Cryo-electron Microscopy Enables Structural Studies of Protein Complexes Relevant to Neurodegeneration

by

Purbasha Nandi

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

Approved April 2022 by the
Graduate Supervisory Committee:

Po-Lin Chiu, Chair
Yuval Mazor
Debra T. Hansen

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

The understanding of protein functions *in vivo* is very important since the protein is the building block of a cell. Cryogenic electron microscopy (cryo-EM) is capable of visualizing protein samples in their near-native states in high-resolution details. Cryo-EM enables the visualization of biomolecular structures at multiscale ranging from a cellular structure to an atomic structure of protein subunit.

Neurodegenerative diseases, like Alzheimer's disease and frontotemporal dementia, have multiple dysregulated signaling pathways. In my doctoral studies, I investigated two protein complexes relevant to these disorders: one is the proNGF- p75 neurotrophin receptor ($p75^{NTR}$) - sortilin neurotrophin complex and the other is the $p97^{R155H}$ mutant complex. The neurotrophins are a family of soluble basic growth factors involved in the development, maintenance, and proliferation of neurons in the central nervous system (CNS) and peripheral nervous system (PNS). The ligand for the neuronal receptors dictates the fate of the neuronal cells. My studies focused on understanding the binding interfaces between the proteins in the proNGF-$p75^{NTR}$-sortilin neuronal apoptotic complex. I have performed the biochemical characterization of the complex to understand how the complex formation occurs.

Single amino-acid mutation of R155H on the N-domain of p97 is known to be the prevalent mutation in 40% patients suffering from neurodegenerative disease. The $p97^{R155H}$ mutant exhibits abnormal ATPase activity and cofactor dysregulation. I pursued biochemical characterization in combination with single-particle cryo-EM to explore the interaction of $p97^{R155H}$ mutant with its cofactor p47 and determined the full-length structures of the $p97^{R155H}$-p47 assemblies for the first time. About 40% $p97^{R155H}$

organizes into higher order dodecamers, which lacks nucleotide binding, does not bind to p47, and closely resembles the structure of p97 bound with an adenosine triphosphate (ATP)-competitive inhibitor, CB-5083, suggesting an inactive state of the $p97^{R155H}$ mutant. The structures also revealed conformational changes of the arginine fingers which might contribute to the elevated $p97^{R155H}$ ATPase activity. Because the D1-D2 domain communication is important in regulating the ATPase function, I further studied the functions of the conserved L464 residue on the D1-D2 linker using mutagenesis and single-particle cryo-EM. The biochemical and structural results suggested the torsional constraint of the D1-D2 linker likely modulates the D2 ATPase activity. Our studies thus contributed to develop deeper knowledge of the intricate cellular mechanisms and the proteins affected in disease pathways.

DEDICATION

To my beloved parents, and my Mathematics teacher

Late Mrs. Shanti D'Souza

ACKNOWLEDGEMENTS

I am extremely grateful for this journey, and it would not be possible without the support and guidance from many mentors and friends. I was fortunate to have amazing teachers and colleagues.

Firstly, I would like to express my sincere gratitude to my Ph.D. supervisor Prof. Po-Lin Chiu for his patient support, constant motivation, immense knowledge, and project guidance. I am very grateful for giving me the opportunity to gain expertise in various aspects of biochemistry and single-particle cryo-EM from the amazing projects you gave me and for believing in my skillsets. His invaluable feedback pushed me to sharpen my thinking and brought my work to a higher level. Po-Lin is a passionate scientist, and I am proud to say my experience in the Chiu lab was intellectually exciting and fun and has energized me to continue in academic research. I sincerely hope I continue to have opportunities to interact with Po-Lin and the lab members for the rest of my research career. It would not be possible without Dr. Yu-Ping Poh's guidance with molecular biology experiments, and various insightful comments which incented me to widen my research from multiple perspectives. I will forever treasure the friendship with all my fellow lab mates, especially Dr. Chloe D. Truong and Ka Yi Chan, with whom I spent the maximum research hours, and everyone else for supporting me on the bad research days and all the fun lab fun trips around Arizona.

A major part of my thesis is possible due to an extremely good collaboration with Prof. Tsui-Fen Chou's lab at Caltech, with her expert guidance and the tremendous help from all her lab members for making my cryo-EM structure determination journey so exciting. All the electron microscopy work and cryo-EM data collection at Arizona State

University would not be possible without the huge support from our EM Scientists, especially Dr. Dewight Williams who trained me initially on the microscopes and helped me troubleshoot experiments by sharing his decades of experience. I also thank Dr. Janet Myers, at Pacific Northwest Center for Cryo-EM for her valuable suggestions.

I sincerely thank my dissertation committee members Dr. Yuval Mazor and Dr. Debra T. Hansen, to propel successful results from their discussion and critical analysis of the experimental designs. I am grateful to have friendly neighboring labs of Prof. Mazor's group during my early part of Ph.D. and Prof. Nannenga's group during the later part, for sharing all equipments upon request and being available for many scientific discussions. I would like to acknowledge all the fantastic project collaborations with great colleagues Dr. Raghu Narayan Pradeep and Dr. Guanhong Bu for being excellent team players.

In addition, I would like to thank my beloved parents for their wise counsel and sympathetic ear, for being my pillars of strength helping me pursue my dreams even during a worldwide pandemic situation. My heartfelt thanks to my genetic counterpart, my sister Kasturi for cheering me and for everything that helped me get to this day. I owe my previous teachers to instill my early curiosity and love in STEM and for their constant motivation, especially my mathematics teacher, *Late* Miss Shanti D'Souza for her undying optimism. My stay at Arizona and was made enjoyable and homely from all my lovely lab mates, friends, and cousins. I would like to thank my Ph.D. colleague and fiancé Subhadeep Dutta for helping me cherish graduate school and always encouraging me. I feel blessed to have strong peer support from my lovely friends for all the happy distractions outside research.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

LIST OF ABBREVIATIONS

2D, two-dimensional

3D, three-dimensional

AAA+, ATPase associated with diverse cellular activities

AD, Alzheimer's disease

ADP, Adenosine diphosphate.

ALS, amyotrophic lateral sclerosis

APP, Amyloid precursor protein

Aß, Amyloid beta peptides

ATP, Adenosine triphosphate

ATPγS, Adenosine 5-(γ-thio) triphosphate

BDNF, Brain-derived neurotrophic factor

BEVS, Bac-to-Bac® Baculovirus Expression vector system

CRD, Cysteine rich domain

Cryo-EM, cryogenic electron microscopy

CT, Cytosolic tail

D1, First ATPase domain of p97

D2, Second ATPase domain of p97

DUB, Deubiquitinating enzyme

E. coli, Escherichia coli

E1, ubiquitin-activating enzyme

E2, ubiquitin-conjugating enzyme

E3, ubiquitin ligase

ECD, Extracellular domain

ER, endoplasmic reticulum

ERAD, endoplasmic reticulum-associated degradation

ESI-MS, Electrospray-ionization mass spectrometry

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HTH, helix-turn-helix

IBMPFD, inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia

IC, Inclusion body

IPTG, Isopropyl β-D-1-thiogalactopyranoside

ISS, inter-subunit signaling

LB, Luria broth

MALDI, Matrix-assisted laser desorption/ionization

MBP, Maltose binding protein

melanogaster eyes closed gene (eyc) and vertebrate p47 domain.

