performed using "OM", a customized version of OnDA (Mariani et al., 2016) for LCLS. Offline calibration and hit finding were performed with a new (unpublished) hybrid version of cheetah (Anton Barty et al., 2014) and OnDA (Mariani et al., 2016) developed by Alexandra Tolstikova at DESY in Germany. Peakfinding was performed using the peakfinder 8 algortihm (parameters: adc_threshold=500, minimum-SNR=5.0, min_pixel_count=1, local_bg_radius=4, min_res=80, max_res=450), and patterns with more than 12 peaks identified were classified as a hit. The hits, along with the recorded peak positions were submitted to the crystal indexing software crystfel version 0.9.1. (White, 2019; White et al., 2012; White et al., 2016), employing the indexing algorithms xgandalf (Gevorkov et al., 2019), dirax, mosflm (Powell et al., 2013), and XDS (Kabsch, 2010) in that order. Due to the indexing ambiguity for serially collected "still" diffraction patterns in pointgroup P63, the list of indexed diffraction intensities was consequently passed through the program ambigator (part of the crystFEL package, based on the clustering algorithm described in Brehm 2014) applying the re-

indexing operator hkl. The reflection intensities were then merged using partialator (from THE CrystFEL package).

The merged list of reflection intensities was converted to ccp4 compatible mtz format (Winn et al., 2011) using xscale (Kabsch, 2010), where they were scaled using aimless (Evans & Murshudov, 2013) for structure solution. The initial search model 6XDH (Dranow et al, unplublished) was downloaded from the PDB and all solvent and ligand atoms removed. The structure was then refined applying alternate cycles of refmac (Murshudov et al., 2011) and manual refinement with coot (Emsley et al., 2010). Ultimately, citrate (and water) molecules were fitted and refined into the electron density in the active site, and refined using POLDER maps in Phenix (Liebschner et al., 2017), omitting bulk solvent around the ligand. The final model was passed through the web interface PDB-redo (Joosten et al., 2014) for a final cycle of refinement prior to submission.

Figures were prepared using Pymol (pymol). Channel sizes and maps through the crystal were calculated using MAP_CHANNELS v.0.5 (Juers & Ruffin, 2014), and ligand plots were produced with LigPlot+ v.2.2 (Laskowski & Swindells, 2011).

2.4 Results

In vivo *Expression and Large-scale Purification of NendoU*

For XFEL studies large protein quantities are required, which was a challenge for the SFX NendoU project as the expression yields were very low. A clone was constructed for the expression of NendoU wildtype (WT) with a C-terminal His$_{10}$ tag by Gene-script (**Figures S1 A-C**). When initially following the published expression protocol for NendoU from Kim et al (Y. Kim et al., 2020), 1 – 3 mg NendoU protein was isolated

from 1 liter of cell culture. Optimization of expression conditions including varying the cell culture media, temperature, time of induction was tested; however, all conditions performed chronologically worse. NendoU toxicity to the expression host cell and low expression has been previously reported (Xu et al., 2006). Fresh transformation weekly was essential to prevent loss of expression. To isolate 300mg of WT NendoU for the RT-SFX experiments, we have scaled up an expression and isolation protocol, where cell culture batches of up to 48 Liters (16 liters x 3 growths/week) were grown from freshly transformed *E. coli* cells, from which NendoU was isolated by metal-affinity chromatography followed by size exclusion chromatography (SEC). NendoU (WT) purified protein (hexamer) was verified by SDS-PAGE with Coomassie stain and Western Blot (**Figure S2; B,D,F**). Dynamic light scattering showed the protein was monodispersed for crystallization (**Figure S2, E**).

In vitro Expression and Purification from Cell Free Expression

To address potential toxicity effects on the in vivo expression yields of NendoU, an *E. Coli*-based cell-free expression platform was implemented. Cell-free expression is a flexible and increasingly cost effective one-pot in vitro expression approach in which the cell's translational machinery in the form of a cell lysate is provided with expression target cDNA and supplemented with amino acids and any necessary additives. This removes cell viability concerns while enabling a highly tunable expression platform (Lu, 2017). Initial small scale expression screening with varied pHis-TEV-Nsp15 plasmid concentration utilized reaction mixtures including a tRNA that incorporates BODIPY labeled lysine into in-vitro translation products. After expression, the reaction mixtures containing whole lysates were analyzed via SDS-PAGE and imaged at 600 nm to
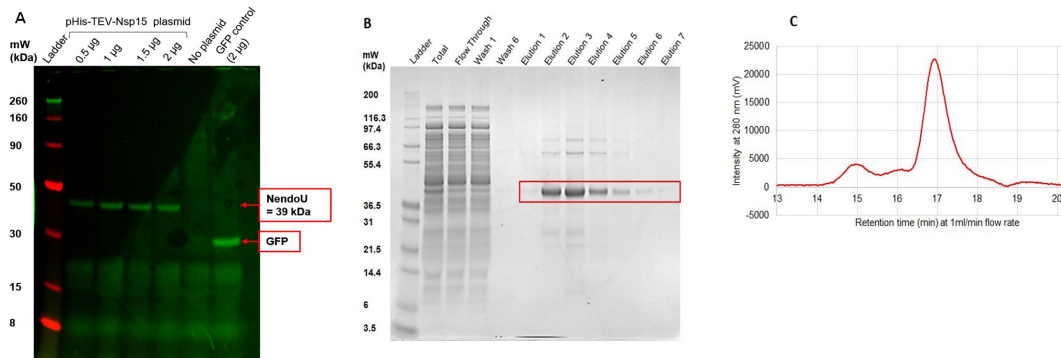
Figure 2.2. Results of in vitro cell-free expression of NendoU **(A)** BODIPY-labeled cell free expression products from expression screening of pHis-TEV-Nsp15 plasmid. 4-12% Bis-Tris SDS-PAGE gel imaged at 700 nm (red) and 600 nm (green) to visualize Li-cor Chameleon 700nm ladder **(B)** Fractions from gravity-flow Ni-NTA purification of 1ml preparative cell free expression of NendoU assayed with SDS-PAGE. Samples were run on a 4-12% Bis-Tris SDS-Page gel, stained with Sypro and imaged at 600nm. **(C)** SEC chromatogram for pooled and concentrated elution fractions from Ni-NTA purification of 1ml preparative cell free expression of NendoU. The Protein was purified with Superdex 200 Semi-prep in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP at 1ml/ml (pictured, 50 µg test injection).

visualize the incorporated BODIPY, which clearly labels a 41 kDa translation product, likely NendoU, in all reactions containing the pHis-TEV-Nsp15 plasmid (**Figure 2A**). Higher molecular weight contaminants are not apparent in this analysis.

Moderate scale-up reactions of 1 ml volume were performed to assess the translation products. Reaction products were purified from the cell-free lysate with a gravity flow Ni-affinity column and purification fractions were assayed with SDS-PAGE (**Figure 2B**). Several elution fractions show a strong band at 41 kDa with some minimal contaminant bands present at approximately 60 kDa and 75 kDa. Protein quantification of the elution fractions estimates yields of NendoU at 0.8mg/ml. Further purification by SEC results in a very similar elution profile to those produced during the purification of NendoU expressed in vivo (Figure 3C). These results confirm that high yields of NendoU can be
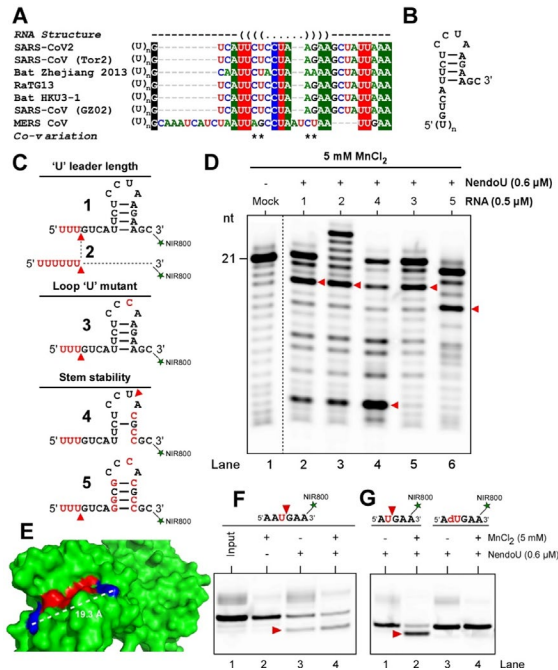
Figure 2.3. Design of RNA substrates for NendoU cleavage that mimic the 5' end of the SARS-CoV-2 anti-sense RNA Determining SARS-CoV-2 NendoU cleavage specificity and the optimal substrate for TR-SFX. **(A)** Multiple sequence alignment of nucleotides immediately downstream of the poly U leader sequence of negative strand RNA of SARS-CoV-2 and related corona virions. Nucleotides are colored in red (U), blue (C), green (A) and black (G). Nucleotides showing 100% conservation are highlighted with white text in colored background. RNA structure is shown as dot bracket notation above the alignment with co-varying residues marked with an asterisk underneath the alignment. **(B)** Secondary structure model of the first 21 nt of SARS-CoV-2 negative strand RNA immediately downstream of the poly U leader. Structure determined based on the multiple sequence alignment shown in (A). **(C)** Secondary structure models of synthetic RNA oligos to test the effect of various sequence or structural properties on NendoU activity. Expected major cleavage sites are shown with red arrowheads. **(D)** NendoU endoribonuclease assay gel for substrates in (C) with bands corresponding to major cleavage sites shown with red arrowheads. **(E)** Surface representation of NendoU crystal structure (PDB – 6VWW) with the active site cleft colored in red and putative RNA interacting amino acids flanking the active site colored in blue. The distance between the blue residues is 19.3 Å as measured in PyMol (Schrödinger, Inc.) shown by a white dotted line. **(F)** NendoU endoribonuclease assay for the 6 nt RNA substrate. RNA sequence is shown above the gel and the major cleavage site indicated both above the sequence and the corresponding band in the gel. **(G)** NendoU endoribonuclease assay for the cleavable and uncleavable 5 nt RNA substrate. Respective RNA sequences are shown above the gel and the major cleavage site indicated both above the sequence and the corresponding band in the gel for the cleavable substrate.

35

expressed in vitro using a cell-free approach, which alleviates concerns about cell viability during induction. Further optimization of expression yields and translation product purity may be achieved through improvement of the expression vector and variation of the expression conditions to promote monomer stability, for instance through addition of a reducing agent to the reaction.

NendoU activity and preferential cleavage sites

To determine if cleavage of the poly-U leader is structure or sequence dependent, a systematic approach was used in designing an optimal RNA substrate. A multiple sequence alignment of the native cleavage site, downstream sequences of SARS-CoV2, and closely related coronavirion negative strand RNAs were performed. While the first nucleotide following the poly-U tract was an invariant G residue, the alignment revealed a conserved 4 bp stem supported by co-variation immediately downstream of this G nucleotide (**Figure 3A, B**). RNA secondary structure modeling studies on 3' UTR negative strand sequence of SARS-CoV-2 did not predict this hairpin likely due to the reduced focus on evolutionary conservation in the prediction method used (Andrews et al., 2020). Thus, to determine if the length of the poly U tract or the presence of the conserved 4 bp hairpin influences the cleavage activity or specificity of NendoU, we designed 5 different substrates 3' end labeled with NIR-800 (Near Infra-Red 800 nm) fluorophore (**Figure 3C**). Substrates 1 and 2 have 3 and 6 'U' residues, respectively. Substrate 3 has a mutated 'U' nucleotides in the loop region and substrates 4 and 5 harbor mutations either disrupting or reinforcing the base pairing of the 4 bp stem (**Figure 3C**).

Endoribonuclease assays were setup with 0.6 µM NendoU and 0.5 µM of respective NIR-800 labeled RNA in the presence of 5 mM $MnCl_2$ (**Figure 3, lanes 2-6**)

including a mock control without NendoU (**Figure 3, lane 1**). Both substrates 1 and 2 are predominantly cleaved 3' of the last U residue of the 5' U tract with similar efficiencies irrespective of the length of the U leader (**Figure 3, lanes 2, 3**) suggesting that the length of the U tract does not play a significant role in NendoU cleavage efficiency. Substrate 4 was predominantly cleaved 3' of the analogous loop U residue (8th U from the 5' end) (**Figure 3, lane 4**).

However, the 4 U residues upstream of this U and downstream of the first U-G junction in substrate 4 do not seem to be preferred. Firstly, the reason that the analogous loop U residue (8th U from the 5' end) is cleaved with higher efficiency in substrate 4 than substrate 1 or 2 is likely due to the destabilization of the 4 bp stem that facilitates better access of NendoU to this residue as substrates 1 and 2 have a stable stem that might restrict binding of the loop U residue to the NendoU active site (**Figure 3C**). Additionally, mutating the loop U residue to a C in substrate 3, completely abolishes cleavage of the bond following this nucleotide (**Figure 3D, lane 5**) consistent with the cleavage occurring 3' of a U residue. Finally, substrate 5 is only cleaved after the 3rd U but not the 5th (**Figure3D, lane 6**) reinforcing that a G or an A is needed 3' of the U for efficient cleavage. Overall, this suggests that the preferred bond cleaved by NendoU is between a U and a purine, with cleavage occurring 3' of the U and the preferred substrate is unstructured RNA as structural elements involving the cleavage site or proximal to the cleavage site seem to have lower cleavage efficiency. Thus, substrate 5 with a single preferred cleavage site with a stable stem is suitable for crystallographic purposes and was chosen to proceed forward with downstream experiments.

Next, we aimed to determine the minimal length of the substrate capable of being cleaved by NendoU. The length of the active site cleft in the published structure of SARS-CoV-2 NendoU (PDB ID – 6VWW) was determined to be 19.3 Å by measurements in PyMol (**Figure 3E**). As the distance between two consecutive residues in RNA is 3-4 Å, a 6 nt RNA would likely be sufficient to be bound and cleaved by NendoU. Thus, 6 nt and 5 nt substrates were designed with a single preferred 'U-G' cleavage site (5'-AAUGAA-3') and (5'-AUGAA-3') respectively. Both substrates were cleaved with high efficiency by NendoU at the expected cleavage site (**Figure 3F, lanes 3, 4 and Figure 3G, lane 2**) while controls without NendoU produce intact substrate bands (**Figure 3F, lane 2 and Figure 3G, lane 1**). Additionally, an identical 5 nt substrate with a 2'-deoxyuridine which does not contain the 2'-hydroxyl on the Uridine preventing enzymatic cleavage was synthesized. As expected, NendoU cannot digest this substrate (**Figure 3G, lane 4**). Thus, either the 5 nt cleavable or uncleavable RNAs were the substrates of choice for further experimentation.

NendoU Microcrystallization

After a broad screening for potential micro-crystallization techniques, two micro-crystallization conditions were developed and optimized for the SFX experiments at SLAC National Accelerator Laboratory at the Linac Coherent Light Source (LCLS) (Menlo Park, CA). Two different crystallization methods were established and are shown in **Figure 4**. In the first method crystals are grown by ultrafiltration, where the protein concentration is increased under low salt conditions. These crystals had a needle morphology increasing the length to 5 – 10 μm. **Figure 4D** shows the phase diagram for this crystallization method, where crystallization takes place in the region of the phase

diagram that is referred to as "reverse of salting in". The solubility is decreased at low

salt concentration, where the surface of the protein becomes depleted from counter ions,
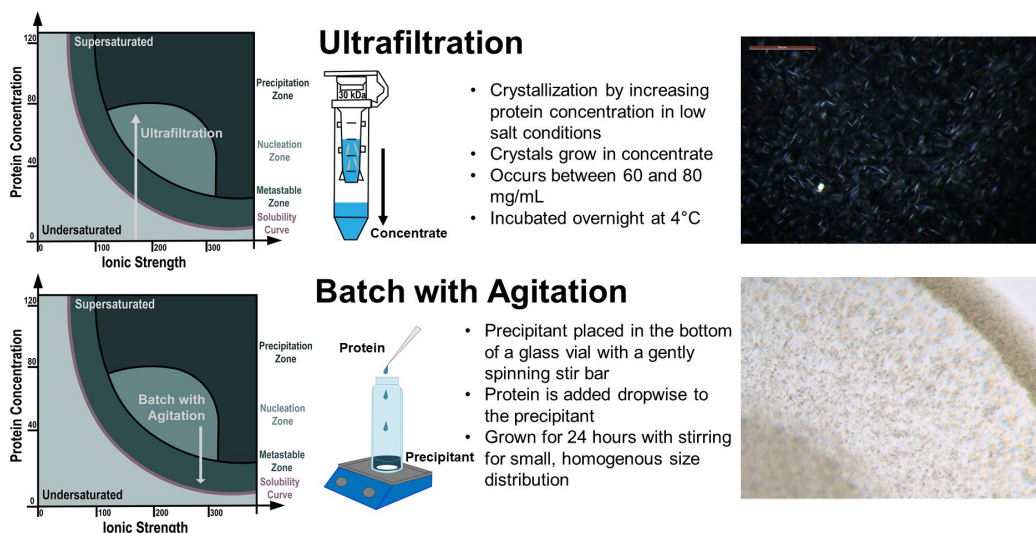
Figure 2.4. Microcrystallization of NendoU for SFX studies. Crystals grown by ultrafiltration use a high protein concentration under low ionic strength. Crystals with a needle morphology, 5 – 10 µm in size were observed. Crystals grown in the batch with agitation method produce 2 µm crystals with a shard morphology.

thereby fostering formation of crystal contacts between protein side chains of opposite

charge. This method has been used previously for crystallization of the large

photosynthetic membrane protein Photosystem I.

In the second method NendoU crystals were grown by a batch crystallization

method with agitation where protein was added dropwise to the precipitant solution under

agitation in a glass vial at room temperature (**Figure 4B**). **Figure 4A** illustrates how the

nucleation zone is reached stepwise in the phase diagram using this method. NendoU

crystals grown in batch with agitation condition a shard morphology and were

approximately 2 µm in length, width, and height (**Figure 4C**). The crystallization is also

referred to as the "citrate condition," in reference to the precipitant in the condition.

As citrate binds to the catalytic site of NendoU we have attempted to remove citrate from the crystals grown with citrate in the crystallization buffer by a 5-step washing procedure. 1 ml of the crystal suspension was centrifuged for 1 minute at 200rpm in a 1.5 ml reaction vessel. The supernatant was completely removed and 1 ml washing buffer was added which contained PEG and MPD but no citrate. The crystals were gently resuspended with pipette mixing and then again centrifuged. This washing procedure was repeated 5 times before the crystals were resuspended in 1 ml of the washing buffer for sample delivery.

NendoU is Active in the microcrystals

To determine if NendoU crystals are enzymatically active, endoribonuclease assays were performed by incubating substrate number 5 with crystals with and without $MnCl_2$. Remarkably, the crystals can efficiently cleave the RNA substrate even in the absence of Mn^{2+} ions suggesting that Mn^{2+} is not necessary for in crystal enzymatic cleavage as shown in **Figure 5A, lanes 1, 2**. We performed additional experiments to verify substrate turnover was due to the protein remaining active within the crystal, as opposed to the substrate being cleaved by the protein upon crystal dissolution. The crystal activity cleavage experiments were centrifuged following the reaction and the supernatant was assayed for solvated NendoU protein (**Figure 5B**). Both in the presence and absence of Mn^{2+}, no significant amount of protein was detected in the supernatant (**Figure 5B**). Further, the pelleted crystals post centrifugation did not show visible loss of crystal integrity due to dissolution under a microscope (**Figure 5C**). Finally, as expected, the crystals can also digest the 5 nt substrate with high efficiency and consistently in the
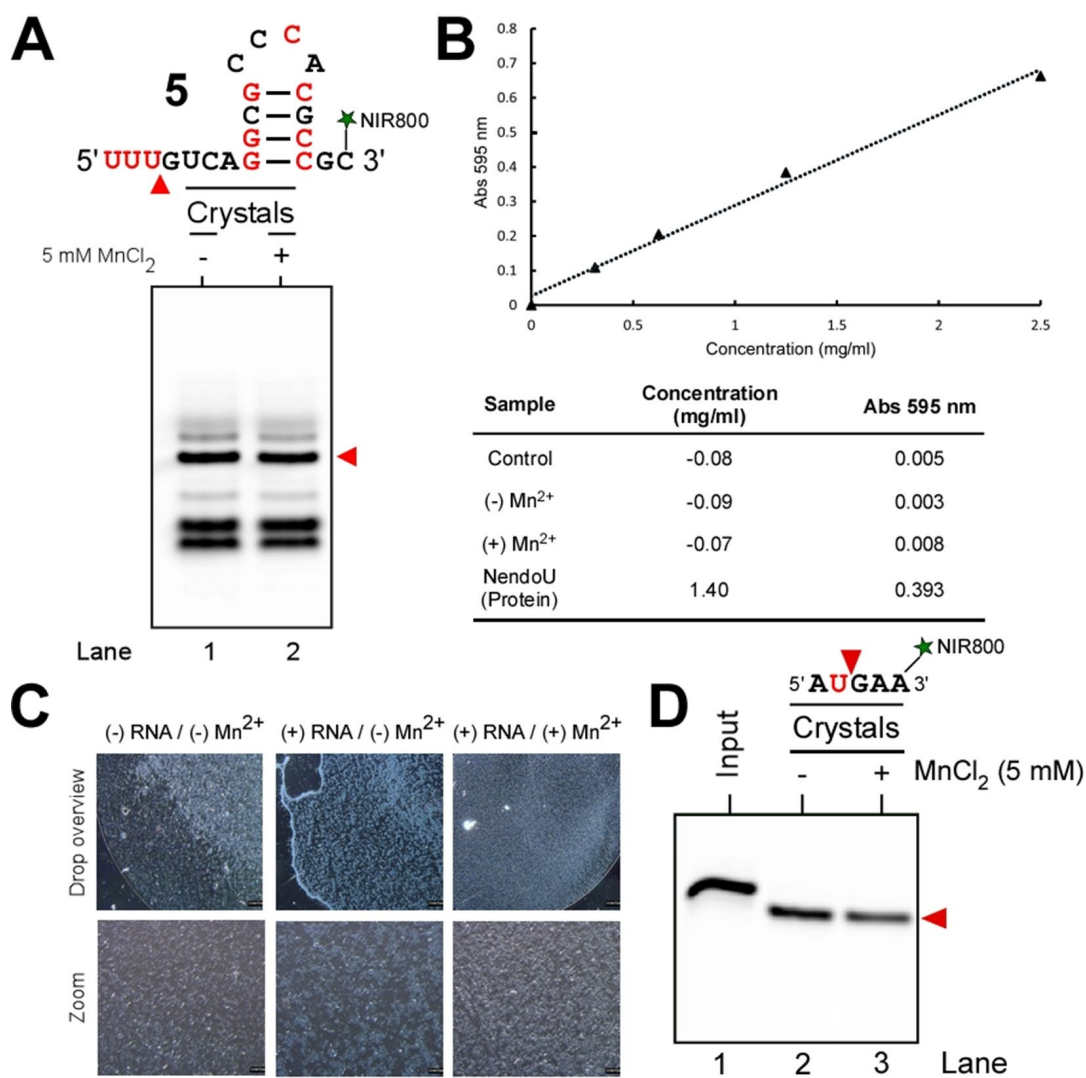
Figure 2.5. Determining catalytic activity of NendoU crystals and crystal stability post catalysis. The crystals grown by the batch method in the presence of citrate were used for this experiment **(A)** NendoU crystals endoribonuclease assays for 21 nt RNA (substrate #5). Secondary structure model shown above RNA gel with the major cleavage site shown with red arrowheads underneath the RNA structure and the gel. **(B)** Bradford assay of supernatant from the crystal enzymatic assay to test crystal dissolution. Bradford assay standard curve (top) with concentration determination of samples or controls shown (bottom). **(C)** Polarized microscopy images of crystals post in-crystals enzymatic cleavage reactions. Drop overview (top row) and zoom (bottom row) shown with corresponding samples labeled above each image column. **(D)** NendoU crystals endoribonuclease assays for 5 nt RNA substrate. Sequence shown above RNA gel with the cleavage site shown with red arrowheads underneath the RNA sequence and besides the gel.

absence of Mn^{2+} ions (**Figure 5D, lanes 2, 3**). Overall, these experiments suggest that the crystals are catalytically active and intact during and post catalysis.

Room Temperature serial fs crystallography data collection on different crystal forms of NendoU

The first room temperature structure of NendoU was determined by serial fs crystallography to 2.6 Å. Data was collected at the Macromolecular Femtosecond Crystallography (MFX) beamline at the LCLS. The WT of NendoU was crystallized in the presence of PEG and citrate by the batch method as well as an ultrafiltration procedure at low ionic strength described above.

Test of different injection system for serial fs crystallography data collection at LCLS data was collected at 30 fs nominal pulse duration with a repetition rate of 120Hz. The ultimate future goal of the XFEL studies of NendoU are time resolved SFX studies, with mix and inject systems to study the dynamics of the enzymatic cleavage of RNA by NendoU. We therefore used two sample injection systems that are designed for future mix and inject studies for our SFX data collection of NendoU crystal diffraction: Co-MESH and GDVN. The Co-MESH system is based on a concentric-flow electrokinetic injector, where jetting is achieved by application of a voltage to the system. The crystals in their mother liquor can be mixed with substrate through a simple T-intersection.

The Co-MESH system has the advantage that it can operate at low flow rates down to 3 ul/min and thereby conserves sample compared to the Gas dynamic virtual nozzle (GVDN) system (Weierstall, 2014; Weierstall et al., 2012) that had been developed by U Weierstall for SFX experiments and routinely uses flow rates of 15-25 µl/min. Another advantage of the co-Mesh system is that it generates a concentric jet that

is 30-50 um diameter in diameter and thereby increases the chances that the crystal is located in the interaction region. This could increase the chance of the crystals to interact with the X-ray beam and thereby increases the "hit rate" which is defined as the percentage of images collected that feature an X-ray crystal diffraction pattern.

However, the co-Mesh system also has challenges: it cannot operate in a Helium atmosphere because of the danger of voltage arching damaging the X-ray detector, therefore data must be collected in air with increased background scattering. The thicker jet also decreases the signal to noise ratio due to the higher background contribution, which is a disadvantage for data collection on very small crystals. Furthermore, the achievable mixing time points for future TR-SFX studies would be longer. Timepoints between 500 ms and multiple seconds could be achieved, therefore capturing only the longer time points of the catalytic cycle.

Data collection and analysis

Representative diffraction patterns from the NendoU protein crystals are shown

from the GDVN system in the Helium chamber and the MESH injector operating in air

Figure 2.6A Representative Diffraction pattern for NendoU-citrate crystals injected with the 3D printed GDVN system in Helium (left) and the co-MESH injector in air (right). The detector counts around a Bragg spot at a similar scattering angle for individual pixels are displayed in the red insert.
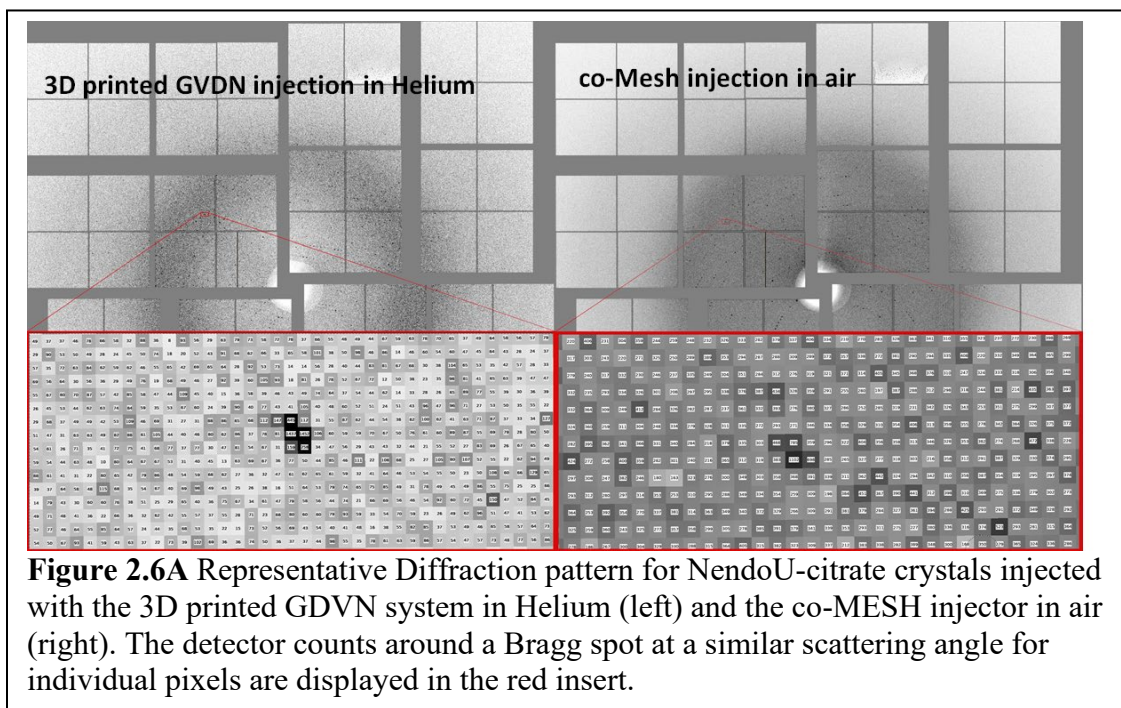
(**Figure 6A**). The background contribution from the 3D printed mix-and-inject injector

system (**Figure 6, left**) is strongly reduced compared to the MESH injection in air

(**Figure 6A, right**). While the maximal peak intensity of the crystal diffraction is similar

with maximal peak intensities of 1000 to 1500 counts for the highest intensity peak

pixels, the backgrounds vary. The average background count/pixel is only 30-50 counts

for the data collected in He with the 3D printed GVDN system, but the average

background pixel count is >300 for the data collected in air with the co-Mesh system.

The MESH injector was selected for subsequent data collection. Data collection

statistics for all datasets are presented in **Table 1**, along with the refinement statistics for

the citrate crystallization condition. Representative diffraction patterns for the different

crystallization conditions comprising the citrate condition, the low salt

(ultracentrifugation) condition and the washed condition are depicted in Figures 7B, C and D.
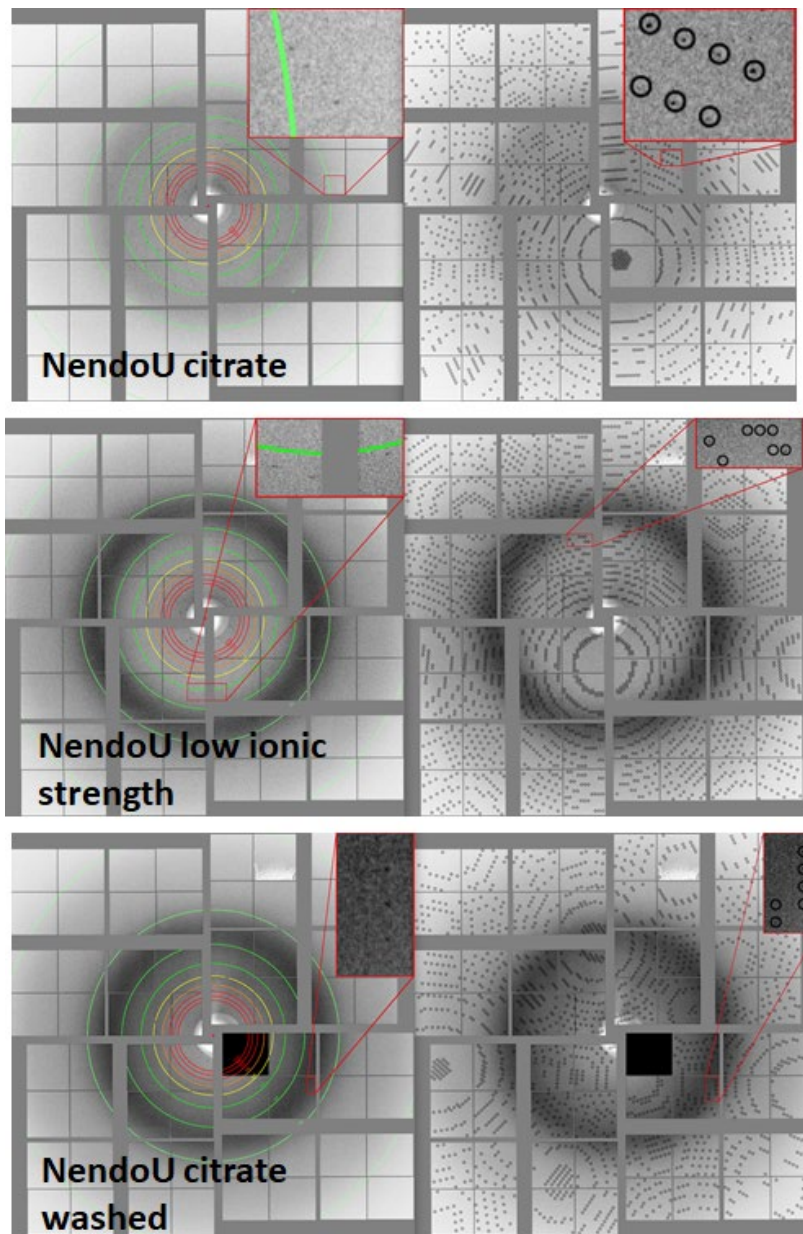
Figure 2.7. B, C, D Representative Diffraction patterns of different NendoU crystals. Representative diffraction patterns with resolution rings are shown on the left and the corresponding indexing solution is shown on the right. **(B)** crystals grown with PEG/citrate/MPD using the batch with agitation method **(C)** Crystals grown at low ionic strength using the ultrafiltation method **(D)** Crystals grown with PEG/citrate/MPD which were washed 5 times in washing buffer that contained PEG and MPD but no citrate.

The data statistic show the crystals grown in the presence of citrate and PEG

diffract to a significantly higher resolution, 2.6 Å, compared to the crystals grown at low

salt, for which maximal diffraction was observed to 4.4 Å. Therefore, data evaluation and structure determination were focused on the crystals grown by batch with agitation.

In preparation for future time-resolved studies where we would aim to mix apo NendoU crystals with substrate, we washed the NendoU crystals grown in the citrate condition 5 times with a citrate-free stabilization buffer before the sample was delivered to the XFEL beam. While we had feared that the washing could impair the diffraction quality of the crystals, the data statistics show that the diffraction quality was not affected by attempts to wash the crystals in a citrate-free buffer.

Room temperature structure of NendoU

Figure 8 A and B show the RT structure of NendoU determined at a resolution of 2.6 A based on the SFX data collected from NendoU crystals grown under the PEG/citrate conditions and delivered with the MESH injector system. NendoU forms a hexamer featuring a dimer of trimers (Fig 8B), which form a double ring consisting of one trimer of monomer A (bottom in Figure 8) and one trimer of monomer B (top in figure 8). One monomer of each trimer is in the asymmetric unit (Figure 8A).

NendoU 7K9P				
Data Collection				
	Citrate MESH 7K9P	Low Salt MESH (pH 7.5)	Citrate GDVN	Citrate MESH washed
X-ray Wavelength [Å]	1.277	1.277	1.277	1.277
Spacegroup	P63	P63	P63	P63
# indexed patterns	25 123	6 843	168	3 365
a,b,c (Å) **α, β, γ (°)**	153.4, 153.4, 116.8 90, 90, 120	153.6, 153.6, 117.9 90, 90, 120	153.8, 153.8, 117.2 90, 90, 120	153.5, 153.5, 116.9 90, 90, 120
Resolution (Å)	33.59 – 2.6 (2.69-2.6)	35.44 – 4.4 (4.49-4.4)	30.56 – 3.0 (3.16-3.0)	33.62 – 3.0 (3.11-3.0)
# unique Reflections	94 549 (9 443)	21 162 (2 645)	11 617 (900)	61 688 (6 204)
Redundancy	100 (70)	312 (236)	2 (2)	17 (12)
SNR	3.1 (0.2)	1.1 (0.2)	10.0 (4.4)	1.6 (0.2)
Completeness (%)	100 (100)	100 (100)	18.7 (10.2)	100 (100)

CC*	0.97 (0.23)	0.94 (0.20)	0.80 (0.55)	0.85 (0.27)
CC1/2	0.88 (0.03)	0.78 (0.02)	0.47 (0.18)	0.57 (0.04)
Structure Refinement				
R_{work}	0.191			
R_{free}	0.201			
Mean isotropic B-factor [Å^2]	47.7			
Wilson plot B factor [Å^2]	86.4			
# Atoms				
Protein	5504			
Solvent	16			
Heterogen	26			
R.m.s.d., bond lengths [Å]	0.006			
R.m.s.d., angles [°]	1.034			
Ramachandran plot [%]	94.94			
Favored	4.19			
Allowed	0.87			

Outliers		

Table 2.1: Data statistics for the SFX data sets collected on NendoU crystals grown under different crystallization conditions and delivered with different sample injection systems: **1st column**, data set of crystals were grown with the PEG/citrate crystallization buffer and delivered with the MESH injection system, this data set is complete and has been evaluated and fully refined to 2.6 A resolution. The Structure has been deposited in the pdb under accession code **7K9P**. Refinement was only attempted for this data set and statistics are given in the table. **2nd column**, data set collected from crystals grown under low salt conditions and delivered with the MESH injector, **3rd column**, data set collected from crystals grown in PEG/citrate and delivered with the MIX GVDN injection system, **4th column** data set of the crystals grown in PEG/citrate which underwent a 5 step washing procedure with stabilization buffer without citrate, which were also delivered with the MESH injector.

The structure determined at room temperature reflects the native state of NendoU, where dynamics of the molecule are not frozen. In Figure 8 RIGTH we compare the RT-SFX and the cryogenic structure determined by Kim et al (Y. Kim et al., 2020) (both with

citrate bound to the active site). Our room temperature structure is slightly more

extended compared to the standard crystallography structure determined under cryogenic
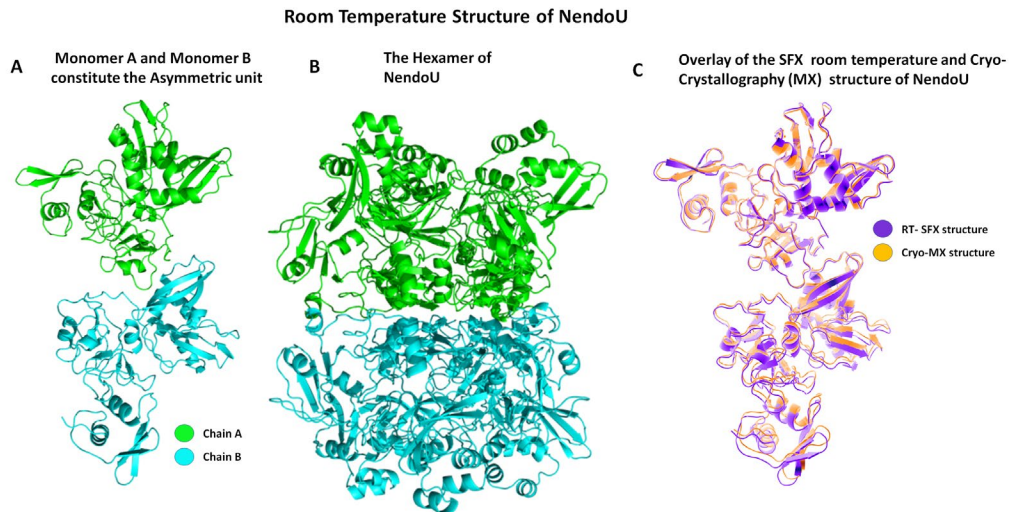
Figure 2.8. Room temperature SFX structure of NendoU **(A)** structure of the two monomers in the asymmetric unit, monomer A (chain A) is depicted in blue and monomer B (chain B) is depicted in green; **(B)** side view of the NendoU hexamer, which is arranged in the form of a dimer of trimers, which form a double ring structure . The bottom ring consists of the trimer of monomers A (blue) and the top ring of the trimer of monomers B (green). (C) Overlay of the RT SFX structure of NendoU (this work pdb entry 7K9P in violet) with the structure determined by Kim et al (Y. Kim et al., 2020) by standard crystallography under cryogenic conditions (PDB entry 6W01 in orange). Please note that a clear relative shift of the monomers is evident.

conditions, which is also reflected in the slightly larger unit cell of the SFX room

temperature structure. This effect is not unexpected and has also been seen by

comparison of other structures where both SFX and cryo-X-ray structures have been

determined (W. Liu et al., 2013). More interesting is that the monomers are shifted

relative to one another. So the room temperature structure features a shift of the

orientation of monomer B versus monomer A. This indicates that there is some flexibility

at room temperature between the two trimers in the hexamer. To investigate differences

in the flexibility we compared the B-factors of the two structures.

The structure determined at room temperature reflects the native state of NendoU, where dynamics of the molecule are not "frozen". We will now compare the RT-SFX and the cryo-MX (both with citrate bound) concerning their B-factors, which gives indications of the flexibility of the protein.

To illustrate these differences, we show the comparison of the structures with the B-factor analysis of the two NendoU molecules in the asymmetric unit in **Figure 9**. This
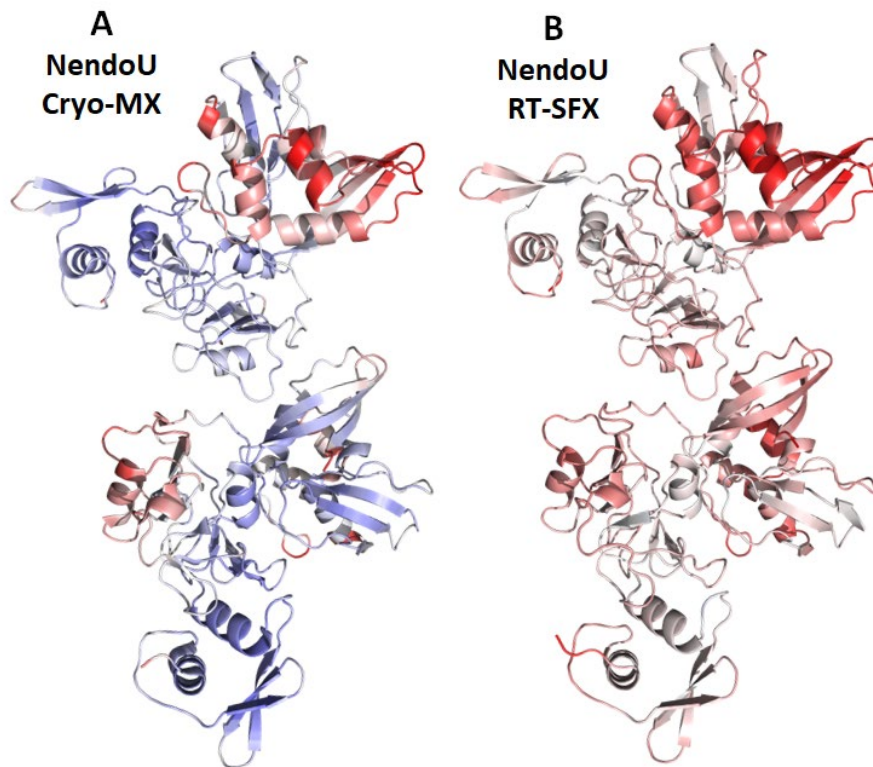
Figure 2.9. Comparison of the B factors of the RT-SFX and Cryo-MX structures of NendoU. Cartoon representation of the protein, colored with the lowest B-factors (i.e. least flexible domains) depicted in blue and highest B-factors (i.e. highest flexibility domains) depicted in red: **LEFT**: structure of NendoU from Kim et al (pdb 6W01) determined by standard macromolecular crystallography under cryogenic condition (Cryo-MX) (Y. Kim et al., 2020), RIGTH: RT SFX structure of NendoU from this work (PDB: 7K9P).

figure features the cryo-MX structure from Kim et al (Y. Kim et al., 2020)(pdb 6W01) on the left and the RT-SFX structure determined in this work on the right (pdb 7k9P).
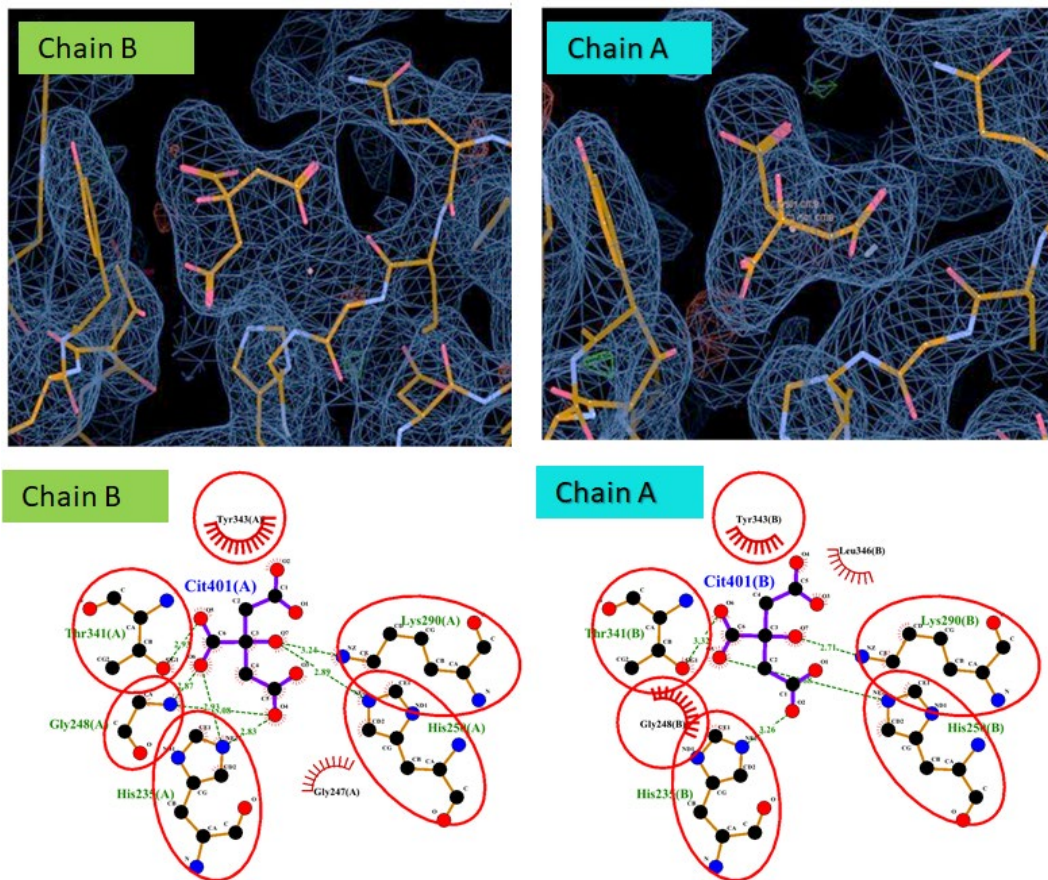
Figure 2.10. Differences in Binding sites of citrate in the RT-SFX structure **TOP** Comparison of electron density (2Fo-Fc map) of citrate in the binding sites of the more rigid molecules A and the more flexible NendoU molecule B. Please note that the electron density for citrate is better defined in the rigid molecule A than in the more flexible molecule B. **BOTTOM:** Ligand plot analyses of the citrate molecules in both binding pockets. Residues that are in equivalent 3D positions when the structural models are superposed are circled in red. **Please note that the** orientation and hydrogen bonding network is different for citrate in chain A and B, with a closer hydrogen bonding network of citrate identified in the low flexibility molecule A compared to the hydrogen bonding network of citrate in the binding site in the NendoU molecule B.

The first imminent difference is that the RT-SFX structure shows overall higher

B-factors than the cryo-MX structure of NendoU. This is expected as molecular motions

are frozen under cryogenic conditions while the molecules have higher flexibility at room temperature, thereby presenting the more native state of NendoU.

The next striking feature is that the NendoU molecules in both the cryo-MX and RT-SFX structure feature a dominant difference in the flexibility of the structures of chain A and B, where the NendoU molecules in the lower ring (molecules A) are much more rigid than in the upper ring (molecules B). The RT-SFX structures shows that the chain B is much more flexible than chain A, which also explains why it has shifted its orientation in the RT structure compared to the structure determined under cryogenic conditions.

Differences in the citrate Binding sites between the two chains of NendoU

Now we will compare the binding of citrate in the two binding sites in the RT-SFX structure of NendoU. **Figure 10** shows the comparison of citrate located in the binding sites of the more rigid NendoU chain A and the flexible Chain B.

The top section shows the comparison of the electron density maps in the two binding sites. Here it is clearly evident that the electron density for citrate in the binding site of the more rigid NendoU molecule A is better defined than the electron density for citrate in the more flexible binding site of monomer B. Furthermore, the binding conformation and hydrogen bonding network are different between the two binding sites, where the less flexible binding site A shows a more tightly connected hydrogen bonding network compared to binding site B as shown in the ligand binding analysis shown in Figure 11 bottom. To confirm the results, we have also calculated omit and POLDER maps that are shown in **Supplementary Figures 3 and 4**.

Citrate can be removed from the flexible Binding site B

Non-refined electron density of the catalytic sites of NendoU crystals after washing
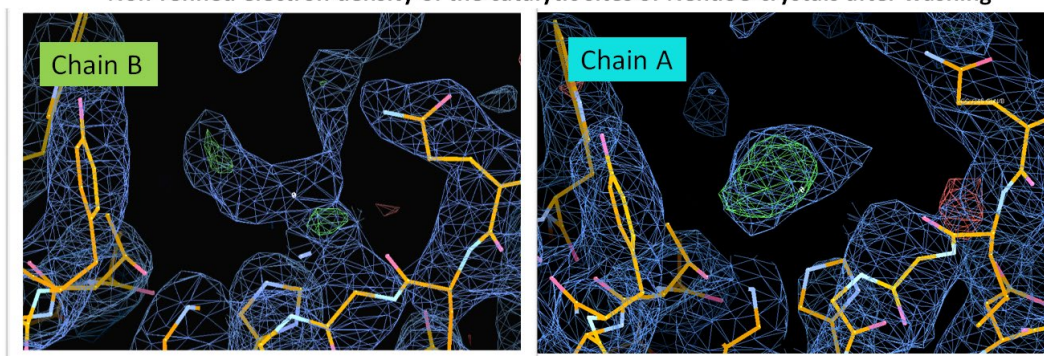
Figure 2.11: Non-refined electron density maps for the washed NendoU crystals in the catalytic sites of NendoU chains A and B. LEFT The electron density in the more flexible B-site shows multiple smaller electron density "blobs" indicating that they may not come from a single larger ligand molecule, but could instead represent a network of water molecules. RIGTH: The single large electron density "Blob" resembles unrefined citrate. All maps are 2Fo-Fc maps contoured at 1.0 σ.

For future time-resolved studies it is very important to know if citrate can be removed from the catalytic site post-crystallization. We implemented a washing procedure where crystals were sedimented by low speed centrifugation and washed five times with citrate-free washing buffer that contained PEG and MPD. The crystals survived the washing steps and data were collected from these washed crystals. The data statistics is shown in Table 2. Unfortunately data collection was restricted to 20 minutes at the end of our last shift at LCLS. Therefore this data set consist only of 3365 indexed images with a multiplicity of 17 overall and 12 for the highest resolution shell, which is borderline for the refinement of an SFX XFEL structure. We have therefore not refined this structure and show the non-refined electron density maps for the catalytic sites in Figure 12.

While the less flexible catalytic site A features a large electron density "blob" that resembles the density of an unrefined citrate in the catalytic site the unrefined electron density in the catalytic site of the more flexible B site features several smaller

electron density "blobs" that could represent a network of water molecules. While this result is preliminary it may indicate that the two catalytic binding sites have different affinity for citrate and that citrate can be removed from the more flexible B binding sites, which is very interesting for future time-resolved studies on NendoU.

2.5 Discussion

One important requirement for future time resolved studies is that the substrates will be able to diffuse into the crystal and that the catalytic sites are accessible to the substrates via solvent channels. We therefore analyzed the crystal packing and the size of the solvent channels in the NendoU crystals as shown in **Figure 12.**
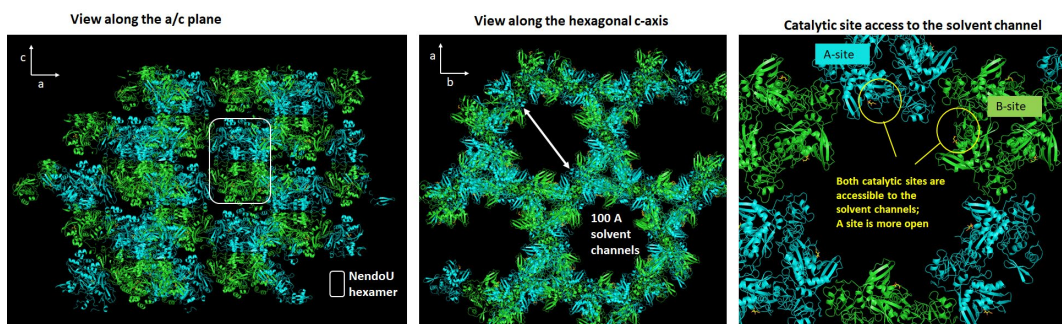
Figure 2.12. Packing of NendoU and solvent channels in the crystals **LEFT**: view along the a/c axis with outline of the location of the hexamer **MIDDLE**: view along the hexagonal c-axis visualizes the large 100 A wide solvent channels along the c-axis of the crystals **RIGTH**: zoom in to the two catalytic sites. Both sites are highly accessible to the large solvent channel. Note that catalytic site A is fully open to the solvent channel while site B is more secluded.

The NendoU crystals contain a wide solvent channel of 100A that will allow even very large substrates to diffuse into the crystals. Even more exciting is the fact that both catalytic sites in the A and B chain are accessible to the solvent channel. This explains why our crystals are highly active in cleaving the large RNA substrates of 5 and 21 nucleotides mimicking the 5' end of the SARS-CoV-2 negative strand RNA. **We therefore conclude that the crystal packing with large solvent channels along the C-**

axis provides full access to the catalytic sites, ideal for future time resolved studies with the final goal of determining a molecular "movie" of the catalytic reaction of NendoU.

Another interesting finding is that the A-site is wide open to the solvent channel, while the B-site is more secluded. It was therefore initially unexpected when we discovered that we can potentially remove the citrate from the more secluded B site by washing of the crystals, while citrate stays tightly bound in the A site, which is wide open to the solvent channel.

The B-factor analysis provides a potential explanation: the B site is much more flexible than the A site and binds citrate less strongly, therefore potentially allowing for an exchange of the citrate by water molecules in the binding site. This finding also has a great impact on the relevance of our findings as it leads to the conclusion that the higher flexibility of the B trimer and its catalytic site compared to the A site is not caused by restrictions in crystal contacts but that this higher flexibility is an intrinsic feature of the NendoU hexamer.

To investigate this point further it is important to compare our structure with the recently determined cryo-EM structure of NendoU with bound RNA (Pillon et al., 2021), (M. N. Frazier et al., 2021). One of the most coveted structures yet to be revealed is NendoU in complex with long RNAs. There is an apparent threshold at three nucleotides, with structures published (M. N. Frazier et al., 2021) that identified densities for 2 of the

Figure 2.13: Flexibility analysis of the RT-SFX trimer of NendoU and comparison to the cryo-EM structure with nucleotides bound in the P1 ,P2 and P3 monomers in the trimer. **(A)** Surface model of the TR-SFX structure of the NendoU Hexamer colored by B-factors (blue, low B factor; pink, medium B-factors; red, high B-factors). **(B)** Backbone depiction of the RT-SFX trimer with view from the top, where the catalytic sites show high B-factors (red arrows). **(C)** View from the bottom, where the catalytic sites show medium B-factors. **(D?)** Cryo-EM structure from Frazier et al (Frazier et al., 2021) with the view from the bottom on the trimer which has nucleotides bound in the 3 catalytic sites P1, P2 and P3.

3 nucleotides in two structures they refer to as pre-and post-cleavage structures. Interestingly only one of the two trimers in the cryo-EM structure contained well defined electron density for the nucleotides in the binding site. While there was some density in the catalytic sites of the other trimer, it could not be positively identified as nucleotides in the second trimer. They therefore applied C3 symmetry and built their structural model with nucleotides only in one of the two trimers. This is a very interesting result of the cryo-EM structure and combined with our results shows that different flexibilities for the

two timers in the NendoU hexamer is an intrinsic and very likely functionally important feature of NendoU. This poses two major questions: Why would only one of the two trimers bind and cleave substrate and why is the hexamer required for catalytic activity?

We want to put forward the new hypothesis that NendoU may catalyze cleavage of the coronaviral negative strand RNA by a binding change mechanism. In this model shown in Figure 14 the affinity of the substrate would alternate between the timers in the hexamer. In this model, substrate would be bound only to the high affinity trimer that would corresponds to the trimer A in our structure, which in this model would correspond to the same trimer that contains a 2 nucleotide RNA fragment in the binding site of the pre and post cleavage cryo-EM structures of Frazier et al. (M. N. Frazier et al., 2021). The B site trimers feature a high flexibility and therefore do not or only weakly bind the substrate, which would explain why the catalytic site is empty in the second trimer in the cryo-EM structure. The cleavage of the RNA in trimer A would provide the energy for the triggering of a conformational change where the catalytic sites in trimer A would become weakly binding, thereby enabling substrate release. Simultaneously the catalytic sites in Trimer B would change their conformation and tightly bind the substrate and catalyze the next round of RNA cleavage. Thereby NendoU would follow a binding change mechanism that is similar to the process of ATP synthesis by the ATP-synthase, which had been proposed by Paul Boyer a long time before any structural information was available from this enzyme (Boyer, 1998). Decades later he received the Nobel Prize in Chemistry for the binding change mechanism.

This functional binding change model would explain the fact that the hexamer is essential for the function, the fact why only one trimer contains RNA in the cryo-EM structure and also explain why we observe strong intrinsic differences in flexibility
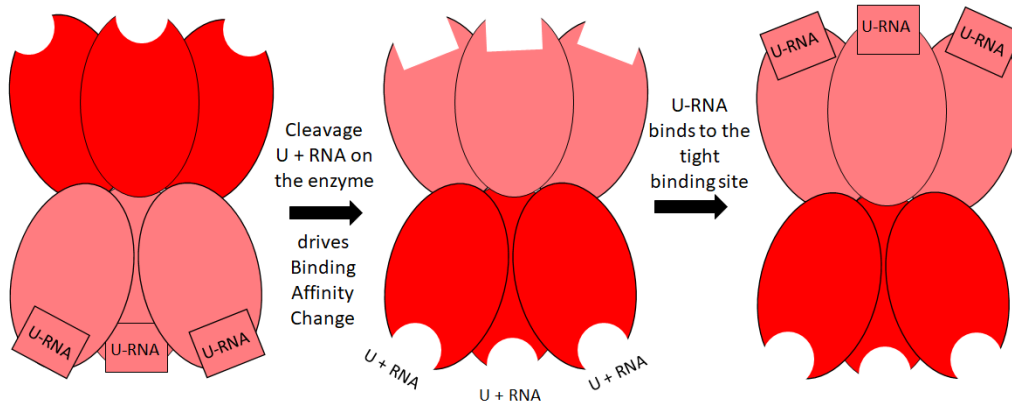
Figure 2.14. Scheme of the binding change mechanism for the function of U-RNA hydrolysis by hexameric NendoU **(**The hexamer is depicted in a side view) **LEFT**: the bottom trimer with the low flexibility (pink) has the U-RNA substrate bound in its tight binding sites. The binding sites of the highly flexible trimer (red) have a lower binding affinity and are empty **MIDDLE:** Cleavage of U-RNA into U + RNA provides the energy for a change in binding affinity, where the bottom trimers become highly flexible (red) lowering the binding affinity leading to the of the products U and RNA. The top trimer becomes less flexible (pink) and its substrate binding site changes to high affinity **RIGTH:** U-RNA binds to the high affinity binding sites of the top trimer initiating the next round of the cycle.

between the two trimers and why we can remove the citrate from the flexible binding site but not from the tightly bound site. While the proposal of a binding change mechanism for NendoU is a pure hypothesis at this time, it will be very exciting to see if it could be confirmed by future time-resolved studies on NendoU,

Conclusions and Outlook

The SARS-CoV-2 endoribonuclease, NendoU, has become a focal point of structural studies in COVID-19 research for its role in viral evasion of the host immune system and applications in antiviral drug development. Structural studies demonstrate the

highly conserved nature between SARS-CoV, MERS-CoV, and SARS-CoV-2 and congruent results between techniques. This work describes the first room temperature structure determination of NendoU by serial fs crystallography at an XFEL, which paves the way for future time resolved SFX studies, which could discover the mode of substrate binding, the catalytic mechanism of cleavage and the substrate release. We have established both in vivo and in vitro expression systems for NendoU and established large scale isolation and purification procedures for NendoU which will be suitable for future time-resolved studies, with notoriously large sample consumption.

Two different mix and inject systems have been successfully established that can be used for future time resolved studies. Based on critical 2D structure prediction of the negative strand RNA of NendoU from SARS-COV-2, we have designed 5 and 21 nucleotide substrates that mimic the 5'end of the RNA before the poly-U header. For both substrates we established functional assays that confirmed that NendoU fast cleaves these large RNA substrates both in solution and in assembled into crystal. The design and functional studies of multiple length RNAs coupled with optimized microcrystallization conditions at an XFEL light source set the foundation for time resolve studies of NendoU. Attempting time resolved studies of NendoU could reveal critical insight into the docking, cleavage, and substrate release performed by NendoU. Any structure of NendoU in complex with larger RNAs could begin to address questions regarding allosteric interactions between longer RNA sequences with areas beyond the active site. Additionally, it may unravel the role of manganese in coordinating the RNA – explaining why Mn has yet to be seen in any structure despite being a manganese-dependent reaction.

TR-SFX experiments may be critical for unlocking these structures for multiple reasons. Firstly, one hypothesized reason for the difficulty in obtaining these elusive structures is the fragile nature of longer RNAs that are subject to degradation. Incubation of NendoU with RNAs allows time for the slightest RNaseA contaminant to degrade the RNA prior to data collection. This may explain why the shorter incubations used in preparing CryoEM samples (1hr) have been more productive than attempts at standard cryogenic crystallography that can require hours or even days depending on the preparation method.

By using a mix-and-inject method, a sterile preparation can be used to house an RNA sample without exposure to other components risking contamination. The short mixing times by the 3D printed nozzle immediately prior to diffraction would circumvent this possible issue. The nature of SFX experiments allows for diffraction of nanocrystals and small microcrystals, which is advantageous for fast diffusion times of the RNA into the crystals.

The first RT structure of NendoU shows strongly different flexibility of the two trimers in the hexamer, with one tight and one more flexible catalytic site. This heterogeneity in the binding affinity is also observed in the cryo-EM structure of NendoU with RNA bound (M. N. Frazier et al., 2021), indicating that the difference in flexibility and binding affinity may be an intrinsic and functionally important feature of NendoU. Based on these findings we present the hypothesis that NendoU may feature a binding change mechanism, however this idea has to be experimentally explored in future studies. We propose to approach this discovery challenge with time-resolve (TR) mix-and-inject serial femtosecond crystallography (SFX) at an X-ray Free Electron Laser (XFEL) light

source. Proposed TR-SFX studies hold the potential to directly observe the NendoU mechanism; however, these studies are also intensely intricate in nature and require extensive preliminary results and preparation. Such preparatory results from LCLS experiments to develop XFEL optimized conditions are presented in this study, culminating in the first room temperature structure of NendoU by SFX. Therefore, this study paves the way for future time resolved serial fs crystallography studies to investigate the mechanism of RNA cleavage by NendoU. It will be very exciting to explore the hypothesis that NendoU may cleave RNA by an allosteric binding change mechanism.

2.6 Supplemental Information

2.6.1 Supplemental Figures
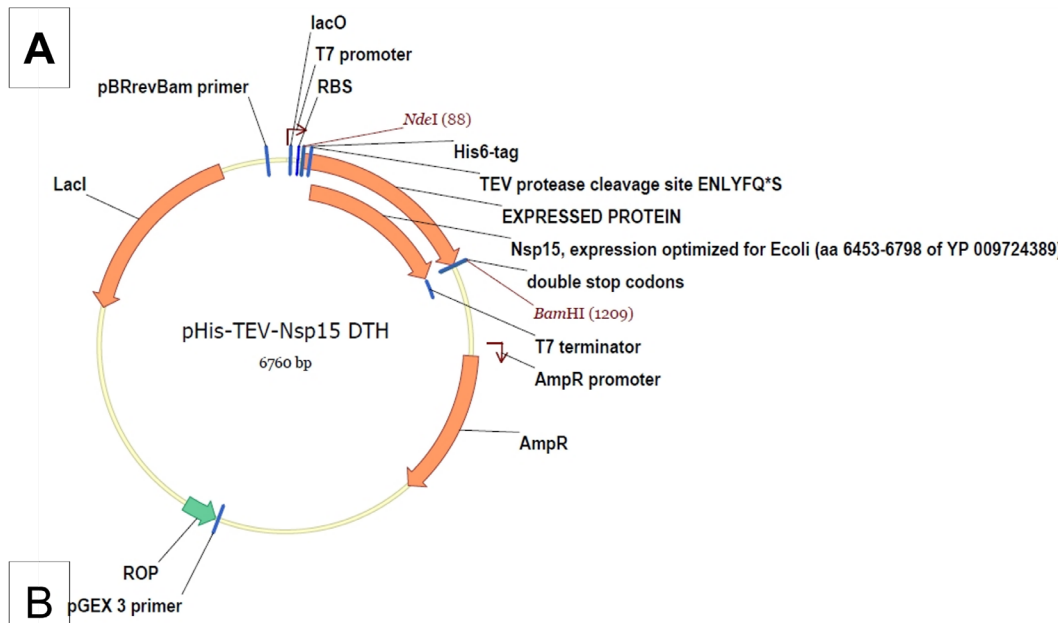

```
LOCUS           Translation\of\pHis-TEV-Nsp15           371 aa          7-MAY-2020
DEFINITION      Translation of a fragment of pHis-TEV-Nsp15.
KEYWORDS        TRANSLATED.
SOURCE
  ORGANISM
FEATURES                  Location/Qualifiers
     Region               26..371
                          /vntifkey="1000"
                          /label=Nsp15\(aa\6453-6798\of\YP_009724389)
     Site                 2..7
                          /vntifkey="264"
                          /label=His6-tag
     Site                 16..22
                          /vntifkey="221"
                          /label=TEV\protease\cleavage\site\ENLYFQ*S
     Region               22..25
                          /vntifkey="1003"
                          /label=extra\N-terminal\residues\following\TEV\cleavage
     Region               1..371
                          /vntifkey="266"
                          /label=complete\protein\sequence\from\PDB\6VWW
ORIGIN
        1 mhhhhhhssg vdlgtenlyf qsnamslenv afnvvnkghf dgqqgevpvs iinntvytkv
       61 dgvdvelfen kttlpvnvaf elwakrnikp vpevkilnnl gvdiaantvi wdykrdapah
      121 istigvcsmt diakkpteti capltvffdg rvdgqvdlfr narngvlite gsvkglqpsv
      181 gpkqaslngv tligeavktq fnyykkvdgv vqqlpetyft qsrnlqefkp rsqmeidfle
      241 lamdefiery klegyafehi vygdfshsql gglhlligla krfkespfel edfipmdstv
      301 knyfitdaqt gsskcvcsvi dlllddfvei iksqdlsvvs kvvkvtidyt eisfmlwckd
      361 ghvetfypkl q
//
```

C

```
LOCUS       pHis-TEV-Nsp15        6760 bp    DNA     circular    7-MAY-2020
SOURCE
  ORGANISM
COMMENT      Protein sequence is identical to 6VWW FASTA file.
FEATURES             Location/Qualifiers
     promoter        1670..1698
                     /vntifkey="30"
                     /label=AmpR\promoter
     CDS             1740..2597
                     /vntifkey="4"
                     /label=AmpR
     primer          3773..3795
                     /vntifkey="27"
                     /label=pGEX_3_primer
     misc_feature    complement(3789..3980)
                     /vntifkey="21"
                     /label=ROP
     CDS             complement(5295..6383)
                     /vntifkey="4"
                     /label=LacI
     primer          6658..6677
                     /vntifkey="27"
                     /label=pBRrevBam_primer
     promoter        1..19
                     /vntifkey="30"
                     /label=T7\promoter
     misc_feature    19..46
                     /vntifkey="21"
                     /label=lacO
     RBS             64..80
                     /vntifkey="32"
                     /label=RBS
     terminator      1216..1344
                     /vntifkey="43"
                     /label=T7\terminator
     misc_feature    1202..1207
                     /vntifkey="21"
                     /label=double\stop\codons
     CDS             89..1201
                     /vntifkey="4"
                     /label=EXPRESSED\PROTEIN
     CDS             164..1201
                     /vntifkey="4"
                     /label=Nsp15,\expression\optimized\for\Ecoli\(aa\6453-6790\of\YP_009724389)
     misc_feature    92..109
                     /vntifkey="21"
                     /label=His6-tag
     misc_feature    134..154
                     /vntifkey="21"
                     /label=TEV\protease\cleavage\site\ENLYFQ*S
BASE COUNT     1611 a     1758 c     1854 g     1537 t
ORIGIN
```

Supplemental Figure 2.1. NendoU (Nsp15) construct design **(A)** Plasmid map of pHis-TEV-Nsp15 based on PDB: 6VWW (Y. Kim et al., 2020). **(B)** Amino acid sequence of pHis-TEV-Nsp15 **(C)** Nucleic acid sequence of pHis-TEV-Nsp15
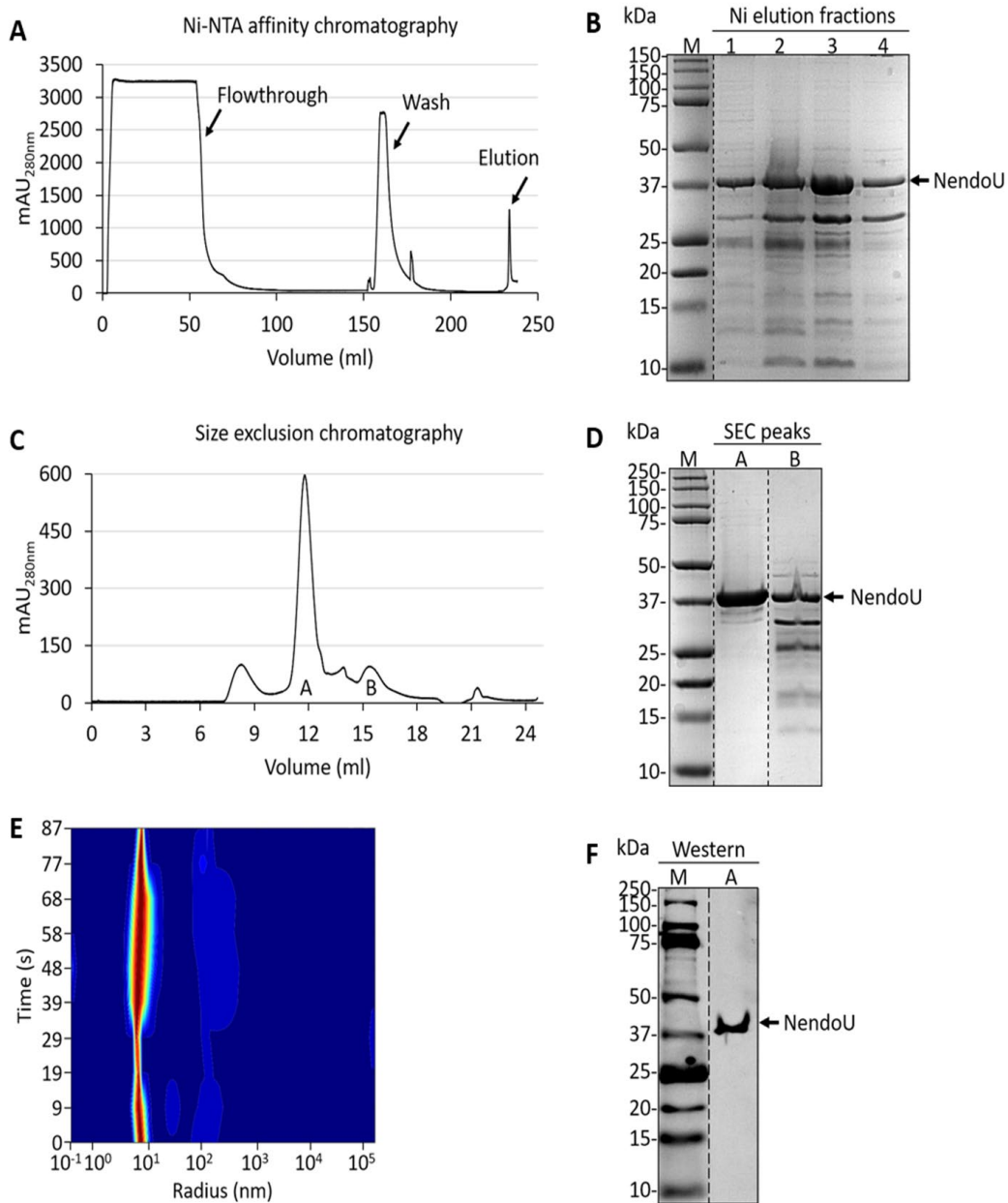
Supplemental Figure 2.2. Isolation, purification and biophysical characterization of the isolated NendoU protein. **(A)** Ni-NTA affinity chromatogram **(B)** SDS gel-electrophoresis of the elution fractions. The gel was stained with Coomassie staining sensitive to 5ng of protein **(C)** size exclusion chromatogram **(D)** SDS gel-electrophoresis of the SEC fractions, **(E)** dynamic light scattering of the purified NendoU shows a monodisperse solution with uniform size distribution corresponding to a hexameric assembly **(F)** anti-His immunoblot of NendoU following SEC purification.

CHAPTER 3

SARS-COV-2 PROTEIN, NENDOU, AS A NON-STANDARD MODEL PROTEIN

FOR MACROMOLECULAR CRYSTALLOGRAPHY

Text and figures in this chapter are printed with permission from co-authors: Dhenu Logeswaran, Nirupa Nagaratnam, Jose Martin-Garcia, Sabine Botha, Thomas Grant, Gihan Ketawala, Michelle Sheikh, Megan Shelby, Mimi Cho Yung, Matt Coleman, Emily K. Kaschner, Matthew R. Goode, Michele Zacks, Debra T. Hansen, Raimund Fromme, Matthias Frank, Julian Chen and Petra Fromme

3.1 Abstract

Model proteins are frequently needed to demonstrate technology and method development in crystallography, prior to wide-spread application and distribution. In doing so, model proteins that more closely resemble user proteins are often desirable to demonstrate feasibility and application. The NendoU protein from SARS-CoV-2 is an exciting target for its role in evasion of the innate immune system, but also fulfills the needs for many different needs in crystallography. The NendoU H234A mutant expresses in large quantities making it suitable for crystallography experiments, such as serial crystallography, which require high sample consumption. Additionally, crystals of NendoU can be grown using different methods like vapor diffusion, batch, batch with agitation, and dialysis and can be adjusted for different size distributions. As an enzyme, NendoU can also be studied in respect to substrate and inhibitor binding experiments for time resolved experiments. This fits the needs of a wide range of testing new technology and devices that require a model protein but wish to use something less common than the

67

standard proteins like lysozyme, which are too easy to crystallize and handle to serve as a good model system.

3.2 Introduction

Model proteins are key for further developments in areas from new methods to novel technologies and education. These proteins are often known for their robust nature and usefulness. Lysozyme is the most classic example of a model protein as it is easy to obtain as it can be purchased and is cheap, it is crystallizable in many conditions and by different methods, it is resilient to denaturation, and it is easy to grow well-diffracting crystals that diffract to high resolution. But as methods advance, "intermediate" model proteins are often desirable to prove methods because what works for a protein like lysozyme may not translate well to other more difficult to isolate and crystallize proteins. These non-standard model proteins are needed as a closer match to the sensitivity of other proteins, but still maintain versatility for use. In this paper, we propose NendoU from SARS-CoV-2 as a non-standard model protein for macromolecular crystallography.

The emphasis of SARS-CoV-2 research yielded many impressive findings towards the study of disease. NendoU is the uridine-specific nidoviral endoribonuclease expressed by the *nsp15* gene of the SARS-CoV-2 virus. Functionally, NendoU assists the SARS-CoV-2 virus in evasion of the innate immune system. In research conducted to determine its catalytic mechanism and the exploration as a potential drug target, NendoU has also demonstrated notable characteristics for crystallographic studies.

NendoU is not an easy target protein. It was first noted that it took five days to develop a crystallization pipeline for the NendoU apoprotein, including expression, purification, crystallization, data analysis, and structural deposition (Youngchang Kim et

al., 2020). Since the beginning of the pandemic, there are 40 crystal structures deposited in the Protein Data Bank (PDB) for NendoU which include mostly cryo-MX, but also one room temperature SFX crystal structure; large, single crystals and serial microcrystals; apoprotein and ligand bound structures; and finally, the first nucleotide bound structures that serve as substrates have been determined and intermediates proposed. In this paper, we explore the many different crystallization conditions and uses of NendoU.

3.3 Methods

Protein Expression

Expression of wildtype NendoU was optimized from prior methodology (Makowska-Grzyska et al., 2014 & Kim et al., 2020). The plasmid for Nsp15-TEV-His$_6$ (plasmid pHis-TEV-Nsp15; Chapter 2) was obtained from GenScript and transformed into commercial competent *Escherichia coli* BL21-Gold (DE3) cells from Agilent Technologies. Starter cultures were prepared by inoculation of a single colony into a 50 mL LB Lennox (10 g Tryptone, 5 g yeast, 5 g NaCl) media (+50 µg/mL carbenicillin) and grown at 37 °C overnight with 180 RPM shaking. Small cell cultures of 100 mL were inoculated with 1 % of an overnight starter culture and grown at 37 °C with 180 RPM shaking. Three types of media were tested: LB Lennox, LB Miller, and TB media under the same antibiotic conditions. The cultures were grown to an OD$_{600}$ of ~0.95 – 1.0 and the expression of NendoU in the cells was induced under the following conditions: 0.2 mM and 0.4 mM IPTG, 0 and 0.1% glucose. Once induced, the cultures were grown at 18 °C overnight. Whole cell contents were analyzed by SDS-PAGE gel with Instantblue™ Coomassie stain and anti-His western analysis. Large scale expression was performed under the same conditions, scaled to 1 L cell cultures using LB Miller media

with 50 µg/mL carbenicillin antibiotic. Cells were induced with 0.4 mM IPTG and 0.1% glucose. Cells were harvested by centrifugation at 7,000 xg for 15 min at 4 °C.

An additional optimization was performed using the small-scale expression using the Kim, et al. 2020 expression condition of LB Lennox, but with reducing agents in the induction step. The positive control of induction based on the Kim, et al., 2020 conditions was performed with 0.1 % glucose, 40 mM K_2HPO_4, and 0.2 mM IPTG. Additional reducing agents were tested in combination with the standard induction conditions: 12 mM dithiothreitol (DTT) and 50 mM reduced glutathione (GSH). A control with no inducing agents was also performed. The 12 mM DTT reducing condition was scaled up to a one liter growth for expression comparison.

Protein Purification

Purification of the NendoU protein was performed based on the Kim et al., 2020 protocols with modifications and were identical to the previously described methods in Chapter 2 (Youngchang Kim et al., 2020).

Batch – microcrystals <20 µm (LCLS Beamtime P182)

Batch crystallization was performed by rapidly mixing the protein with precipitant for a total 1:6 protein to precipitant ratio; however, the precipitant is added in 12 separate steps. Protein (50 µL of 40 mg/mL) was placed on the bottom of an Eppendorf 1.5 mL reaction vessel (tube). Precipitant (0.2 M calcium acetate, 0.1 M HEPES pH 7.5, 10 %(w/v) PEG 8000 or 16%(w/v) PEG 4000, 100 mM Tris-HCl pH 8.5, 200 mM sodium acetate) based on (Y. Kim et al., 2021) was added 25 µL at a time with rapid mixing with 15 seconds in between each step. The crystallization solution was incubated at 4 °C overnight. Mixing all at once will still result in a dense crystallization pellet.

Batch – microcrystals 20 – 40 μm

NendoU protein at 10 mg/mL in size exclusion buffer (defined in chapter 2) was diluted to 5 mg/mL by addition of a 20 mM HEPES pH 7.5 solution. The protein was mixed in a 1:1 ratio with precipitant (0.2 M Calcium Acetate, 0.1 M HEPES pH 7.5, 7.5% PEG 8000).

Low Ionic Strength (Desalting)

A desalting column was equilibrated with 20 mM HEPES, pH 7.5, 20 mM NaCl, and 20 mM MnCl$_2$ on an AKTA FPLC. The NendoU protein in size exclusion buffer was loaded onto the column, desalted into the equilibration buffer, and placed on ice where it becomes turbid. Pipette homogenization of solution increases the turbidity.

Vapor Diffusion Crystallization

Precipitant (0.2 M Calcium Acetate, 0.1 M HEPES pH 7.5, 7.25 - 7.5 % (w/v) PEG 8000) with microcrystalline seeds is prepared by adding 5 μL of the supernatant of mature batch microcrystals (<20 μm) per 100 μL of precipitant. NendoU protein (10 mg/mL) was mixed 1:1 (2 μL + 2 μL) with the precipitant with seeds and crystals are grown by hanging drop vapor diffusion in 18 mm 24-well plates with 0.5 mL reservoir volume.

Additional crystallization conditions were established to grow large (100+ μm) crystals under the 16% (w/v) PEG 4000, 100 mM Tris-HCl pH 8.5, 200 mM sodium acetate crystallization condition described in (Y. Kim et al., 2021). Co-crystallization with 6-azauridine was performed by incubating the protein with 50 mM of the compound for 30 min prior to crystallization.

Fixed Target Crystallization

The batch microcrystal (<20 µm) protocol was used with the PEG 8000

concentration increased to 7.5 – 7.75 % (w/v). Once the crystallization solution was

prepared (prior to incubation), 10 µL was dispensed into cyclo-olefin-copolymer (COC)

microfluidic chips provided by the Kuhl research group from UC Davis (Gilbile et al.,

2021).

3.4 Results

NendoU wildtype (WT) expresses in low yields in vitro with only 1 – 3 mg of

protein (Chapter 2). Expression of NendoU WT was optimized from the original growth

conditions in LB Lennox media with 0.2 mM IPTG/0.1% glucose (Youngchang Kim et

al., 2020). This expression was optimized for media type (LB Lennox, LB Miller, TB
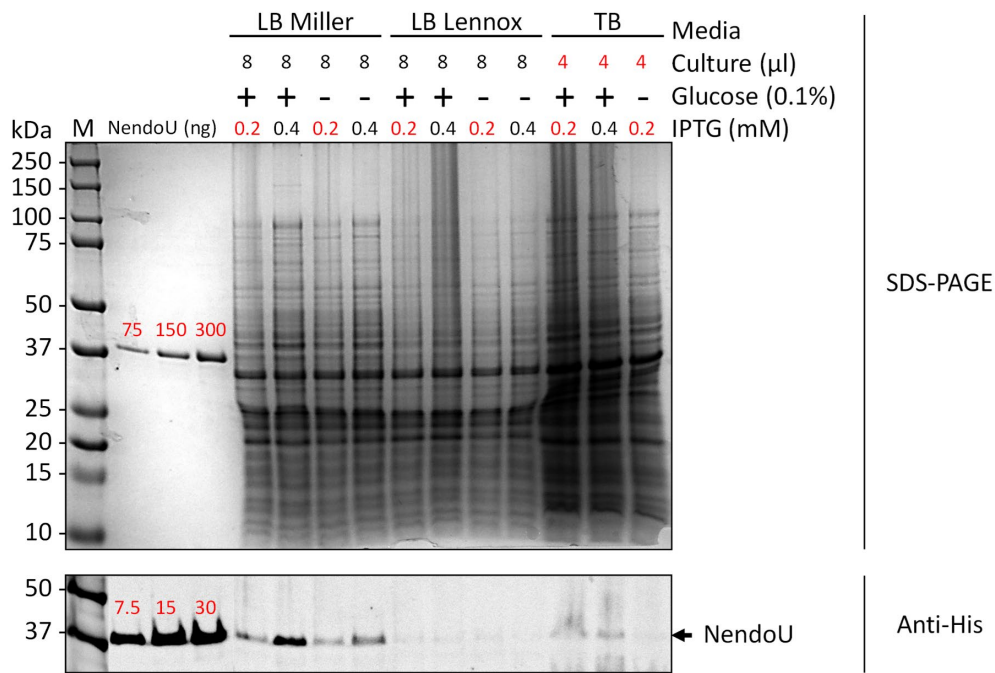
Figure 3.1. Optimized growth conditions for NendoU WT. Positive controls of purified NendoU were loaded (ng of protein in red).