MOI, Multiplicity of infection

MRW, mean residue weight

MSP1, multisystem proteinopathy type 1

MW, molecular weight

NF-kB, nuclear factor kappa-light-chain-enhancer of activated B cells

NGF, Nerve growth factor

Ni-NTA, Nickel-nitrilotriacetic acid agarose resin

NMR, Nuclear magnetic resonance

NMR, nuclear magnetic resonance

NSF, N-ethylmaleimide-sensitive factor

NTD, N terminal domain

NU, Npl4–Ufd1

$p75^{NTR}$, p75 neurotrophin receptor

PCR, Polymerase chain reaction

PDB, Protein Data Bank

PMSF, Phenylmethyl sulfonyl fluoride

PNGaseF, Peptide N-glycosidase F

PP1, protein phosphatase 1

PTM, Post-translational modification

SEC, Size exclusion chromatography

SEP, Saccharomyces cerevisiae suppressor of high-copy PP1 protein (shp1), Drosophila

Sf9, Spodoptera frugiperda 9

SNR, Signal to noise ratio

SPA, Single particle analysis

TEM, Transmission electron microscope

TGN, trans-Golgi network

TM, Trans-membrane

TNF, Tumor necrosis factor

TRIC, temperature-related intensity change

Trk, Tyrosine Kinase receptor

UBA, ubiquitin-associated

UBX, ubiquitin-regulatory X.

UPS, ubiquitin-proteasome system

VCP, valosin-containing protein

Vps10p, Vacuolar protein sorting 10 protein domain

CHAPTER 1

SINGLE PARTICLE CRYO-EM AND ITS APPLICATIONS IN STRUCTURAL BIOLOGY

1.1 Introduction

It is central to understand biomolecular functions by gaining deep insights into their three-dimensional (3D) structures[1]. The commonly employed tools to study protein structures is an integrated approach of light microscopy, scanning electron microscopy, transmission electron microscopy (TEM), X-ray crystallography, nuclear magnetic resonance (NMR) and X-ray free electron lasers (X-FEL)[2]. In X-ray crystallography, X-rays are used to generate the diffraction patterns of proteins crystals[3]. The major difficulties are that some crucial protein complexes are too flexible to crystallize[4]. Particularly, membrane proteins are not stable in an aqueous solution and usually have a lower level of heterologous overexpression, which make the crystallization and further structural study not trivial. Because membrane proteins are solubilized in detergents, the crystal structure may not reveal the structure in their native conditions. NMR spectroscopy provides an alternative approach to probe membrane protein structures in solution, however, it becomes cumbersome for solving macromolecular structures since it uses radio magnetic waves[5] to probe the distances between nearby atoms, producing overlapping spectra in case of larger proteins.

Transmission electron microscopy (TEM) is versatile to visualize smaller membrane proteins or larger structures, like cellular organelles. The electron imaging of

the proteins does not require a large amount of purified protein samples. Because the particles may exhibit multiple states in an EM grid, the TEM imaging can capture the images of multiple states of the target proteins. Cryogenic electron microscopy (Cryo-EM) uses electron microscopy to study a cryogenic sample, mostly radiation-sensitive material as shown in Figure 1-1 (A). The frozen sample is usually embedded in a thin layer of vitreous ice after rapidly freezing at cryogenic temperatures[6-8]. Dr. Richard Henderson, Dr. Jacques Dubochet and Dr. Joachim Frank were awarded the 2017 joint Nobel prize for "developing cryo-EM for the high-resolution structure determination of biomolecules in solution". Negative-stain electron microscopy (EM) helps to primarily assess the protein quality as shown in Figure 1-1(B) and discussed in section 1.2.1. In micro-crystal electron diffraction (Micro-ED), a parallel beam of electrons is used to collect diffractions patterns of 3D crystals in various tilt angles, as shown for a specimen in Figure 1-1(C). Cryo-electron tomography (cryo-ET) is a technique where images are collected by tilting the specimen at different angles as shown in Figure 1-1 (D). The application of various techniques used depends on the desired resolution range as discussed in Figure 1-1 (E). Cryo-EM gained immense importance over the past decade for solving biological structures of viruses, understanding of protein-ligand interactions and structure-based drug design in pharmaceuticals. It allows minimum radiation damage to the protein samples and typically produces low contrast images, and once the density maps can be reconstructed, additional software help to build the atomic model of the target protein samples[9-11].