Miller), glucose concentration (0 – 0.1 %), and IPTG concentration (0.2 – 0.4 mM).

Whole cell expression was analyzed by SDS-PAGE Coomassie and Western blot (**Figure**

3.1). Purified NendoU WT protein (75, 150, and 300 ng for SDS-PAGE and 7.5, 15, 30

ng for Western blot) was used as a positive control (**Figure 3.1**). NendoU WT protein

was only weakly visible in the SDS-PAGE at 41 kDa (**Figure 3.1**). Anti-His Western blot

showed NendoU WT expression more clearly (**Figure 3.1**). Expression in LB Miller

produced the most intense bands compared to TB and LB Lennox media (**Figure 3.1**).

Within the LB Miller condition, expression performed best with the addition of glucose

(0.1 %) and higher concentration of IPTG inducing agent (0.4 mM; **Figure 3.1**). The

most optimal condition for expression was therefore identified as LB Miller media with

0.4 mM IPTG/0.1 % glucose (**Figure 3.1**).

Reducing conditions during induction were tested based on (Šiurkus & Neubauer,
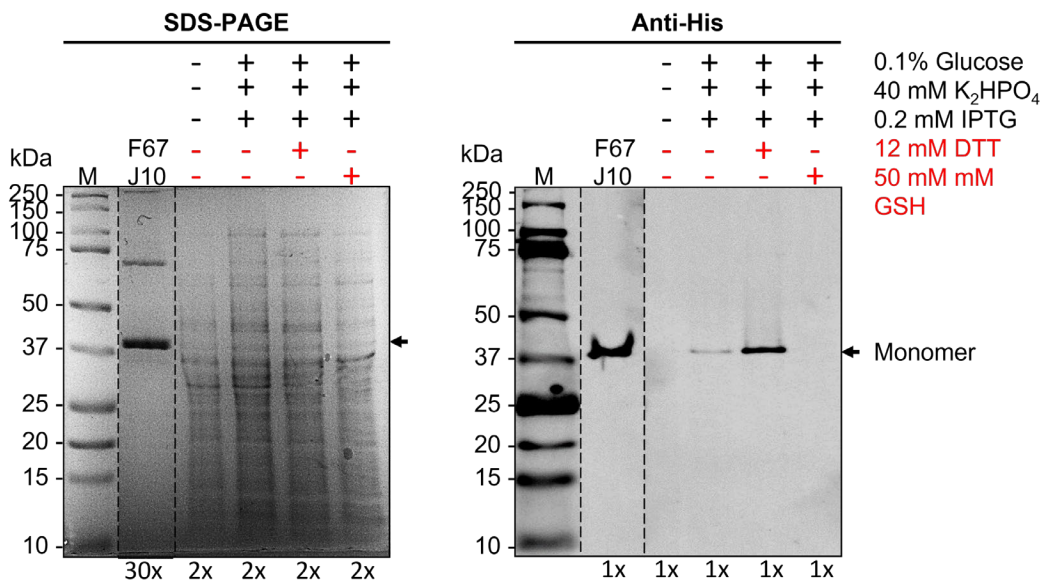
Figure 3.2. Whole cell expression of NendoU WT in reducing condition

2011). In small culture volumes (100 mL), whole cell protein expression increased under

conditions where 12 mM dithiothreitol (DTT) showed increased expression, but 50 mM

L-glutathione (GSH) showed decreased expression levels (**Figure 3.2**).

Expression was scaled up to 1 L for a side-by-side comparisons of NendoU WT with and without DTT reducing conditions during induction. The wet weight of cells grown was 9.5 g for non-reducing and 4.3 g for reducing conditions. Protein was purified and the total yield for both was 1.19 mg. Size exclusion chromatography and SDS-PAGE Coomassie of the NendoU in reducing conditions had the same characteristics and purity compared to NendoU grown in standard conditions (**Figure 3.3**). The yield of protein (mg) per gram of cell pellet totaled 0.125 mg for non-reducing and 0.273 mg for the
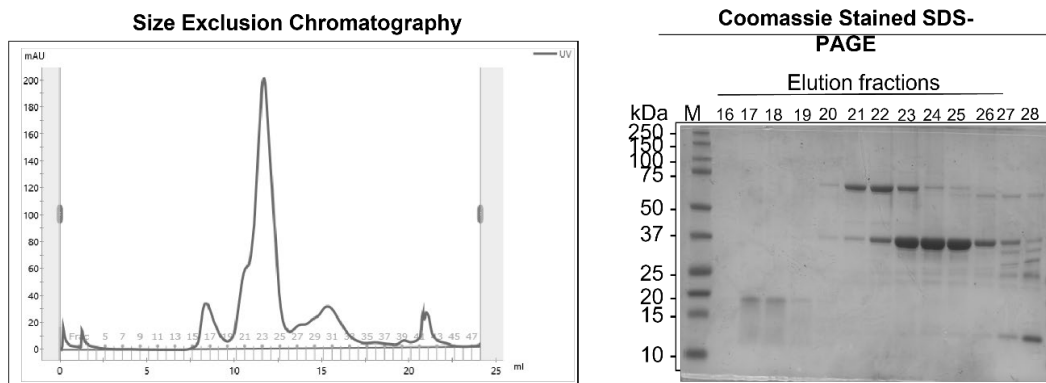
Figure 3.3. NendoU wildtype expression under reducing conditions

reducing conditions. While the reducing conditions produced more protein per gram of cells, the reduced amount of cell growth per liter resulted in approximately the same yield of protein per liter of cell culture.

The catalytically inactive NendoU H234A mutant, which binds substrate but cannot cleave it, was previously described (Bhardwaj et al., 2006) and characterized for this study. This mutant expressed in large quantities as compared to the NendoU WT protein. The purification of one liter of NendoU H234A cell culture yielded between 60 and 80 milligrams of purified protein. Comparison of expression during the size exclusion chromatography (SEC) showed a 25x fold increase of the maximum of the absorbance peak relative to the culture volume (**Figure 3.4**). The NendoU hexamer

protein peak was between 11.5 and 12 mL on a Superdex 200 10/300 column. The

NendoU WT protein from 4 L of cell culture (**Figure 3.4, dashed line**) peaked at ~800

mAu for the hexamer. The NendoU H234A protein from one-third of a liter of cell
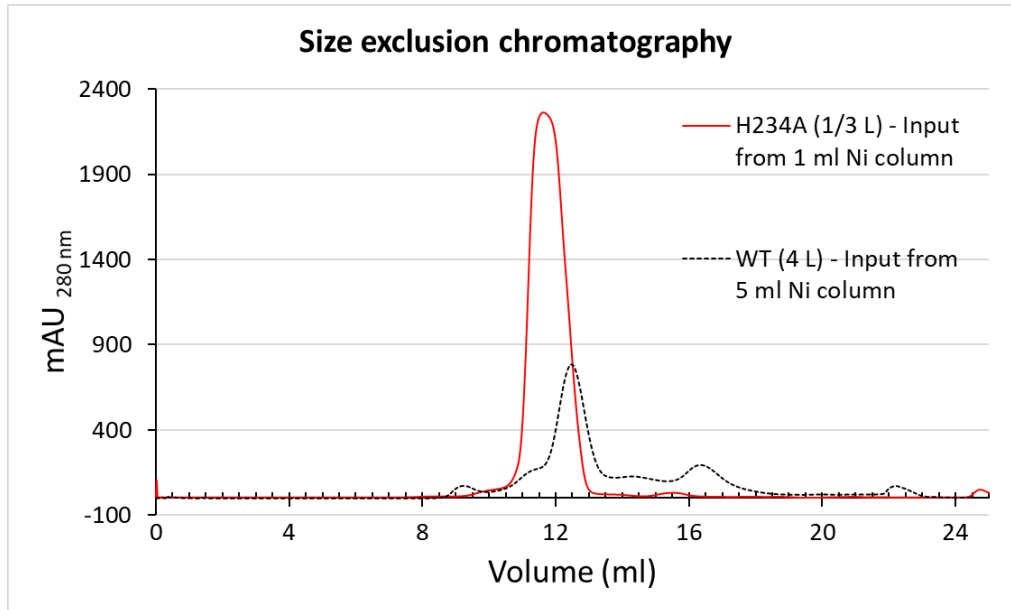
Figure 3.4. Size Exclusion Chromatography (SEC) comparison of NendoU H234A
from 0.33 L cell culture (red) and wildtype NendoU from 4 L of cell culture (dashed).

culture peaked at ~2300 mAu for the hexamer (**Figure 3.4, red line**). The equivalent for

1 L of culture was ~266 mAu for wildtype NendoU and ~6,900 mAu for H234A

NendoU.

Our first attempts in crystallization were to produce microcrystals for diffraction

at the Linac Coherent Light Source (LCLS) for serial femtosecond crystallography (SFX)

experiments. The first microcrystallization conditions established were by batch with

agitation under PEG/MPD citrate conditions and by ultrafiltration of the protein in low

salt conditions as described in **Chapter 2**. Subsequent microcrystallization conditions

were established using a sodium potassium phosphate-based condition from (Y. Kim et

al., 2021) and a calcium acetate-based condition from (Youngchang Kim et al., 2020).

Produced under standard (no agitation) batch conditions, crystals grew larger than batch

with agitation to around 10 – 15 μm (**Figure 3.5**). SFX X-ray diffraction data was
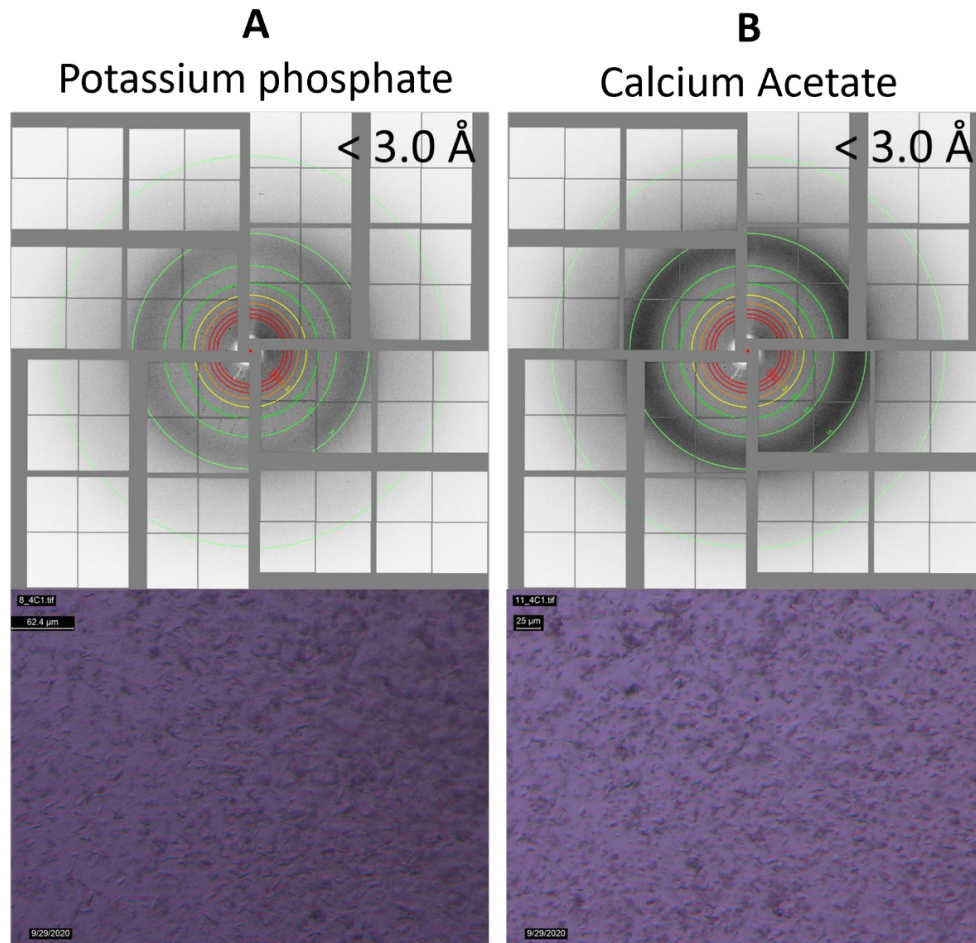
Figure 3.5. NendoU batch crystallization without citrate. **(A)** Potassium Phosphate crystals. **(B)** Calcium acetate crystals.

collected from these crystals at the Linac Coherent Light Source (LCLS), and they

showed X-ray diffraction beyond 3 Å (**Figure 3.5**).

For serial millisecond crystallography (SMX) data collection at synchrotron

sources, the size distribution of the crystals was increased by reducing the salt in the

protein buffer and the change of the protein: precipitant ratio to 1:1. Crystals were grown

in the 20 – 40 μm range (**Figure 3.6**). Additionally, a new octahedral morphology was

observed in crystals. For standard, single crystal, cryogenic crystallography at a

synchrotron, large crystals were grown based on the conditions reported in (Youngchang
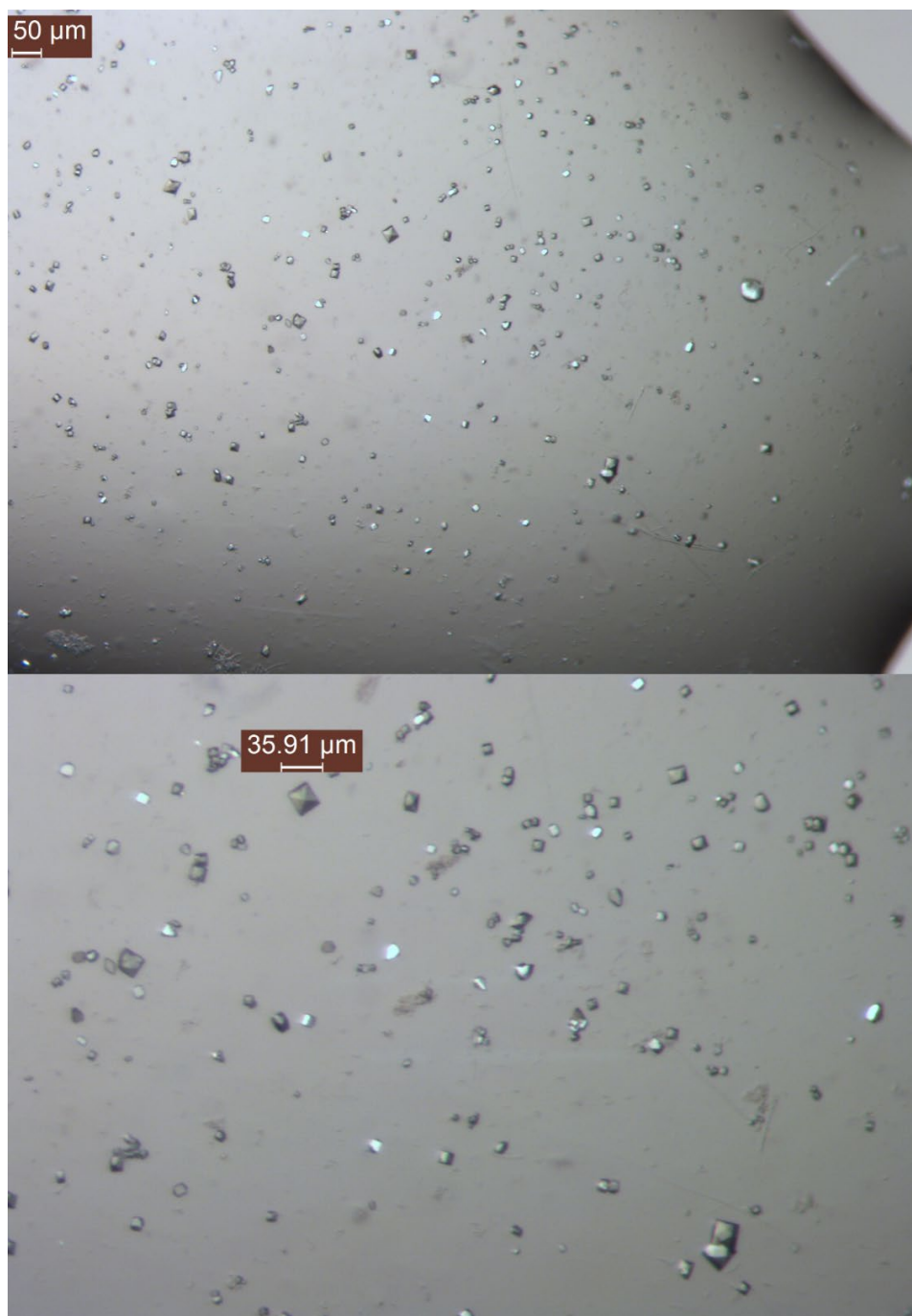
Figure 3.6. Large microcrystals for SMX experimentation. TOP: drop overview BOTTOM: zoomed in view of crystals.

Kim et al., 2020; Y. Kim et al., 2021). Large, needle-like crystals were grown under

conditions containing calcium, sodium acetate, and Tris pH 8.5, but with frayed edges.

Figure 3.7. Changed crystal morphology by seeding to reduce fraying and hollow characteristics in large crystals.
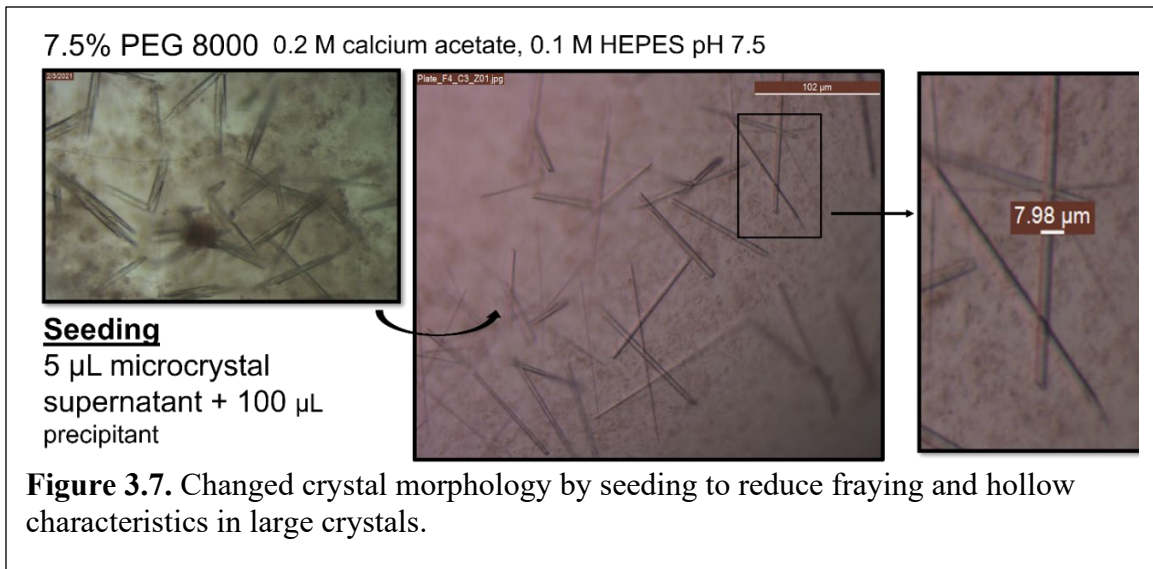
Splitting into multiple crystals, a layered or joined, hollow areas, and cracks were

noticeable in many of the large crystals (**Figure 3.7-9**). Large crystals grown with seeds

Figure 3.8. NendoU apoprotein crystals compared to co-crystals with an NendoU inhibitor.
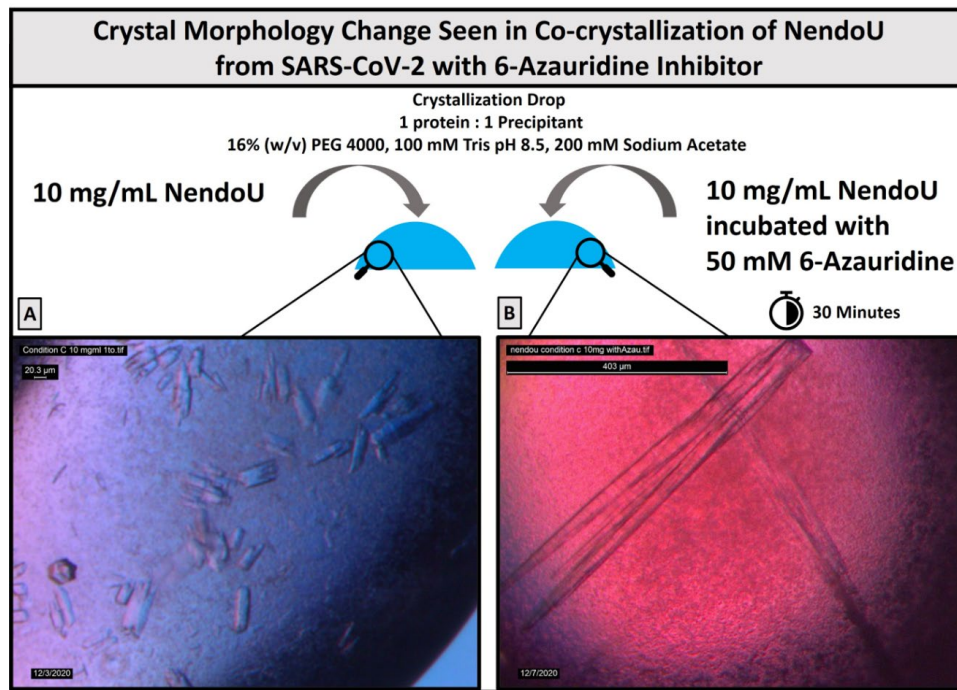
were observed to have a less frayed or layer appearance with sharper edges and a clear rectangular needle morphology (**Figure 3.7**).

Addition of a ligand for NendoU co-crystallization also produced a notable change in morphology. Crystals that were originally shorter (40 – 50 µm) with frayed edges grew longer (400+ µm) and more needle like (**Figure 3.8**). Varying the precipitating agent resulted in more hexagonal and plates-like crystals; however, these often formed together from a single point and the crystals grew together (**Figure 3.9**).
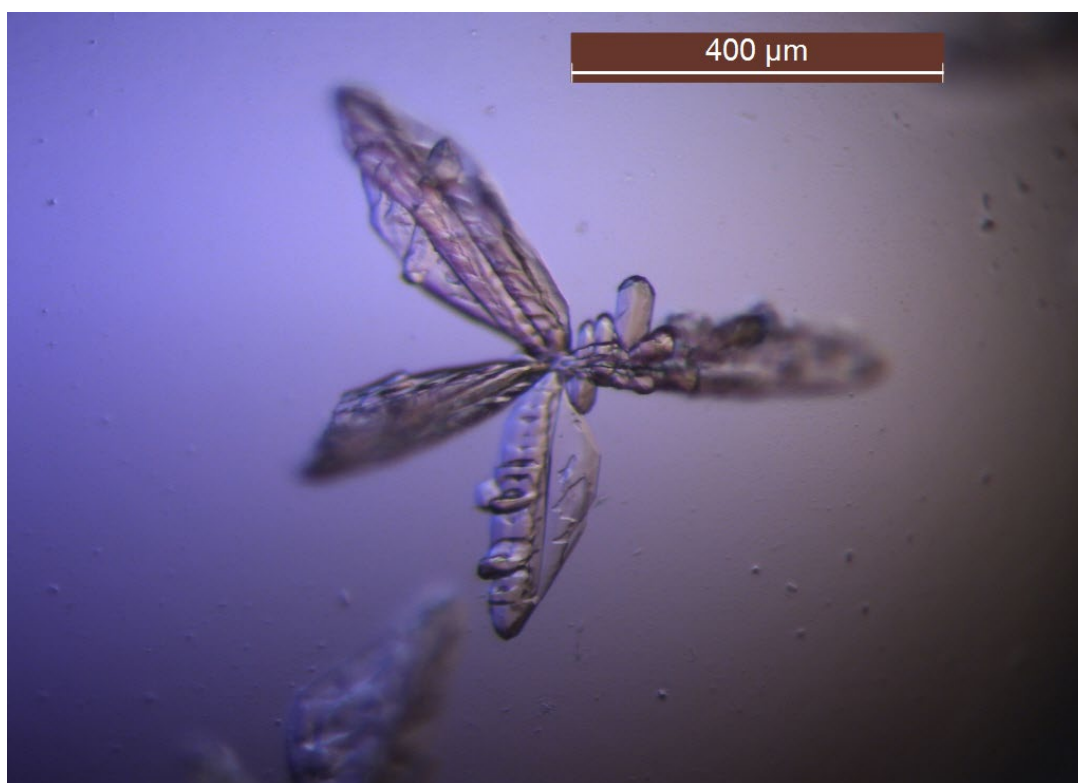
Figure 3.9. NendoU co-crystallized with 6-Azauridine produces large crystals. Tris

These large protein crystals were fragile, and the multiple crystals were weakly joined. During fishing of the large crystals, the multi-crystalline structures broke apart into smaller, individual crystals. Despite the ragged or layered appeared of the large crystals, these smaller single crystals that formed had sharp edges (**Figure 3.10**).

Figure 3.10. NendoU co-crystallized with 6-Azauridine break apart during fishing into smaller crystals with sharper edges.

Similarly, the addition of RNA accelerated crystallization under low ionic strength conditions. The solution of NendoU protein (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP) desalted into a buffer at lower ionic strength (20 mM HEPES, pH 7.5, 20 mM sodium chloride, 20 mM manganese chloride) turned turbid over the course of a half hour. With addition of RNA, protein placed under low ionic strength turned turbid immediately. The sample was observed to have a well-defined, dense pellet the following day after incubation at 4 °C. This pellet was resuspended and observed under the microscope as <2 μm crystals (**Figure 3.11**). The crystals were noted to have slight birefringence and similar morphology to the crystals grown under batch with agitation
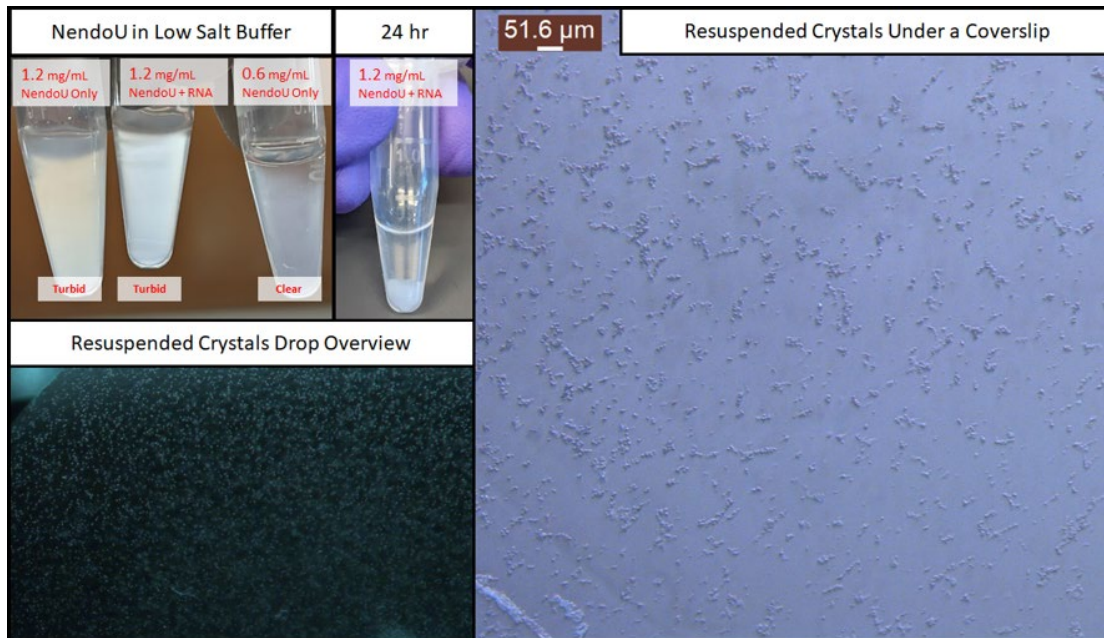
Figure 3.11. Crystallization of NendoU at low ionic strength by desalting and increased density with the addition of RNA. Note: Eppendorf tubes did not look brown on inspection; however, the brown background affected the lighting slightly.

conditions. Crystals were subjected to powder diffraction at the Arizona State University home source; however, diffraction was not observed. Crystallization occurred at low protein concentrations, with the lowest concentration noted as 1.2 mg/mL but was not seen in the 0.6 mg/mL sample (**Figure 3.11**). While the 0.6 mg/mL protein sample was not as turbid, it was also not entirely transparent.

Room temperature synchrotron X-ray diffraction data was collected from large crystals grown in fixed target COC crystallization chips provided by Kuhl laboratory at UC Davis. The chips were based on the previously published generation (Gilbile et al., 2021), and the final chip used at the beamtime was optimized for the area of diffraction. Large crystals (300+ μm) were grown inside the chip using batch conditions with higher PEG concentrations (7.5 – 7.75 %). As a control, the remaining batch experiment not loaded into a chip was grown at room temperature. The crystals grown in the COC chips

had lower density than the relative condition (**Figure 3.12**). Small rotation series (15 –

20°) were collected from multiple crystals. Data collection was limited by the "shadow"
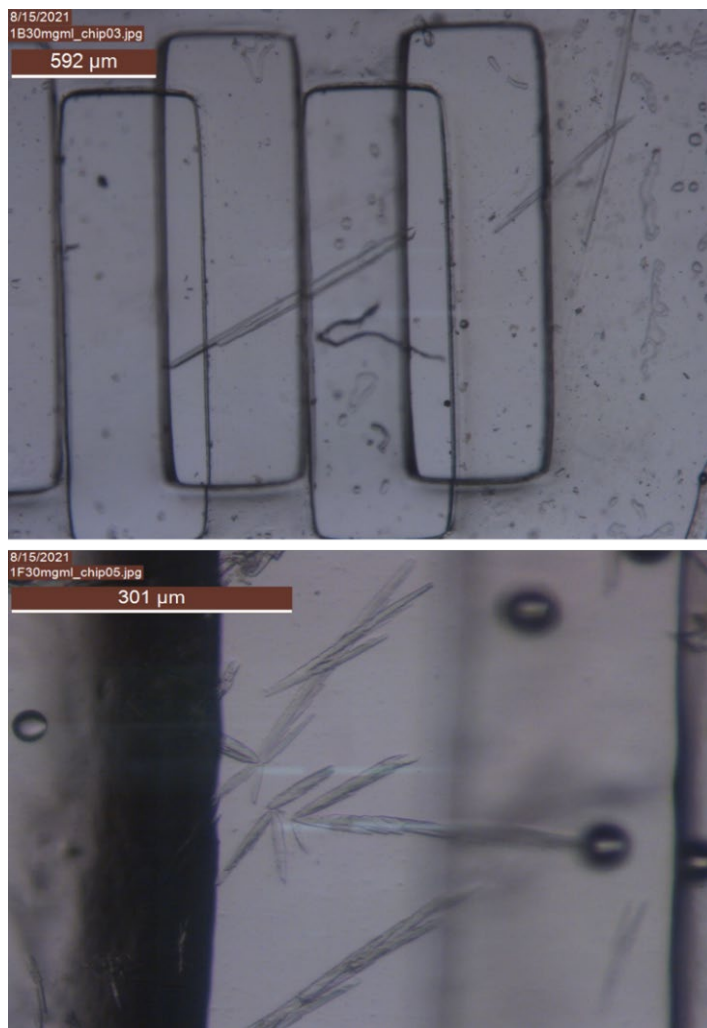
Figure 3.12. NendoU crystallization in fixed target COC crystallization chips.

from the rest of the chip and goniometer's sole rotation around the X-axis. An area of

difficulty in data collection was also that some chips did not survive transportation (not

humidity controlled) and showed signs of desiccation.

3.5 Discussion

Model proteins in crystallography are needed for many experiments to generate

X-ray diffraction data including detector calibration, prove methodology and technology,

and to support education. In this study we describe the features of the crystallization of the catalytically inactive NendoU H234 mutant protein from SARS-CoV-2 that make it an ideal test subject for these three categories as it is easy to produce and crystallize through many different approaches and under very different conditions.

The NendoU WT protein expression was optimized for media type and induction conditions from the expression by Kim and co-authors (Youngchang Kim et al., 2020) to improve expression (**Figure 3.1**). Preliminary screening of expression under reducing conditions (12 mM DTT) based on (Šiurkus & Neubauer, 2011) also showed positive results in whole cell expression, but showed a decrease in the amount of cells that can be grown (**Figure 3.2-3**). Overall, the total NendoU WT produced per liter is approximately the same for non-reducing and reducing conditions.

The generation of a catalytically inactive H235A NendoU mutant was described for SARS NendoU by (Bhardwaj et al., 2006). The SARS-CoV-2 NendoU H234A expressed here shows a 25-fold increase in protein expression compared to the wild-type to produce between 60 and 80 milligrams of protein per liter cell culture (**Figure 3.3**). For serial femtosecond crystallography (SFX) experiments performed at X-ray Free Electron Lasers (XFELs) and serial millisecond crystallography (SMX) experiments performed at synchrotrons, hundreds of milligrams or even one gram of protein can be required to allow for time-resolved studies and replenish the sample between the shots at megahertz repetition XFELs like the European XFEL. NendoU H234A is easily expressed and purified in quantities for serial crystallography studies including the challenging cases of time resolved XFEL studies and studies at high repetition XFELs, as 8 x 1 L of cell culture will produce roughly half a gram of protein.

The NendoU H234A protein can be easily crystallized and crystals across a wide range of sizes can be grown through a variety of different crystallization methods. Microcrystals of NendoU H234A grown by batch with agitation and ultrafiltration have been shown in **Chapter 2** and showed diffraction between 2.5 – 4.5 Å (**Chapter 2**). Additional microcrystals of xxx have been grown by sequential batch methods in two new conditions designed based on precipitants established for growth of large crystals (Youngchang Kim et al., 2020; Y. Kim et al., 2021). The microcrystals appear quickly and mature over the course of 12 – 24 hours. These crystals have been demonstrated to feature <3 Å diffraction at LCLS, supporting the use of NendoU in SFX experiments (**Figure 3.5**). These conditions were developed based on the need for a condition that diffracts to higher quality than the crystals grown by ultrafiltration. Additionally, the sodium citrate used in the batch with agitation condition (**Chapter 2**) occupies the catalytic site, which is nonoptimal for time-resolved SFX studies where substrate/ligand has to bind in the binding site. These two new microcrystallization conditions where crystals grow by the sequential batch method grow high quantity of homogenous microcrystals without citrate for future time-resolved studies.

While SFX experiments have many advantages, the use of X-ray Free Electron Lasers is highly competitive and restrictive as there are only five large hard X-ray XFELs operational worldwide. Synchrotrons are a more available light source for users and with newer microfocus beamlines, enable diffraction of smaller crystals and also beamlines have been developed for serial millisecond crystallography (SMX). Modification of the protein concentration and protein to precipitant ratio results in slightly larger crystals, designed to be more optimal for SMX diffraction (**Figure 3.6**). This development

supports the development of SMX at more accessible synchrotron light sources for increased access to serial crystallography data collection and better preparation for SFX experiments.

Large crystals are easily obtained by vapor diffusion experiments. In the publication of the first SARS-CoV-2 structure of NendoU, it was noted that the process from expression (with conditions established previously over a 4 year time course) to structure deposition took four days (Youngchang Kim et al., 2020). Large crystals also grow quickly in under 12 hours; however, crystal morphology can be a challenge. NendoU crystals frequently grow in a needle morphology where the needle splits into multiple crystals, appears layered, or is hollow at the ends, indicative of long range disorder by too fast crystal growth. We found in our data from crystals grown under conditions including the PDB: 6VWW calcium acetate condition and PDB: 6X1B sodium/potassium phosphate conditions that many crystals were twinned with an estimated twin fraction from the L/H-test to be 0.215 and 0.37, respectively. However, not all structures deposited using these conditions were twinned, indicating that the individual crystal chosen may play a role. Large crystal morphology can be improved using seeding (**Figure 3.7**) and ligand binding (**Figure 3.8-9**). Seeding results in a more uniform crystal morphology with sharper, rectangular needles, but the crystals are narrower in the dimension of the long needle axis. Ligand binding increases the size of the crystals. Large thin crystals frequently broke into smaller, more rectangular crystals with sharper edges during harvesting. This is a useful note for working specifically with NendoU crystals to address this concern, should it arise.

Another application of this information lays in the education of future biochemists and crystallographers. As NendoU can be produced in large quantities, students can set up their own experiments using different techniques (batch, batch with agitation, vapor diffusion, seeding, and more) to learn the methods and experience how these methods explore different path through the phase diagram to create different types and sizes of crystals. The large crystals are easy targets to practice fishing and also practice attempting to break them into smaller, well-defined crystals (**Figure 3.9-10**).

Similar to ligand binding, addition of an RNA substrate (with either the inactive H234 mutant protein or the active WT protein) produced changes in crystal morphology and density. Microcrystals of NendoU (sub two microns) were observed when lowering the ionic strength of the protein buffer. The addition of RNA improved crystal density, presumably by helping to stabilize the catalytic site conformation. The turbidity of the solution increased upon addition of RNA, whereas the solution of the NendoU protein without RNA addition took longer (~ half hour) to reach the same turbidity level. While the nano/microcrystal diffraction limit can only be tested at an XFEL light source due to their small size, the concentration (1.2 mg/mL) needed to create a dense crystal pellet for 0.5 mL is noteworthy (**Figure 3.11**).

Many experiments relating to the field of structural biology focus on important aspects other than determining a protein structure. A good example is technology development of new sample delivery devices, like the next generation fixed target COC crystallization chips (Kuhl Laboratory) (Gilbile et al., 2021). NendoU protein was shown to be successfully crystallized and diffracted at room temperature in next generation COC Chips (**Figure 3.12**). Additionally, experiments focusing on other factors, such as

commissioning of new light sources, require sample to support the commissioning experiments. NendoU has a strong biological case for study, but the inactive H234A mutant is also easy to work with because it can be produced in large quantities, crystallizes readily under different conditions, methods, and sizes, and has not shown sensitivities to changes in environmental factors such as temperature and transportation. For students learning crystallography, NendoU H234 provides a good sample protein as it can be used to demonstrate a variety of different techniques where one can produce noticeable changes in size and morphology in respect to the phase diagram. Overall NendoU H234A may be the focus of COVID-19 research, but it could hold additional uses within the field of crystallography.

CHAPTER 4

HIDING COVID: PRELIMINARY FINDINGS ON OMICRON BA.2 NENDOU AND

POTENTIAL THEREPUTIC, VALLEY FEVER DRUG, NIKKOMYCINZ

Text and figures in this chapter are printed with permission from co-authors: Dhenu
Logeswaran, Nirupa Nagaratnam, Manashi Sonowal, Eranjalee Ranaweera, Sabine
Botha, Thomas Grant, Gihan Ketawala, Michelle Sheikh, Megan Shelby, Mimi Cho
Yung, Matt Coleman, Emily K. Kaschner, Matthew R. Goode, Michele Zacks, Debra T.
Hansen, Raimund Fromme, Matthias Frank, Julian Chen and Petra Fromme.

4.1 Abstract

The uridine-specific endoribonuclease, NendoU, from SARS-CoV-2 enables the
evasion of the innate immune system by degrading a pathogen associated molecular
pattern (PAMP) for RNA virus infection. While the NendoU protein is high conserved, in
the Omicron subvariant BA.2 the so called, "stealth variant," NendoU contains a lineage
defining mutation, T112I. In structural characterization of this mutation, we observe that
the change from a polar threonine to a hydrophobic isoleucine is positioned in the
hypothesized groove for RNA binding and creates favorable binding conditions. As
NendoU hides the virus from the immune system, inhibitors become a more important
role in anti-COVID therapeutics. Most antivirals target the virus' ability to replicate,
either by targeting entry into the host cell or viral machinery for genomic processing.
NendoU inhibitors could provide a complementary path where the virus cannot hide as
effectively, and the host immune system is activated, which would lead to less severe
COVID cases. Using structure determination by crystallography and molecular dynamics,
we also propose the uridine analogs 6-Azauridine (approved anti-cancer therapeutic) and

NikkomycinZ (antifungal therapeutic) as potential inhibitors. NikkomycinZ has been demonstrated as a well-tolerated and effective therapeutic against valley fever, a fungal disease endemic to the US Southwest and northern Mexico. Lack of funding has limited continued development during phase II clinical trials. In molecular dynamic simulations, NikkomycinZ can dock to the NendoU active site in the uridine position, blocking the dual histidine mechanism of RNA cleavage. We propose NikkomycinZ as a potential antiviral therapeutic that, if further developed, could also lead to multiple avenues in the treatment for other orphan diseases like valley fever.

4.2 Introduction

Where most non-structural proteins for SARS-CoV-2 enable the virus to replicate and promote transcription of its genome, Nsp15 helps the virus to stay hidden from the innate immune system (X. Deng et al., 2017). Also known as the uridine specific endoribonuclease, NendoU, the protein degrades the poly-uridine leader sequence of the viral anti-sense RNA (Hackbart et al., 2020). The poly-U sequences form intermediate structures with poly-A sequence with double-stranded RNA (dsRNA), which is recognized by the innate immune system as a pathogen associated molecular pattern (PAMP) for RNA viruses. Proteins like MDA5 and RIG-I recognize these dsRNA segments and initiate the signal cascade for the human immune response including macrophages, natural killer (NK) cells, and type I interferon response (Brisse & Ly, 2019; Chow et al., 2018). However, because NendoU degrades part of this PAMP, the innate immune response is significantly impaired.

This makes NendoU a target for rational drug design for COVID therapeutics but does rely on the same approach as other antiviral therapies. Most antiviral therapies focus

on blocking the entry of the virus into the host cell or inhibit the replication/transcription of genetic material (Frediansyah et al., 2021). In both cases, the virus is prevented from replicating. Alternatively, an NendoU inhibitor would keep dsRNA intact and allow for strong larger host immune system response so that the viral infection is efficiently stopped by the human immune system (Xufang Deng & Baker, 2018). The goal of NendoU-targeted therapeutics would be to provide complementary treatment – while other antivirals prevent the virus from continuing to replicate, NendoU inhibitors would allow for the host immune system response to detect and eliminate the virus. Prophylactic use is an additional explored avenue for NendoU-targeted therapies. If a person is exposed to COVID-19, taking a NendoU inhibitor would interrupt the mechanism for immune system evasion and allow a coronavirus infection to be recognized by the host immune system (and responded to) much sooner than if left untreated, thereby preventing spread from infected but not yet symptomatic individuals and lowering the number of cases of severe forms of COVID-19 as the immune system may eliminate the virus before it causes severe lung infection and inflammation.

In this paper we propose and explore NikkomycinZ as a potential inhibitor of NendoU and its potential as a COVID-19 therapeutic. NikkomycinZ (NikZ) is a proposed antifungal compound that has notably shown efficacy in mice and dogs in the treatment of coccidioidomycosis, or valley fever (Larwood, 2020; Sass et al., 2021; Shubitz et al., 2013). Two primary fungal strains cause valley fever, *Coccidioides immitis* and *Coccidioides posadasii,* and infections occur primarily in the Southwestern United States when fungal spores in the soil are inhaled. While the majority of valley fever cases resolve independently of treatment, in disseminated cases of valley fever, the disease

spreads to other parts of the body. This can result in draining lesions, osteomyelitis (infection of the bone, **Figure S4.6.**2), and neurological symptoms than can be life threatening (Graupmann-Kuzma et al., 2008; Shubitz, 2007). Cases of disseminated coccidiomycosis (DCM) account for <1% of human cases, but 25% of canine cases, which also present in higher rates (Davidson et al., 2019). The estimated lifetime cost of Valley Fever as of 2019 was $736 million and there are growing concerns of the affected demographic spreading with climate change (Matlock et al., 2019; Wilson et al., 2019). The azole class of drugs are primarily used to treat infection; however, they only prevent fungal replication and can have serious (sometimes black box) side effects. Conversely, NikkomycinZ is well tolerated and has been shown to kill fungal infection in murine models (Sass et al., 2021; Shubitz et al., 2014). NikkomycinZ entered phase II clinical trials; however, it was discontinued due to lack of funding. The NikZ compound is a uridine-based nucleoside-peptide that inhibits chitin synthesis, needed to form the fungal cell wall. As a uridine analog, NikZ was selected for our study as a proposed SARS-CoV-2 NendoU inhibitor.

COVID therapeutics are important for the long-term perspective of the fight against the disease. COVID will become endemic and thereby is here to stay and may re-appear in a seasonal pattern. While vaccinations may help prevent serious disease and even infection, treatments will still be needed for those infected. Variants of COVID-19 are also an important factor as accrued mutations can change factors such as transmissibility and severity along with the effectiveness of vaccination. NendoU is a highly conserved protein in the *Coronaviridae* family, but has a lineage defining mutation, T112I, in the Omicron subvariant BA.2, nicknamed the "stealth variant," for its

difficulty to be detected in PCR-test kits. In this paper, we explore the T112I NendoU Omicron BA.2 variant and use of NikkomycinZ as a potential anti-NendoU COVID-19 therapeutic.

4.3 Methods

Expression of Nsp15 mutant T112I

Nsp15 T112I mutant constructs were transformed into *Escherichia coli* BL21-Gold (DE3) strain (Agilent cat. 230132). A starter culture was prepared using LB Miller with 50 µg/mL carbenicillin, inoculated with a single colony, and grown overnight at 37 °C with 200 RPM shaking. A 1 L cell culture was grown in LB Media containing 50 µg/mL carbenicillin at 37 °C with the addition of 1 % starter culture. Expression of the T112I mutant NendoU was induced as follows: The cell culture was grown to a cell density of an OD_{600} between 0.9 and 1.0, subsequently expression was induced by addition of 0.4 M IPTG and 1% glucose and the temperature was reduced to 18 °C (see **Chapter 3** for more details). Cells were harvested after 18 h by centrifugation at 5000xg for 15 minutes. Cells were stored at -80 °C until use.

Protein Purification

Cells were resuspended in 10% w/v lysis buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 20 mM imidazole) and lysed by sonication (2 s on/2 s off for 2 minutes, 5 cycles with 2 minutes rest on ice between). The cell lysate was isolated by centrifugation at 45,000 xg for 30 min at 4 °C to remove unbroken cells and the membrane fractions. The supernatant was filtered with a 0.2 µm syringe filters. Immobilized metal affinity chromatography was performed using Ni-NTA (5 mL) HP His-Trap columns from GE on an AKTA FPLC. Filtered lysate was loaded onto the column, washed with 20 CV of lysis

buffer and 20 CV of wash buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 50 mM imidazole), and eluted with 2 CV of elution buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 500 mM imidazole). The eluted protein was concentrated by with Amicon™ 30 kDa spin concentrators. Size exclusion chromatography was performed with a Superdex 200 10/300 increase column on an AKTA FPLC in 20 mM HEPES, pH 7.5, 75 mM NaCl, 1 mM TCEP. The primary peak at 11.8 mL was collected and concentrated to 5 mg/mL using the 30 kDa spin concentrators and stored at -80°C.

Endoribonuclease Inhibition Assay

NendoU (wildtype) protein was incubated with NikkomycinZ for 1 h at 30 °C. A NIR800 labeled 21 nt RNA substrate was then incubated with the protein and the reaction was quenched by the addition of loading dye at 5 min, 15 min, 30 min, 1 h, 3 h, and overnight timepoints. The overall reaction contained 12 nM protein, 2 mM NikZ, 20 mM manganese chloride, and 0.1 µM RNA substrate in assay buffer containing 50 mM Tris-HCl, pH 7.5, 50 mM KCl, 1 mM DTT. A control sample was prepared in parallel using no NikkomycinZ along with two input RNA samples, one with loading dye added after all samples were set up and one incubated at 30 °C overnight. Samples containing RNA were kept in dark conditions using aluminum foil to avoid degradation of the far infrared fluorescent label. Results were analyzed by 10% Urea-PAGE and run for 1 hr at 150 V/8 mA. The gel was imaged with the Typhoon imager scanning at 783 nm lasing option.

Crystallization

The NendoU T112I protein was prepared at 5 mg/mL protein concentration in 20 mM HEPES, pH 7.5, 75 mM NaCl, 1 mM TCEP. Vapor diffusion hanging drop experiments were prepared for crystallization. Protein was mixed rapidly at room

temperature in a 1:1 ratio with the precipitant: 0.1 M HEPES, pH 7.5, 0.2 M calcium

acetate, 7.5% PEG 8000. The crystallization cocktail was incubated at room temperature

for 1 day. Once crystals 2 – 30 μm appeared, settled crystals were resuspend and mixed

1:20 with the precipitant to create a seed stock. The experiment was then repeated using

the seed stock in place of the precipitant and crystals were grown for one day at room

temperature. Crystals harvested were cryoprotected by placing them into a drop of the

precipitant containing 30% glycerol prior to freezing in liquid nitrogen.

Data Collection

Data collection was performed at the Advanced Photon Source (APS) in Chicago,

IL, at beamlines 23 ID-D and 23 ID-B. X-ray diffraction data were collected at a

wavelength of 1.03 Å, with 0.2 seconds exposure/image and 20% attenuation. The

detector was placed at 400 mm from the sample/X-ray interaction point and 900

diffraction images were collected under cryogenic cooling. The crystals were rotated

between 30 and 120 degrees with 0.1-degree rotation/image.

Data Analysis

Data analysis was performed through the JBluIce automated data analysis pipeline

(McPhillips et al., 2002). The Dimple pipeline in CCP4 (Winn et al., 2011) was used to

solve and refine the structure with PDB: 6VWW as a reference for molecular

replacement. Figures were made in PyMOL (DeLano, 2002) and COOT (Emsley et al.,

2010).

4.4 Results

The T112I mutation in NendoU was listed in the Stanford University Coronavirus

Antiviral and Resistance Database (https://covdb.stanford.edu/page/mutation-viewer/) as

lineage defining for the Omicron subvariant BA.2. The position of the T112 residue in

NendoU and highlighted in blue on the NendoU hexamer, showing it is a surface exposed

mutation in the middle domain (**Figure 4.1**). The position of the T112 residue is in the
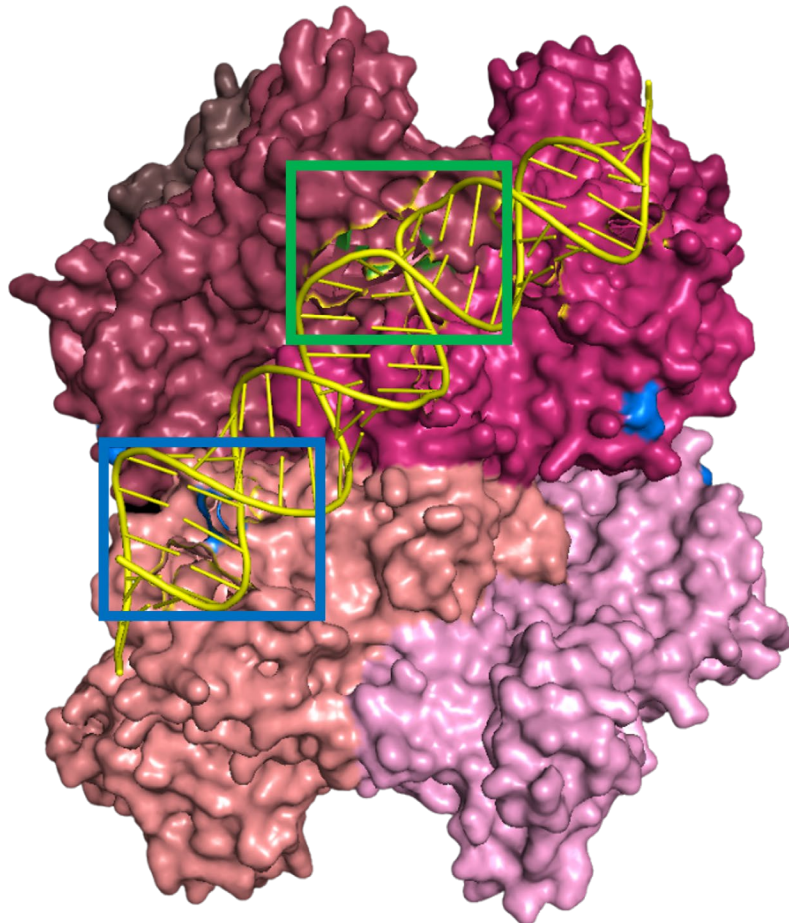
Figure 4.1. NendoU hexamer with 52 nt DNA bound (PDB: 7TJ2) with the T112 residue highlighted in blue and active site highlighted in green.

groove for RNA binding, demonstrated in the recently released NendoU/RNA complex

(PDB:7TJ2) on March 23, 2022 (**Figure 4.1**) (Frazier et al., 2022). The RNA binding

grove was hypothesized originally in 2006 (Bhardwaj et al., 2006) and stretches across

the two dimers. The RNA bound to an active site in the top dimer stretches across the

groove and interacts with the middle domain of NendoU on the lower dimer (**Figure 4.1**).
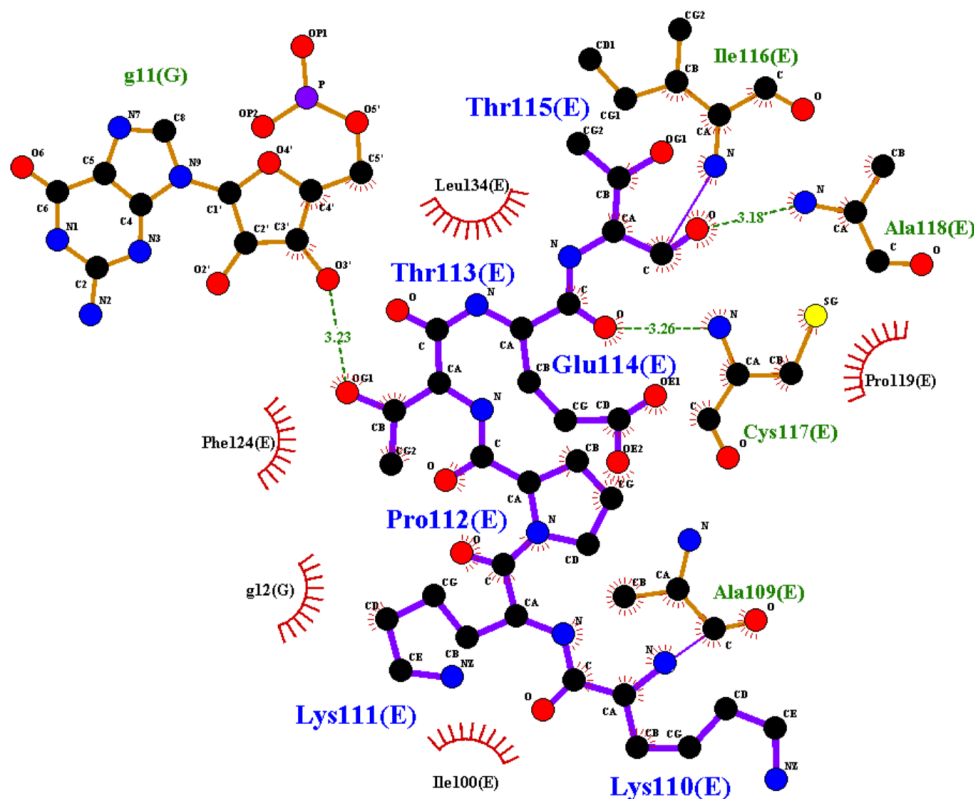
Figure 4.2. LigPlot+ of the NendoU/RNA complex (PDB 7TJ2) of residues 110 – 115 of NendoU. Threonine 113 of this construct corresponds to the T112 of the wildtype construct.

LigPlot+ 2.2 (Laskowski & Swindells, 2011) was used to analyze the interactions

of residues 110 to 115 in the NendoU/RNA complex released on the PDB. The

catalytically inactive construct, NendoU H234A, was used and the threonine 112 residue

identified for the point mutation in Omicron BA.2 corresponds with the threonine 113 in

the construct. The threonine 113 was shown to from a 3.23 Å hydrogen bond between its

hydroxyl group and the 3' hydroxyl of the guanine ribose (**Figure 4.2**).

The T112I NendoU mutant was expressed in *E. coli* and purified based on the

developed protocol for wildtype expression (**Chapter 3**). Purification of NendoU T112I

indicated higher yields than the NendoU wildtype protein. The A_{280} absorption peak for

NendoU T112I in the size exclusion chromatogram (SEC) for the hexamer neared ~1100

mAu and a broadening of the SEC peak was observed (**Figure 4.3 Right**). The activity of

NendoU was confirmed using the negative controls H234A inactive NendoU for no
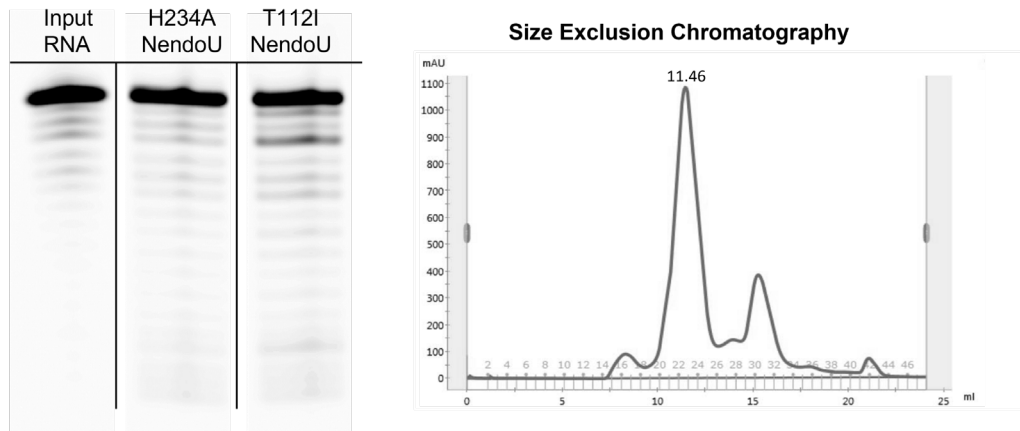
Figure 4.3. Activity RNA cleavage essay and purification of the T112I mutant. **LEFT** activity assay for NendoU T112I compared to the catalytically inactive H234A and input RNA. **RIGHT** Chromatogram for the SEC purification of NendoU T112I from 2 L of cell culture.

protein cleavage and RNA input only for no degradation (**Figure 4.3 Left**) in an

endoribonuclease activity assay.

The T112I mutant was crystalized by vapor diffusion hanging drop. X-ray

diffraction data was collected at the Advanced Photon Source in Chicago, IL. The

structure was solved to 2.4 Å with R_{work}/R_{free} values of 0.1986 and 0.2347, respectively,

by molecular replacement with PDB: entry 6VWW and refinement in the Dimple

pipeline (Winn et al., 2011) (**Figure 4.4**). The NendoU monomer is show in blue with the

active site highlighted in red and the T112I mutation in the middle domain in yellow

(**Figure 4.4**) (DeLano, 2002). The electron density of isoleucine 112 is shown in **Figure 4.5**.
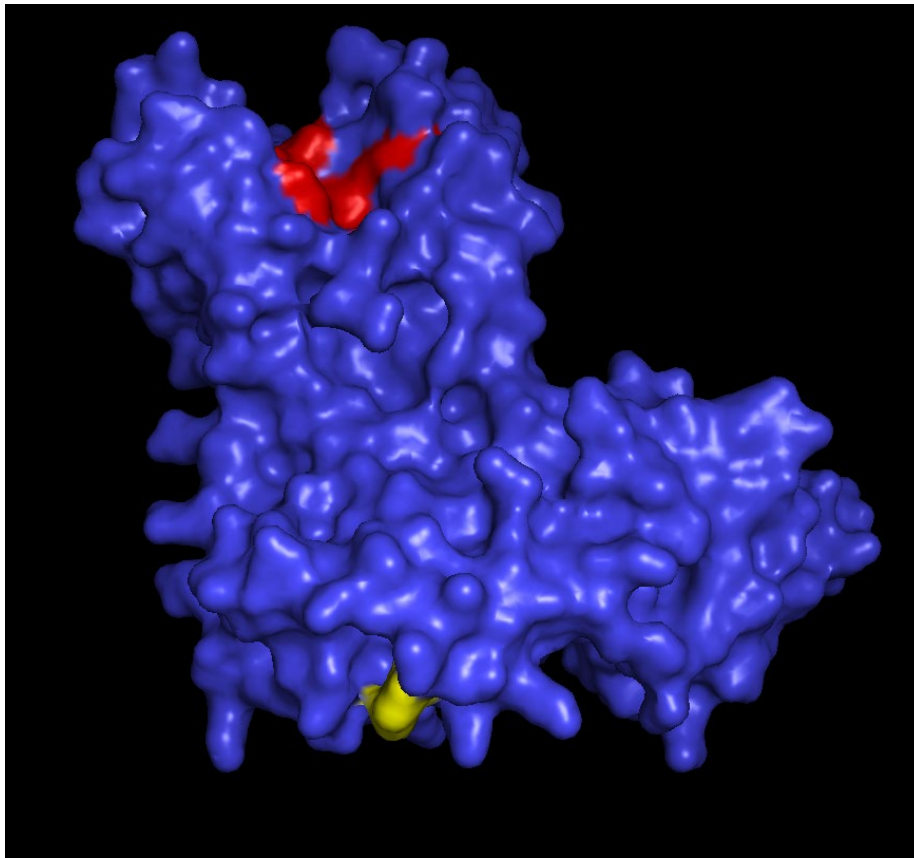
Figure 4.4. Crystal structure of T112I NendoU monomer (Blue). The active site is highlighted in red and the T112I mutation is highlighted in yellow.
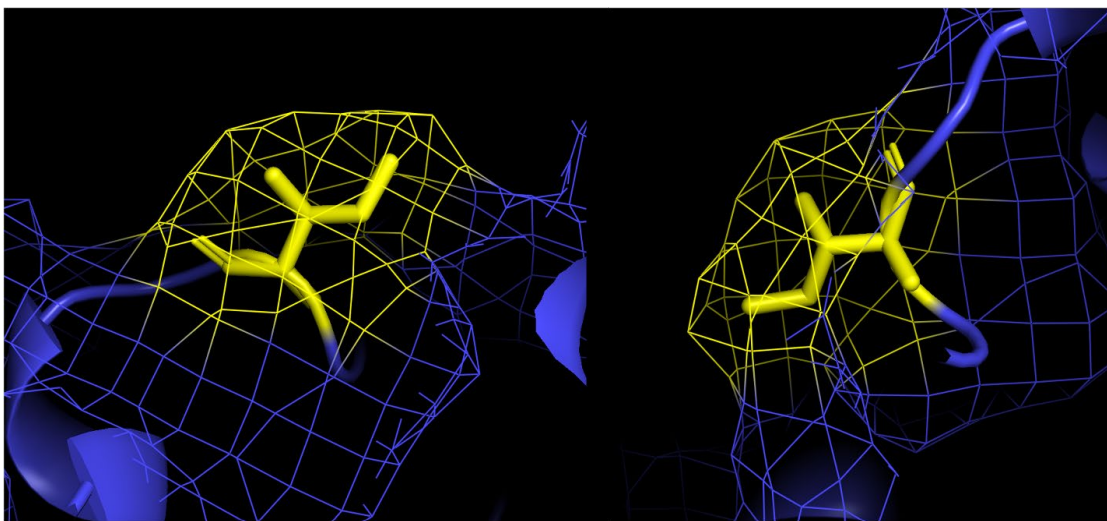
Figure 4.5. Electron density of the T112I mutation (yellow) in NendoU.
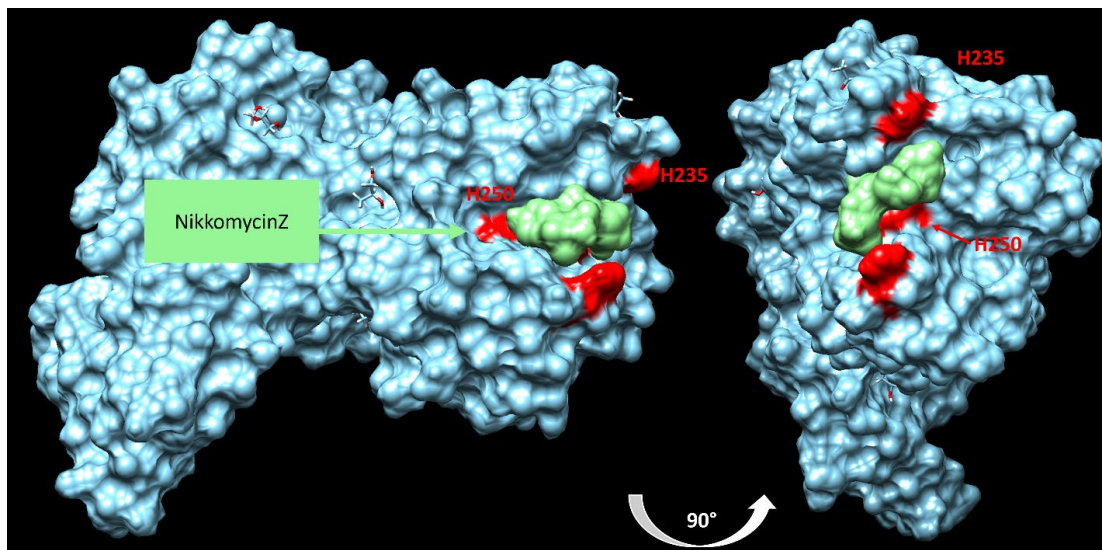
Figure 4.6. Molecular dynamics docking of NikkomycinZ in the NendoU catalytic site by AutoDock Vina (Trott & Olson, 2010) in Chimera (Pettersen et al., 2004).

Molecular dynamic (MD) simulations were used to explore NikkomycinZ as a potential inhibitor for NendoU. NikZ was modeled into the NendoU active site by AutoDock Vina (Trott & Olson, 2010) and a conformation similar to uridine binding was selected at less than -6 kcal/mol (**Figure 4.6**). The ligand was able to dock to NendoU in the active site between the two primary active site histidine residues (H235 and H250), which are the basis for the hypothesized mechanism (Y. Kim et al., 2021). In a subsequent activity assay, 3 mM NikkomycinZ was incubated with NendoU wildtype protein for one hour at 30 °C prior to RNA incubation. Near infrared (NIR) labeled RNA substrate was incubated with wildtype and wildtype with inhibitor protein and samples of the reaction were stopped over the course of three hours. Differences between the control and the inhibitor were observed starting at the thirty-minute time point. A cleavage product of NendoU (indicated by the yellow arrows in **Figure 4.7**) was visible in the sample without NikZ, but not in the NikZ/NendoU condition (**Figure 4.7**). Additionally,
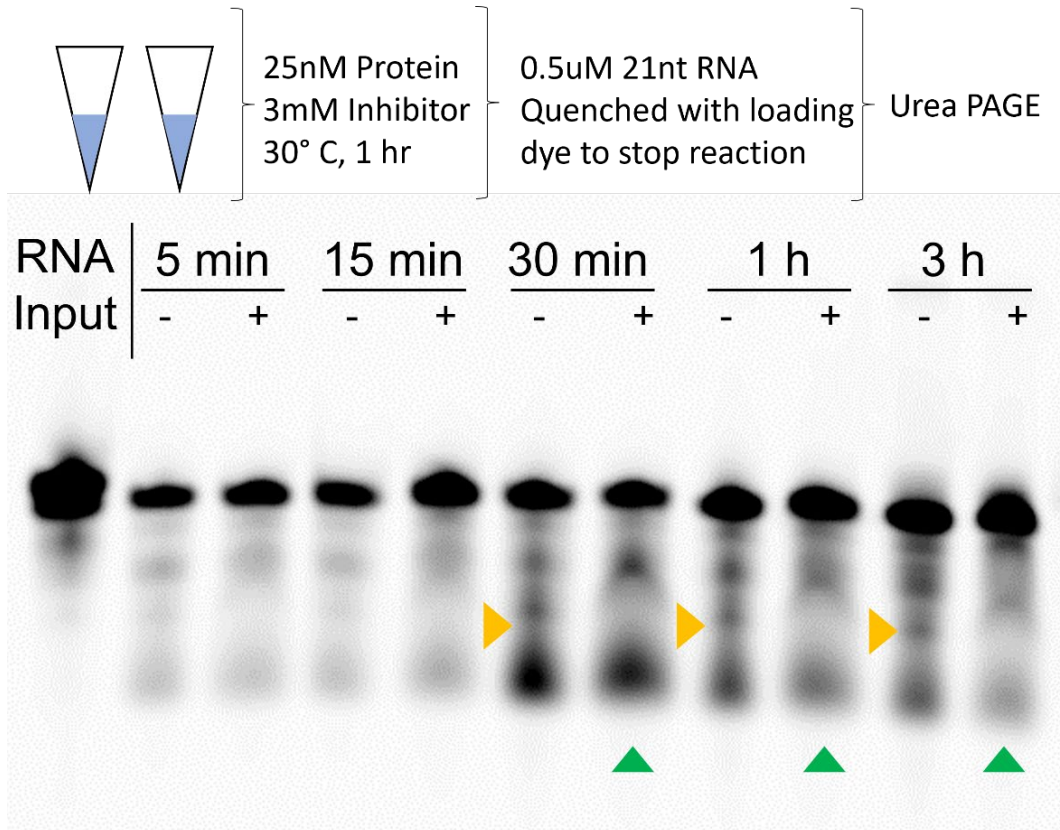
Figure 4.7. Inhibition of NendoU wildtype activity and specificity by NikkomycinZ. The green arrow highlights the reduced activity in the inhibitor condition. The yellow arrow indicated highlights a cleavage product that is seen in the uninhibited condition.

the NendoU with inhibitor condition cleavage products showed less intensity (indicated by the green arrows) than the wildtype cleavage alone (**Figure 4.7**).

4.5 Discussion

The position of the T112I mutation in NendoU from the Omicron BA.2 subvariant was identified as exposed to the surface and in the middle domain of NendoU. The change from an uncharged, polar threonine to a hydrophobic isoleucine is hypothesized to be favorable for substrate coordination due to the hydrophobic nature of RNA. The position of the T112I mutation is located in a groove, predicted for RNA binding first by Bhardwaj in Figure 9E and later by Perry in Figure 2CD and Figure 3D

(Bhardwaj et al., 2006; Perry et al., 2021). The RNA binding grove was later confirmed in the NendoU/52 nt RNA complex solved by cryoEM (Frazier et al., 2022).

The purified T112I NendoU is shown to be catalytically active and expressed in larger quantities than the wildtype. The higher expression levels are advantageous to the study of catalytically active NendoU in time-resolved studies which were previously WT sample was limited. While the activity has not been studied in comparison to the wildtype, it has been previously hypothesized that there is a relationship between expression level and activity as some hyperactive mutants have shown less expression and protein yield while the inactive mutant H234 showed very strong expression and high protein yields of up to 80 mg purified NendoU per liter cell culture (**Chapter 3**). We were able to produce … mg /liter cell culture of the T112I mutant of NendoU. However, because the T112I mutation is hypothesized to affect specificity, this may be beneficial for expression as it may reduce non-specific cleavage of the host mRNAs and eventually also its own mRNA, as indicated by the low expression yields in the in vitro expression systems described in **Chapter 2**. Information regarding the T112I mutation is currently very limited as it is a recent development as part of the Omicron subvariant BA.2 and is only acknowledged as being present with no documented research conducted on this mutant as of March 2022 (Abbas et al., 2022), so our studies are to our knowledge the first studies on this mutant.

The NendoU T112I protein structure was solved to 2.4 Å based on X-ray diffraction data sets collected at the Advanced Photon Source in Chicago, Illinois (**Figure 4.4**). This is the first structure of this NendoU mutant and shows the position and altered electron density as a result in the change of a threonine to an isoleucine at residue 112.

This apo structure is specifically important for future binding studies to show ligand binding, but will be also a great target for future structural studies of NendoU in complex with RNA. It will be especially exciting to discover how this mutation affects RNA docking compared to the wildtype from the Wuhan SARS-Cov-2 strain.

NendoU plays an important role in the coronavirus family including the SARS-CoV-2 virus. Monitoring mutations in NendoU for activity could be a useful tool in monitoring larger effects on disease outcomes as NendoU is responsible for hiding coronaviral infection from the host immune system. If NendoU is more effective at hiding the virus, this could contribute to factors such as asymptomatic or pre-symptomatic transmission as the virus continues replicating without the infected patients knowing that they are sick. Cases of severe disease can occur as the result of the virus replicating out of control without the host immune system responding until it's too late. Overreaction of the immune system in the form of massive inflammation once the virus has been detected late is the main cause of death of COVID patients. NendoU activity and hyperactivity may play a role in this aspect of disease and therefore would be an important target for therapeutics. Just as the viral machinery uses the host to produce more viruses, an NendoU inhibitor would help the immune system to early on detect the viral infection by SARS-CoV-2 do what it does best – clear the infection.

NikkomycinZ was hypothesized to inhibit the NendoU protein activity as NikZ is a uridine-analog. NikkomycinZ is an antifungal therapeutic that inhibits chitin synthase that catalyzes an essential step in the formation of fungal cell walls. Studies of NikZ for the treatment of fungal disease coccidioidomycosis (Valley fever), which is endemic to the Southwestern United States, show that it is well tolerated in high doses. The

substantial previous research makes NikZ a good novel drug target as it has already progressed to phase II clinical trials and shows excellent response. So the development is currently not continued only because of lack of funding for a niche disease. As a well-tolerated drug, this also fits the goal of NendoU as a prophylactic drug target. NendoU impacts the innate immune system's ability to respond, so inhibition early in infection may help (i.e., if taken after exposure) to reduce the duration and severity of illness.

Docking of NikkomycinZ to NendoU was performed in Chimera with Autodock Vina (**Figure 4.6**). This demonstrated that the uridine-like component of NikkomycinZ can align with the uridine-specific active site of NendoU. It's hypothesized positions in the NendoU binding site also interferes with the H235 and H250 active site residues, which are the main catalytic side chains of the hypothesized mechanism and critical for NendoU activity (Y. Kim et al., 2021). Activity of NendoU incubated with NikZ showed reduced activity and reduction in the number of total cleavage products after a half hour as compared to the active NendoU protein (**Figure 4.7**), so it could be also a very good initial start for further optimization of the drug binding inhibition of the activity of NendoU. The results of NikkomycinZ as an inhibitor and role of mutations in NendoU are still emerging; however, it is clear that both are important aspects of COVID research going into the future of monitoring emerging variants and treating the disease.

2.6 Supplementary Information

2.6.1 Supplemental Figures
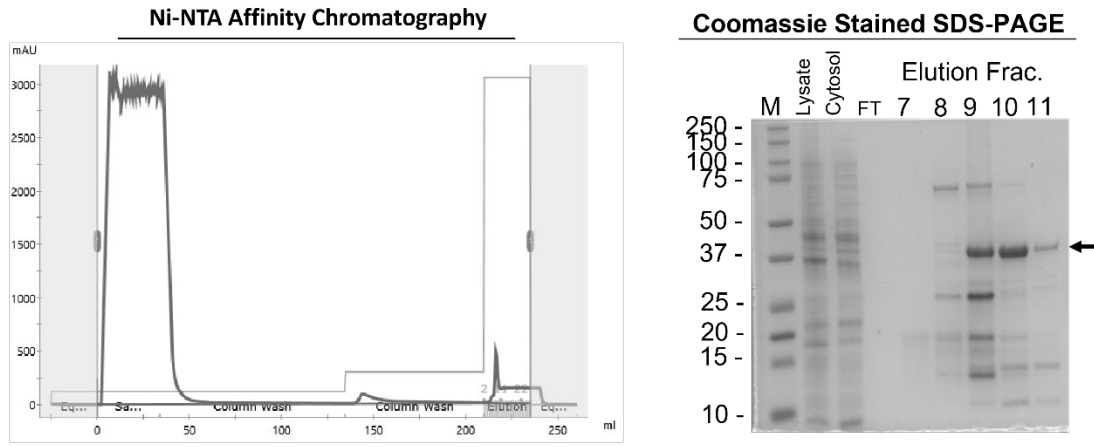
Supplementary figure 4.1: LEFT: Chromatogram from Nickle Immobilized Metal Affinity Chromatography of the T112I NendoU mutation RIGTH SDS-PAGE stained with Coomassie blue.
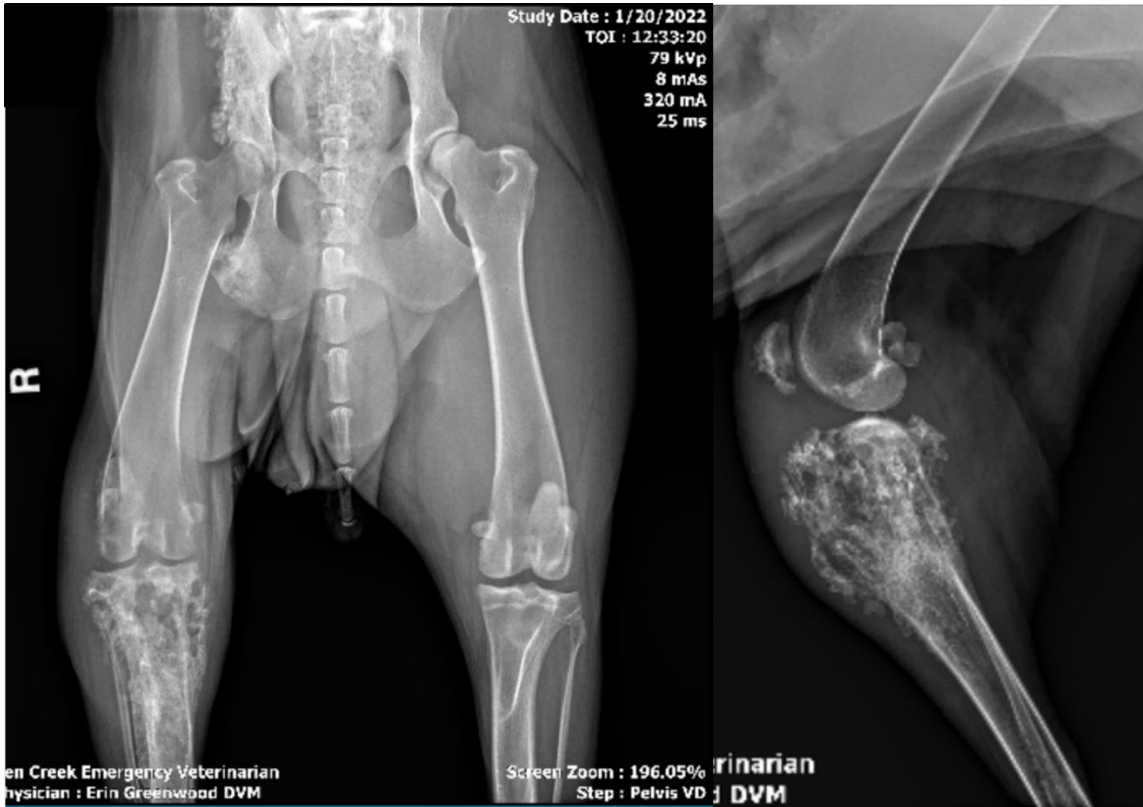
Supplementary Figure 4.2. Osteomyelitis in a case of disseminated coccidiomycosis (canine). Image provided by Dr. Erin Greenwood from Queen Creek Emergency Veterinary clinic.

CHAPTER 5

SYNCHROTRON AND ROOM TEMPERATURE SERIAL FEMTOSECOND

STUDIES OF TASPASE1: STRATAGEMS FOR ANISOTROPY CORRECTION

Sections of text and figures in this chapter were reprinted with permission from **Rebecca Jernigan**, Nirupa Nagaratnam, Gihan Ketawala, Jay-How Yang, Silvia L. Delker, Thomas E. Edwards, Janey Snider, Darren Thifault, Dewight Williams, Brent L. Nannenga, Mary Stofega, Lidia Sambucetti, James J. Hsieh, Andrew J. Flint, Petra Fromme, and Jose M. Martin-Garcia

5.1 Abstract

Taspase1 (threonine aspartase 1) is an endopeptidase that is overexpressed in primary human cancers. It has been identified as a potentially potent anticancer drug target because loss of Taspase1 activity disrupts proliferation of human cancer cells in vitro and in mouse tumor xenograft models of glioblastoma. It functions as a non-oncogene addiction protease that coordinates cancer cell proliferation and apoptosis. Based on its promise as a novel anti-cancer therapeutic target, Taspase1 was accepted into the NCI Experimental Therapeutics (NExT) Chemical Biology Consortium (CBC) drug discovery and development pipeline. Several promising Taspase1 inhibitors have already been identified in high throughput screens. A high-resolution structure of the enzymatically active full-length Taspase1, and Taspase1 in complex with substrate and inhibitors, would greatly enable optimization (through rational design) of pharmacologically potent Taspase1 inhibitors. Taspase1 is a highly conserved 50 kDa proenzyme. The biologically active form of Taspase1 is a hetero-tetramer that displays an overall αββα structure. Crystallographic structures of a truncated version of Taspase1

have been determined as apo enzyme and in complex with covalent inhibitors; however, this truncated form of the enzyme is catalytically inactive. The full-length protein contains an additional sequence that was originally anticipated to be disordered. Until recently, no structure of active, full-length Taspase1 nor complexes of the catalytically active enzyme with inhibitors existed. Recently, first crystals of the full-length Taspase1 have been grown and data collected at the Advanced Photon Source (Chicago, IL). These crystals diffracted to ~ 3.5 Å but show a strong diffraction anisotropy. A crystal structure of the full length Taspase1 has been solved by our team and shows that the protein forms a double ring structure, where the previously missing amino acids from an alpha helix that is located close to the catalytic site that connects the two rings. Taspase1 micro-crystals have been recently obtained and first SFX experiments were conducted at the PAL XFEL in S. Korea. Improved crystals will be screened at LCLS during PCS beamtime P114 for serial femtosecond crystallography (SFX) where we aim to determine the first high-resolution structure of the full length apo-protein and complexes with inhibitors of the enzyme. These structures should greatly benefit the structure-guided medicinal chemistry development of Taspase1 inhibitors as anti-cancer therapeutics.

5.2 Introduction

Study of proteases are important for understanding their roles, which are diverse and important for regular functioning of the body including tissue remodeling, metabolism, and migration. In the context of cancer, apoptosis and cell proliferation are important, normal functions but they have to be tightly regulated to avoid premature aging or cancer. The problem of cancer frequently involves the mis-regulation of genes involved in these pathways. Failure for unhealthy cells to stop growing, or for the cells to

108

continually proliferate can lead to tumor growth. Enzymes can be implicated in these tight regulations and misregulations as the result of changes in expression of key proteins like Taspase 1 that are themselves no oncogenes but their expression fosters cancer proliferation (Rakash et al., 2012; Y. Yang et al., 2009).

Taspase1 (threonine aspartase 1) is an anti-cancer therapeutic drug target that is overexpressed in many primary human cancers, including glioblastoma, breast cancer, infant leukemias, and colon cancer (Wünsch et al., 2016). Taspase1 is considered a non-oncogene addiction protease for its indirect regulations of genes associate with cancer (Chen et al., 2010). Taspase1 activates target proteins such as mixed linage leukemia 1 protein (MLL1), mix lineage leukemia 2 protein (MLL2), TFIIA, and AKF (Niizuma et al., 2015). Activation of MM1 can then induce oncogenic initiation and cell proliferation (Hsieh et al., 2003; H. Liu et al., 2008). The role of Taspase1 also extends to regulating cell-cycle genes and the expression of *HOX* genes, further emphasizing the role of Taspase1 in cancer (Hsieh et al., 2003; Takeda et al., 2006).

The activation of these proteins is performed by the nucleophilic T234 residue to the aspartate at the cleavage site of the target IXQL(V)D//G motif. This is *trans-activity*; however, Taspase1 may also perform a *cis*-cleavage during spontaneous self-cleavage (Hsieh et al., 2003). The self-cleavage behavior is characteristic of the N-terminal nucleophile (Ntn) hydrolase superfamily. Within the family, Taspase1 is in the type 2 asparaginase subfamily. These enzymes are expressed as proenzymes that are catalytically inactive. It isn't until the N-terminal peptide bond is hydrolyzed that the enzyme becomes catalytically active (Oinonen & Rouvinen, 2000).

109

Taspase1 is a 50 kDa, 420 amino acid protein that consists of an α subunit (28 kDa, N-terminal) and a β subunit (22 kDa, C-terminal). They assemble into a αββα structure (Guo et al., 1998; Oinonen & Rouvinen, 2000). In pursuit of structure-based rational drug design, Taspase1 has proven to be a challenging target, likely due to the conformation of the full-length protein. Original structures of Taspase1 include the proenzyme and two-chain truncated form of Taspase1 (J. A. Khan et al., 2005). While it was previously proposed that the truncated residues 207 – 233, which are critical for the activity of Taspase 1, are flexible and disordered, we have recently shown that they form a long flexible helix as part of a larger flexible region between residues 183 – 233 (Nagaratnam et al., 2021). The first structure of catalytically active human Taspase1 (containing the previously truncated residues) is described here in a short summary below as a partial reprint with permission as a co-first author of the study (Nagaratnam et al., 2021).