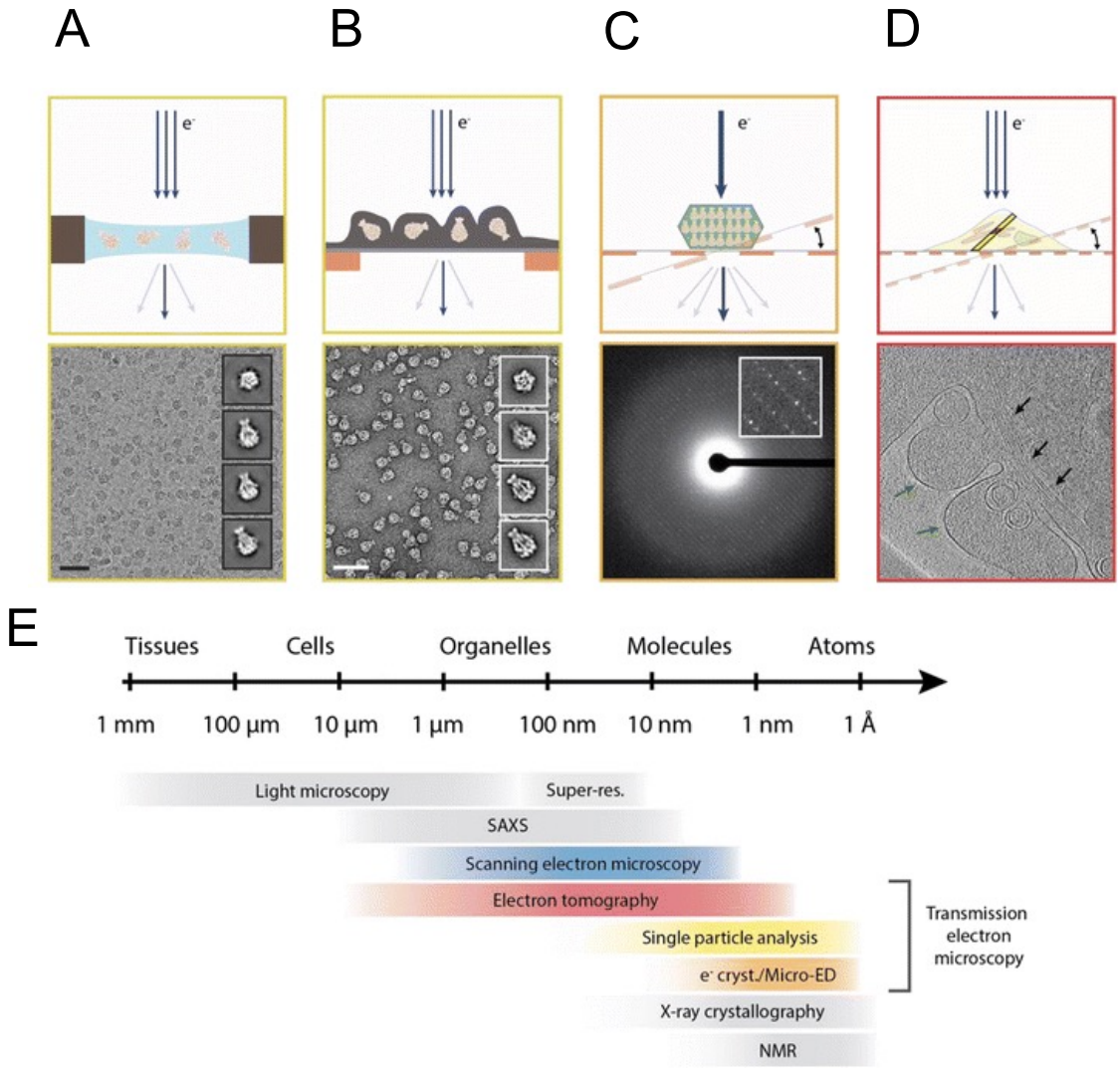


Figure 1-1. The most common techniques in structural biology[12].
Upper panel A-D describes the principles of the common transmission electron microscopy (TEM) techniques, and lower panel is the application of integrative techniques to study biomolecular structures (A) Single-particle cryo-EM: proteins are embedded in a thin layer of vitreous ice and the representative 2D class averages are shown as insets. Scale bar indicates 50 nm. (B) Single-particle negative-stain EM: particles are embedded in a layer of heavy metal salts to improve the contrast of target molecules. Resulting representative class averages are shown as insets on the right. Scale bar indicates 50 nm. (C) Micro-ED: A parallel beam of electrons generate diffraction pattern of a small 3D crystals. Inset shows a small section of the diffraction pattern[13]. (D) Cryo-ET: The images are collected at different angles by tilting the specimen at different angles. Inset shows a tomographic slice showing the cellular periphery with microtubule bundles (black arrows) and plasma membrane (green arrows). (E) Use of various structural biology methods for different resolution range, and the TEM based

methods (A)–(D)are color coded, Yellow: single-particle analysis; orange: electron crystallography/Micro-ED; red: electron tomography

1.2 Specimen Preparation for Cryo-EM

1.2.1 Negative Staining

Before high-resolution cryo-EM study, the quality of the sample can be assessed using negative-stained EM. Negative-stain EM is a straightforward technique of bright-field imaging, which involves the use of a small quantity of the sample and is primarily used for screening the particle composition to evaluate the various conformations present in the microscopic scale as shown in Figure 1-1(B). With a relatively small dataset of hundred images, it is a relatively quick technique to assess how the 2D class averages present the various views of the protein. Commonly used stains are prepared with heavy metals, like uranium, tungsten, molybdenum, and lead. The advantage of using uranium salt stains is that it provides high image contrast and less grainy background and are versatile to bind to proteins, lipids and nucleotides. The procedure of the negative can result in sample flattening by distorting the macromolecular specimen[14] and also limits resolution in the range of 15-20 Å[15]. Sometimes, the acidic nature of the stain can destabilize proteins. If the stain poorly interacts with the sample, it leads to under staining or overstaining, preventing proper electron scattering and resulting in blob-like features lacking identity.

1.2.2 Chemical Crosslinking

Successful structure determination by single-particle cryo-EM largely depends on the sample homogeneity, since it relies on averaging the identical views. Although the

computational algorithms can classify the different protein populations in different 3D volumes, it is crucial to improve the protein biochemistry techniques to achieve highly pure sample. One of the common methods that can help to stabilize the potentially transient protein complexes and prevent dissociation is the addition of a chemical crosslinker like glutaraldehyde[16] or BS3, which I have used to stabilize the p97-p47 complexes in Chapter 3. To perform controlled crosslinking with high reproducibility, GraFix (gradient fixation)[17] is another technique which exposes macromolecules to a density gradient of varying crosslinker concentrations. It is important to use mild crosslinking conditions (0.01%(w/v)) to avoid artifacts and non-specific intermolecular interactions.

1.3 Cryo-EM Grid Preparation by Optimizing Common Challenges

After screening the sample homogeneity of purified protein samples by negative-stain EM, we prepare the frozen sample for cryo-EM imaging. The vitrification is well established and involves application of a small volume of aqueous sample to the holey-carbon EM grid in the protein concentration range from 50 nM to 5 μM, followed by blotting away excess solution, and then rapidly plunge freezing in liquid ethane to achieve vitreous ice. Rapid plunge-freezing in liquid ethane prevents formation of crystalline ice and the protein molecules are kept hydrated in their native-like conditions. The target ice thickness should be thin enough to avoid dynamical electron scattering; however, it is difficult to achieve an appropriate ice thickness. Additionally, an ideal situation should result in homogeneous distribution of the protein samples giving rise to

all possible orientations. There are many factors which play an important role as discussed below.

An optimal cryo-EM grid has a high concentration of protein particles spread evenly over the grid, distributed in various orientations. However, proteins encounter many challenges like aggregation and denaturation, particles preferentially stick to the carbon layer failing to enter the holes, and protein complexes may disassemble during the vitrification process. Depending on the protein, particles can adopt a preferred orientation when they are exposed to the air-water interface, thereby restricting the unique views. To overcome these obstacles commonly encountered, research involves thorough optimization of cryo-grids depending on the target protein samples. Proteins samples are often biased to adopt a particular orientation and limiting its unique views resulting in poor 3D reconstruction (Figure 1-2). The air-water interface plays a major role. The well-established practice of plunge freeing involves application of a few microliters of the aqueous sample to the grid, followed by blotting and then rapidly plunging. Theoretically, particles will diffuse ~10-1000 times to the air-water interface within the timeframe of blotting and plunging and get denatured or result in diffusion mediated adsorption. There is a possibility of formation of a new phase at the air-water interface due to adherence and this process nucleates further protein adhering and denaturation. Sometimes the effect is seen only in specific stretches of sample (like exposed β-sheets). The major drawbacks are anisotropy or missing information in a direction-dependent manner and decrease of the global resolution.

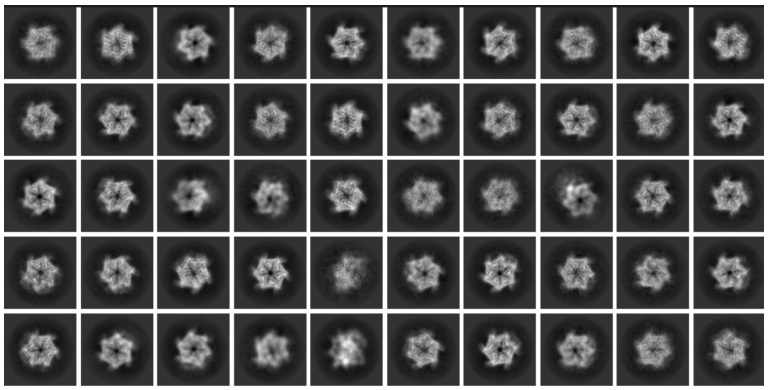


Figure 1-2. p97, an ATPase suffers from strong preferred orientation from the air-water interface. In one p97 dataset collected by me showing the representative two-dimensional (2D) class averages are depicting top views and lacking information about side-views (unpublished data). These datasets attenuate global resolution and lead to poor three-dimensional (3D) reconstruction.

There are a few preventive treatments like introducing mild concentrations of detergents on the grids, addition of trehalose, using solid support for the holes, and using tilted stage for data collection as discussed below.

1.3.1 Addition of Detergents

Detergents have been widely studied to prevent orientation bias of samples by reducing surface tension of the solution, thereby affecting how the proteins interact with the air-water interface. Often application of mild detergent concentrations needs to be accompanied by higher protein concentration and requires optimization of the ice thickness. Grids need to be screened to avoid effects of detergents interacting with the proteins and affecting the stability. Some of the example where detergent doping enabled better protein structures are bacterial RNA polymerase at 3.8 Å using 0.5% CHAPSO (EMD-7002)[18], CRISPR-Cas13d at 3.3 Å using 0.1% of amphipol A8-35 (EMD-9014)[19], PRC2 complex at 3.5 Å using 0.01% NP-40 (EMD-7335)[20] and TRiC-PFD complex at 6.5 Å using 0.1% Octyl glucoside (EMD-0496)[21].

1.3.2 Addition of Trehalose

Sugar molecules, like glucose or trehalose, can protect biomolecules from the air-water interface by using the hydroxyl moieties as a shield and hence can be effective to prevent orientation bias in cryo-EM datasets[22]. The starting concentration for sugar embedding is dependent on the protein samples, ranging between 1-7% (w/w). Trehalose has been tested in single-particle cryo-EM samples during vitrification, and provides additional protection from beam damage and prevents sample denaturation by dehydration[23]. There are various theoretical hypothesis suggesting the protective mechanism of biomolecules by trehalose, a disaccharide. Trehalose is proposed to create a viscous glass-like layer protecting the proteins in a native-like environment, using the advantage of highest glass transition temperature among sugars[24]. Trehalose, being a free radical scavenger, can prevent oxidative stress and in the "water replacement" theory the hydroxyl ions in trehalose interacts with the hydrophilic head groups of proteins and lipids to shield them[25,26]. The "preferential exclusion" theory suggests that trehalose drives water molecules away reducing the hydration diameter and thus stabilizes the proteins[27,28]. The chemical structure of trehalose makes it superior since it has the hydration shell radius at least 2.5 times higher than other sugars. I have tested the benefits of adding 1% trehalose to achieve higher proportion of side-views for the ATPase p97 vitrification. Since addition of sugars like trehalose will raise the buffer density surrounding the protein samples, the image contrast in the cryo-EM datasets gets reduced, with the exception of electron-dense samples like ferritin[29].

1.3.3 Solid Support Layer on Grids

Introduction of a thin solid support layer, typically a thin layer of amorphous carbon in the thickness range of 3-5 nm is deposited on the holey carbon grids. This results in protein particles adhering to the carbon layer and prevents damage from the air-water interface. The most common technique is to coat an ultrathin layer of graphene[30] applied as a monolayer on the carbon, making the grid sturdy, and contributing negligible noise to cryo-EM images. This work further branched out to other alternatives like functionalized graphene[31], graphene oxide[32] and functionalized graphene oxide[33]. Additional grid treatments can produce "affinity grids" by coating with specific antibodies, layering with streptavidin crystals or attaching functional moieties and are particularly helpful to increase particle concentration on grids, when working on low yield proteins. Some examples where structure determination benefitted from solid supports are transcription factor IIH at 3.7 Å using ultrathin carbon coating (EMD-0452)[34], Streptavidin at 2.6 Å using graphene coating (EMD-0689)[35], Tulane virus at 2.6 Å using antibody coated affinity grids (EMD-8252)[36].

1.3.4 Avoiding Particle Partitioning

For certain proteins, the particles have a strong adherence to the carbon or align around the edges of the hole, thereby reducing the particle concentration inside the holes, and the center of the hole with the thin layer of ice lacks any particles. Grids can be treated with solid support, or the glow-discharge time can be varied to adjust the hydrophobicity, as well as chemically treated grids and multiple rounds of sample application to the grids[37] help to mitigate these problems.

From this discussion, it becomes obvious that it requires iterative trial-and-error experiments to perfect protein biochemistry and to achieve the ideal ice thickness and particle distribution on cryo-EM grids.

1.4 Single-particle Cryo-EM Data Collection

1.4.1 Anatomy of a Transmission Electron Microscope

The ability to form atomic structures to understand the structure and morphology of samples make transmission electron microscopy (TEM) a unique tool. This method was first discovered in 1931 by German scientists Max Knoll and Ernst Ruska, and later gained vast popularity over light microscope because electrons have much smaller de Broglie wavelength than photons, enabling higher resolution determination. The anatomy of the transmission electron microscope is discussed in Figure 1-3A, it has an electron source as either (i) a heated tungsten filament (ii) LaB_6 crystal or (iii)a field emission gun (FEG). The FEG releases coherent, near monochromatic and high current density electron beams that get modulated by electromagnetic fields in their trajectory under 100-200 kV, as they travel under vacuum. The high vacuum ensures the electrons don't lose energy by colliding among themselves or gas molecules in their path. These fast-moving electrons interact with the sample and can either penetrate the biological specimen or get scattered. The electron microscope is comprised of condenser, objective, intermediate and projector lenses. The electrons are focused to a narrow parallel beam by condenser lenses, with large apertures eliminating the effect of electrons at higher deflection angles. The transmitted electrons after passing through the sample are focused on the image plane by the objective lens. The combination of intermediate and projector lenses

modulates the image magnification before electrons are recorded at the detector. The micrographs are recorded after projection on the fluorescent screen using a charge coupled device (CCD) camera. This technology is suitable to image various samples ranging from proteins, viruses, bacteria, or geological and materials science and forensic samples.