To achieve this structure, a circularly permutated construct of Taspase1 (cpTaspase1$_{\alpha41-233/\beta}$) was designed to switch the α and β subunits using a GSGS linker to successfully express the complete protein. Truncations of the protein in the flexible domain (**Supplementary Figure 5.1**) were also created to probe the effect of this domain on of the catalytic function, given the proximity to the active site (**Figure 5.1**). The activity of the expressed proteins was tested using a FRET-based assay (**Supplementary Figure 5.2**). The wildtype (WT) Taspase1 and substrate only were used as controls. The cpTaspase1$_{\alpha41-233/\beta}$ containing the complete flexible domain showed little to no change in activity. Truncation of residues 207 – 233 significantly reduced activity and complete truncation of residues 183 – 233 eliminated all activity (**Figure 5.2**). The major finding is

that the flexible domain is essential for catalytic activity and study of the mechanism for rational drug development.

The structure of catalytically active Taspase1 (cpTaspase1$_{\alpha41-233/\beta}$) was solved by X-ray diffraction to 3.04, 3.04, 5.4 Å in the A, B, and c direction (**Supplementary Figure 5.3**). Crystals of Taspase1 were difficult to obtain, with only one condition creating quality, large crystals for synchrotron diffraction. The structure was solved by molecular

A

```
        10         20         30         40         50              10         20         30         40         50
MTMEKGMSSG EGLPSRSSQV SAGKITAKEL ETKQSYKEKR GGFVLVHAGA        ☆ MTVGAVVVDH EGNVAAAVSS GGLALKHPGR VGQAALYGCG CWAENTGAHN

        60         70         80         90        100              60         70         80         90        100
GYHSESKAKE YKHVCKRACQ KAIEKLQAGA LATDAVTAAL VELEDSPFTN          PYSTAVSTSG CGEHLVRTIL ARECSHALQA EDAHQALLET MQNKFISSPF

       110        120        130        140        150             110        120        130        140        150
AGMGSNLNLL GEIECDASIM DGKSLNFGAV GALSGIKNPV SVANRLLCEG          LASEDGVLGG VIVLRSCRCS AEPDSSQNKQ TLLVEFLWSH TTESMCVGYM

       160        170        180        190        200             160        170        180        190        200
QKGKLSAGRI PPCFLVGEGA YRWAVDHGIP SCPPNIMTTR FSLAAFKRNK          SAQDGKAKTH ISRLPPGAVA GQSVAIEGGV CRLEGSGSGG FVLVHAGAGY

       210        220        230        240        250             210        220        230        240        250
RKLELAERVD TDFMQLKKRR QSSEKENDSG TLDTVGAVVV DHEGNVAAAV          HSESKAKEYK HVCKRACQKA IEKLQAGALA TDAVTAALVE LEDSPFTNAG

       260        270        280        290        300             260        270        280        290        300
SSGGLALKHP GRVGQAALYG CGCWAENTGA HNPYSTAVST SGCGEHLVRT          MGSNLNLLGE IECDASIMDG KSLNFGAVGA LSGIKNPVSV ANRLLCEGQK

       310        320        330        340        350             310        320        330        340        350
ILARECSHAL QAEDAHQALL ETMQNKFISS PFLASEDGVL GGVIVLRSCR          GKLSAGRIPP CFLVGEGAYR WAVDHGIPSC PPNIMTTRFS LAAFKRNKRK

       360        370        380        390        400             360        370        380
CSAEPDSSQN KQTLLVEFLW SHTTESMCVG YMSAQDGKAK THISRLPPGA          LELAERVDTD FMQLKKRRQS SEKENDSGTL DLEHHHHHH

       410        420
VAGQSVAIEG GVCRLESPVN
```

B

WT Taspase1	N	α (1–234)	β (234–420) C

cpTaspase1 α1-206/β, cpTaspase1 α41-183/β, cpTaspase1 α41-206/β, cpTaspase1 α41-233/β

Figure 5.1 Sequence and domain organization of circularly permuted constructs. A) Comparison of the sequence of the circularly permuted construct, cpTaspase1$_{\alpha41\text{-}233/\beta}$ (right), with the sequence of the WT Taspase1 (left). The catalytic site nucleophile Thr234 is highlighted by the yellow star. The α and β subunits are colored in magenta and blue, respectively. In the circularly permuted construct, α and β subunits are swapped and linked through a GSGS linker (green). The GSGS residues replace the SPVN residues in the WT Taspase1 (green, boxed). The residues comprising the long C-terminal helix fragment are underlined. The first 40 residues of the α subunit were deleted except for the N-terminal initiating methionine. The two extra LE residues and the His (H) tag residues at the C-terminus (highlighted in red) are for cloning and for affinity purification purposes, respectively. B) Schematic representation of Taspase1 constructs. Amino acid positions in all constructs are numbered according to the wild type Taspase1 (WT Taspase1). The α and β subunits are highlighted in magenta and blue, respectively. Taspase1$_{\alpha1\text{-}206/\beta}$ construct (PDB 2A8J) (J. A. Khan et al., 2005) shows the two subunits are co-expressed with the missing fragment highlighted by the dashed lines. The three circularly permuted constructs (cp-Taspase1$_{\alpha41\text{-}183/\beta}$, cp-Taspase1$_{\alpha41\text{-}206/\beta}$, cp-Taspase1$_{\alpha41\text{-}233/\beta}$) used in our study are shown: both subunits are swapped and linked through a GSGS tetrapeptide (green). These circularly permuted constructs differ from each other only in the length of the α subunit sequence.
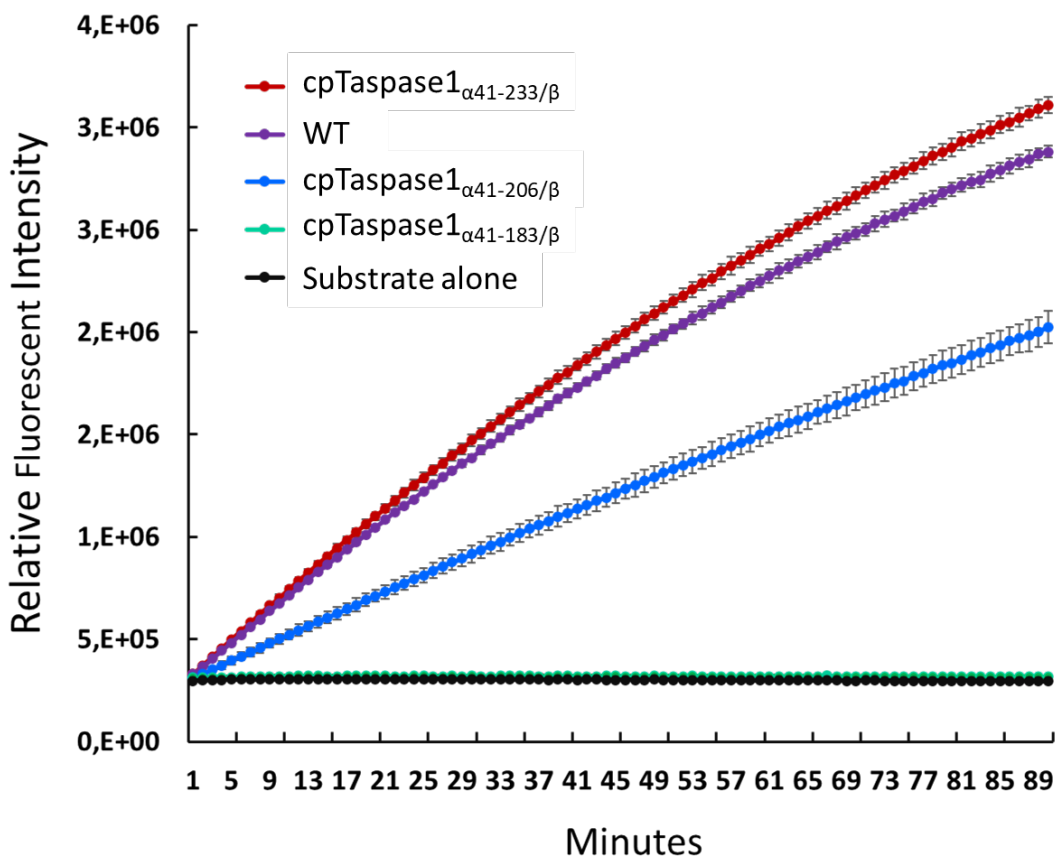
Figure 5.2 Activity assays of the circularly permuted constructs (cpTaspase1$_{\alpha41\text{-}233/\beta}$ (red), cpTaspase1$_{\alpha41\text{-}206/\beta}$ (blue), and cpTaspase1$_{\alpha41\text{-}183/\beta}$ (green)), and the WT Taspase1 (purple). Reaction progress curves of the proteolytic cleavage of the labeled peptidic substrate MCA-Lys-Ile-Ser-Gln-Leu-Asp↓Gly-Val-Asp-Asp-Lys-(DNP)-NH$_2$ by various Taspase1 proteins. The experiment with the substrate alone (control) is represented in black. Error bars have been included.

replacement and underwent anisotropy correction as described in Nagaratnam 2021 [ref]. The structure was refined to values of R$_{work}$ 23.7% and R$_{free}$ 30.7%. Despite the severe anisotropy (**Supplementary Figure 5.6.4**), the 2mF-DF electron density maps of the long helical fragment became already visible during the early stages of data analysis and show high levels of structural details after anisotropy correction (**Supplementary Figure 5.6.5**). The structure of the Taspase1 is shown in **Figure 5.3**. The structure of Taspase1 in combination with the activity assays highlight new considerations for drug development
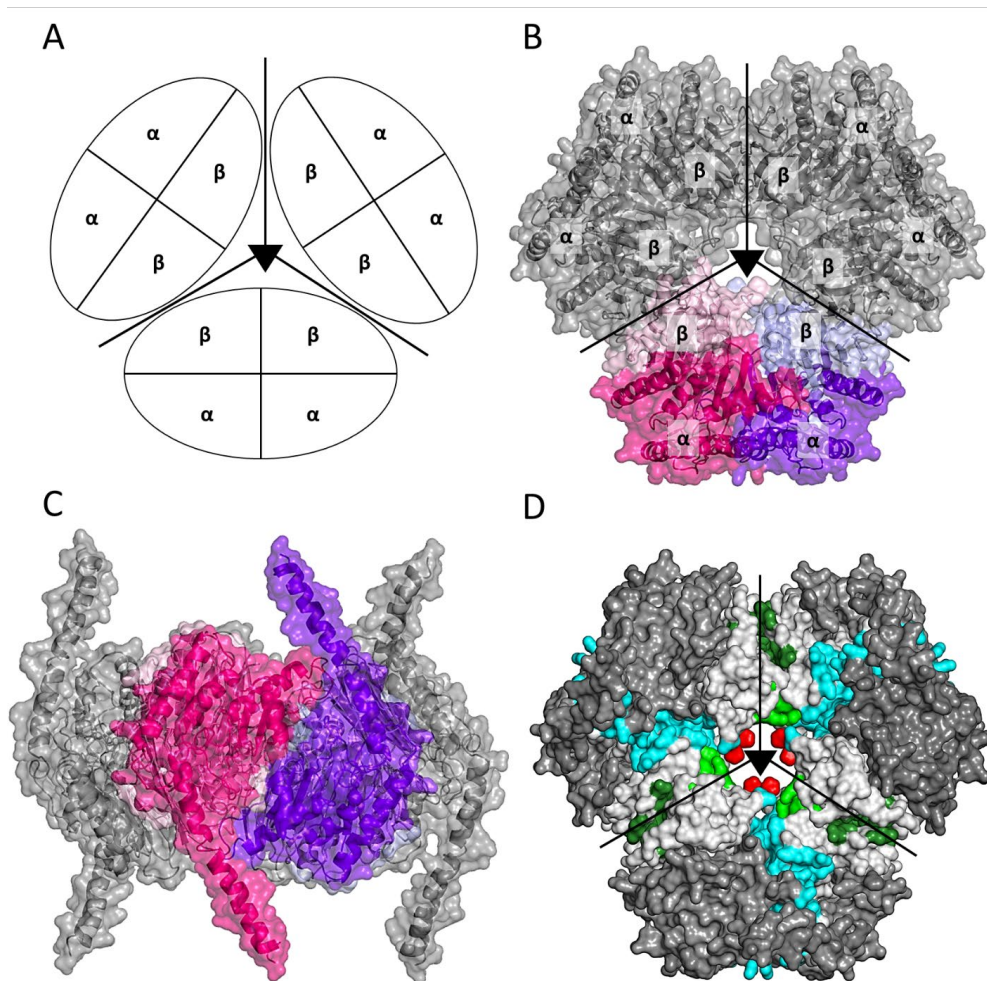
Figure 5.3 Subunit interfaces that maintain the single-ring assembly of cpTaspase1$_{\alpha41\text{-}233/\beta}$ in the crystals. A) Schematic view of the single-ring structure. B) Top view of the surface representation of the cpTaspase1$_{\alpha41\text{-}233/\beta}$ of the single-ring structure. C) Side view of the surface representation of the cpTaspase1$_{\alpha41\text{-}233/\beta}$ of the double-ring structure. D) Interfaces at the single-ring structure. The interfaces are colored as follows: cyan for the dimer interface, green for the intra-ring interface, and red for the histidine interface. The cpTaspase1$_{\alpha41\text{-}233/\beta}$ dimers are represented with the same color as in Figure 3 and the dimers generated by symmetry are shown in a grey and black.

for Tapsase1 as an anti-cancer therapeutic target. The flexible region was identified to a

be a long helix and important for catalytic activity. This region provides an opportunity

for inhibition by drug targeting that is alternative to the catalytic site.

To continue structural studies of Taspase1 to explore the mechanism and ligand

binding, the anisotropy in the Taspase1 crystals must be addressed. Anisotropy in

macromolecular crystallography has been frequently addressed by several methodologies (DeLucas et al., 1989; Van Driessche et al., 2008). First, optimization of the crystallization conditions that may lead to new morphologies, different space groups, or crystals that just have higher quality diffraction may be applied. Barring the improvement of the crystals individually, mutations in the protein altering aspects such as crystal contacts may lead to new crystallization conditions. Finally, anisotropic data correction methods have been established that optimize hit finding parameters for improved structures. Here we describe the continued studies and approaches for resolving anisotropy in Taspase1.

5.3 Methods

Mutagenesis for Stable Crystal Contacts

Robetta (protein structure prediction service through CAMEO) was used to identify key interacting residues between Taspase1 monomers in the alpha helix. Additional molecular dynamic simulations were performed with the program to analyze the identified key residues in the crystal contacts (Layton et al., 2020). Subsequent bioinformatic analysis was performed by aligning the Taspase1 proteins across species to eliminate mutations in conserved residues. Conserved residues were not altered in order to avoid altering protein function.

Protein Expression and Purification

Protein expression and purification was performed as described in (Nagaratnam et al., 2021). As this is continuing work, there are currently no changes to the protocols established by our team. Constructs for the mutants were ordered though ~~GenScript~~ and were expressed and purified through the same methods.

Microcrystallization- Taspase1

Microcrystals of Taspase1 were grown on-site at the PAL XFEL using the free interface diffusion method (FID). Concentrated protein (30 mg/mL) was added to the crystallization buffer (0.1 M sodium citrate, pH 4.0, 1.4 M ammonium sulfate, 15% MPD, 0.4 M sodium chloride) in a 1:3 protein to precipitant ratio. Microcrystals appeared after incubating the mixture for approximately 16 h at 20 °C. The crystals pellet was resuspended at 12 h and allowed to settle for 4h to increase crystal density and pellet volume.

Microcrystallization- D360T

Microcrystallization was performed with the D360T mutant using the same methods as the standard Taspase1 microcrystals and large Taspase1 D360T crystals were grown with slight alterations. For the PAL beamtime, microcrystals of the full length Taspase1, and its mutants, with and without the presence of inhibitors were grown on-site at the laboratories at the PAL XFEL using the free interface diffusion method (FID). The following full length Taspase1 proteins were used: Taspase1, D360T, S371D, and S371E. Additional conditions were set up by incubating each Taspase1 protein with one of three native peptides—NSC 805003, NSC 805004, NSC 805006— or with one of two peptide inhibitors—7238, 7240—for 2 h. Crystallization included two different conditions with variations on pH (4 – 8) and protein to precipitant ratio. In the pH 4.0 condition, 0.1 M sodium citrate pH 4.0, 1.4 M ammonium sulfate, 15% MPD, 0.4 M sodium chloride, was used as the precipitant, the protein concentration was 25 mg/mL, and a protein to precipitant ratio of 1:3 was used. In the pH 7.0 condition, 0.1 M HEPES, pH 7.0, 2.1 M ammonium sulfate, 15% MPD, 0.4 M sodium chloride was used as the precipitant, the

protein concentration was 25 mg/mL, and a protein to precipitant ratio of 1:2 was used. Crystallization experiments for Taspase1, S371E, and S371D were set up at pH 4.0 and crystallization experiments for each of the mutants D360T, S371E, and S371D were set up at pH 7.0. A volume of 100 µL of concentrated protein was placed at the bottom of a 1.5 mL Eppendorf tube. Then a total volume of either 300 µL for a 1:3 protein to precipitant ratio or 200 µL for a 1:2 protein to precipitant ratio was added drop wise in steps of 100 µL with buffer vortexing between each step.

Microcrystals appeared after incubating the mixture for approximately 16 h at 20 °C. The crystal pellet was resuspended at 12 h and allowed to settle for 4 h to increase crystal density and pellet volume. A homogenous sample containing crystals up to 10-20 µm in size was obtained and used for SFX experiments on days one, two, and three of our PAL XFEL beamtime. Microcrystal size, homogeneity and density were further increased in a second round of FID experiments with seeds. Seeds approximately 10-20 µm in size were obtained from the first FID experiments in the crystal pellet. Then, 10 – 20 µm seeds were used to feed the precipitant solution prior to the FID experiment. A highly dense crystal pellet containing crystals up to 50 and 75 µm in size was obtained and used for SFX experiments on days two and three of our beamtime.

SFX Experiments at LCLS

Microcrystals (Taspase1 standard) were grown on site to 15-20 µm in size with the addition of 10% PEG 400 to the crystallization buffer. Samples were delivered by double flow-focusing nozzle (DFFN) with an inner sheath of ethanol and outer jet of Helium gas in vacuum. Experiments were performed in the back chamber of CXI with a

pulse length of 40 fs, repetition rate of 120 Hz, photon energy of 9.5 keV, transmission of 6%.

LCP Mixing of Sample

Microcrystals were subsequently reconstituted into lipidic cubic phase (LCP) by using the conventional LCP setup for SFX experiments [reference] with modifications. Briefly, molten monoolein (MAG 9,9) was firstly mixed with precipitant solution prior to adding microcrystals to form an LCP phase to host microcrystals. The microcrystal slurry was then added and gently mixed with the LCP phase to form a more liquid-ish phase also called the sponge phase. This phase was brought to the lipidic cubic phase (or LCP) by adding more lipid in subsequent steps. Monoolein MAG 7,9 was also added to prevent lamellar phase formation during the experiment at CXI in vacuum. The mixing process was carried out at 20°C at all times.

SFX Experiments at PAL

Samples were delivered by LCP injection with a flow rate of … nL/min. Serial fs crystallography diffraction data were collected at a pulse length of 20 fs, repetition rate of 30 Hz, photon energy of 9.7 keV, with $2x10^{11}$ photons/pulse with 100% transmission. X-ray diffraction images were recorded with a Rayonix MX225-hs detector.

Data Analysis

Data analysis was performed on the *cpTaspase1*$_{α41-233/β}$ crystals for data collected at PAL XFEL with sample delivery in LCP. Hit finding and pre-processing of the data to identify potential diffraction patterns for indexing was performed by Cheetah (Anton Barty et al., 2014). The images were then further processed with CrystFEL (White, 2019; White et al., 2012; White et al., 2016). Auto indexing was performed with the following

programs: *MOSFLM, XDS, DIRAX, TAKETWO,* and XGANDALF. Montecarlo

integration was performed in the *CrystFEL* suite (White, 2019; White et al., 2012; White

et al., 2016). The structure was solved by molecular replacement with *Phenix.phaser* and

refinement was done in *Phenix.refine* (Afonine et al., 2012; Liebschner et al., 2019).

5.4 Results

Molecular dynamics studies were performed on the long, flexible alpha helix

(residues 332 – 381) of Taspase1 to identify residues that, if mutated, could improve the

stability of the crystal contacts and decrease anisotropy. Ten residues were identified

Figure 5.4. Sequence of Taspase1 (native and circular permutated construct) with the alpha helical region highlighted. The arrows indicate the position of identified residues that form crystal contacts. Green arrows indicate conserved residues and red arrows indicate non-conserved residues. (difficult to see the difference between green and red arrow can you remove the black border around the arrows which will allow for easier discrimination, also for publication we would need to choose a different color scheme as colorblind people cannot discriminate between red and green)

(listed with the numbering scheme of the circular permutated Taspase1 construct) to

form interactions in crystal contacts: 201, 202, 350, 353, 357, 360, 361, 364, 371, and

376 (**Supplementary Table 5.6.3**). Sequence analysis of Taspase1 across 41 difference

species was performed (**Supplementary Figure 5.6**) to identify non-conserved residues for mutation, leaving D360 and D371 as possible candidates for mutation (**Figure 5.4**).

Seven mutants were designed: D360T, D360Y, D360F, S371I, S371D, D360T/S271D, and D360Y/S371D. Of the five single mutants, the S371E and D360T expressed in the greatest quantities (**Figure 5.5**). Mutation of the D360T residue enabled new crystal morphologies under physiological pH values of 6.6, 7.0, and 7.5 (**Figure 5.6**); We collected X-ray diffraction data on these crystals at 3.0 Å however, anisotropy in was still observed in the X-ray diffraction data.

Figure 5.5. Expression of Taspase1 mutants as compared to the wildtype (*cpTaspase1*$_{\alpha 41\text{-}233/\beta}$)
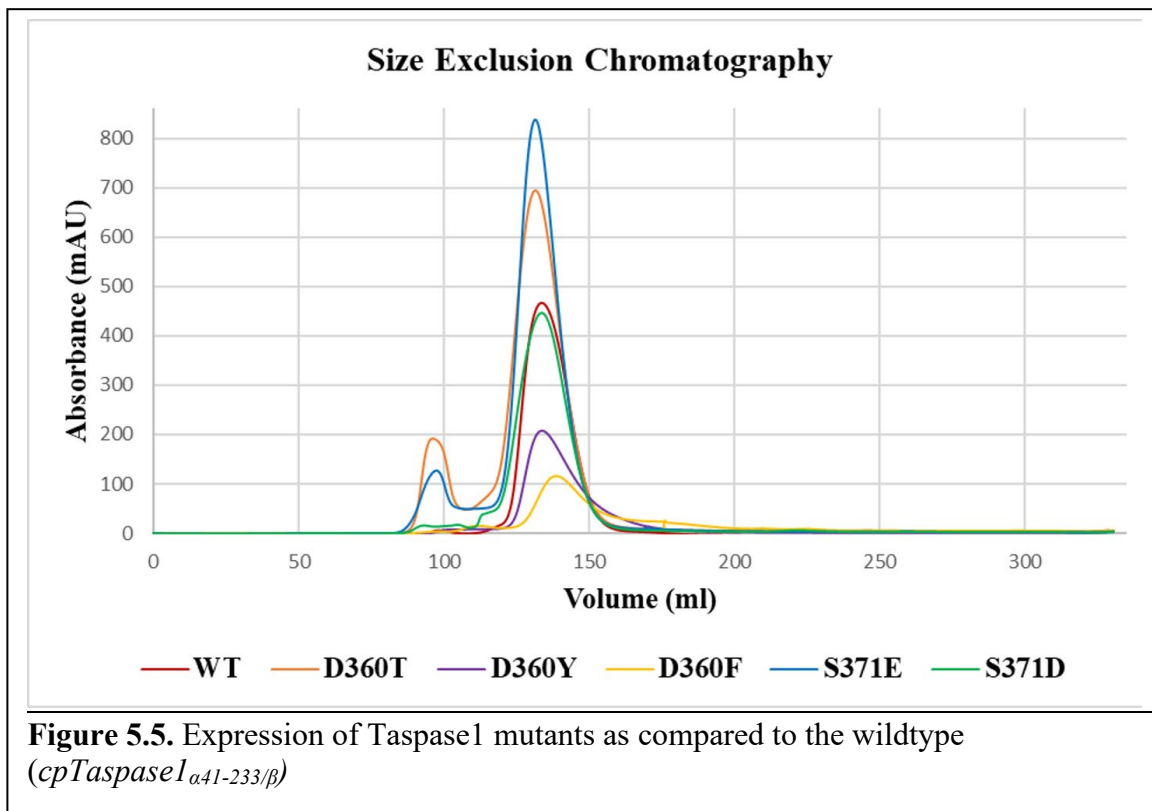
Serial fs crystallography data of crystals of the wildtype (WT) Taspase1 mircocrystals collected at LCLS during a protein crystallization screening beamtime in the back chamber of CXI (**Figure 5.7**). Microcrystals were used to minimize potential crystal imperfections leading to the poor diffraction of the large crystals. Sample was

delivered with the doble flow-focusing nozzle (DFFN) in vacuum conditions. Crystals were screened for optimal conditions and size distribution. Diffraction fraction was observed to ~4 Å and indexed in the H32 space group (**Figure 5.7**).

Subsequent SFX experiments were performed at PAL-XFEL in S. Korea. The

Figure 5.6. Crystallization of Taspase1 D360T under physiological conditions (pH 6.6, 7.0, and 7.5).
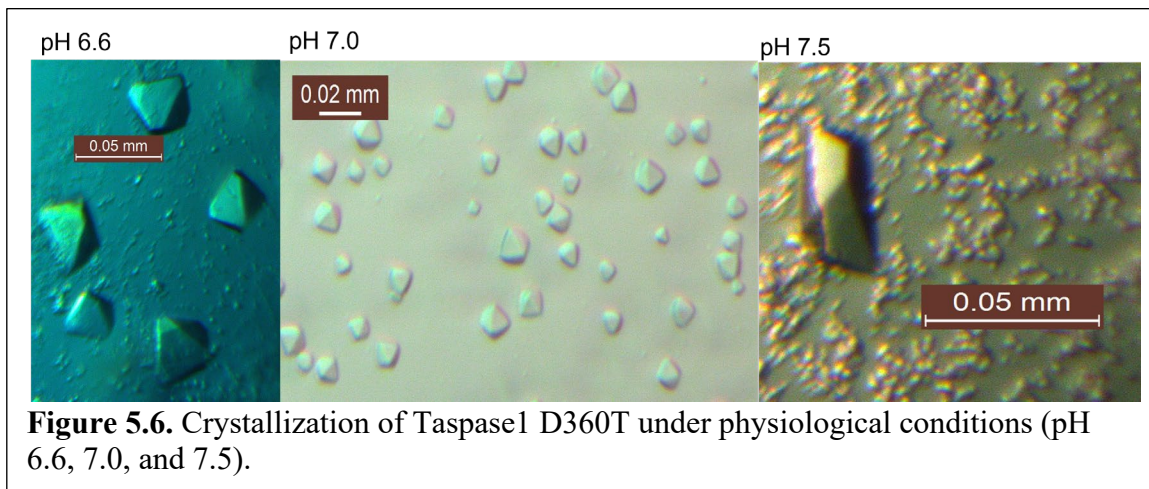
sample delivery method was changed to viscous injection in LCP due to difficulties in jetting the crystals in a liquid jet in the crystallization buffer as upon evaporation in the He atmosphere salt crystals of ammonium sulfate formed on the nozzle tip. We implemented an automated "nozzle-washing" procedure where every minute the outside of the nozzle was flushed with water during data collection. The experiments were conducted in an enclosure of the beamline extra built by the beamline scientists for our experiments to allow for sample delivery at 20°C, as the crystals were very temperature sensitive dissolved at 25°C, which is the temperature of the experimental CXI PAL-XFEL crystallography endstation (**Figure 5.8**). However, very few "hits" were detected in the liquid jet, and we could not increase the crystal density due to clogging. We therefore decided to test a new approach where crystals were embedded in LCP. The

crystals had a 5–10 μm size distribution (**Figure 5.8C**) and were grown on site in the

PAL XFEL Labs with monodispersed Taspase1 crystal seeds (<1 μm) (**Figure 5.8B**).

Figure 5.7. LCLS Taspase 1 microcrystals serial crystallography diffraction experiment: Pulse length 40 fs, repetition rate 120 Hz, photon energy 9.5 keV, transmission 6%. A) Micro-crystals (15-20 μm) of the active full-length Taspase1 protein grown onsite at 4°C using a combination of free interface diffusion (FID) (Kupitz et al., 2014) and batch with seeds. The crystallization precipitant solution was 0.1 M sodium citrate pH 4.0, 0.4 M NaCl, 1.4 M Ammonium SO4, 10% MPD, 10% PEG 400. Protein concentration was 20 mg/mL and protein to precipitant ratio 1:3. B) Diagram of the SFX experiment conducted at LCLS. A double flow-focusing nozzle (DFFN) was used to deliver the crystals in their crystallization buffer (inner green stream) into the sheath jet formed by ethanol (middle blue stream) through a 50 μm ID capillary. The sheath liquid was accelerated by Helium gas (outer blank channels). C) and D) Non-indexed and indexed diffraction patterns, respectively, of a single micro-crystal of the active full-length Taspase1. Bragg spots were seen to ~ 4Å. E) Unit cell size distribution plot of 771 indexed patterns into the H32 space group.

Figure 5.8. PAL-XFEL with details on the SFX experiments conducted at the microcrystallography CXI endstation. X-ray diffraction data were collected with the following X-ray parameters: Pulse length 20 fs, repetition rate 30 Hz, photon energy 9.7 keV, 2×10^{11} photons/pulse, transmission 100%, X-ray diffraction images were recorded with a Rayonix MX225-hs detector. A) PAL-XFEL facility. B) DLS experiment of Taspase1 crystal seeds (<1µm). C) Micro-crystals (5-10 µm) of Taspase1 grown onsite at 20°C using FID method with seeds (Kupitz et al., 2014) in 0.1 M sodium citrate pH 4.0, 0.4 M NaCl, 1.4 M AmSO4, 15% MPD. Protein concentration 30 mg/mL and protein to precipitant ratio 1:3. D) Sample chamber at PAL-XFEL NCI beamline. E) ASU LCP injector holder assemble view (inside the

After embedding Taspase crystals in LCP, diffraction data were collected using the viscous injector developed by the Weierstall group at ASU (Nogly et al., 2016; Weierstall, 2014). Taspase1 diffraction was still anisotropic. The structure was solved to 4.5 Å with R-work and R-free values of 0.3513 and 0.3633, respectively (**Figure 5.9**) (**Table S5.6.4**).

Protein Crystal Growth in Microgravity

Large Taspase1 and Taspase1 D360T crystals were grown under low (pH 4.0) and physiological pH conditions as part of the protein crystal growth 18 (PCG-18) mission for growing crystals in microgravity. Conditions for large crystals were set in 96-well

Figure 5.9. Room temperature structure of Taspase1. **(A)** Anti-settling device for crystals. **(B)** Diffraction pattern of Taspsae1 at PAL-XFEL. **(C-D)** 2Fo-Fc maps of room temperature Taspase1 at 1σ level. **(C)** Active site threonine (blue) of Taspase1 with neighboring flexible loop (green). **(D)** View of flexible loop.
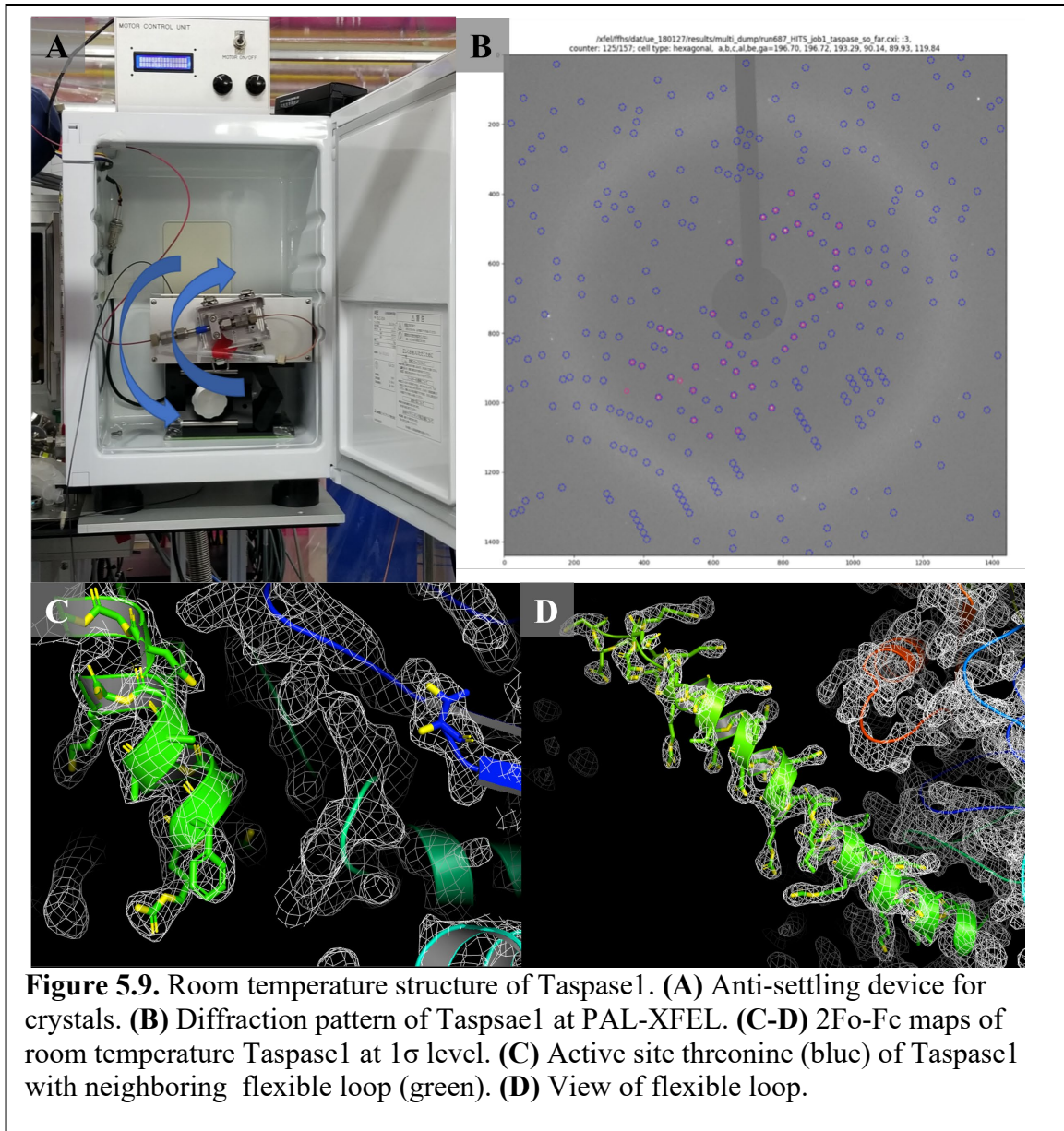
plates and vacuumed sealed in bags at the Space Life Science Lab (SLSL) at Kennedy

Space Center (KSC) in Merritt Island, Florida. Plates were delivered to the International

Space Station (ISS) National Laboratory via Commercial Resupply services mission 17

(CRS-17) seven days prior to set up. Plates were returned after 30 days and crystals were

observed and diffracted at the Advanced Photon Source (APS) in Chicago, IL. Some

crystals had a larger morphology; however, anisotropic diffraction was still observed.

5.5 Discussion

Taspase 1 is an important novel cancer target as loss of Taspase1 activity has been demonstrated to disrupt proliferation of human cancer cells *in vitro* and in mouse tumor xenograft models of glioblastoma. By functioning as a non-oncogene addiction protease, Taspase1 regulates key proteins and genes for cancer cell proliferation, invasion, and more[refs]. Inhibition of Taspase1 by an inhibitory drug may provide a mechanism for an anti-Taspase1, anti-cancer therapeutic, for which there are no known competitors in the drug development arena. The importance for the flexible domain in catalytic activity and corresponding structure was established in earlier studies (Nagaratnam, 2021). However, the anisotropy seen in diffraction current limits continuing drug development studies. Methods for resolving this challenge for future studies were explored here in more detail.

Mutagenesis of Taspase1 was performed along the flexible alpha helix. The goal of these mutations was to enhance crystallization contacts and change residues that weaken crystal contact such as changing a negatively charged residue to a polar, uncharged residue as seen in D360T. Residues were identified based on the crystal contacts and narrowed down to non-conserved residues to maintain the native function of the Taspase1 protease. Mutations of these residues were also based on molecular dynamic simulations from our collaborator Abhishek Singharoy (Layton et al., 2020). Five mutants containing single point mutations were expressed and purified: D360T, D360Y, D360F, S371E, and S371D. Crystallization of the mutants were explored for each protein. Only the D360T protein produced large, single crystals for diffraction. The pH of the D360T during crystallization could also be increased from a maximum of pH 4 for the wildtype to pH 7.5 for D360T for near physiological conditions. While new

crystallization conditions were established, diffraction of the mutants at a synchrotron light source still showed severe anisotropy similar the standard Taspase1 protein.

Microcrystals of Taspase1 were explored to see if smaller crystals were more well-ordered due to decreased long range disorder, which is often observed in large crystals. The first microcrystals of Taspase1 WT were tested for their diffraction by SFX at a PCS beamtime at LCLS with the DDFN injector [ref] in vacuum. With the crystals further optimized, the experiment advanced to PAL-XFEL for two beamtimes of three shifts each. Microcrystals were also produced using the D360T and S371E mutants, including co-crystallization of the mutants with inhibitors. However, without the vacuum conditions, the sample had difficulty jetting compared to the LCLS beamtime, which was performed in vacuum. The sample injection was then switched to high viscosity LCP injection. The crystals could be resuspended carefully in a "sponge phase," but were prone to dissolving due to temperature increases during embedding of the crystals into the LCP with a syringe coupler. In both injection methods, buffer component MDP (an alcohol) caused problems. In jetting the crystals in the mouther liquor, it was hypothesized that the gas layer and low humidity in the chamber was altering the MPD concentration of the buffer which may have led to the dissolving of the crystals. Phase separation was also observed in the sample and may have dissolved the crystals during the mixing of LCP in addition to prior observation of temperature sensitivity. Crystals could be successfully resuspended in LCP during the first PAL beamtime, where a very slow mixing and embedding procedure was used by Dr. Martin-Garcia for embedding and data was collected. However, he could not join the second beamtime in person and even with his remote participation, we could not accomplish the LCP embedding on site

at PALXFEL, without destroying the crystals. When we are comparing the exact times for each step after the beamtime, we realized that the key might be extremely slow mixing of the crystals with the LCP as this step took 45 minutes in his procedure, while we performed mixing in 15 minutes.

The collected data was still noted to be anisotropic and a could not be corrected for with the anisotropy servers due to compatibility of the software. This issue has since been resolved and data evaluation is currently underway to explore if the use of serial femtosecond crystallography improved diffraction quality.

In continued efforts to resolve the anisotropic diffraction of Taspase1 crystals, Taspase1 crystallization plates were sent to the International Space Station (ISS) as a part of the Commercial Resupply Service (CRS-17) launch with SpaceX. Protein crystal growth under the microgravity conditions of the ISS allow for multiple advantages. First, crystals do not settle to the bottom of the experiment, which can restrict crystal growth on the contacting surface. Additionally, the reduced kinetics of crystal growth under microgravity can allow crystals to growth with fewer defects than protein crystals grown on Earth (DeLucas et al., 1989). The crystallization plates were prepared at Kennedy Space Center (KSC) two days prior to launch at 4 °C. The temperature was selected to slow protein crystal growth to appear in one week versus in two days at room temperature. This was an important consideration as the payload specialists need two days to transfer the experiments between completion and launch. However, the launch was delayed over a week due to power grid issues aboard the ISS and permit issues for recovery of a booster rocket on the drone ship, "Of Course I Still Love You." The launch delay is hypothesized to allow for initial crystal growth under Earth's gravity.