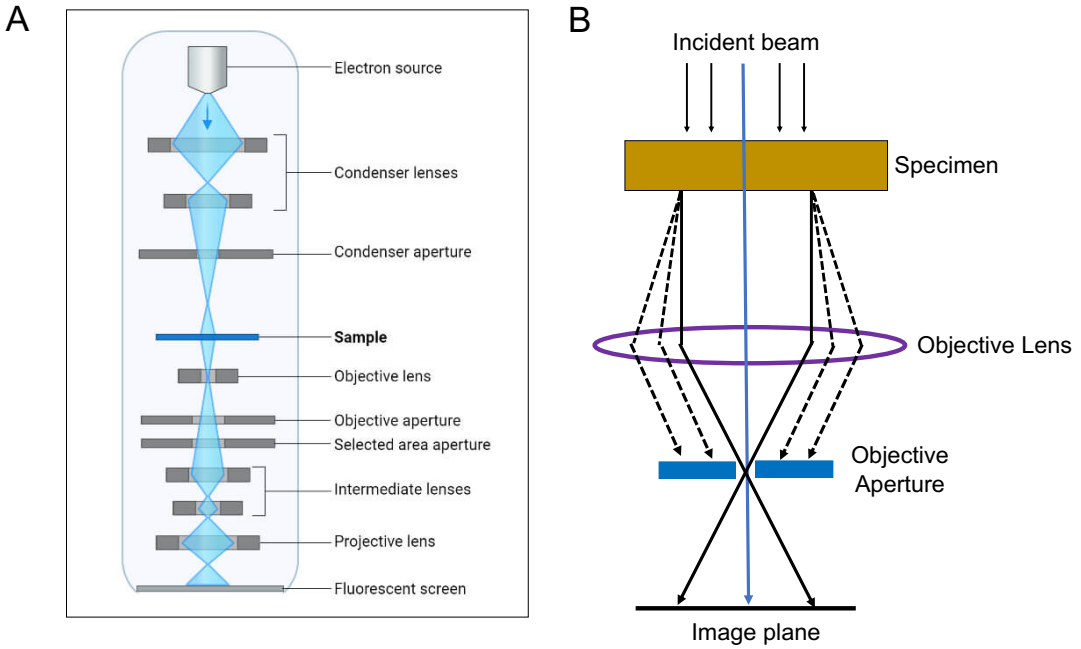


Figure 1-3. Transmission electron microscope. (A) Anatomy of a transmission electron microscope maintained in high vacuum consisting of the (source) electron gun, electromagnetic coils and lenses, condenser and objective aperture, specimen holder and the fluorescent screen where the CCD camera converts the electron signal to images (B) Mechanism of image formation in bright field imaging mode of TEM. Dashed black lines indicate scattered electrons.

TEM can develop defects of the electromagnetic lenses, resulting in spherical and chromatic aberration and axial astigmatism. Spherical aberrations result from electrons refracted from the edge of the lens that cannot be focused on a common focal plane and hence blurs the images. Chromatic aberrations arise from wavelength dependent problems that arise when electrons travel with variable energy. Modern microscopes are

equipped with stable accelerating voltages and controlled lens current to evade the influence of chromatic aberrations. Axial astigmatism occurs when the electrons experience non-uniform magnetic field, and thus can be stabilized by using stigmatism correctors below the objective lens[38].

1.4.2 Image Formation in the Transmission Electron Microscope

The incident beam of accelerated electrons can interact with the sample in the following ways: (i) electrons that don't interact but pass through the sample, which form the unscattered beam, (ii) electrons that can pass through the sample with its energy unaltered, resulting in the elastically scattered electrons (iii) electrons that scatter inelastically and transfer some energy to the atoms in the path as shown in Figure 1-3B. Only the inelastically scattered electrons contribute to image formation due to appropriate phase shift. The interference of the elastically scattered and unscattered electrons causes contrast in the image while the inelastically scattered electrons contribute to the noise. When the electrons interact with the atoms in the biological specimen, their trajectory is deflected. The contrast can be improved by adjusting the objective lens to exclude the highly deflected electrons. However, they contain high resolution details and hence the images close to the contrast are important.

The biological specimens are considered "weak phase objects" since they are usually made up of light atoms such as hydrogen, carbon, oxygen and nitrogen. This results in weak scattering of electrons with minimal phase shift and negligible amplitude shift. In TEM, the potential model is followed, where the phase of the electron wave is

modulated by the atomic potentials and can be defined by a phase object approximation as

$$q(r) = e^{-i\sigma V_t(r)}$$

σ is the interaction constant, V_t is the projected potential of the sample and r is a 2D vector. For thin specimens, σV_t is close to 1, and thus the equation simplifies as

$$q(r) \approx 1 + i\sigma V_t(r)$$

Scattered electrons from weak-phase objects experience phase shift, but amplitude stays unaltered. Since electrons passing through various parts of the thin specimen cannot create an amplitude difference or contrast, the defocus is modulated to adjust the phase contrast in images. Thus, a systematic alteration of the image data occurs by applying defocus, and this is known as contrast transfer function (CTF). Image formation is therefore described as an action of the contrast transfer function (CTF) H(k). In Fourier space it is expressed as

$$I(k) = O(k)H\,(k)$$

where I (k) represent the Fourier transform of the image formed, O(k) is the Fourier transform of the object and H(k) is the CTF dependent on spatial frequency k. The term k in the above equation is a direct conversion of the 2D vector x and y, k= (*kx, ky*).

The transmitted electrons through the sample are focused on the image plane, where the detector can recognize the resultant amplitude after constructive and

destructive interference. The contrast at the image plane is modulated from the phase difference between the scattered and unscattered electrons. The defocus setting can enhance or repress the features of the object. This is because of the oscillation of the CTF between -1 (negative contrast transfer) and +1 (positive contrast transfer) as the spatial frequency changes from low to high. This can be visualized in the Fourier space to be inverted, with the value of zero contrast transfer leading to a loss of data, the value with negative contrast makes black pixels white and vice versa for positive contrast transfer. In 2D, this phenomenon of modulating the contrast transfer in a defocus-dependent way is visualized as Thon rings[38], which are concentric circles. The shape of the rings is dependent on electron-optical parameters like astigmatism and defocus. The mathematical equation to calculate the contrast transfer function in the Fourier space depends on equation

$$CTF\ (k) = A * \cos\Phi(k) - \sqrt{1 - A^2}\ * \sin(\Phi(k))$$

Here A is the amplitude contrast fraction and $\Phi(k)$ is given by following equation:

$$\Phi(k) = -z\lambda k^2\ /2 + C_s\ \lambda^3 k^4/4$$

Here, k is the spatial frequency, λ is the relativistic wavelength of the electron wave, z is the defocus of the objective lens (applying the conventional understanding that positive relates to under focus and negative is overfocus), and C_s is the spherical aberration of the objective lens. Thus, the defocus of the objective lens decides the CTF. At higher values of defocus, the contrast at low frequencies is high as the number of zero crossing increases, while low values of defocus decrease the contrast at low frequencies. This

makes the images at low defocus impart important high-resolution information, and using high defocus is useful to understand the overall shape and features of the biomolecules.

1.4.3 Electron Recording and Image Acquisition

Detectors play a key role in the improved data resolution. In the early days, electron microscopes had photographic films, while now they are replaced with digital devices like charge-coupled devices (CCD) and direct detector devices (DDD), which decreased the time lag between exposure and observation of image. The measurable quantities to determine performance of the detector are the detective quantum efficiency (DQE) and modulation transfer function (MTF). The DQE is a ratio of SNR of the output image to input image to measure the degradation by noise[39]. The MTF is the ratio of image contrast by object contrast to measure if spatial frequencies outside the Nyquist are recorded. The CCDs have semiconductors which can record an electric charge proportional to the intensity of electromagnetic radiation and can convert them to a digital signal. Often CCDs are coupled to scintillators in TEM. Thus, the incoming signal has a two-fold conversion from electron to photon and photon to electron, with decrease in spatial resolution. However, CCDs have good DQE and have rapid read out times. DDD's avoid the second conversion step in CCDs and are made of complementary metal-oxide-semiconductors (CMOS). The advantage of CMOS is it gets ionized when the beam is incident, and the resultant electron (or hole) is accelerated, leading to very high DQE and SNR in the produced images. The thin layer of CMOS prevents spatial charge spread, and thus high-resolution details are better retained[40,41]. DDD's also have

an exceptional framerate and thus can be useful for beam-induced motion correction and image corrections.

1.5 Single-particle Analysis for cryo-EM

Cryo-EM involves collection of projection images also known as micrographs, that represent the electron density recorded in 2D. The images can be reconstituted to obtain desired information like single particle reconstruction to obtain various views of the proteins, look at the cellular components using electron tomography or obtain regularly arranged structures as seen in 2D crystallography. The characteristics of cryo-EM images have low signal-to-noise ratio (SNR), leading to poor contrast since proteins are primarily composed of light atoms like carbon, hydrogen, oxygen, and nitrogen. This is also due to the low dose imaging conditions, to avoid any beam induced radiation damage of sensitive biological specimens. Thus, the data processing aims to improve the SNR, implement CTF corrections and finally reconstruct the three-dimensional volume using the information from two-dimensional projection images.

Single particle reconstruction performs computational averaging of the various 2D projection views from the images collected for the protein samples. It relies on particles oriented in all possible views with sufficient representation of each conformation. If the target object is symmetric, it benefits from lesser number of particle images. Generally, a large dataset is collected to compensate for the poor SNR ratio and low-dose imaging in cryo-EM[42]. The workflow involves specimen preparation, low dose image collection and model building as described in Figure 1-4.

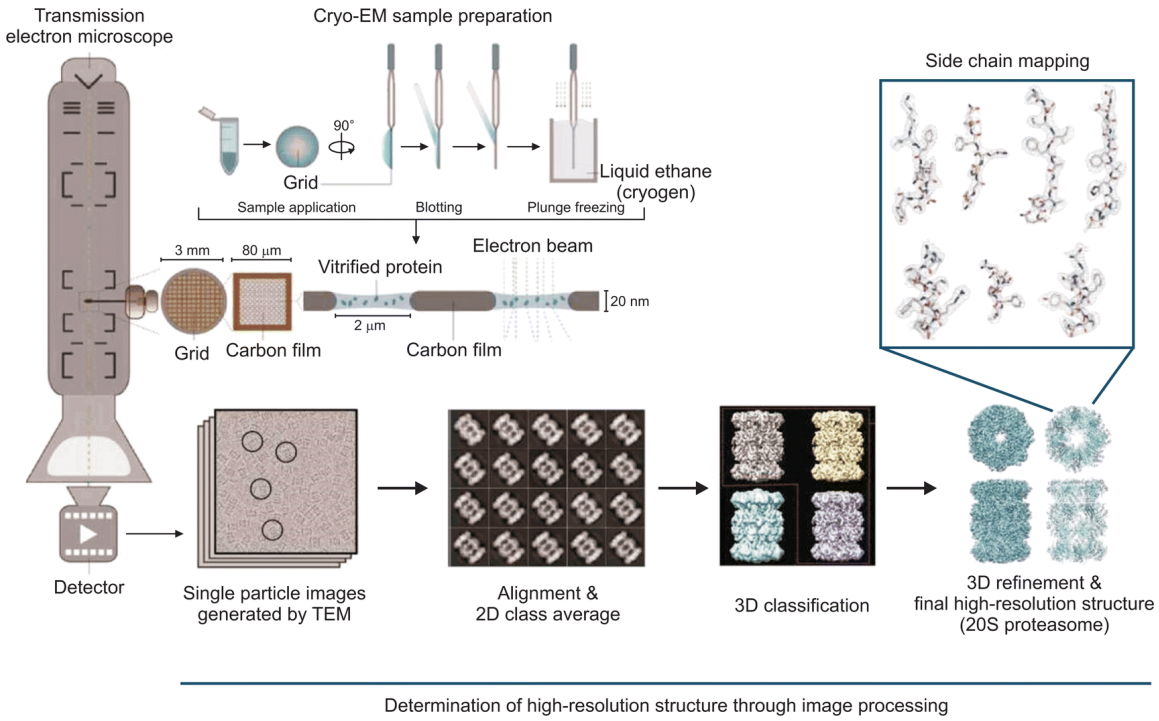


Figure 1-4. Workflow for the single-particle analysis of cryo-EM[43]. A highly pure protein sample is applied to the cryo-EM grid and rapidly plunge frozen in liquid ethane to form a vitreous layer of ice. The protein particles are held in the thin layer of ice distributed in all random orientations, and the grid is imaged using cryogenic conditions in a cryo-TEM. The electron beam is radiated on the grid, and two-dimensional (2D) projections also known as micrographs are collected. After motion correction and CTF estimation, the computational software picks the individual protein particles, which are then aligned to calculate two-dimensional (2D) class averages. Further three-dimensional (3D) refinement and reconstruction helps to calculate the high resolution final cryo-EM structure.

1.5.1 Movie Alignment

The direct detector devices (DDD) with high-speed image acquisition enable collection of multi-frame movies. However, they have low exposure per frame (0.8 to 3 electron/ Å /frame) resulting in poor contrast and low SNR in individual frames. These movies further reveal the electron "beam-induced" sample motion, which blurs the images due to small independent motion on every ice-embedded protein particles[44]. To achieve optimal

high-resolution structures, all the frames are precisely aligned to eliminate the effect of these relative movements and increase the SNR.