5.6 Supplemental Information

5.6.1 Supplementary Figures
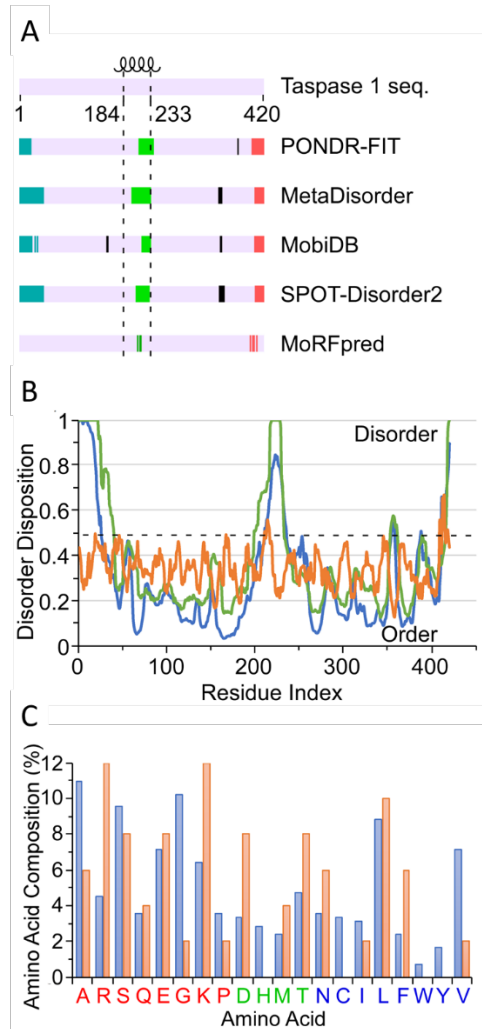
Supplementary Figure 5.1. Prediction of intrinsically disordered regions (IDRs) of Taspase1. A) Disorder predictions in three discrete regions of Taspase1 (N-terminal (cyan), middle (green) and C-terminal (red)) by PONDR-FIT(Xue et al., 2010), MetaDisorder(Kozlowski & Bujnicki, 2012), MobiDB(Piovesan et al., 2018), MoRFpred(Disfani et al., 2012), and SPOT-Disorder2(Hanson et al., 2018). Short disordered regions are represented in solid black. The location of the long helix fragment of Taspase1 is delimited by the black dotted line. B) Disorder disposition of each residue in Taspase1 sequence predicted by PONDR-FIT (blue), MetaDisorder (green) and MoRFpred (dark orange). Residues with disorder probability higher than 0.5 (black dotted line) are predicted to be disordered. C) Comparison of the composition of each amino acid in the long helix (coral orange) with full-length Taspase1 (blue). Disorder-promoting

residues are red, order-promoting residues are blue and disorder-order neutral residues are green.
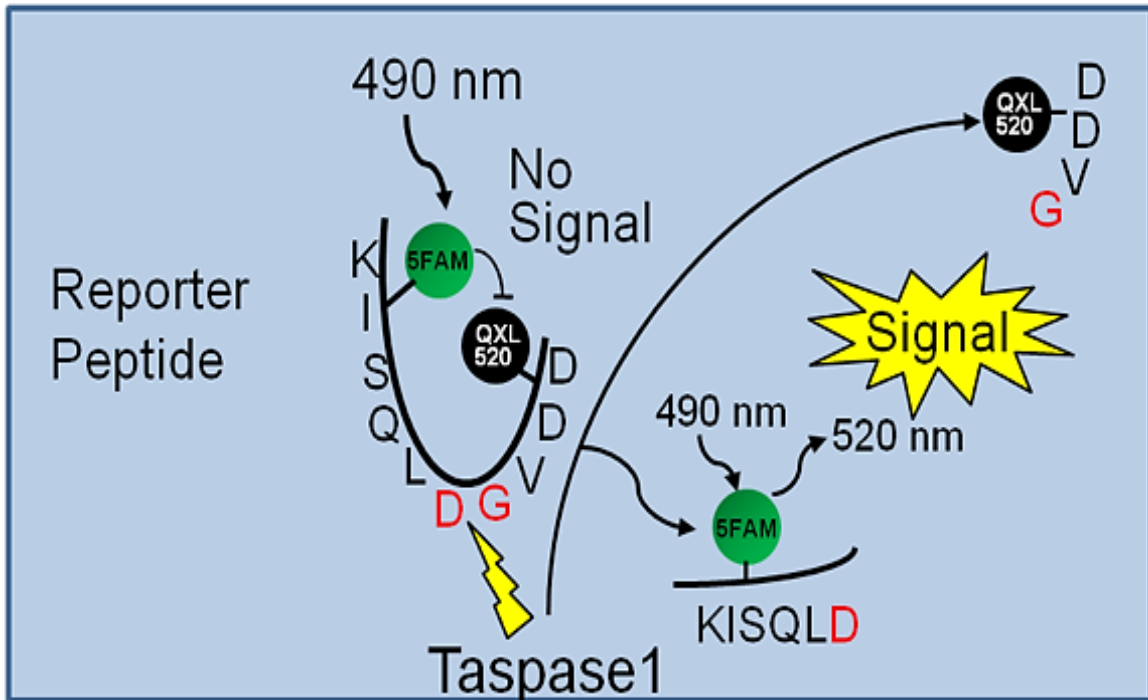
Supplementary Figure 5.2. Schematic picture of the 12 point dose response assay for Taspase1 full length split (FL split) enzymatic activity tests. The assay was performed using a quenched FRET based assay that measures proteolysis of a labeled Taspase1 substrate by purified Taspase1 FL split enzyme prepared and purified by Beryllium. The substrate consisted of the peptide substrate derived from the second Taspase1 cleavage site (KISQLD↓GVDD) of the endogenous Taspase1 substrate MLL, flanked by a 5 FAM fluorophore and a QXL520 quencher. Proteolysis of the peptide substrate releases the quencher from the fluorophore and generates a fluorescence signal. The fluorescence signal was read on an EnVision multi-label plate reader using a FITC filter set.
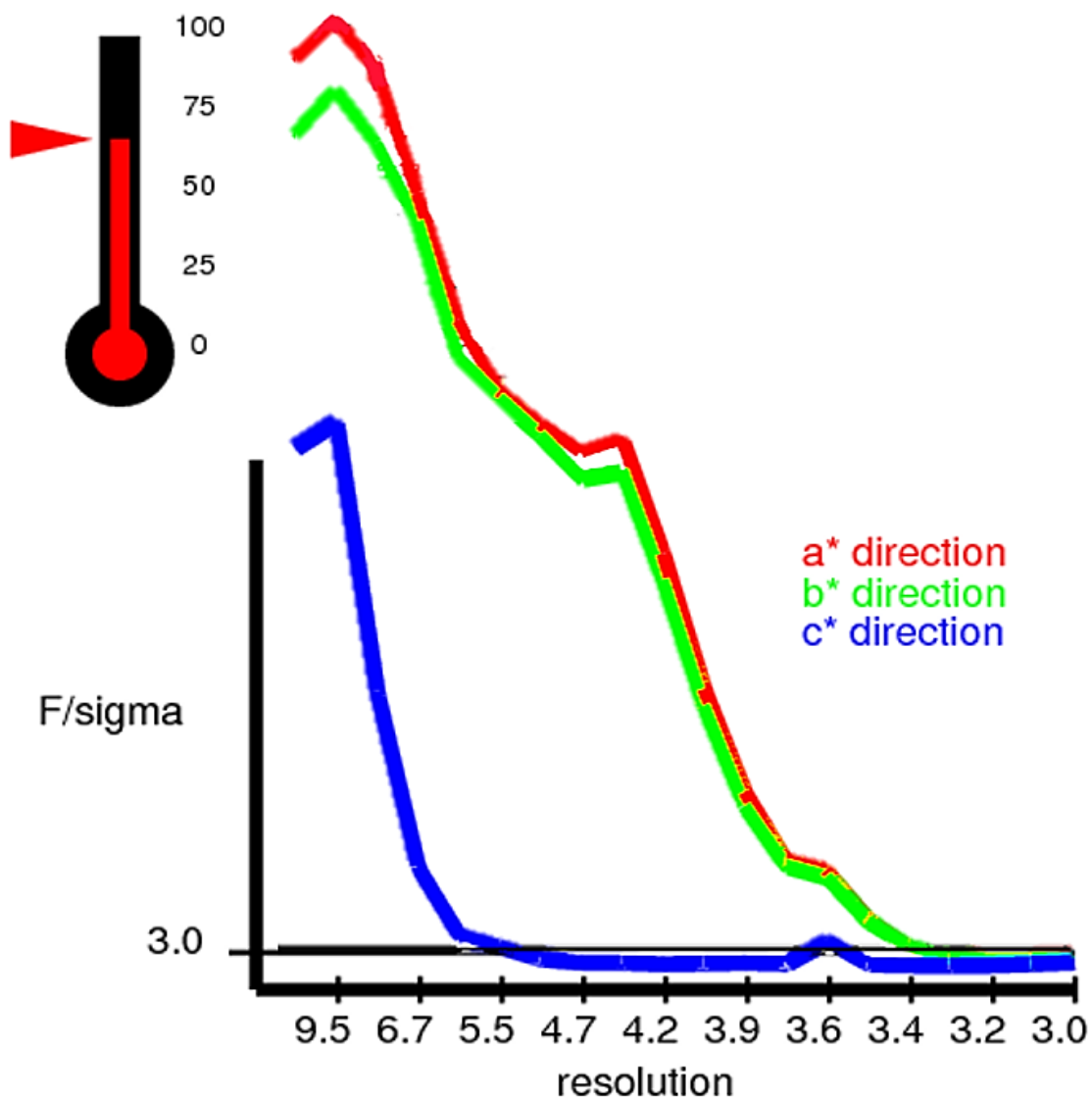
The recommended resolution limits along/near to a*, b*, c* are

3.1 Å   3.1 Å   5.4 Å

These are the resolution limits at which the F/sigma drops below an arbitrary 3.0 value

Supplementary Figure 5.3. Resolution limits and B factor of the structure, showing anisotropy in the C direction.
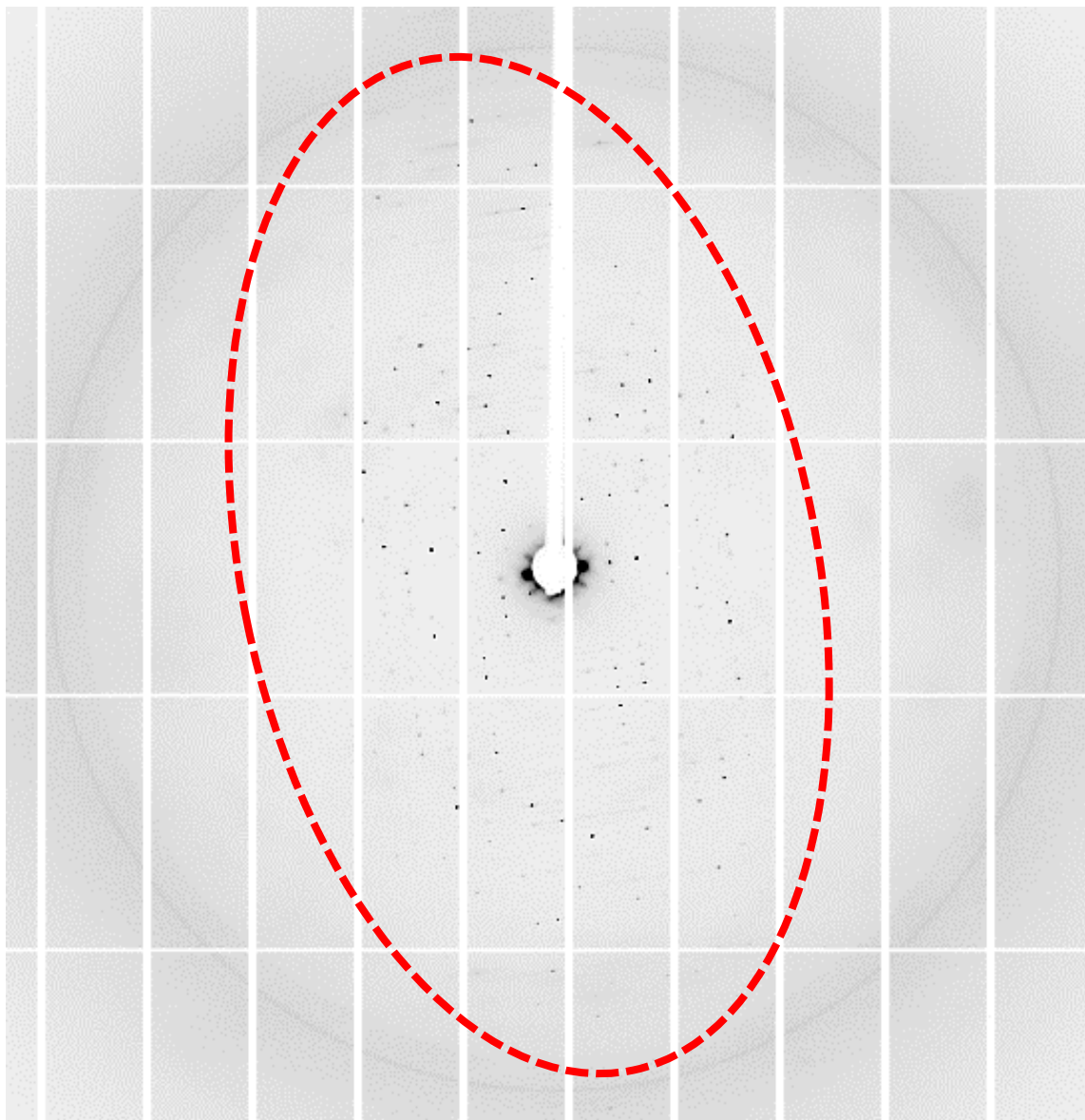
Supplementary Figure 5.4 Sample diffraction pattern showing anisotropy of Taspase1 diffraction.
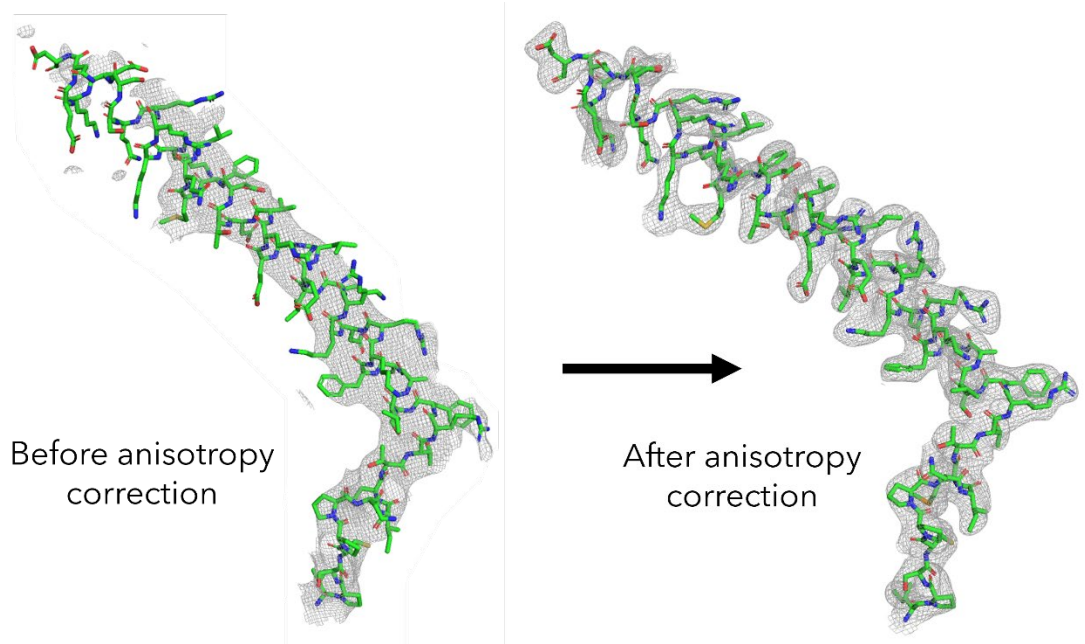
Supplementary Figure 5.5 Electron density maps (countered to 1σ) of the long helical fragment (residues Arg377-Asp415) before and after anisotropic correction. Anisotropy analysis was done using the diffraction anisotropy server at STARANISO available online

```
                                                                    β1
                         1        10        20            30        40        50
Human                 MTMEKGMSSGEGLPSRSSQVSAGKITAKE.......LETKQSYKEKRGGFVLVHAGAGY
Chimpanzee            MTMEKGMSSGEGLPSRSSQVSAGKITAKE.......LETKQSYKEKRGGFVLVHAGAGY
Rhesus-macaque        MTMEKGMSSGEGLPSRSSQVLAGKITAKE.......LEAKQSYKEKRGGFVLVHAGAGY
Tarsier               MTMEKGMSSGEGLPSRSSQVSAGKTAVKE.......LEAKWTHKEKRGGFVLVHAGAGY
Naked-mole-rat        MTMEKGMSSGQGLPSRSPQASAVKLTAKE.......LEAKQSHKEKRGGFVLVHAGAGY
American-beaver       MTMEKGMSSGEGLPSKSSQASAVKITAKE.......LETKQSYKEKRGGFVLVHAGAGY
Chinese-hamster       ..MEKGMSSGEGLPSRSPQASAAKVTVKE.......SETRQPCEEKRGGFVLVHAGAGY
Treeshrew             MTMEKGMSSGEGLPCRSSQVSAVKITVKE.......LETKQSHKEKRGGFVLVHAGAGY
Wedell-Seal           MTMEKGMSSGEGLPSRSSQVSAVKITVKE.......LETKQSHKEKRGGFVLVHAGAGY
Walrus                MTMEKGMSSGEGLPSRSSQVSAVKITVKE.......LETKQSHKEKRGGFVLVHAGAGY
Domestic-yak          MTMEKGMSSGEGLPSRSSQVSAVKTAVKE.......LEAKQSHREKRGGFVLVHAGAGY
Domestic-pig          MTMEKGMSSGEGLPSRSSQVSAVKITAKE.......LETKQAHKEKRGGFVLVHAGAGY
Dolphin               MTMEKGMSSGEGLPSRSSQVSAVKITAKE.......LETKQPHKEKRGGFVLVHAGAGY
Brandt_s-bat          MTMEKGMSSGEGLPSRSSQVSAVKITAKE.......LETKQSHKEKRGGFVLVHAGAGY
Horse                 .........................................................GFVLVHAGAGY
Hedgehog              MTMEKGMNSGEGLPSRSSQFSPVKVTTKE.......LEAKQSYKEKRGGFVLVHAGAGY
Emperor-penguin       MIMEKGISSVEVLPSKSSQVTAVKVVVKE.......PETKQTQRGKRVGFVLVHAGAGY
Parrot                MIMEKGLSSVEVLPSESSQVTAVKVVVKE.......PETKQTQRGKRVGFVLVHAGAGY
Hummingbird           MIMEKGISPVEVLPSKSSQVTAVKVLVKE.......PEAKQTQRGKRVGFVLVHAGAGY
King-cobra            MIMEKGISSVEVP.PVSAHATAVRMMVKE.......HEAKPAQKVRRVGFVLVHAGAGY
Green-sea-turtle      ...MS......GDCSIATLTSALLLMAAL.......PSELGGRRAVAAGGREPRCEGS.T
American-alligator    ..MEKGMSSIEVLPSNSSQITAVKVVKE.......HETKQTQRGKRVGFVLVHAGAGY
Japanese-puffer-fish  ..............K..KE...ESSKNQKAKPVGGFVLVHAGAGY
Turquoise-killifish   ..MESLISYAEEQPPSSLMTNSSSL.KQ...ESNESGSPKCVGGFVLVHAGAGY
Japanese-rice-fish    ..MESVMNCADEQQPSS.MANLWSN..KQ...ESSRTQKAK..GGFVLVHAGAGY
Nile-tilapia          ..MESLMNSAEEKQPSSPLTNSWAK..KQ...ESSKSQKPNGVGGFVLVHAGAGY
Zebrafish             ..MENLENSSDSKETVSPLNNS.GF..KL...NSTISHRDGKNMGFVLVHAGAGY
Phytophthora-plant-pathogen                           MSDWLLAVHCGAGR
Achlya-plant-pathogen                                 MWAVGVHIGAGR
Pichia-yeast          .....MYFPFPQFPPSSNQFNLVIDVEQQRSYGIYVLYEVNRNMNGKR..IVIVHCGAGG
Clavispora-yeast                                      ..MLVVHLGAGR
Penicillium                                  MPQTSDNDRVCAIFLHAGAGF
Aspergillus                                  MSRPR.RDNVSAIFVHAGAGF
S.pombe-fission-yeast                                 MSNGLVILHAGAGL
White-beech-mushroom                                  MSLRGVYIAVHGGAGV
Hydra-fresh-water-polyp  M.......................TQRCSGFVCLHIGAGY
Lamp-shell            M.......................ADDGIEGFVAVHAGAGF
Pacific-oyster                                        MANYVIAVCH....
Japanese-scallop                                      MTGVFIAVHAGAGY
Chinese-liver-fluke                                   MEKRICVAVHAGAGY
tr|A0A094ZUG7|A0A094ZUG7_SCHHA                        MACVVVHAGAGY
consensus>70          ..................................................gfv.vhagagy
```

```
                          α1                              α2
                        lll..llllllllllllllllllll      lllllllllllll          TT
                                 60        70        80        90        100
Human                 HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Chimpanzee            HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Rhesus-macaque        HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Tarsier               HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Naked-mole-rat        HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
American-beaver       HSESKAK..EYKHVCRRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Chinese-hamster       HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Treeshrew             HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Wedell-Seal           HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Walrus                HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Domestic-yak          HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Domestic-pig          HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Dolphin               HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Brandt_s-bat          HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Horse                 .....................................................................
Hedgehog              HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALVELEDSPFTNAGMG
Emperor-penguin       HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALIELEDSPFTNAGLG
Parrot                HSESKAR..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALIELEDSPFTNAGLG
Hummingbird           HSESKAK..EYKHVCKRACQKAIEKLQ......DGALATDAVTAALIELEDSPFTNAGLG
King-cobra            HSESKAK..EYKHVCKRACRKAIEKLQ......AGALVTDAVTAALVELEDSPFTNAGMG
Green-sea-turtle      ASSNIKK..KFHRFCPGECGQSND.........TEFRIYEVMDSPFTNAGMG
American-alligator    HSESKAK..EYKHVCKRACQKAIEKLQ......AGALATDAVTAALIELEDSPFTNAGMG
Japanese-puffer-fish  HSESKAK..EYKHVCKRACQRAVDQLN......AGALAVEAVAAALVELEDSPFTNAGMG
Turquoise-killifish   HSESKAK..EYKHVCKRACQRAVDQLR......AGALAVEAVTAALVELEDSPFTNAGLG
Japanese-rice-fish    HSESKAK..EYKHVCKKACQRAVDQLK......AGALAVEAVAAALVELEDSPFTNAGMG
Nile-tilapia          HSESKAK..EYKHVCKRACQRAVDQLK......AGALAVEAVAAALVELEDSPFTNAGMG
Zebrafish             HSESKAK..EYKHVCKRACKAVDRLR......AGGLALEAVTAALVELEDSPFTNAGTG
Phytophthora-plant-pathogen  YGSSHEA..AYLSLLRSALESAHQVIASS....RSAPTATQIAVHLLQCFERSSLTNAGLG
Achlya-plant-pathogen HQSAPAAVAAAETSMRQALAVAASSLRD.....GASAVEAAAQAVMVLEDAACTNAGSN
Pichia-yeast          KLPSSVE..RRLV...RVVKKAFEA.G......VGDGHNDRFANIARVLEDSPLTNTGYG
Clavispora-yeast      HSSTFNE..KYRKLLRRALGSDLR.........SAARCIEHSALTNTGYG
Penicillium           HSHENEH..KHLKACEVAALKAMAFLKS.....GGTAIDAVEAAIMVLEDNPITNAGFG
Aspergillus           HSPHNEK..LHLETCENAAKVAIHMLKN.....GGSAVDAVEIAIMLLEDSEITNAGYG
S.pombe-fission-yeast YSGQREI..QAKKTCSDACKAAIQALKV.....GQSALSAAIQAAKIMEDSPVTNAGVG
White-beech-mushroom  HSHTRED..EVKQALRVACTEGLKCIGITTKAQRDGTRLQMVENAVSCLEDNPNLNAGYG
Hydra-fresh-water-polyp  HSINKEN..LYINLCKKACMLGNKLLK......EGNTCIDIVEKVCCFLEDSDLTNAGFG
Lamp-shell            HSPNKTQ..LYSEVCQLACKKAAEALH......DKKSALDAVVMAVTQLENSACTNAGIG
Pacific-oyster        ...............MLE......KGSEAVECVTEAVKLLEDSKLTNAGRG
Japanese-scallop      HSRDRES..EYKKACAEACSKAHGVLT......TGASSLEAVMAAVISLENCPLTNAGTG
Chinese-liver-fluke   HARSKEQ..LYNSLCSEACKIASSLLV......KGSNAVDAVAAVQYLENSDLTNAGYG
tr|A0A094ZUG7|A0A094ZUG7_SCHHA  HSKSKEK..KYNKLCSEACSLAGKLLSV......DHKSAIDAVEAAVAYLEDCPLTNAGIG
consensus>70          hs.sk.....y...c...ac..a.e.l....g...a.dav.aa...ledsp.tnag.g
```

```
                                 TT      β2      TTT      β3             α3
                                 110       120        130         140       150

Human                           .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Chimpanzee                      .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Rhesus-macaque                  .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Tarsier                         .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Naked-mole-rat                  .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
American-beaver                 .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Chinese-hamster                 .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Treeshrew                       .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Wedell-Seal                     .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Walrus                          .SN..LNLIGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Domestic-yak                    .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Domestic-pig                    .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Dolphin                         .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Brandt_s-bat                    .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Horse                           ...........................................................
Hedgehog                        .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLYEGQKGK.L..
Emperor-penguin                 .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Parrot                          .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Hummingbird                     .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLYEGQKGK.L..
King-cobra                      .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSG....
Green-sea-turtle                .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSG....
American-alligator              .SN..LNLLGEIECDASIMDGKSL.NFGAVGALSGIKN.PVSVANRLLCEGQKGK.L..
Japanese-puffer-fish            .SN..LNLLGEIECDASIMDGKSL.QYGAVGSISGVKN.PVLVANRLLSEAQRGK.L..
Turquoise-killifish             .SN..LNLSGEIECDASIMDGKSL.HFGAVGSISGIKN.PVLVANCLLNEAQKGK.L..
Japanese-rice-fish              .SN..LNLSGEIECDASIMDGKSL.HFGAVGAISGIKN.PVSVANRLLKEGQKGK.L..
Nile-tilapia                    .SN..LNLSGEIECDASIMDGKSL.HFGAVGAISGIKN.PVMVANRLLSEAQKGK.L..
Zebrafish                       .SN..LNLSGEVECDASIMDGKSL.NYGAVGALSGIKN.PVLVSRRLLSETQKGK.L..
Phytophthora-plant-pathogen     .SN..LTERGQVECEASVVCGETH.LVSCCGAVRGVKD.PSALALKLLEQNQDCD.TNS
Achlya-plant-pathogen           GPCVNLTEAGVVQTDVSLVDGASG.SMGCCGAVGVRN.PIQLAKALLLEAADG....
Pichia-yeast                    .SN..ICSDGTVRCDSSVVQCTPRGMKMGSVVC.NSSRY.PMEDALSVMNESLRV..S..
Clavispora-yeast                .AA..LDVSGAVACDATVLQGARG..VCTSLSLLAVVDTCPTQVCLEEMARQAEE....C
Penicillium                     .SN..LNAKGVVEGDASIVDH.HG..LSGAVGAVPNIKN.PIMLARKIYDKANVQM....
Aspergillus                     .SN..LTIDGTVECDATIVNH.LG..RSGAAGAVSQVKN.PISLARVVLEASTRPL....
S.pombe-fission-yeast           .SN..LNIDGKVECEAGVMDSESG..LTASVACCNCCRH.PSEACLYILNKRKVM.....
White-beech-mushroom            .SN..LTLHGTVECDAAIMDGRSG..DFGSVGAVSGIKN.PIRVARAILDYSKVPD....
Hydra-fresh-water-polyp         .SN..ICQDGSIECEASIMESKSK..RFSAVVCVKDLKN.PIMLAKRLLCNQSDEH....
Lamp-shell                      .SN..LTLDGTVECDASVMDGASL..KFGAVGAVESIKN.PIKLASCLLTEQKKGY.L..
Pacific-oyster                  .SS..LNIRGCVECDASIMEGKKL.LFGAVGAISGVQN.PIEVSAMLINRQLGEK.L..
Japanese-scallop                .SS..LTLEGEVECDASIMDGRSL.LYGGAGCVTGVKN.PILVARSLLDEQKKGC.L..
Chinese-liver-fluke             .SN..LTLNGLNECDSGIMCGRTM.RFSAVGAVSRIRN.PILLSHLLLKEQLRNSVT..
tr|A0A094ZUG7|A0A094ZUG7_SCHHA  .SN..LSINGYVECDAGIMSGSTL.QFAGIGAVRNMKN.PVKVARLLLQSQIDHPLG..
consensus>70                    .sn..lnl.g.iecdasimdg.sl.fgavga.sgikn.pv.va.ll.e.........
```

```
                                     β4    α4
                                 160     170        180       190       200

Human                           ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Chimpanzee                      ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Rhesus-macaque                  ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Tarsier                         ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Naked-mole-rat                  ..SAGRIPPCFLVGEGAYRWALDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
American-beaver                 ..SAGRIPPCFLVGEGAYRWAVDHGI....PTCAPNI.MTTRFS.....LAAFKRNKRKL
Chinese-hamster                 ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPSI.MTTRFS.....LAAFKRNKRKL
Treeshrew                       ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Wedell-Seal                     ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Walrus                          ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Domestic-yak                    ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPSI.MTTRFS.....LAAFKRNKRKL
Domestic-pig                    ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Dolphin                         ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Brandt_s-bat                    ..SAGRIPPCFLVGEGAYRWAVDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Horse                           ............................................................
Hedgehog                        ..SAGRIPPCFLVGEGAYKWAIDHGI....PSCPPNI.MTTRFS.....LAAFKRNKRKL
Emperor-penguin                 ..SAGRIPPCFLVGEGAYRWAVDHGI....PACPPGI.MATRFS.....LAAFKRNKRKL
Parrot                          ..SAGRIPPCFLVGEGAYRWAVDHGI....PACPPGI.MATRFS.....LAAFKRNKRKL
Hummingbird                     ..SAGRIPPCFLVGEGAYRWAVDHGI....PACPPGI.MATRFS.....LAAFKRNKRKL
King-cobra                      ......SFLVAEGAFRWAVDHGI....SACPPGV.MATRLS.....LAAFKRNKRKL
Green-sea-turtle                ..............F................S.....LAAFKRNKRKL
American-alligator              ..SAGRIPPCFLVGEGAYRWAVDHGI....PACPPGI.MATRFS.....LAAFKRNKRKL
Japanese-puffer-fish            ..SAGRIPPCFLVGEGAYRWAVSHGI....APCPS....EFS.....LSAYKRNKRKM
Turquoise-killifish             ..SAGRIPPCFLVGPGAHDWAVSHGI....PPCPSEK.MATRPS.....LSAYKRNKRKM
Japanese-rice-fish              ..SAGRIL.SFLVGRGAHDWAVCHGI....PSCPSEK.MATKFS.....LSAYKRNKRKM
Nile-tilapia                    ..SAGRIPPCFLVGRGAHDWAVSHGI....PPCPSDK.MATKPS.....LSAYKRNKRKM
Zebrafish                       ..SAGRIPPCFLVGKGAEQWAISHGI....PACPTEK.MTTKPS.....LSGYKRNKRKM
Phytophthora-plant-pathogen     TFAFGRQPPLVVVGEHARQLARDFGL....ETATEEEKEVENYQVIDKTREHWNKWYRRF
Achlya-plant-pathogen           ...DRAPPVVLVGATARAKELGL....PLVDYTAT......PIDTEAVAKHERQRA.
Pichia-yeast                    ..KEGIVSPMVKVGD.............LVPDPQL.VSKLQ.....QNVYER....
Clavispora-yeast                .AQTGLSRPSVMI........  ...AP....PASSGEAL......VSPATRALYKKYIQST
Penicillium                     ..GMSRVPPNFLAGEGAKDFAWEHGL....VIIPEEA.LISPVA.....LERYQVWAAEV
Aspergillus                     ..TLQRVPPNFLVGQGATNFAWEQGL....IVMPHDG.LISEEA.....KGRWLRWQQDL
S.pombe-fission-yeast           .SQHGLVPPAMLVGNGIEKLLLHSNI...KLVPESHL.....ITERSMKTQIKWKEI.
White-beech-mushroom            ..PLGRIPPLTLVSEGAHAFAVAHCADGKSQSVPAES.LIASRS.....KREWEEWKARL
Hydra-fresh-water-polyp         ..KLGLSPPMIWAGKGTKLKAVELGM....KLIKNKSL.RSMKA.....WEDFLSYQKNA
Lamp-shell                      ..SLGRVPPCFLVGRGAVAWAVEHNI....EVVEKKD.LITDSS.....LRSYRDYQRRL
Pacific-oyster                  ..SLGRLHPSVLVGEGASQWAVANGI....HKVDPRL.LIT.........
Japanese-scallop                ..SLGRLPPSVLVGEGARKWAQDNGV....ELIENKQ.LKTDSS.....LKTYASHKRKL
Chinese-liver-fluke             ..RLGRVQPTLLSGSGAVQWAALKKP....ALVEPYN.LVSHRS.....RTEWRKYRRWL
tr|A0A094ZUG7|A0A094ZUG7_SCHHA  ..ELGRVPPSVLVGDGAHLWAASKG....YPVDNSN.LVTKHS.....RVSWEKYKSWL
consensus>70                    ....gripp..lvgega..wa...gi.......p...m....s.....l..f...kr.l
```

```
                                        210                                      220
Human                   ELA.......E.R..V.D.......T...............DFMQLKKRR
Chimpanzee              ELA.......E.R..V.E.......T...............DFMQLKKRR
Rhesus-macaque          ELA.......E.R..V.E.......T...............DFIQLKKRR
Tarsier                 ELA.......E.R..V.E.......T...............DFIQLKKRR
Naked-mole-rat          ELA.......E.R..V.E.......T...............DFIQLKKRR
American-beaver         ELA.......E.R..V.E.......T...............DFIQLKKRR
Chinese-hamster         ELA.......E.R..V.E.......T...............DFIQLKKRR
Treeshrew               ELA.......E.R..V.E.......T...............DFIQLKKRR
Wedell-Seal             ELA.......E.R..V.E.......T...............DFIQLKKRR
Walrus                  ELA.......E.R..V.E.......T...............DFIQLKKRR
Domestic-yak            ELA.......E.R..V.E.......T...............DFIQLKKRR
Domestic-pig            ELA.......E.R..V.E.......T...............DFIQLKKRR
Dolphin                 ELA.......E.R..V.E.......T...............DFIQLKKRR
Brandt_s-bat            ELA.......E.R..V.E.......T...............DFIQLKKRR
Horse                   ..A.......................................... ..
Hedgehog                ELA.......E.R..V.E.......T...............DFIQLKKRR
Emperor-penguin         ELA.......E.K..V.E.......T...............DLIQLKKRR
Parrot                  ELA.......E.K..V.E.......T...............DLIHLKRRR
Hummingbird             ELA.......E.K..V.E.......T...............DLIQLKKRR
King-cobra              ELA.......E.K..V.E.......T...............NLFQLKKRR
Green-sea-turtle        ELA.......E.K..V.E.......T...............DLIQLKKRR
American-alligator      ELA.......E.K..V.E.......T...............DLIQLKKRR
Japanese-puffer-fish    ELA.......E.K..M.E.......S...............GHNQIKKRR
Turquoise-killifish     ELA.......E.G..T.V.......P...............GQNRTKKRR
Japanese-rice-fish      ELA.......E.R..M.D.......T...............GHSQTKKRR
Nile-tilapia            ELA.......E.K..M.D.......T...............GHNQTKKRR
Zebrafish               ELA.......E.Q..V.E.......................KRRI
Phytophthora-plant-pathogen QAKTEETKD..................................
Achlya-plant-pathogen   ...M..YS...................................
Pichia-yeast            ...........................................
Clavispora-yeast        ...NLQH.HL......LTQESEENGGT..............DFVRAASPK
Penicillium             KD.............YE..N.........ETLG.EV......TDPWIRRPV
Aspergillus             EAA.......E.LKEAQ..Q.........HPARYER......HKASVRRPV
S.pombe-fission-yeast   ...L..YQ..................NPI..............
White-beech-mushroom    DRG.......D.Y..G.T.......E..................
Hydra-fresh-water-polyp QRL.......LGS.RDT..E.........DLHSEDSVTENGLEKLYTSKKKEDI..KVGN
Lamp-shell              DRA.......E.M..G.L.......V...............EPKKMKKDV
Pacific-oyster          ...........................................
Japanese-scallop        EQH.......E.E..RMRR......K....RA..........KWLTVPSSK
Chinese-liver-fluke     SEV.......E.S..K.P.........................T.....
tr|A0A094ZUG7|A0A094ZUG7_SCHHA FNG.......N.E..M.D.......A...............TTKDSHFNH
consensus>70            e.........e.......e........................
```

β5

```
Human                   ........................QSSE...K.............
Chimpanzee              ........................QSSE...K.............
Rhesus-macaque          ........................QSSE...K.............
Tarsier                 ........................QSSE...K.............
Naked-mole-rat          ........................QSSE...K.............
American-beaver         ........................QSSE...K.............
Chinese-hamster         ........................QSSA...K.............
Treeshrew               ........................QSSE...K.............
Wedell-Seal             ........................QSSE...K.............
Walrus                  ........................QSSE...K.............
Domestic-yak            ........................QSSE...K.............
Domestic-pig            ........................QSSE...K.............
Dolphin                 ........................QSSE...K.............
Brandt_s-bat            ........................QSNE...K.............
Horse                   ...........................................
Hedgehog                ........................QSSE...K.............
Emperor-penguin         ........................QSNE...K.............
Parrot                  ........................QSNE...KIRSFPEQ......
Hummingbird             ........................QSNE...K.............
King-cobra              ........................QTSE...K.............
Green-sea-turtle        ........................QSSE...K.............
American-alligator      ........................QSSE...K.............
Japanese-puffer-fish    ........................QSTE...TANPTMLE......
Turquoise-killifish     ........................QSSE...N.............
Japanese-rice-fish      ........................QSNE...N.............
Nile-tilapia            ........................QSSE...K.............
Zebrafish               ........................QSCG...HEN...........
Phytophthora-plant-pathogen ...........................................
Achlya-plant-pathogen   ...........................................
Pichia-yeast            ...........................................
Clavispora-yeast        ........................SQIRVKPM.............
Penicillium             TPLDTRLERLERASQPGAHPGSDT...AD.......EETGTLPDPTLFADVKGRQGAS.
Aspergillus             SVNPTHLLSTPSSIRPASSLSSSL...GD.TRLRDSGSSSPVGSNDTLMLPPRARDASYM
S.pombe-fission-yeast   ...........................................
White-beech-mushroom    ........................SQV.................
Hydra-fresh-water-polyp ........................QKDL...Q.............
Lamp-shell              ........................KMVE.................
Pacific-oyster          ...........................................
Japanese-scallop        ........................ACNE...EDTPL.........
Chinese-liver-fluke     ........................EPPP...K.............
tr|A0A094ZUG7|A0A094ZUG7_SCHHA ..................HENVPEL...K.............
consensus>70            ...........................................
```

```
                                                                              230
Human                          ..........................END....SGTLDTVG
Chimpanzee                     ..........................END....SGTLDTVG
Rhesus-macaque                 ..........................END....SGTLDTVG
Tarsier                        ..........................END....SGTLDTVG
Naked-mole-rat                 ..........................END....SGTLDTVG
American-beaver                ..........................END....SGTLDTVG
Chinese-hamster                ..........................END....SGALDTVG
Treeshrew                      ..........................END....SGTLDTVG
Wedell-Seal                    ..........................END....SGTLDTVG
Walrus                         ..........................END....SGTLDTVG
Domestic-yak                   ..........................END....SGPLDTVG
Domestic-pig                   ..........................END....TGTLDTVG
Dolphin                        ..........................END....SGTLDTVG
Brandt_s-bat                   ..........................END....SGTLDTVG
Horse                          ...........................ND....SGTLDTVG
Hedgehog                       ..........................END....SGTLDTVG
Emperor-penguin                ..........................END....SGTLDTVG
Parrot         HANFSIFKDIRPYVLAPESCAALVGEPQRPMVVGKGGDQIELYSPEND....SGTLDTVG
Hummingbird                    ..........................END....SGTLDTVG
King-cobra                     ..........................ENE....SETLDTVG
Green-sea-turtle               ..........................END....SGTLDTVG
American-alligator             ..........................END....SGTLDTVG
Japanese-puffer-fish           .................TCTLLSRGNSVNRKRARVCLSSHEDG....SGCLDTVG
Turquoise-killifish            ..........................GNG....SGCLDTVG
Japanese-rice-fish             ..........................GNG....SGCLDTVG
Nile-tilapia                   ..........................ENG....SGCLDTVG
Zebrafish                      .................DFGCLEPVEGEENN....SACLDTVG
Phytophthora-plant-pathogen    ...........................E..EEFLDTVG
Achlya-plant-pathogen          ...........................T....SELLDTVG
Pichia-yeast                   ..........................:.....AGDTIG
Clavispora-yeast               ......................SLA.EEETSLIFPVSDTVG
Penicillium    ..LEHQTRKKAKLANSHLPVQPASPVSSSRDGSPDDSPDTADSEQDEE....DSITDTVG
Aspergillus    DGIMSQTRKSSSVPGST.SFEASHVLGRNMDVETTPPSVATDIHYAKM....DQISDTVG
S.pombe-fission-yeast          ...........................N....LSSQDTIG
White-beech-mushroom           ..........................ENI....RDIQDTVG
Hydra-fresh-water-polyp        ..........................EKE....SERLDTVG
Lamp-shell                     ..........................EDD....STLLDTVG
Pacific-oyster                 ...........................
Japanese-scallop               ..........................PCIS....EDVQDTVG
Chinese-liver-fluke            ..........................KLC....TSPLDTVG
tr|A0A094ZUG7|A0A094ZUG7_SCHHA ..........................TSR....VHRFDTVG
consensus>70                   end....s...ldtvg
```

```
                β6      TT      β7       TT  TT     η1         β8      TT
                   240      250       260     270       280
Human                          AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Chimpanzee                     AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Rhesus-macaque                 AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Tarsier                        AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Naked-mole-rat                 AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
American-beaver                AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Chinese-hamster                AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Treeshrew                      AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Wedell-Seal                    AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Walrus                         AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Domestic-yak                   AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Domestic-pig                   AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Dolphin                        AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Brandt_s-bat                   AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Horse                          AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHNP..Y
Hedgehog                       AVV...VDHEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGVHNP..Y
Emperor-penguin                AVV...VDQEGNVAAAVSSGGLALKHPGRVGQ......AALQGCGCWAENTGAHTP..Y
Parrot                         AVV...VDQEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHTP..Y
Hummingbird                    AVV...VDQEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGAHTP..Y
King-cobra                     AVV...IDQEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENMGFQNP..Y
Green-sea-turtle               AVV...VDQEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENIGVHNP..Y
American-alligator             AVV...VDQEGNVAAAVSSGGLALKHPGRVGQ......AALYGCGCWAENTGVLNP..Y
Japanese-puffer-fish           AVV...VDLEGNVAAAVSSGGLAMKHPGRVGQ......AAHYGCGCWAENACNMSP..Y
Turquoise-killifish            AVV...VDLEGNVAAAVSSGGLAMKHPGRIGQ......AAHYGCGCWAENSGNMNP..Y
Japanese-rice-fish             AVV...VDLEGNVAAAVSSGGLAMKHPGRIGQ......AAHYGCGCWAENARNIDS..Y
Nile-tilapia                   AVV...VDLEGNVAAAVSSGGLAMKHPGRVGQ......AAHYGCGCWAENACNMNS..Y
Zebrafish                      AVV...VDWEGNVAAAVSSGGLAMKHPGRVGQ......AAHYGCGCWAENARDVSL..Y
Phytophthora-plant-pathogen    AVC...MDPMGNVAAALSSGGVAYKVPGRLGL......AGCPRMGCNASNAKSASK..T
Achlya-plant-pathogen          VVC...MDASGSMAAGVSSAGTALKASGRVGH......AGCPHMGCVASGGTSG.....
Pichia-yeast                   VLLFEEVDKEWIITTGSSSGGNIMRDWRRIGS......AGIPSGGCWTETMDPDEGGTN
Clavispora-yeast               YIE...IGTE..TTAAVSSGGHWLRPKGRVSC......AGIPGAGVALAGRGSV.....
Penicillium                    AIA...IDKYGNIAAGSSSGGIGMKHRGRIGP......AALIGIGTHVIPQDPTDADGT
Aspergillus                    AIA...VDSHGNIAAGSSSGGIGMKHSGRIGP......AALVGIGTAVIPVDPNDDPQT
S.pombe-fission-yeast          VIC...VDKNGRIAVVSSSGGLLLKPAGRIGS......SPIPGHGFWIESFDNKSH...
White-beech-mushroom           AVV...CEGTVDGISAGVSSGGILLKHPGRIGEPTYGFSKAAMFGAGCWAQYFPKAK...L
Hydra-fresh-water-polyp        VLC...IDVYGEMASAVSSGGIAFKVPGRLGH......TGQYGASCWAFQYSPCL...
Lamp-shell                     AVC...LDCQGNMAAAVSSGGIALKCPGRVGQ......AAVFGCGCWADNIKDVTP...
Pacific-oyster                 ...........
Japanese-scallop               AVA...MDSEGNMSAAVSSGGISLKHSGRLGP......AAMYGAGCWAHNWKSDKK..A
Chinese-liver-fluke            AVC...IDKGGSVAAAISSGGIALKQEGRIGQ......ASIYGCGCWAEES.QLG....
tr|A0A094ZUG7|A0A094ZUG7_SCHHA AVC...IDIEGKIAAASSSGGIAMKYEGRLGQ......ACTYGCGCWAEELSSFS....
consensus>70                   av....vd.egnvaaavssgg.alkhpgrvgq.......aa.ygcgcwaen........
```

```
                                        β9            α5           α6
                                        ━━━━━━▶    ℓℓℓℓℓℓ      ℓℓℓℓℓℓ·ℓℓℓ
Human                                   290       300          310

Human                       .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Chimpanzee                  .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Rhesus-macaque              .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Tarsier                     .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Naked-mole-rat              .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
American-beaver             .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.TE...........D
Chinese-hamster             .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Treeshrew                   .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Wedell-Seal                 .........STAVSTSGIC.......................................
Walrus                      .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Domestic-yak                .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Domestic-pig                .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Dolphin                     .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Brandt_s-bat                .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Horse                       .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Hedgehog                    .........STAVSTSGCGEHLVRTILAREC.SHAL....Q.AE...........D
Emperor-penguin             .........STAVSTSGICYFYKYFKKLLSF.SVIN....L.Q...........I
Parrot                      .........STAVSTSGCGEHLVRTILAREC.SCAL....Q.TE...........D
Hummingbird                 .........STAVSTSGICYFYKYFRKLLEC.SCAL....Q.TE...........D
King-cobra                  .........STAVSTS.........................................
Green-sea-turtle            .........STAVSTSAQ.QHVE......P.HTQL....G.CL...........N
American-alligator          .........STAVSTSGCGEHLVRTILAREC.SRAL....Q.TE...........D
Japanese-puffer-fish        .........STAVSTSGCGEHLIRTMLAREC.SAAM....R.SE...........D
Turquoise-killifish         .........STAVSTSGCGEHLIRTMLAREC.AAAM....Q.SE...........D
Japanese-rice-fish          .........STAVSTSGCGEHLIRTMLAREC.SAAM....Q.SE...........D
Nile-tilapia                .........STAVSTSGCGEHLIRTMLAREC.SAAM....Q.SE...........D
Zebrafish                   .........STAVSTSGSGEHLIRTMLAREC.STAM....Q.AE...........N
Phytophthora-plant-pathogen RKRKRSGQNKVENAPAIACTGRGEHFIRSNFVGNL.T....RRLK...............
Achlya-plant-pathogen       ...........DDGYAWSATGRGEDIIRSRLLQQI.E....YRSR.RQ.........SLA
Pichia-yeast                ...........VVSIMVSGQGEELIKRDFARYI.SKNM....LKYI...........N
Clavispora-yeast            ......Q..VCCLCSGQGDDIVRMTLASHVCDQLSREALD.AP...........AL.
Penicillium                 ...........TCAVVTSGTGELIASTLAASTC.AQRM....YYGQKMGNNGMFTQVM
Aspergillus                 ...........CVATVTSGTGEHIATSMAASTC.ASRV....YYNQRKCEDGSFEEVT
S.pombe-fission-yeast       ........SSTCAVATSGTGEHISNTCFACRS.S...QLLV.SE.........DN.
White-beech-mushroom        ...........GIACSVSGAGESIVRATIARAI.GDGF....KDCIS......KEEEVD
Hydra-fresh-water-polyp     ...........KVASCTSGCGELLTQTNLAKSA.ADYS....A.NN...............
Lamp-shell                  ...........GVAVSTSGCGEHLIYTQLAREC.ALEV....RGAD...........D
Pacific-oyster              ................GSGEHIIRTVLAREC.AMSI....QREN...........N
Japanese-scallop            ...........GVAVATSGSGEHLMRTLFAQKC.AECV....QTQD...........S
Chinese-liver-fluke         ...........KVAVVTSGTGEQLIRTQLAQKA.AERL....LEPSP.........A
tr|A0A094ZUG7|A0A094ZUG7_SCHHA ........SIGVVTSGTGEQLIKTQLAQRS.AERF....NGDS...........I
consensus>70                .............avstsg.geh.vrt.la.............e..............


                            ℓℓ.ℓℓℓℓℓℓℓℓ   ℓℓℓ          ━━━━━━━━▶          ━━.━
                               α7          η2            β10                β11
Human                          320         330          340         350        360

Human                       .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Chimpanzee                  .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Rhesus-macaque              .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Tarsier                     .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Naked-mole-rat              .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
American-beaver             .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Chinese-hamster             .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSSESDSSQN..KQT.LLV.EF
Treeshrew                   .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQD..KQT.LLV.EF
Wedell-Seal                 ........CFHDHFS.............................................
Walrus                      .AH.QALLETMQNKFI............................................
Domestic-yak                .AH.QALLETMQNKFIS...........................................
Domestic-pig                .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPNSPQN..KQT.LLV.EF
Dolphin                     .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPNSSPN..KQT.LLV.EF
Brandt_s-bat                .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSAEPDSSQN..KQT.LLV.EF
Horse                       .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSTDPDSSQN..KQT.LLV.EF
Hedgehog                    .AH.QALLETMQNKFISSPFLASEDGVLGGVIVLRSCRCSMEPDSTHN..KQT.LLV.EF
Emperor-penguin             .PN.KVFLQ.VSIYLCSSPFLANEDGVLGGVIVLRSCRCSAEPESSQE..KPT.LLV.EF
Parrot                      .AH.QALLETMQNKFIGSPFLANEDGVLGGVIVLRSCRCSAEPESSQE..KPT.LLV.EF
Hummingbird                 .AH.QALLETMQNKFIGSPFLANEDGVLGGVIVLRSCRCSAETESSQG..KPT.LLV.EF
King-cobra                  .........................................................RF
Green-sea-turtle            .AH.KATHESHRERHQLIPPGPQ.........PCTPELYRLAL.LRS.LIR.RF
American-alligator          .AH.QALLETMQNKFIGSPFLANEDGVLGGVIVLRSCRCSAEPDSSQE..KQT.LLV.EF
Japanese-puffer-fish        .AH.QALLEAMQNKFISSPFLAGEDRVLGGVIVLRGCRC.VEAPSSQN..IQG.ILV.EF
Turquoise-killifish         .AH.QALLEAMQNKFISSPFLTSEDRVLGGVIVLRCCRC.VEAQPSPN..IQG.ILV.EF
Japanese-rice-fish          .AH.QALLEAMQNKFISSPFLASEDRVLGGVIVLRCCTC.VEAQPSQN..IQG.KLV.EF
Nile-tilapia                .AH.QALLEAMQNKFISSPFLASEDRVLGGVIVLRCCRC.VEAQPSQN..FQG.ILV.EF
Zebrafish                   .AH.QALLEAMQNKFISSPFLAGEDCVLGGVIVLRCCTC.GEAQRSED..IQA.LLV.EF
Phytophthora-plant-pathogen .TS.TDLDRGLRKVFHETRTENGGVGIEGGVLAFLSDATEG........DKSRR.I.QL
Achlya-plant-pathogen       .DA.ASLDRAVSRAFTDARERNNDVGIEGGVLGLVRAKMSSHKKRKRSQTEKS.A.L.HF
Pichia-yeast                .NEGMFVSDALVMSLNDMGRETRQLQW..GVAGLAV................IDG.RPI.GF
Clavispora-yeast            ..G.PLLVKTIEERAKSAGIESAYVGVLATVAI.................EG.R.V.QL
Penicillium                 .EE.EALRAWMMREFNEHAAVSNSVLF..GALGVVVV......KKTN..Y....GI.EL
Aspergillus                 .ED.EALRGMIASEFMGHPGVRDSHCR..GAIGIMAV......KKTV..D.....GV.YL
S.pombe-fission-yeast       .VV..SSLNKLINDFHEHPSATLYSDLQVGIIPAK......VETSNSHNK..R....I.IF
White-beech-mushroom        .PH.SIISTVVVDRFCGPCRDRGEVNLDVGIILL........VTED...IEN.KITARL
Hydra-fresh-water-polyp     .EI.LNFNSYLLNEFVHSQYISSSQMRLCGLLVAKFV......NE...DSD.CFV.DF
Lamp-shell                  .IS.QAVTDVFSRQFIESDMLSDQVAKYGGAIVLKYSSSAAERRE...........V.EL
Pacific-oyster              .AS.LGLSQGFKQSLLESEFLSEEKEKYAGVLALVH..........NED.NTA.EV
Japanese-scallop            .GS.LGVSTAFRDHFLESDYLASVTQKLGGVLALRQ........DKEE..GSA.IML.EV
Chinese-liver-fluke         PTP.ETLQSVITTDFIDSRLLANEEHKFAGIAGIYCP........D..ADS.GSI.EV
tr|A0A094ZUG7|A0A094ZUG7_SCHHA SVS.EIVHSSLTDGFLNSKYLSCDMEKFAGVAGVYANFE........R..IDF.PNV.EV
consensus>70                ......l.e.m.n.f.......e.....gv................v.e.
```

138

```
Human                                        β12                    β13      TT          β14
                                    370       380       390       400       410
Human                            LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Chimpanzee                       LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Rhesus-macaque                   LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Tarsier                          LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPM
Naked-mole-rat                   LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
American-beaver                  LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Chinese-hamster                  LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Treeshrew                        LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Wedell-Seal                      ...................................................................
Walrus                           ...................................................................
Domestic-yak                     ...................................................................
Domestic-pig                     LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Dolphin                          LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Brandt_s-bat                     LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLQ.....SPV
Horse                            LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPV
Hedgehog                         LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPE
Emperor-penguin                  LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPE
Parrot                           LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPE
Hummingbird                      LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPE
King-cobra                       ...................................................................
Green-sea-turtle                 LGFRLNS.V..KIFT...KG.KSTLQRYIS...AAACLVKTHYADGR...TL.....SC.
American-alligator               LWSHTTESMCVGYMS...AQ.DGKAKTHISRLPPGAVAGQSVAIEGGVCRLE.....SPA
Japanese-puffer-fish             LWSHTTESMCVGYMS...AQ.DSKAKTHISRLPPGTVPGQCLAIEGGVCRLT.....SIP
Turquoise-killifish              LWSHTTESMCVGYMS...AQ.DSKAKTHISRLPPGTVPGQSLAIEGGVCRLK.....GVV
Japanese-rice-fish               LWSHTTESMCVGYMS...VQ.DSKAKTHISRLPSGTVPGQSLAIEGGVCRLM.....EPT
Nile-tilapia                     LWSHTTESMCVGYMS...AQ.DSKAKTHISRLPPGTVPGQSLAIEGGVCRLM.....GVV
Zebrafish                        LWSHTTESMCVGYMS...AQ.DSKARTHISRLPPGAVAGQSLAIEGGVCRLG.....TAS
Phytophthora-plant-pathogen      GAAFTTPCMGVGYMQCHASA.EAKAKVQILRQPGTRTSISDTDKQELSIHVSFIQFPLPS
Achlya-plant-pathogen            VAAFTTPSMGIGVCSS...Y.DAEPRGP..................N
Pichia-yeast                     F.RGNSRSFVGG........SNSRVRLVQNSGYGEILL...
Clavispora-yeast                 VFCHSTESFYFGFRAA.....DGRPRIVLSRAEAGFVSGEYSLGRSQ...........
Penicillium                      YFAHNTDSFAIASMS...SR.LDKPSCLMSRGSR.GTIAQGGCRYRFKDNFV.....S..
Aspergillus                      CFGHNTDSFALASMS...SE.DKKPVSVMSRSNGNGSIAQGGRAYRSRSVVA.....FHV
S.pombe-fission-yeast            GLAHSSPDMVFGFMKG...D.HSKPTTEISRKGSKRSSVQLY..........AE.....R
White-beech-mushroom             WCAFTTPSMAIAYAS...SE.TPKPKAQILRRPKSKPSSLTSEPHIFITA.......IS
Hydra-fresh-water-polyp          HIAHTTDSFCIAFTTG...E.MKKPQVVMSRLNLK.TKGP..VSQGFLLKL....
Lamp-shell                       VWAHTTESMCIGYMK...TG.EHKPKVHISRMPSTSTVGQTTSVMGKYFHLR....
Pacific-oyster                   LWGHTTDSMCVGYMS...SS.LKKPKVVMSRLPTGGVPGKSFTMLGMP...V.......
Japanese-scallop                 IWGHTTESMCVGYWG...SD.THKPKVLVSRLPENSEAGKSFKMEGKVFSVT....
Chinese-liver-fluke              FFGHTTRSMVAYFV..EDAMPIPMAYVSRLSPEKPLRVDAFAHRLQRS......
tr|A0A094ZUG7|A0A094ZUG7_SCHHA   FFGHSTRSMSVGHYL...PSVMKKPEVRVSRKQIDRPTYIEVYAYT...I..........
consensus>70                     ...httesm.vgy..........k....isr...........
```

Human

                                              4 2 0
                                                 .
Human   . . . . N
Chimpanzee   . . . . N
Rhesus-macaque   . . . . N
Tarsier   . . . . N
Naked-mole-rat   . . . . N
American-beaver   . . . . N
Chinese-hamster   . . . . N
Treeshrew   . . . . N
Wedell-Seal   . . . .
Walrus   . . . .
Domestic-yak   . . . .
Domestic-pig   . . . . N
Dolphin   . . . . S
Brandt_s-bat   . . . . N
Horse   . . . . N
Hedgehog   . . . . N
Emperor-penguin   . . . . S
Parrot   . . . . S
Hummingbird   . . . . S
King-cobra   . . . .
Green-sea-turtle   . . . .
American-alligator   . . . . N
Japanese-puffer-fish   . . . . D
Turquoise-killifish   . . . . E
Japanese-rice-fish   C T P P . . L . L
Nile-tilapia   D A L T . . G L P
Zebrafish   . . . . D
Phytophthora-plant-pathogen   H G I
Achlya-plant-pathogen   L
Pichia-yeast   . . . .
Clavispora-yeast   . . . .
Penicillium   . . . . H
Aspergillus   P S A I N Q P . M
S.pombe-fission-yeast   L
White-beech-mushroom   L . . . .
Hydra-fresh-water-polyp   . . . .
Lamp-shell   . . . .
Pacific-oyster   . . . .
Japanese-scallop   . . . .
Chinese-liver-fluke   . . . .
tr|A0A094ZUG7|A0A094ZUG7_SCHHA . . . .
consensus>70   . . . .

Supplementary Figure 5.6 Sequence alignment of Taspase1 across 41 different species.
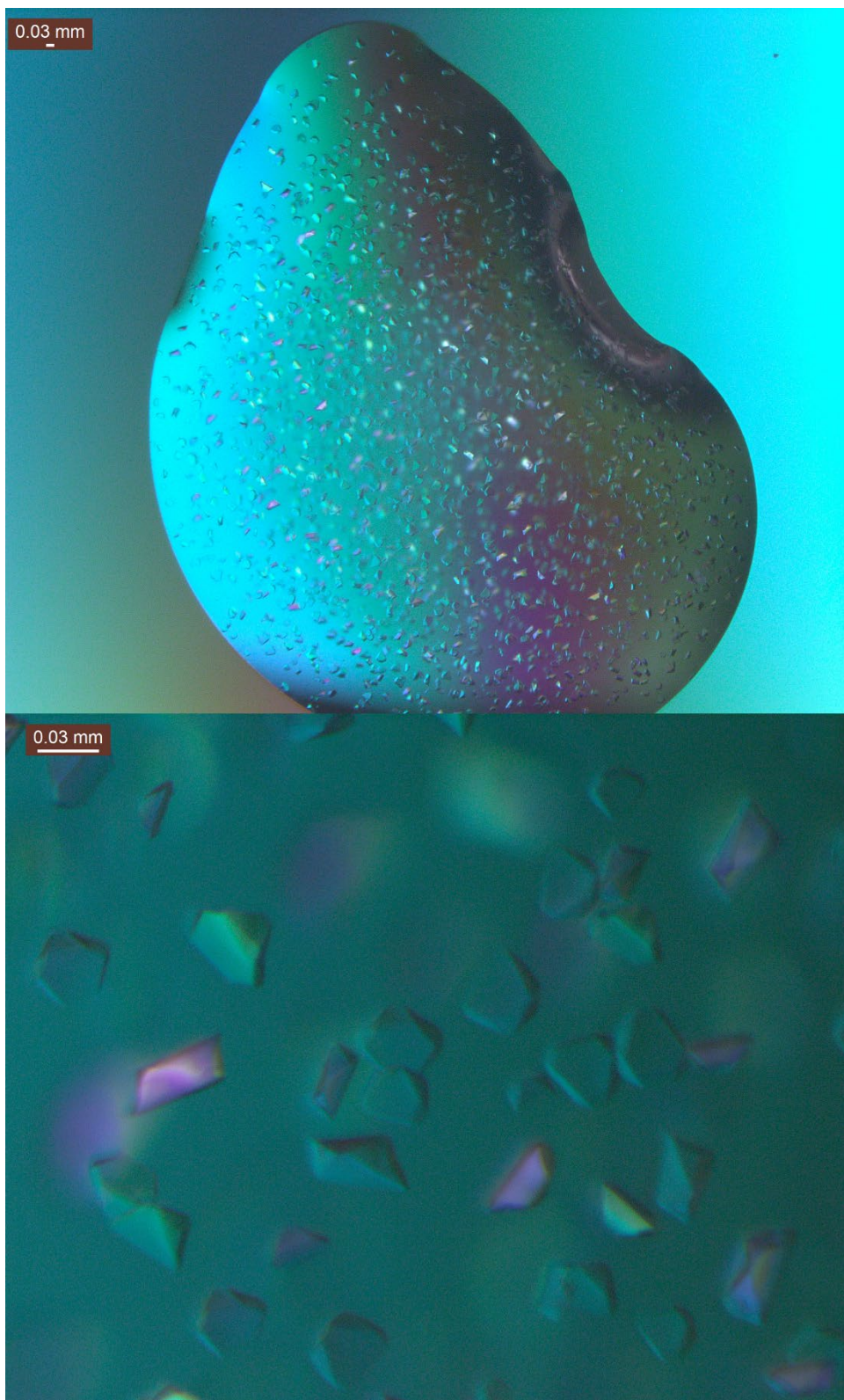
140
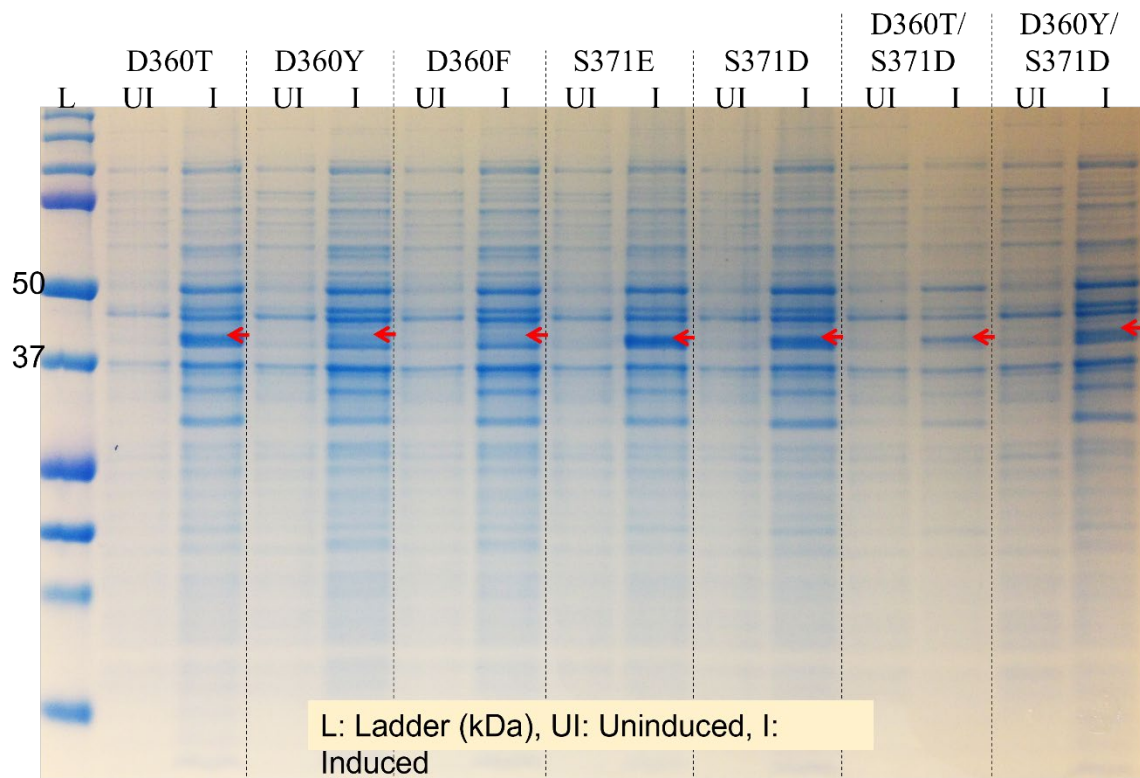
Supplementary Figure 5.7. Expression of mutants in whole cell contents induced versus uninduced.
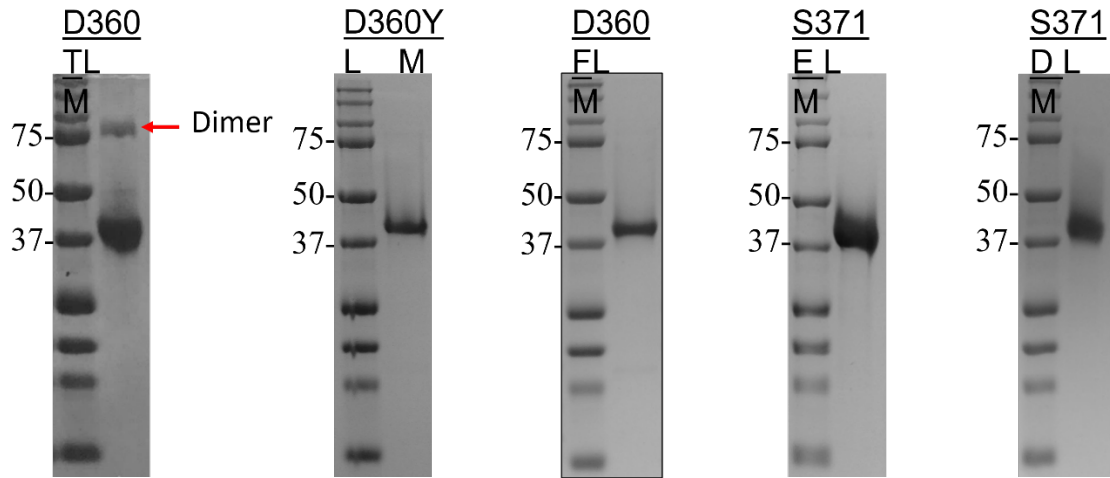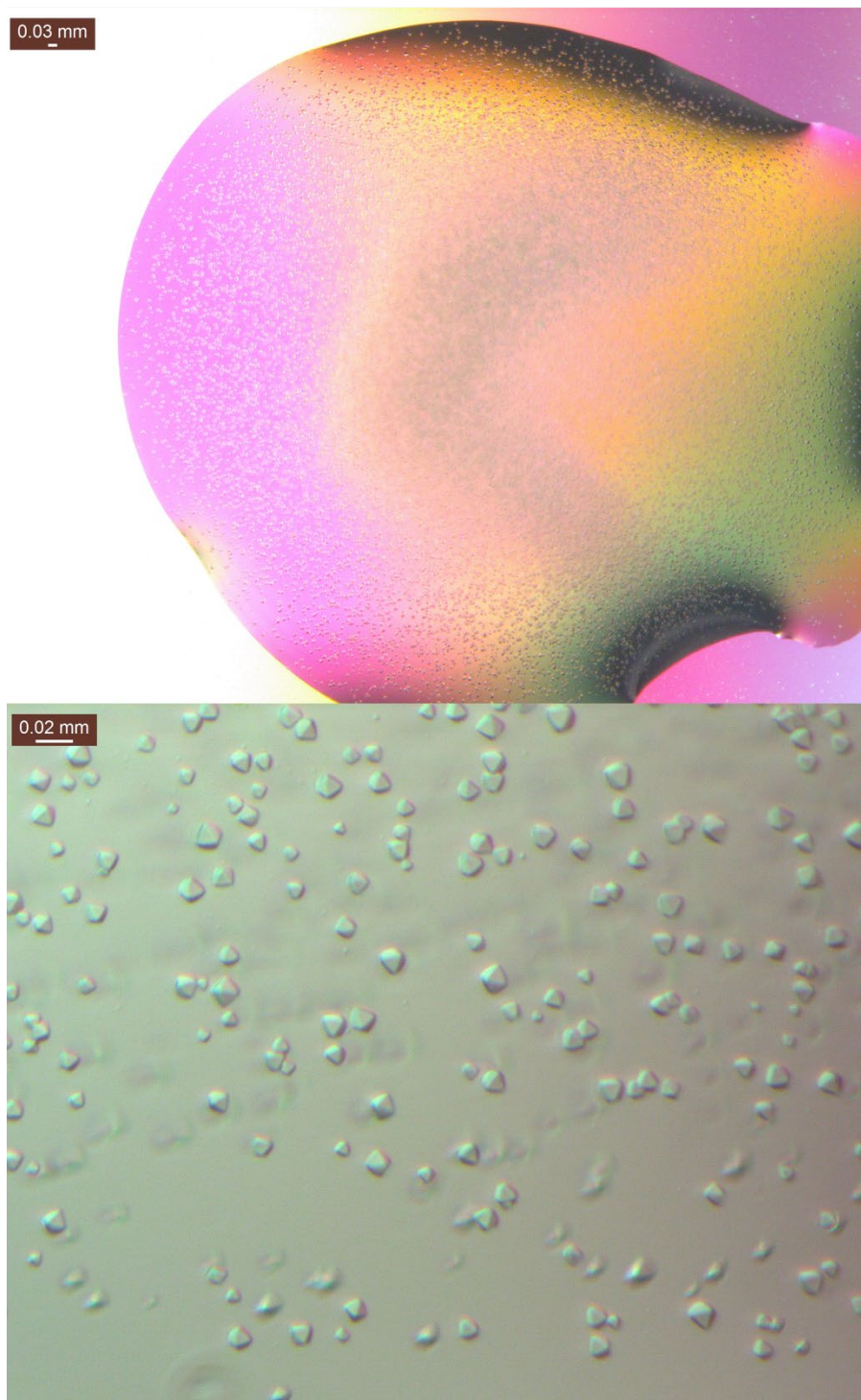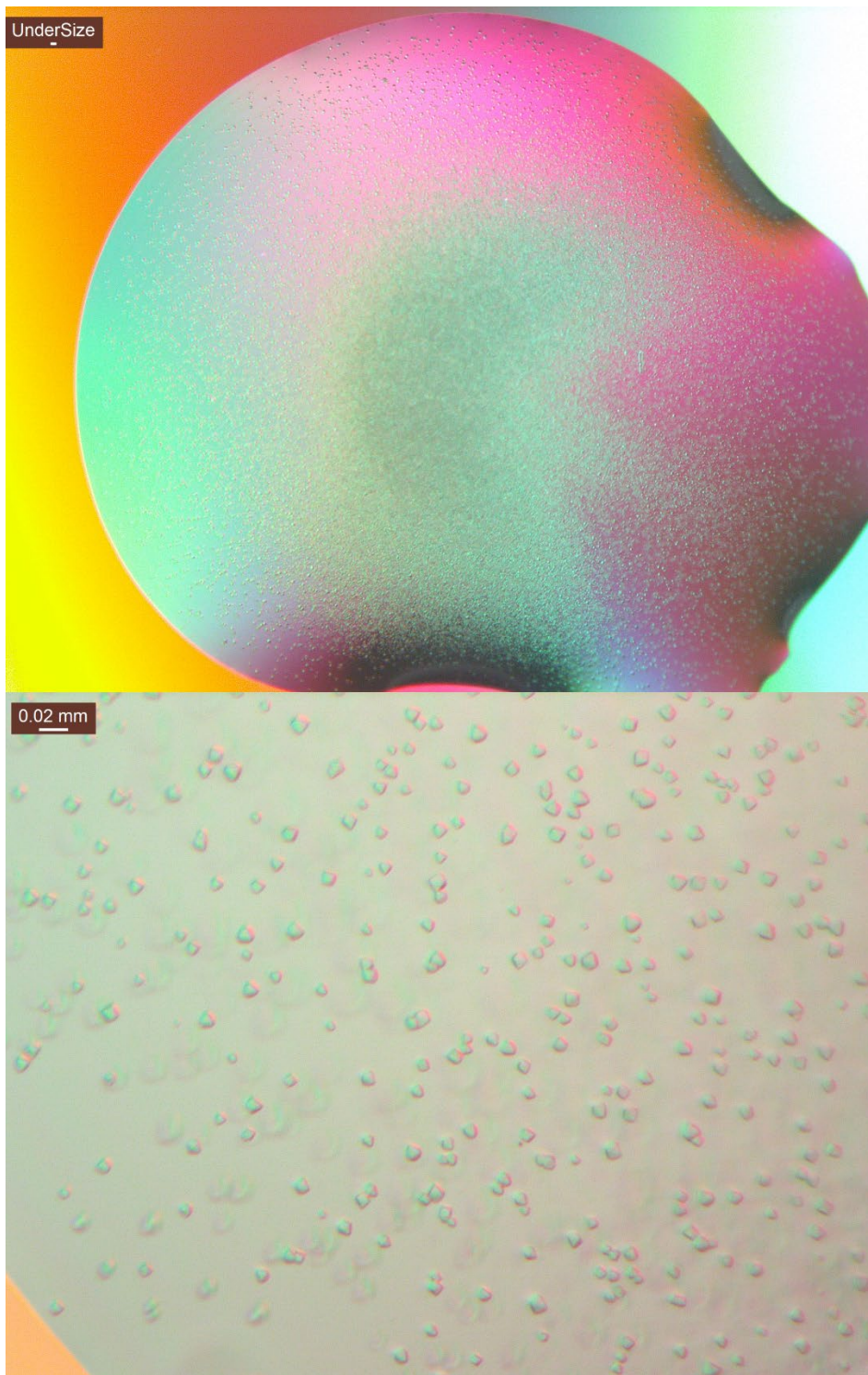
L: Ladder (kDa), M: Mutant

Supplementary Figure 5.8 Purification of the single point mutations of Taspase1.

Supplementary Figure 5.9. Crystallization of Taspase1 D360T at pH 6.6.

Supplementary Figure 5.10. Crystallization of Taspase1 D360T at pH 7.0.

Supplementary Figure 5.11. Crystallization of Taspase1 D360T at pH 7.5.

Supporting Tables 5.6.2

Supporting Tables

Supporting Table 5.1. Crystallographic data collection and refinemer

	cp-Taspase1$_{\alpha41-233/\beta}$	cp-Taspase1$_{\alpha41-183/\beta}$
Data collection statistics		
Radiation source	APS 23-ID-D	APS 21-ID-F
Wavelength (Å)	1.04	0.98
Detector	PILATUS3 6M	RAYONIX MX300HE
Temperature (K)	100	100
Space group	H32	*P*6$_5$
Unit cell dimensions		
a/c (Å)	196.00/196.91	60.30/319.01
Resolution range (Å)	3.04-49.05 (3.04-3.11)	2.15-40.41 (2.15-2.20)
I/δ(I)	17.0 (1.9)	15.24 (3.74)
Multiplicity	19.6 (20.4)	5.8 (5.8)
Completeness (%)	99.8 (99.4)	99.8 (85.8)
CC$_{1/2}$ (%)	100 (91.1)	99.8 (85.8)
R$_{merge}$ (%)	6.8 (187.5)	7.8 (52.0)
Total number reflections	343,892	206,406
No. of unique reflections	17,964	35,513
Ellipsoidal truncation		
Ellipsoidal[a] resolution (Å) (direction)[b]	2.9 (0.894a* – 0.447b*) 2.9 (b*) 4.8 (c*)	N/A
Best/Worst diffraction limits after cut-off (Å)	3.04 / 5.3	N/A
I/δ(I) (ellipsoidal[a])[c]	21.2 (1.5)	N/A
Completeness (%) (spherical[a])[c,d]	54.4 (13.7)	N/A
Completeness (%) (ellipsoidal[a])[c,d]	93.9 (65.5)	N/A
Total no. of reflections (ellipsoidal[a])[c]	299,919	N/A
No. of unique reflections (ellipsoidal[a])[c]	15,315	N/A
Data refinement statistics for		
truncated data		

Number of reflections used in refinement	13,749	35,417
Number of free reflections used in refinement	1,568	1,709
R_{work} / R_{free} (%)	23.7 / 30.7	15.9 / 19.7
Wilson B-factor (Å^2)	152.0	43.5
Solvent content (%) / Matthews coefficient (Å^3/Dalton)	70 / 4.3	48 / 2.4
No. of atoms		
Protein	10,510	4,368
Water and others (ligands or ions)	0	189
Overall B factor (Å^2)	165.0	43.48
RMS deviations from ideal values		
Bonds lengths (Å)	0.007	0.007
Bond angles (°)	1.744	0.827
Ramachandran plot statistics (%)		
Favored	85	96.55
Allowed	12	2.80
Outliers	3	0
Molprobity score / percentile	3.5 / 38th	1.55 / 99th
PDB	6VIN	6UGK

[a]These statistics are for data that were truncated by STARANISO to remove poorly measured reflections affected by anisotropy.

[b]The resolution limits for three directions in reciprocal space are indicated here. To accomplish this, STARANISO computed an ellipsoid postfitted by least squares to the cutoff surface, removing points where the fit was poor. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

[c]Values in parentheses are for the highest-resolution shell. Note that data collection and refinement statistics have different highest-resolution shells. [d]The anisotropic completeness was obtained by least squares fitting an ellipsoid to the reciprocal lattice points at the cutoff surface defined by a local mean I/σI threshold of 1.2, rejecting outliers in the fit due to spurious deviations (including any cusp), and calculating the fraction of observed data lying inside the ellipsoid so defined. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

Supporting Table 5.2. Interfaces in Cp-Taspase1$_{\alpha41\text{-}233/\beta}$ calculated by PDBePISA (Krissinel & Henrick, 2007)

Interface #	Name of Interface	Interface residues component 1	Interface residues component 2	BSA* (Å²)	Hydrogen bonds	Salt bridges	ΔG Kcal/mol	CSS#
1	Dimer	60	59	2039.1	21	7	-23.8	1.000
2	Inter-single ring	42	43	1355.1	13	2	-9.5	1.000
3	Helix	19	19	648.2	2	0	-6.6	0.069
4	Intra-double ring	16	16	535.7	0	0	-9.7	0.000
5	Linker	8	8	168.0	0	2	-1.7	0.026
6	Histidine	1	1	36.1	0	2	0.1	0.006
7		1	1	33.6	0	2	0.1	0.006

*Buried surface area (BSA).
#Complex formation significance score (CSS).

Supporting Table 5.3. Interacting residues identified by Robetta as potential areas for mutation. Non-conserved residues identified by sequence analysis are indicated in bold.

Interacting residues of the Taspase1 flexible loop as identified by Robetta	
Amino acid position in Taspase1 <u>native</u> sequence (as shown in sequence alignment)	Amino acid position in Taspase1 circular <u>permutated</u> sequence (as shown in crystal structure)
H 53	H 201
S 54	S202
K 202	K350
L 205	L353
V 209	V 357
D 212	**D 360**
F 213	F 361
L 216	L364
S 223	**S 371**
D 228	D 376

Supporting Table 5.4. Data statistics of the room temperature structure of Taspase1 by SFX.

Data Analysis	
Resolution Range (Å)	30 – 4.50
Total reflections	27,497,959 (609,072)
Unique reflections	16,565 (535)
Redundancy	1660 (1138)
SNR (I/σ)	1.70 (1.11)
$CC_{1/2}$	0.9525 (0.3027)
CC*	0.9877 (0.6817)
R_{split}	44.57% (112.06%)
R_{work}	0.3513 (0.3765)
R_{free}	0.3633 (0.3864)
RMS* (Bonds) Å	0.004
RMS (angles) deg	0.825
Mean B factor	Factor 28.7

5.7 Funding Acknowledgements

CHAPTER 6

CONCLUSIONS AND OUTLOOK

Direct observation of substrate docking to the NendoU protein from the SARS-CoV-2 virus is a step towards understanding the biological mechanism and advancing rational drug design for COVID therapeutics. The focus of the National Science Foundation RAPID grant for COVID-19 research was time-resolved studies of the NendoU mechanism. The goal was to use mix-and-inject technologies to diffuse RNA (5 and 21 nucleotide) substrate into catalytically active NendoU microcrystals in order to observe the mechanism at different time points. The combination of these time points in chronological order would build a "molecular movie." A molecular movie would function as a guide to understanding substrate docking, cleavage, and release. By having a dynamic understanding of NendoU in action, rational drug design can be more effective and identify more points than with static structures alone. Lastly, understanding the biology of this unique enzyme would be scientifically relevant as there are few known homologs.

The methods developing mix-and-inject experiments to perform time resolved crystallography were successful; however, in addition to advances the team has also worked to overcome specific challenges. First, microcrystallization conditions were developed with catalytically active NendoU, which maintained its activity in crystalline form. However, at the first beamtime, it was discovered that the citrate from the buffering component also occupied the active site. While it was not enough to inhibit the protein, it was not favorable for observing high RNA occupancy in the crystals. Multiple crystallization conditions have since been identified without the use of citrate. Overall,

mixing systems, crystals, and RNA substrates have all been identified to move forward with time-resolved experiments.

A challenge limiting the study, that is currently being addressed, is the formation of the RNA/protein complex structure. While the conditions for mix and inject studies have been met, demonstration of an RNA protein complex structure is an important step in understanding substrate binding and feasibility of time-resolved studies. Recently, the first structure of NendoU in complex was released in the PDB (March 23rd, 2022) with a supporting manuscript in pre-print publication (Frazier et al., 2022). With this structure, hopefully the bottleneck of NendoU-Substrate complexes will be undone and help advance structural studies of NendoU with its substrate. Previously the closest anyone has come to this has been 2 double nucleotide structures, that have been named pre- and post-cleavage structure (M. N. Frazier et al., 2021). In addition, hypothesized intermediate structures have been proposed based on crystallography results (Y. Kim et al., 2021). Even if cryo-trapped structures of RNA-NendoU complexes have recently been determined, no reports on time resolved studies have been published to date.

Currently we propose two approaches to the challenges of time resolved studies. One potential explanation for why RNA is so difficult to form a complex structure is the salt concentration. RNA processing proteins can be sensitive to salt, and salt can even inhibit activity [ref]. While the concentration of salt isn't high enough to inhibit the activity of NendoU, the disassociation constant maybe too high to observe stable binding of the RNA, instead a short-lived transient complex may form during catalysis. This is consistent with why some density for RNA is observed in cryoEM studies which are performed in the proteins' size exclusion buffer, as compared to crystallography which is

153

in most cases performed at much higher salt concentrations. The current goal is to establish conditions for cryoEM and crystallography in low salt conditions to test this hypothesis. Some of these crystallization methods have been described in **Chapter 3**, and currently are in the process of data collection and analysis.

Additionally, preliminary cryoEM results show conformational changes in the protein upon RNA binding but could not resolve yet the RNA binding to NendoU. One potential reason for this is the low occupancy of RNA in the binding sites; however, this may also be an intrinsic feature of the protein to bind one trimer of the hexamer at a time (**Chapter 2**). In the case that only one trimer is bound to RNA at a given time in our Cryo-EM and X-ray structure, it is possible that in combination with low occupancy, the RNA is being averaged out in data analysis. Another solution to the RNA occupancy issue may also be lowering the salt concentration of the buffer. When the salt concentration in the buffer was reduced to 20 mM HEPES, pH 7.5, 20 mM sodium chloride, 20 mM manganese chloride, negative staining of the sample with RNA featured very large structural changes. Currently these experiments are being scaled up to cryoEM. Overall, by controlling ionic strength we hope to impact what we can see with RNA binding to NendoU.

While the complex structure of NendoU is being explored at low salt conditions, another strategy for time resolved studies is using a substrate that is known to bind to NendoU and result in a conformational change. The substrates 5' uridine monophosphate (5'UMP) and 3' uridine monophosphate (3' UMP) have been shown in complex with NendoU in crystal structures. For preliminary time-resolved studies, we propose to use these substrates to show substrate binding and conformational changes upon binding to

NendoU. These experiments would be a practical intermediate to the initially proposed time-resolved studies because they yield new information that meet our goals for studying the mechanism of NendoU and would also provide important information for future time-resolved studies with larger RNAs. Additionally, by performing these studies with the 5'UMP and 3'UMP substrates, higher concentrations can be diffused into the crystals as the substrate is easier and cheaper to obtain. Millimolar concentrations of large RNA (5 to 21 nt) cost thousands of dollars whereas the single nucleotide substrate costs hundreds of dollars per gram. These experiments have been designed and proposed to run at the European Synchrotron Radiation Facility (ESRF) new beamline ID29 dedicated to serial macromolecular crystallography during commissioning beamtime and subsequent user beamtime during the Summer and Fall of 2022.