1.5.2 CTF Estimation and Correction

The CTF estimates the effect of microscope aberrations on image quality and hence needs to be accurately determined and corrected. The parameters required are the acceleration voltage, spherical aberration constant, defocus during data collection, axial astigmatism, magnification and amplitude contrast. There are several software which perform accurate CTF estimation as EMAN[45], CTFFIND3[46], CTFFIND4[47] and GCTF[48], and only the CTF-corrected micrographs are carried to the next steps.

1.5.3 Particle Picking and Extraction

After motion-correction and CTF-correction, the micrographs are curated where micrographs with poor defocus, protein aggregation, crystalline ice contamination or improper ice thickness are discarded. The next steps involve identifying and extracting the particles representative of the biological samples from the images. A semi-automated method using a combination of manual and automated particle picking results in the most efficiently picked particle coordinates. The commonly used particle picking software packages available are EMAN[49], DoG Picker[50], RELION[51] and cryoSPARC[52,53]. The particles can be picked with or without a reference. In the reference-based approach, a 2D template is prepared from the particles in the dataset, and software trains to pick more similar particles[50]. In the reference-free approach, the shape and features of the particles serve as parameters to identify particles.

In the cryo-EM studies of proteins like the neuronal apoptotic receptor (Chapter 2) and p97 (Chapter 3 and 4), I have used a combination of Manual picker, Template picker and Blob picker in cryoSPARC and RELION. The manual picker allows the flexibility to select particles in different orientations from the dataset, to create templates after running a 2D classification which can be further used for template picking. The template picking is automated and efficient. Another method is to use a gaussian blob of a definite radius determined based on the particle, useful for spherical samples. After particle picking, they are extracted from the original micrographs to generate a particle stack.

1.5.4 Two-dimensional (2D) Classification

The semi-automated particle picking algorithms often exhaustively over-pick particles, regardless of the type of picker, and picked particles should be inspected before moving to the successive steps. The 2D classification helps to eliminate the invalid "junk" particles from the extracted particle set to quantitatively evaluate the angular distribution of the particle views and roughly estimate the resolution before proceeding to the three-dimensional (3D) classification. The 2D classification scores the differences in the X and Y translation and in-plane rotations of the particle images. This is followed by averaging, leading to increased effective SNR and reveals enhanced particle features, by reinforcing consistent features and averaging noise out. The 2D classification is often based on a simple unsupervised-machine learning method called *k*-clustering, where *k* refers to the number of centroids in the dataset, which is the representative center of the cluster[54]. The cluster means a group of data points aggregated together based on similar features. The

k-means helps to collect similar data-points and identifies a fixed number (*k*) of clusters in a dataset. This clustering algorithm has three steps. The first step starts with initialization which means the number of groups *k* is assigned and the *k* centroids are decided randomly in the dataset. The second step assigns particles to the centroids, based on the closest resemblance to a centroid[55]. The final step updates the centroid, using the modified centroid location based on the average coordinates of the objects in the said centroid. This method of *k* clustering is preferred for its simplicity and speed and relies on appropriate initial assignment of the *k* value as it influences the final results. One drawback of this method involves the convergence leading to a local minimum, which reduces the scope of global classification. This can be reduced using the Maximum likelihood[56-58] method, which allows particle assignment to more than one class using a weighting factor. Multiple rounds of 2D classification help to select the best particles for downstream processes.

1.5.5 Three-dimensional (3D) Classification

The three-dimensional volume can be reconstructed from 2D classifications using Fourier based approaches or the back-projection, which are based on the underlying "projection matching theory". This is used because the relative orientation of the particle images are unknown and becomes the rate-limiting step to reconstruct the 3D volumes. Thus, a reference low-resolution 3D model is generated, also known as the *ab-initio* model. The cryo-EM images are compared with this to determine relative orientation and then back projected to get the 3D reconstruction. The projection theorem postulates that a central section of 3D Fourier transform of the object is provided from each projection of the

object. To reconstruct, we can obtain the 3D inverse Fourier transform of the 3D Fourier domain, where the 2D Fourier sections can supply the relevant information. For an n-dimensional object, each projection is an n-1 dimensional sum of its density along the projection axis. Thus, the resulting projections from a sphere are circles and from a cube are squares. These 2D projections are inverted in back projection, which allows to reconstruct the final 3D original object. The process often results in "smearing" as the 2D images are converted to 3D volumes, yielding a blurry object due to lowering of sampling points in the Fourier space as the spatial frequency increases. Applying a low-pass filter before back projection or having reduced weighing on low frequency components can help to overcome any issues. Typically, the 3D reconstruction is an iterative process, where the current map serves to be the reference map for the next cycle, thus improving accuracy of orientations. Euler angles provide the orientation of the objects in 3D and generate the "initial model", and then are iteratively corrected, until the Euler angles converge resulting in the final 3D reconstructed map[59]. Additionally, the particles in the 2D class averages can be initially sorted in 3D volumes to classify compositional and resolve trace conformational states in a mixed population.

Like 2D classification, 3D classification also relies on the maximum likelihood approach, where several in-plane values can be assigned to each image with weights, integrating the probability over many possible values. This is slightly different to the Bayesian approach[51,60], in terms of the regularization parameter which decides the weight from prior 3D reconstruction knowledge:

$$P(\Theta|X,Y) \propto P(X|\Theta,Y)P(\Theta|Y),$$

here $P(X|\Theta, Y)$ is the probability of data with the parameter Θ. $P(\Theta|Y)$ tells the likeliness of parameter Θ from prior knowledge. X represents the particles set = $\{X_1, X_2, X_3, \ldots X_N\}$, with total number of particles being N, an $\Theta = \{V_1, V_2, \ldots V_k\}$, is a set of volumes with given number of classes as k.

This Fourier based approach is similarly used in X-ray crystallography. The map of the object in real space is generated from the inverse Fourier transform, and interpolation from the central section aide the inversion methods for filling required information. Various software packages have built in programs for single-particle cryo-EM 3D reconstruction like SPIDER[61], EMAN2[49], RELION[51,62] and cryoSPARC[52,53], with cryoSPARC being most widely used in our cryo-EM data processing studies discussed in Chapters 2-4.

1.5.6 Structure Refinement and Resolution Evaluation

The reconstructed 3D volumes can be iteratively refined to obtain a final map with improved details. It involves estimating the view angle that best matches the reconstructed 3D object from 2D images and uses this angle to further improve the quality of the 3D volume. However, it is important to avoid any bias from overfitting, which creates false information from the noisy pixels. The commonly used software like RELION and cryoSPARC are equipped to use the Bayesian approach[51,63] to avoid overfitting. The resolution of the final 3D map is crucial to evaluate the results, which is the numerical value suggesting the resolvability of the protein molecules reconstructed in 3D. Commonly researchers follow the standard “gold standard” Fourier shell correlation

(FSC)=0.143 cutoff to report the nominal resolution[64], and I have used the same in our cryo-EM studies as discussed in Chapters 2-4. It compares the dataset by splitting it into two independent halves based on odd or even numbered images and then the half-datasets are reconstructed and refined independent of each other, and the generated half-maps are cross correlated as a function of spatial frequency. FSC is a function of spatial frequency as

$$FSC\ (k) = \frac{\sum_{k\Delta k} F_1(k)F_2(k)^*}{\sqrt{\sum_{k,\Delta k} |F_1\ (k)|^2 + \sum_{k,\Delta k} |F_2\ (k)|^2}}$$

Where $F_1(k)$ and $F_2(k)$ are 3D Fourier transforms made from two independent subsets of data. $F_2(k)^*$ is the complex conjugate of F2(k). The $\sum_{k,\Delta k} F\ (k)$ represent the summation over the small range Δk around k. FSC (k) can have allowed values between -1 and +1 for every spatial frequency k. FSC (k) values close to +1 indicates higher reliability of the reconstructed 3D structure.