The development of novel techniques for time-resolved cryo-EM culminating into future studies to conduct TR cryoEM experiments of NendoU have also been proposed in the NSF COVID grant, Using the mix-and-inject technology developed for time resolved crystallography experiments at X-ray free electron lasers in combination with pico-drop needle picoliter droplet deposition, the sample could be frozen on pre-cooled cryo-EM grids at specific timepoints for time-resolved studies. This could make the best of both techniques – combining the commissioned time-resolved components of crystallography with the low salt and high-resolution structures of cryoEM.

Over the past two years of the pandemic, an unprecedented level of research has been done on the subject of coronaviruses. The scientific community has taken the incredible advancements of sciences accumulated over the past century to apply them to this situation. As a result, the discoveries and technologies developed with exponential speed.

One factor for this included expedited efforts outside of the experimentation itself: grants, proposals, publication, and more. While not every scientific issue or disease can have the full force of the scientific community behind it or be moved to the front of the line, it does highlight that lengthy role of administrative processes that have to be followed to make scientific discoveries can be significantly shortened. What might have taken years alone in applying (and receiving funding), getting experimental time at facilities, undergoing review, would take week to months with RAPID access systems in place for COVID-related research.. This is especially true for larger scale developments that require government review and approval like the development of vaccines.

This rapid access is not without drawbacks. As a result of being pushed to the front of the line, research for many other important issues may now wait longer before grants or beamtimes are awarded. Issues of scientific integrity have been brought into question from rapidly published COVID research that was not peer-reviewed and some studies were not conducted to the standards of the community or were submitted fraudulently for acclaim. These incidents are not unique to COVID, but the rapid nature did create a circumstance for more issues to arise. Ultimately, these issues are proportionally small in comparison to the overall success of the scientific response in understanding, preventing, and treating COVID-19. The advancement of mRNA vaccines and COVID therapeutics are two of the most notable achievements. All of it begs the question, "what if?" What progress could be made to address such challenges when given this level of support? To me, it shows the capability, strength, and importance of modern science in today's world along with the outstanding potential for future research.

REFERENCES

Abbas, Q., Kusakin, A., Sharrouf, K., Jyakhwo, S., & Komissarov, A. S. (2022). Follow-up investigation and detailed mutational characterization of the SARS-CoV-2 Omicron variant lineages (BA. 1, BA. 2, BA. 3 and BA. 1.1). *bioRxiv*.

Abola, E., Kuhn, P., Earnest, T., & Stevens, R. C. (2000). Automation of X-ray crystallography. *Nature structural biology, 7*(11), 973-977.

Abraham, E. (1987). Sir Robert Robinson and the early history of penicillin. *Natural product reports, 4*, 41-46.

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B., & Sheat, S. (1969). Structure of rhombohedral 2 zinc insulin crystals. *Nature, 224*(5218), 491-495.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography, 68*(4), 352-367.

Andersson, R. (2020). Lipidic Cubic Phase Microcrystallization and its Application in Serial Crystallography.

Andrews, R. J., Peterson, J. M., Haniff, H. S., Chen, J., Williams, C., Grefe, M., Disney, M. D., & Moss, W. N. (2020). An in silico map of the SARS-CoV-2 RNA Structurome. *bioRxiv : the preprint server for biology*, 2020.2004.2017.045161. doi:10.1101/2020.04.17.045161

Barty, A., Caleman, C., Aquila, A., Timneanu, N., Lomb, L., White, T. A., Andreasson, J., Arnlund, D., Bajt, S., Barends, T. R., Barthelmess, M., Bogan, M. J., Bostedt, C., Bozek, J. D., Coffee, R., Coppola, N., Davidsson, J., Deponte, D. P., Doak, R. B., Ekeberg, T., Elser, V., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Fromme, P., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, R., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Hunter, M. S., Johansson, L., Kassemeyer, S., Kimmel, N., Kirian, R. A., Liang, M., Maia, F. R., Malmerberg, E., Marchesini, S., Martin, A. V., Nass, K., Neutze, R., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Scott, H., Schlichting, I., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R. G., Soltau, H., Spence, J. C., Stellato, F., Stern, S., Struder, L., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Wunderer, C. B., & Chapman, H. N. (2012). Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat Photonics, 6*, 35-40. doi:10.1038/nphoton.2011.297

Barty, A., Kirian, R. A., Maia, F. R., Hantke, M., Yoon, C. H., White, T. A., & Chapman, H. (2014). Cheetah: software for high-throughput reduction and analysis of serial

femtosecond X-ray diffraction data. *Journal of applied crystallography, 47*(3), 1118-1131.

Bhardwaj, K., Sun, J., Holzenburg, A., Guarino, L. A., & Kao, C. C. (2006). RNA recognition and cleavage by the SARS coronavirus endoribonuclease. *Journal of molecular biology, 361*(2), 243-256.

Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G., & Vogel, H. (2011). Overcoming barriers to membrane protein structure determination. *Nature biotechnology, 29*(4), 335-340.

Blundell, T. L., Jhoti, H., & Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nature Reviews Drug Discovery, 1*(1), 45-54.

Bourgeois, D., & Royant, A. (2005). Advances in kinetic protein crystallography. *Current opinion in structural biology, 15*(5), 538-547.

Boyer, P. D. (1998). ATP synthase—past and future. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1365*(1-2), 3-9.

Bragg, W. H., & Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character, 88*(605), 428-438.

Bragg, W. L. (1913). The structure of some crystals as indicated by their diffraction of X-rays. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character, 89*(610), 248-277.

Brisse, M., & Ly, H. (2019). Comparative structure and function analysis of the RIG-I-like receptors: RIG-I and MDA5. *Frontiers in immunology, 10*, 1586.

Brosey, C. A., & Tainer, J. A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current opinion in structural biology, 58*, 197-213.

Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica Section D: Biological Crystallography, 56*(3), 328-341.

Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallographica Section F: Structural Biology Communications, 71*(1), 3-18.

Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., & Weierstall, U. (2011). Femtosecond X-ray protein nanocrystallography. *Nature, 470*(7332), 73-77.

Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnel, K. U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schroter, C. D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., & Spence, J. C. (2011). Femtosecond X-ray protein nanocrystallography. *Nature, 470*(7332), 73-77. doi:10.1038/nature09750

Chen, D. Y., Liu, H., Takeda, S., Tu, H.-C., Sasagawa, S., Van Tine, B. A., Lu, D., Cheng, E. H., & Hsieh, J. J. (2010). Taspase1 functions as a non-oncogene addiction protease that coordinates cancer cell proliferation and apoptosis. *Cancer research, 70*(13), 5358-5367.

Chow, K. T., Gale Jr, M., & Loo, Y.-M. (2018). RIG-I and other RNA sensors in antiviral immunity. *Annual review of immunology, 36*, 667-694.

Congreve, M., Langmead, C. J., Mason, J. S., & Marshall, F. H. (2011). Progress in structure based drug design for G protein-coupled receptors. *Journal of medicinal chemistry, 54*(13), 4283-4311.

Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A., & Riekel, C. (1998). Small is beautiful: protein micro-crystallography. *Nature structural biology, 5*(8), 634-637.

Davidson, A. P., Shubitz, L. F., Alcott, C. J., & Sykes, J. E. (2019). Selected clinical features of coccidioidomycosis in dogs. *Medical mycology, 57*(Supplement_1), S67-S75.

de Sanctis, D. (2021). *The ID29 upgrade project: a new serial crystallography beamline for time resolved studies at ESRF-EBS.* Paper presented at the ACTA CRYSTALLOGRAPHICA A-FOUNDATION AND ADVANCES.

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr, 40*(1), 82-92.

DeLucas, L. J., Smith, C. D., Smith, H. W., Vijay-Kumar, S., Senadhi, S. E., Ealick, S. E., Carter, D. C., Snyder, R. S., Weber, P. C., & Salemme, F. R. (1989). Protein crystal growth in microgravity. *Science, 246*(4930), 651-654.

Deng, X., & Baker, S. C. (2018). An "Old" protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology, 517*, 157-163.

Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., Kao, C. C., & Baker, S. C. (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci U S A, 114*(21), E4251-E4260. doi:10.1073/pnas.1618310114

Dessau, M. A., & Modis, Y. (2011). Protein crystallization for X-ray crystallography. *JoVE (Journal of Visualized Experiments)*(47), e2285.

Disfani, F. M., Hsu, W. L., Mizianty, M. J., Oldfield, C. J., Xue, B., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2012). MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics, 28*(12), i75-83. doi:10.1093/bioinformatics/bts209

Doppler, D., Rabbani, M. T., Letrun, R., Cruz Villarreal, J., Kim, D. H., Gandhi, S., Egatz-Gomez, A., Sonker, M., Chen, J., Koua, F. H. M., Yang, J., Youssef, M., Mazalova, V., Bajt, S., Shelby, M. L., Coleman, M. A., Wiedorn, M. O., Knoska, J., Schon, S., Sato, T., Hunter, M. S., Hosseinizadeh, A., Kuptiz, C., Nazari, R., Alvarez, R. C., Karpos, K., Zaare, S., Dobson, Z., Discianno, E., Zhang, S., Zook, J. D., Bielecki, J., de Wijn, R., Round, A. R., Vagovic, P., Kloos, M., Vakili, M., Ketawala, G. K., Stander, N. E., Olson, T. L., Morin, K., Mondal, J., Nguyen, J., Meza-Aguilar, J. D., Kodis, G., Vaiana, S., Martin-Garcia, J. M., Mariani, V., Schwander, P., Schmidt, M., Messerschmidt, M., Ourmazd, A., Zatsepin, N., Weierstall, U., Bruce, B. D., Mancuso, A. P., Grant, T., Barty, A., Chapman, H. N., Frank, M., Fromme, R., Spence, J. C. H., Botha, S., Fromme, P., Kirian, R. A., & Ros, A. (2022). Co-flow injection for serial crystallography at X-ray free-electron lasers. *J Appl Crystallogr, 55*(Pt 1), 1-13. doi:10.1107/S1600576721011079

Eckert, M. (2012). Max von Laue and the discovery of X-ray diffraction in 1912. In (Vol. 524, pp. A83-A85): WILEY-VCH Verlag Berlin.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography, 66*(4), 486-501.

Engelman, A., & Cherepanov, P. (2012). The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology, 10*(4), 279-290.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography, 69*(7), 1204-1214.

Frazier, M. N., Dillard, L. B., Krahn, J. M., Perera, L., Williams, J. G., Wilson, I. M., Stewart, Z. D., Pillon, M. C., Deterding, L. J., & Borgnia, M. J. (2021). Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic acids research, 49*(17), 10136-10149.

Frazier, M. N., Dillard, L. B., Krahn, J. M., Perera, L., Williams, J. G., Wilson, I. M., Stewart, Z. D., Pillon, M. C., Deterding, L. J., Borgnia, M. J., & Stanley, R. E. (2021). Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic Acids Res, 49*(17), 10136-10149. doi:10.1093/nar/gkab719

Frazier, M. N., Wilson, I. M., Krahn, J. M., Butay, K. J., Dillard, L. B., Borgnia, M. J., & Stanley, R. E. (2022). Flipped Over U: Structural Basis for dsRNA Cleavage by the SARS-CoV-2 Endoribonuclease. *bioRxiv*.

Frediansyah, A., Tiwari, R., Sharun, K., Dhama, K., & Harapan, H. (2021). Antivirals for COVID-19: A critical review. *Clinical Epidemiology and global health, 9*, 90-98.

Friedrich, W., Knipping, P., & Laue, M. (1913). Interferenzerscheinungen bei roentgenstrahlen. *Annalen der Physik, 346*(10), 971-988.

Fromme, P. (2015). XFELs open a new era in structural chemical biology. *Nature chemical biology, 11*(12), 895-899.

Fromme, P., Graves, W. S., & Martin-Garcia, J. M. (2020). Serial Femtosecond Crystallography: A Decade at the Forefront in Structural Biology. *eLS*, 1-17.

Garman, E. F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D: Biological Crystallography, 66*(4), 339-351.

Garman, E. F., & Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography, 62*(1), 32-47.

Gevorkov, Y., Yefanov, O., Barty, A., White, T. A., Mariani, V., Brehm, W., Tolstikova, A., Grigat, R.-R., & Chapman, H. N. (2019). XGANDALF–extended gradient descent algorithm for lattice finding. *Acta Crystallographica Section A: Foundations and Advances, 75*(5), 694-704.

Gilbile, D., Shelby, M. L., Lyubimov, A. Y., Wierman, J. L., Monteiro, D. C., Cohen, A. E., Russi, S., Coleman, M. A., Frank, M., & Kuhl, T. L. (2021). Plug-and-play polymer microfluidic chips for hydrated, room temperature, fixed-target serial crystallography. *Lab on a Chip, 21*(24), 4831-4845.

Gouaux, J. E., Stevens, R. C., & Lipscomb, W. N. (1990). Crystal structures of aspartate carbamoyltransferase ligated with phosphonoacetamide, malonate, and CTP or ATP at 2.8-. ANG. resolution and neutral pH. *Biochemistry, 29*(33), 7702-7715.

Graupmann-Kuzma, A., Valentine, B. A., Shubitz, L. F., Dial, S. M., Watrous, B., & Tornquist, S. J. (2008). Coccidioidomycosis in dogs and cats: a review. *Journal of the American Animal Hospital Association, 44*(5), 226-235.

Guarino, L. A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., & Kao, C. (2005). Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. *J Mol Biol, 353*(5), 1106-1117. doi:10.1016/j.jmb.2005.09.007

Guo, H.-C., Xu, Q., Buckley, D., & Guan, C. (1998). Crystal Structures of FlavobacteriumGlycosylasparaginase: AN N-TERMINAL NUCLEOPHILE HYDROLASE ACTIVATED BY INTRAMOLECULAR PROTEOLYSIS. *Journal of Biological Chemistry, 273*(32), 20205-20212.

Gusach, A., Maslov, I., Luginina, A., Borshchevskiy, V., Mishin, A., & Cherezov, V. (2020). Beyond structure: emerging approaches to study GPCR dynamics. *Current opinion in structural biology, 63*, 18-25.

Hackbart, M., Deng, X., & Baker, S. C. (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc Natl Acad Sci U S A, 117*(14), 8094-8103. doi:10.1073/pnas.1921485117

Hajdu, J., & Johnson, L. N. (1990). Progress with Laue diffraction studies on protein and virus crystals. *Biochemistry, 29*(7), 1669-1678.

Hall, T., Biosciences, I., & Carlsbad, C. (2011). BioEdit: an important software for molecular biology. *GERF Bull Biosci, 2*(1), 60-61.

Hanson, J., Paliwal, K., & Zhou, Y. (2018). Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. *J Chem Inf Model, 58*(11), 2369-2376. doi:10.1021/acs.jcim.8b00636

Hasegawa, T., Hara, T., Hosoda, N., Maesaka, H., Matsui, S., Otake, Y., Tanaka, H., & Ishikawa, T. (2016). Commissioning Status of the Extreme-Ultraviolet FEL Facility at SACLA.

Hodgkin, D. C., Kamper, J., Mackay, M., Pickworth, J., Trueblood, K. N., & White, J. G. (1956). Structure of vitamin B12. *Nature, 178*(4524), 64-66.

Hsieh, J. J.-D., Cheng, E. H.-Y., & Korsmeyer, S. J. (2003). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell, 115*(3), 293-303.

Ishigami, I., Lewis-Ballester, A., Echelmeier, A., Brehm, G., Zatsepin, N. A., Grant, T. D., Coe, J. D., Lisova, S., Nelson, G., Zhang, S., Dobson, Z. F., Boutet, S., Sierra, R. G., Batyuk, A., Fromme, P., Fromme, R., Spence, J. C. H., Ros, A., Yeh, S. R., & Rousseau, D. L. (2019). Snapshot of an oxygen intermediate in the catalytic reaction of cytochrome c oxidase. *Proc Natl Acad Sci U S A, 116*(9), 3572-3577. doi:10.1073/pnas.1814526116

Ishikawa, T. (2019). Accelerator-based X-ray sources: synchrotron radiation, X-ray free electron lasers and beyond. *Philosophical Transactions of the Royal Society A, 377*(2147), 20180231.

Johansson, L. C., Wöhri, A. B., Katona, G., Engström, S., & Neutze, R. (2009). Membrane protein crystallization from lipidic phases. *Current opinion in structural biology, 19*(4), 372-378.

Joosten, R. P., Long, F., Murshudov, G. N., & Perrakis, A. (2014). The PDB_REDO server for macromolecular structure model optimization. *IUCrJ, 1*(4), 213-220.

Juers, D. H., & Ruffin, J. (2014). MAP_CHANNELS: a computation tool to aid in the visualization and characterization of solvent channels in macromolecular crystals. *Journal of applied crystallography, 47*(6), 2105-2108.

Kabsch, W. (2010). Xds. *Acta Crystallographica Section D: Biological Crystallography, 66*(2), 125-132.

Kendrew, J. C., & Parrish, R. (1957). The crystal structure of myoglobin III. Sperm-whale myoglobin. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 238*(1214), 305-324.

Khan, J. A., Dunn, B. M., & Tong, L. (2005). Crystal structure of human Taspase1, a crucial protease regulating the function of MLL. *Structure, 13*(10), 1443-1452. doi:10.1016/j.str.2005.07.006

Khan, R. J., Jha, R. K., Singh, E., Jain, M., Amera, G. M., Singh, R. P., Muthukumaran, J., & Singh, A. K. (2022). Identification of promising antiviral drug candidates against non-structural protein 15 (NSP15) from SARS-CoV-2: an in silico assisted drug-repurposing study. *J Biomol Struct Dyn, 40*(1), 438-448. doi:10.1080/07391102.2020.1814870

Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., & Wu, R. (2011). High-throughput protein purification and quality assessment for crystallization. *Methods, 55*(1), 12-28.

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., & Joachimiak, A. (2020). Crystal structure of Nsp15

endoribonuclease NendoU from SARS-CoV-2. *Protein Sci, 29*(7), 1596-1605. doi:10.1002/pro.3873

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., & Joachimiak, A. (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein science, 29*(7), 1596-1605.

Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., Kang, S., Nicolaescu, V., Randall, G., Michalska, K., & Joachimiak, A. (2021). Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Commun Biol, 4*(1), 193. doi:10.1038/s42003-021-01735-9

Kindler, E., Gil-Cruz, C., Spanier, J., Li, Y., Wilhelm, J., Rabouw, H. H., Zust, R., Hwang, M., V'Kovski, P., Stalder, H., Marti, S., Habjan, M., Cervantes-Barragan, L., Elliot, R., Karl, N., Gaughan, C., van Kuppeveld, F. J., Silverman, R. H., Keller, M., Ludewig, B., Bergmann, C. C., Ziebuhr, J., Weiss, S. R., Kalinke, U., & Thiel, V. (2017). Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLoS Pathog, 13*(2), e1006195. doi:10.1371/journal.ppat.1006195

Ko, I. S., Kang, H.-S., Heo, H., Kim, C., Kim, G., Min, C.-K., Yang, H., Baek, S. Y., Choi, H.-J., & Mun, G. (2017). Construction and commissioning of PAL-XFEL facility. *Applied Sciences, 7*(5), 479.

Koshland Jr, D. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 44*(2), 98.

Kozlowski, L. P., & Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics, 13*, 111. doi:10.1186/1471-2105-13-111

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol, 372*(3), 774-797. doi:10.1016/j.jmb.2007.05.022

Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N. A., Rendek, K. N., Hunter, M. S., Shoeman, R. L., White, T. A., & Wang, D. (2014). Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature, 513*(7517), 261-265.

Kupitz, C., Grotjohann, I., Conrad, C. E., Roy-Chowdhury, S., Fromme, R., & Fromme, P. (2014). Microcrystallization techniques for serial femtosecond crystallography using photosystem II from Thermosynechococcus elongatus as a model system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1647), 20130316.

Kupitz, C., Olmos, J. L., Jr., Holl, M., Tremblay, L., Pande, K., Pandey, S., Oberthur, D., Hunter, M., Liang, M., Aquila, A., Tenboer, J., Calvey, G., Katz, A., Chen, Y., Wiedorn, M. O., Knoska, J., Meents, A., Majriani, V., Norwood, T., Poudyal, I., Grant, T., Miller, M. D., Xu, W., Tolstikova, A., Morgan, A., Metz, M., Martin-Garcia, J. M., Zook, J. D., Roy-Chowdhury, S., Coe, J., Nagaratnam, N., Meza, D., Fromme, R., Basu, S., Frank, M., White, T., Barty, A., Bajt, S., Yefanov, O., Chapman, H. N., Zatsepin, N., Nelson, G., Weierstall, U., Spence, J., Schwander, P., Pollack, L., Fromme, P., Ourmazd, A., Phillips, G. N., Jr., & Schmidt, M. (2017). Structural enzymology using X-ray free electron lasers. *Struct Dyn, 4*(4), 044003. doi:10.1063/1.4972069

Kupitz, C., Olmos Jr, J. L., Holl, M., Tremblay, L., Pande, K., Pandey, S., Oberthür, D., Hunter, M., Liang, M., & Aquila, A. (2017). Structural enzymology using X-ray free electron lasers. *Structural Dynamics, 4*(4), 044003.

Larwood, D. J. (2020). Nikkomycin Z—ready to meet the promise? *Journal of Fungi, 6*(4), 261.

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. In: ACS Publications.

Layton, J., Nagaratnam, N., Jernigan, R. J., Schneider, J., Flint, A., Mroczkowski, B., Fromme, P., Garcia, J. M., & Singharoy, A. (2020). Molecular Dynamics Simulations for Improving Crystal Quality and Illuminating the Function of Taspase1: A Therapeutic Target. *Biophysical Journal, 118*(3), 507a.

Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R., & Oza, J. P. (2019). Escherichia coli-based cell-free protein synthesis: protocols for a robust, flexible, and accessible platform technology. *JoVE (Journal of Visualized Experiments)*(144), e58882.

Lewis, C. M., Flory, J. D., Moore, T. A., Moore, A. L., Rittmann, B. E., Vermaas, W. F., Torres, C. I., & Fromme, P. (2022). Electrochemically Driven Photosynthetic Electron Transport in Cyanobacteria Lacking Photosystem II. *Journal of the American Chemical Society*.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., & McCoy, A. J. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology, 75*(10), 861-877.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., & Adams, P. D. (2017). Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallographica Section D: Structural Biology, 73*(2), 148-157.

Liu, H., Takeda, S., Cheng, E. H., & Hsieh, J. J. (2008). Biphasic MLL takes helm at cell cycle control: implications in human mixed lineage leukemia. *Cell cycle, 7*(4), 428-435.

Liu, W., Wacker, D., Gati, C., Han, G. W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., Zatsepin, N. A., Li, D., Messerschmidt, M., Boutet, S., Williams, G. J., Koglin, J. E., Seibert, M. M., Wang, C., Shah, S. T., Basu, S., Fromme, R., Kupitz, C., Rendek, K. N., Grotjohann, I., Fromme, P., Kirian, R. A., Beyerlein, K. R., White, T. A., Chapman, H. N., Caffrey, M., Spence, J. C., Stevens, R. C., & Cherezov, V. (2013). Serial femtosecond crystallography of G protein-coupled receptors. *Science, 342*(6165), 1521-1524. doi:10.1126/science.1244142

Lu, Y. (2017). Cell-free synthetic biology: Engineering in an open world. *Synth Syst Biotechnol, 2*(1), 23-27. doi:10.1016/j.synbio.2017.02.003

Madey, J. M. (1971). Stimulated emission of bremsstrahlung in a periodic magnetic field. *Journal of Applied Physics, 42*(5), 1906-1913.

Mariani, V., Morgan, A., Yoon, C. H., Lane, T. J., White, T. A., O'Grady, C., Kuhn, M., Aplin, S., Koglin, J., & Barty, A. (2016). OnDA: online data analysis and feedback for serial X-ray imaging. *Journal of applied crystallography, 49*(3), 1073-1080.

Martin-Garcia, J. M. (2021). Protein Dynamics and Time Resolved Protein Crystallography at Synchrotron Radiation Sources: Past, Present and Future. *Crystals, 11*(5), 521.

Martin-Garcia, J. M., Zhu, L., Mendez, D., Lee, M.-Y., Chun, E., Li, C., Hu, H., Subramanian, G., Kissick, D., & Ogata, C. (2019). High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. *IUCrJ, 6*(3), 412-425.

Matlock, M., Hopfer, S., & Ogunseitan, O. A. (2019). Communicating risk for a climate-sensitive disease: A case study of valley fever in central California. *International journal of environmental research and public health, 16*(18), 3254.

McKenzie, E. A., & Abbott, W. M. (2018). Expression of recombinant proteins in insect and mammalian cells. *Methods, 147*, 40-49.

McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., & Phizackerley, R. P. (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of synchrotron radiation, 9*(6), 401-406.

Meents, A., Roedig, P., Ginn, H. M., Pakendorf, T., Sutton, G., Harlos, K., Walter, T. S., Meyer, J., Fischer, P., & Duman, R. (2017). Sample Preparation and Data Collection for High-Speed Fixed-Target Serial Femtosecond Crystallography.

Miles, A. J., & Wallace, B. A. (2016). Circular dichroism spectroscopy of membrane proteins. *Chemical society reviews, 45*(18), 4859-4872.

Milne, C. J., Schietinger, T., Aiba, M., Alarcon, A., Alex, J., Anghel, A., Arsov, V., Beard, C., Beaud, P., & Bettoni, S. (2017). SwissFEL: the Swiss X-ray free electron laser. *Applied Sciences, 7*(7), 720.

Moffat, K. (2019). Laue diffraction and time-resolved crystallography: a personal history. *Philosophical Transactions of the Royal Society A, 377*(2147), 20180243.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography, 67*(4), 355-367.

Nagaratnam, N., Delker, S. L., Jernigan, R., Edwards, T. E., Snider, J., Thifault, D., Williams, D., Nannenga, B. L., Stofega, M., & Sambucetti, L. (2021). Structural insights into the function of the catalytically active human Taspase1. *Structure, 29*(8), 873-885. e875.

Neidle, S. (2011). *Cancer drug design and discovery*: Elsevier.

Nelson, G., Kirian, R. A., Weierstall, U., Zatsepin, N. A., Faragó, T., Baumbach, T., Wilde, F., Niesler, F. B., Zimmer, B., & Ishigami, I. (2016). Three-dimensional-printed gas dynamic virtual nozzles for x-ray laser sample delivery. *Optics express, 24*(11), 11515-11530.

Neutze, R., Wouts, R., Van der Spoel, D., Weckert, E., & Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature, 406*(6797), 752-757.

Niizuma, H., Cheng, E. H., & Hsieh, J. J. (2015). Taspase 1: a protease with many biological surprises. *Molecular & cellular oncology, 2*(4), e999513.

Nogly, P., Panneels, V., Nelson, G., Gati, C., Kimura, T., Milne, C., Milathianaki, D., Kubo, M., Wu, W., & Conrad, C. (2016). Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography. *Nature communications, 7*(1), 1-9.

Nogly, P., Weinert, T., James, D., Carbajo, S., Ozerov, D., Furrer, A., Gashi, D., Borin, V., Skopintsev, P., & Jaeger, K. (2018). Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science, 361*(6398), eaat0094.

Oinonen, C., & Rouvinen, J. (2000). Structural comparison of Ntn-hydrolases. *Protein science, 9*(12), 2329-2337.

Olmos, J. L., Jr., Pandey, S., Martin-Garcia, J. M., Calvey, G., Katz, A., Knoska, J., Kupitz, C., Hunter, M. S., Liang, M., Oberthuer, D., Yefanov, O., Wiedorn, M., Heyman, M., Holl, M., Pande, K., Barty, A., Miller, M. D., Stern, S., Roy-Chowdhury, S., Coe, J., Nagaratnam, N., Zook, J., Verburgt, J., Norwood, T., Poudyal, I., Xu, D., Koglin, J., Seaberg, M. H., Zhao, Y., Bajt, S., Grant, T., Mariani, V., Nelson, G., Subramanian, G., Bae, E., Fromme, R., Fung, R., Schwander, P., Frank, M., White, T. A., Weierstall, U., Zatsepin, N., Spence, J., Fromme, P., Chapman, H. N., Pollack, L., Tremblay, L., Ourmazd, A., Phillips, G. N., Jr., & Schmidt, M. (2018). Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biol, 16*(1), 59. doi:10.1186/s12915-018-0524-5

Olmos, J. L., Pandey, S., Martin-Garcia, J. M., Calvey, G., Katz, A., Knoska, J., Kupitz, C., Hunter, M. S., Liang, M., & Oberthuer, D. (2018). Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC biology, 16*(1), 1-15.

Palmer III, A. G. (2001). NMR probes of molecular dynamics: overview and comparison with other techniques. *Annual review of biophysics and biomolecular structure, 30*(1), 129-155.

Pandey, S., Bean, R., Sato, T., Poudyal, I., Bielecki, J., Cruz Villarreal, J., Yefanov, O., Mariani, V., White, T. A., Kupitz, C., Hunter, M., Abdellatif, M. H., Bajt, S., Bondar, V., Echelmeier, A., Doppler, D., Emons, M., Frank, M., Fromme, R., Gevorkov, Y., Giovanetti, G., Jiang, M., Kim, D., Kim, Y., Kirkwood, H., Klimovskaia, A., Knoska, J., Koua, F. H. M., Letrun, R., Lisova, S., Maia, L., Mazalova, V., Meza, D., Michelat, T., Ourmazd, A., Palmer, G., Ramilli, M., Schubert, R., Schwander, P., Silenzi, A., Sztuk-Dambietz, J., Tolstikova, A., Chapman, H. N., Ros, A., Barty, A., Fromme, P., Mancuso, A. P., & Schmidt, M. (2020). Time-resolved serial femtosecond crystallography at the European XFEL. *Nat Methods, 17*(1), 73-78. doi:10.1038/s41592-019-0628-z

Pandey, S., Calvey, G., Katz, A. M., Malla, T. N., Koua, F. H. M., Martin-Garcia, J. M., Poudyal, I., Yang, J. H., Vakili, M., Yefanov, O., Zielinski, K. A., Bajt, S., Awel, S., Doerner, K., Frank, M., Gelisio, L., Jernigan, R., Kirkwood, H., Kloos, M., Koliyadu, J., Mariani, V., Miller, M. D., Mills, G., Nelson, G., Olmos, J. L., Jr., Sadri, A., Sato, T., Tolstikova, A., Xu, W., Ourmazd, A., Spence, J. C. H., Schwander, P., Barty, A., Chapman, H. N., Fromme, P., Mancuso, A. P., Phillips, G. N., Jr., Bean, R., Pollack, L., & Schmidt, M. (2021). Observation of substrate diffusion and ligand binding in enzyme crystals using high-repetition-rate mix-and-inject serial crystallography. *IUCrJ, 8*(Pt 6), 878-895. doi:10.1107/S2052252521008125

Pellegrini, C. (2012). The history of X-ray free-electron lasers. *The European Physical Journal H, 37*(5), 659-708.

Pellegrini, C., & Stöhr, J. (2003). X-ray free-electron lasers—principles, properties and applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 500*(1-3), 33-40.

Perry, J. K., Appleby, T. C., Bilello, J. P., Feng, J. Y., Schmitz, U., & Campbell, E. A. (2021). An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *Journal of Biological Chemistry, 297*(4).

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., & North, A. (1960). Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature, 185*(4711), 416-422.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry, 25*(13), 1605-1612.

Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., Perera, L., Hayne, C. K., Gordon, J., Stewart, Z. D., Sobhany, M., Deterding, L. J., Hsu, A. L., Dandey, V. P., Borgnia, M. J., & Stanley, R. E. (2020). Cryo-EM Structures of the SARS-CoV-2 Endoribonuclease Nsp15. *bioRxiv*. doi:10.1101/2020.08.11.244863

Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., Perera, L., Hayne, C. K., Gordon, J., Stewart, Z. D., Sobhany, M., Deterding, L. J., Hsu, A. L., Dandey, V. P., Borgnia, M. J., & Stanley, R. E. (2021). Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat Commun, 12*(1), 636. doi:10.1038/s41467-020-20608-z

Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Micetic, I., Camilloni, C., Davey, N., Dosztanyi, Z., Meszaros, B., Monzon, A. M., Parisi, G., Schad, E., Sormanni, P., Tompa, P., Vendruscolo, M., Vranken, W. F., & Tosatto, S. C. E. (2018). MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res, 46*(D1), D471-D476. doi:10.1093/nar/gkx1071

Powell, H. R., Johnson, O., & Leslie, A. G. (2013). Autoindexing diffraction images with iMosflm. *Acta Crystallographica Section D: Biological Crystallography, 69*(7), 1195-1203.

Quiocho, F. A., & Richards, F. M. (1966). The enzymic behavior of carboxypeptidase-A in the solid state. *Biochemistry, 5*(12), 4062-4076.

Rakash, S., Rana, F., Rafiq, S., Masood, A., & Amin, S. (2012). Role of proteases in cancer: A review. *Biotechnology and Molecular Biology Reviews, 7*(4), 90-101.

Ravelli, R. B., & McSweeney, S. M. (2000). The 'fingerprint'that X-rays can leave on structures. *Structure, 8*(3), 315-328.

Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Srajer, V., & Stoddard, B. L. (1999). Laue crystallography: coming of age. *Journal of synchrotron radiation, 6*(4), 891-917.

Ren, Z., & Moffat, K. (1994). Laue crystallography for studying rapid reactions. *Journal of synchrotron radiation, 1*(1), 78-82.

Röntgen, W. C. (1895). *Ueber eine neue Art von Strahlen*: Phys.-med. Gesellschaft.

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein science, 28*(8), 1412-1422.

Saibil, H. R. (2022). Cryo-EM in molecular and cellular biology. *Molecular Cell, 82*(2), 274-284.

Saramago, M., Costa, V. G., Souza, C. S., Barria, C., Domingues, S., Viegas, S. C., Lousa, D., Soares, C. M., Arraiano, C. M., & Matos, R. G. (2022). The nsp15 Nuclease as a Good Target to Combat SARS-CoV-2: Mechanism of Action and Its Inactivation with FDA-Approved Drugs. *Microorganisms, 10*(2). doi:10.3390/microorganisms10020342

Sass, G., Larwood, D. J., Martinez, M., Chatterjee, P., Xavier, M. O., & Stevens, D. A. (2021). Nikkomycin Z against Disseminated Coccidioidomycosis in a Murine Model of Sustained-Release Dosing. *Antimicrobial Agents and Chemotherapy, 65*(10), e00285-00221.

Schotte, F., Lim, M., Jackson, T. A., Smirnov, A. V., Soman, J., Olson, J. S., Phillips Jr, G. N., Wulff, M., & Anfinrud, P. A. (2003). Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science, 300*(5627), 1944-1947.

Shubitz, L. F. (2007). Comparative aspects of coccidioidomycosis in animals and humans. *Annals of the New York Academy of Sciences, 1111*(1), 395-403.

Shubitz, L. F., Roy, M. E., Nix, D. E., & Galgiani, J. N. (2013). Efficacy of nikkomycin Z for respiratory coccidioidomycosis in naturally infected dogs. *Medical mycology, 51*(7), 747-754.

Shubitz, L. F., Trinh, H. T., Perrill, R. H., Thompson, C. M., Hanan, N. J., Galgiani, J. N., & Nix, D. E. (2014). Modeling nikkomycin Z dosing and pharmacology in

murine pulmonary coccidioidomycosis preparatory to phase 2 clinical trials. *The Journal of infectious diseases, 209*(12), 1949-1954.

Sierra, R. G., Gati, C., Laksmono, H., Dao, E. H., Gul, S., Fuller, F., Kern, J., Chatterjee, R., Ibrahim, M., Brewster, A. S., Young, I. D., Michels-Clark, T., Aquila, A., Liang, M., Hunter, M. S., Koglin, J. E., Boutet, S., Junco, E. A., Hayes, B., Bogan, M. J., Hampton, C. Y., Puglisi, E. V., Sauter, N. K., Stan, C. A., Zouni, A., Yano, J., Yachandra, V. K., Soltis, S. M., Puglisi, J. D., & DeMirci, H. (2016). Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nat Methods, 13*(1), 59-62. doi:10.1038/nmeth.3667

Silverman, A. D., Karim, A. S., & Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics, 21*(3), 151-170.

Šiurkus, J., & Neubauer, P. (2011). Reducing conditions are the key for efficient production of active ribonuclease inhibitor in Escherichia coli. *Microbial cell factories, 10*(1), 1-15.

Šrajer, V., & Schmidt, M. (2017). Watching proteins function with time-resolved x-ray crystallography. *Journal of physics D: Applied physics, 50*(37), 373001.

Stagno, J. R., Liu, Y., Bhandari, Y. R., Conrad, C. E., Panja, S., Swain, M., Fan, L., Nelson, G., Li, C., Wendel, D. R., White, T. A., Coe, J. D., Wiedorn, M. O., Knoska, J., Oberthuer, D., Tuckey, R. A., Yu, P., Dyba, M., Tarasov, S. G., Weierstall, U., Grant, T. D., Schwieters, C. D., Zhang, J., Ferre-D'Amare, A. R., Fromme, P., Draper, D. E., Liang, M., Hunter, M. S., Boutet, S., Tan, K., Zuo, X., Ji, X., Barty, A., Zatsepin, N. A., Chapman, H. N., Spence, J. C., Woodson, S. A., & Wang, Y. X. (2017). Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature, 541*(7636), 242-246. doi:10.1038/nature20599

Stevens, R. C. (2000). High-throughput protein crystallization. *Current opinion in structural biology, 10*(5), 558-563.

Stoddard, B. L. (2001). Trapping reaction intermediates in macromolecular crystals for structural analyses. *Methods, 24*(2), 125-138.

Su, X.-D., Zhang, H., Terwilliger, T. C., Liljas, A., Xiao, J., & Dong, Y. (2015). Protein crystallography from the perspective of technology developments. *Crystallography reviews, 21*(1-2), 122-153.

Takeda, S., Chen, D. Y., Westergard, T. D., Fisher, J. K., Rubens, J. A., Sasagawa, S., Kan, J. T., Korsmeyer, S. J., Cheng, E. H.-Y., & Hsieh, J. J.-D. (2006). Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes & development, 20*(17), 2397-2409.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry, 31*(2), 455-461.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology, 19*(3), 155-170.

Van Driessche, A. E., Otálora, F., Gavira, J. A., & Sazaki, G. (2008). Is agarose an impurity or an impurity filter? In situ observation of the joint gel/impurity effect on protein crystal growth kinetics. *Crystal Growth and Design, 8*(10), 3623-3629.

Wang, D., Weierstall, U., Pollack, L., & Spence, J. (2014). Double-focusing mixing jet for XFEL study of chemical kinetics. *Journal of synchrotron radiation, 21*(6), 1364-1366.

Weierstall, U. (2014). Liquid sample delivery techniques for serial femtosecond crystallography. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1647), 20130337.

Weierstall, U., Spence, J., & Doak, R. (2012). Injector for scattering measurements on fully solvated biospecies. *Review of Scientific Instruments, 83*(3), 035108.

Weinert, T., Skopintsev, P., James, D., Dworkowski, F., Panepucci, E., Kekilli, D., Furrer, A., Brünle, S., Mous, S., & Ozerov, D. (2019). Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography. *Science, 365*(6448), 61-65.

Weise, H., & Decking, W. (2017). *Commissioning and first lasing of the European XFEL.* Paper presented at the Proc. 38th Int. Free Electron Laser Conf.(FEL'17).

White, T. A. (2019). Processing serial crystallography data with CrystFEL: a step-by-step guide. *Acta Crystallographica Section D, 75*(2), 219-233.

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A., & Chapman, H. N. (2012). CrystFEL: a software suite for snapshot serial crystallography. *Journal of applied crystallography, 45*(2), 335-341.

White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., & Nakane, T. (2016). Recent developments in CrystFEL. *Journal of applied crystallography, 49*(2), 680-689.

Wilson, L., Ting, J., Lin, H., Shah, R., MacLean, M., Peterson, M. W., Stockamp, N., Libke, R., & Brown, P. (2019). The rise of valley fever: prevalence and cost burden of coccidioidomycosis infection in California. *International journal of environmental research and public health, 16*(7), 1113.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., & McCoy, A. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography, 67*(4), 235-242.

Wünsch, D., Hahlbrock, A., Jung, S., Schirmeister, T., Van Den Boom, J., Schilling, O., Knauer, S., & Stauber, R. (2016). Taspase1: a'misunderstood'protease with translational cancer relevance. *Oncogene, 35*(26), 3351-3364.

Wyckoff, H. W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L. N., Hardman, K. D., Allewell, N. M., Kelly, D. M., & Richards, F. M. (1967). Design of a diffractometer and flow cell system for X-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S. *Journal of molecular biology, 27*(3), 563-578.

Xu, X., Zhai, Y., Sun, F., Lou, Z., Su, D., Xu, Y., Zhang, R., Joachimiak, A., Zhang, X. C., & Bartlam, M. (2006). New antiviral target revealed by the hexameric structure of mouse hepatitis virus nonstructural protein nsp15. *Journal of virology, 80*(16), 7909-7917.

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., & Uversky, V. N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta, 1804*(4), 996-1010. doi:10.1016/j.bbapap.2010.01.011

Yang, H., & Rao, Z. (2021). Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology, 19*(11), 685-700.

Yang, Y., Hong, H., Zhang, Y., & Cai, W. (2009). Molecular imaging of proteases in cancer. *Cancer growth and metastasis, 2*, CGM. S2814.

APPENDIX A

CO-AUTHOR APPROVAL

I verify that the following co-authors have approved of my use of our publication/s in my dissertation

Alexander Batyuk

Alexandra Ros

Alexandra Tolstikova Michelle Sheikh

Andrew J. Flint

Brent L. Nannenga

Christopher Kupitz

Darren Thifault

Debra Hansen

Dewight Williams

Dhenu Logeswaran

Diandra Doppler

Emily Kaschner

Eranjalee Ranaweera,

Gihan Ketalawa

James J. Hsieh

Janey Snider

Jay-How Yang

Jose M. Martin-Garcia

Julian Chen

Lidia Sambucetti

Manashi Sonowal

Marc Messerschmidt

Marius Schmidt

Mark Hunter

Mary Stofega

Matt Colemann

Matthew Goode

Matthias Frank

Megan Shelby

Meng Liang,

Michele Zacks

Michelle Sheikh

Mimi Cho Yung

Mukul Sonker

Nirupa Nagaratnam

Petra Fromme

Raimund Fromme

Raymond Sierra

Richard Kirian

Sabine Botha

Sahba Zahre

Sebastien Boutet

Silvia L. Delker

Stella Lisova

Thomas E. Edwards

Thomas Grant

Valerio Mariani