Moreover, it is crucial to visually inspect the 3D reconstructed volumes to evaluate structural details. If the resolution is beyond 8 Å, the overall features of the protein subunits cannot be discerned from the volume. With a resolution of ≤4 Å, the secondary structures like the α-helices and β-sheets of the protein backbone can be clearly assigned or identified, suggesting information about the subunits and domain organization. Near 2 Å near-atomic resolution, the protein sidechains can be reliably modelled.

1.6 Cryo-EM to Solve Disease Relevant Protein-complexes: Scope of this Thesis

The aim of this doctoral thesis is to use this robust and impressive technique of cryo-EM to structurally capture the different conformational states of protein-complexes in their active states to investigate disease pathophysiology. The studies involved looking at the unfavorable neuronal apoptotic complex proNGF-p75^{NTR}-sortilin, which lacks structural characterization to date. In another project, we investigated the effects of disease-linked mutation on a cellular ATPase p97 and performed structural studies in presence of its cofactor p47.

1.6.1 Neurotrophin Signaling and Diseases

One of the prominent causes of neurodegenerative diseases is the dysregulation of downstream signaling of the neurotrophin receptor p75^{NTR} in complexation with various neurotrophin ligands like Nerve Growth factor (NGF), Brain Derived neurotrophic factor (BDNF) and/or its receptor Trks (Tyrosine Kinase receptors) (Figure 1-5). All the possible interactions and binding partners of the neurotrophins and their receptors, along with the neuronal outcomes are discussed in Figure 1-5. In healthy neurons, p75^{NTR} upon binding to ligands like mature NGF and TrkA assists the proliferation of neuronal cells leading to cell survival[65-67]. On the contrary, in aging cells or diseased conditions, the level of the precursor of NGF (proNGF) is dramatically high, and its interaction with p75^{NTR} aided by another transmembrane protein, sortilin, has neurotoxic effects, leading to neuronal atrophy or fatal cell apoptosis. To elucidate the governing principles of this unfavorable signaling, we seek to identify the molecular mechanisms underlying the preferential affinity of proNGF over NGF for p75^{NTR}, activating the apoptotic pathway in presence of co-receptor sortilin.

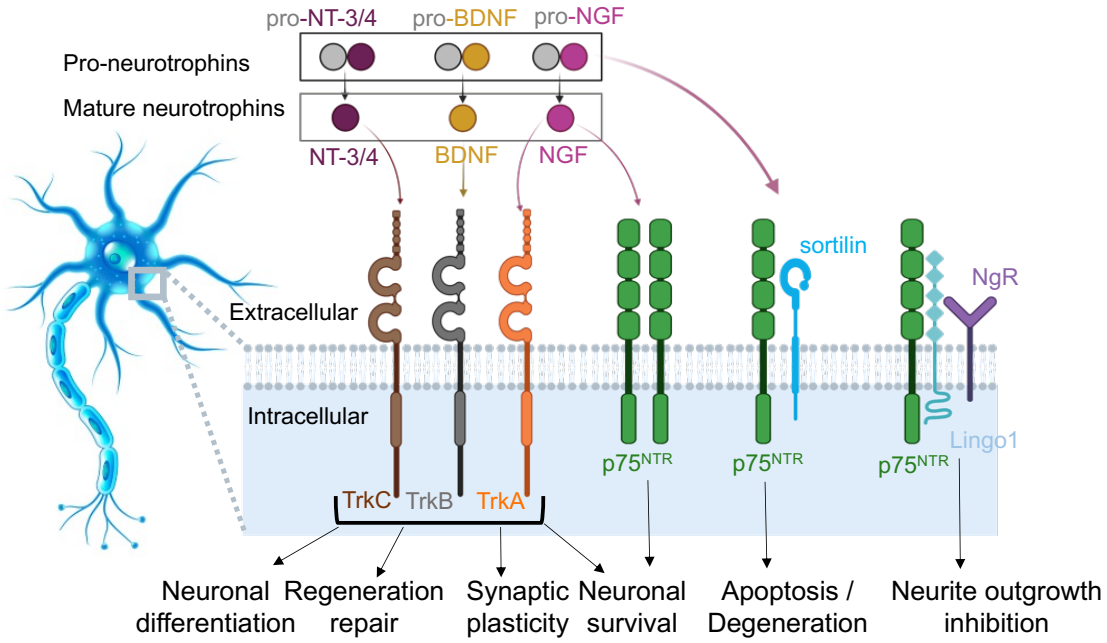


Figure 1-5. Signaling in neurons is dependent on the interaction between the ligands and neuro-receptors[65-69].

In Chapter 2, the protein constructs of the neurotrophin receptor $p75^{NTR}$ was prepared, which propagates the neuronal death pathway by preferential binding to the precursor of nerve growth factor proNGF, along with the co-receptor sortilin. Initially each of these recombinant proteins were purified individually by using the *Sf9* insect cell system for overexpression, followed by analysis of sample purity using size-exclusion chromatography (SEC), Western immunoblotting, and negative-staining EM for single-particle analysis. Since the extracellular domain of sortilin is capable of binding to ligands using the ten-bladed β-propeller motif at the C-terminus, I have further explored the possible models of the interaction of the proNGF-$p75^{NTR}$-sortilin ternary complex by co-expressing the proteins involving the extracellular domains. I have discussed the purification results of each of the three proteins followed by the co-transfection results

which generated the transient ternary complex of proNGF-p75^{NTR}-sortilin. I have collected a small cryo-EM dataset to assess the particle populations, which need additional modifications to achieve homogeneous assembly for cryo-EM data analysis. Our results indicate dimers of the sortilin facilitate the interaction with the binding partners of p75^{NTR} and proNGF at physiological pH. Due to the small, flexible, and highly disordered structure of the pro-domain of NGF, traditional X-Ray Crystallography or NMR remains inconclusive. Hence, single-particle cryo-EM is a promising tool to characterize the structural and mechanistic details of protein-protein interactions at near atomic resolution. Identification of the critical amino acid residues can provide useful insight into potential drug designing for related neurological disorders.

1.6.2 p97-p47 Mediated Golgi-membrane Fusion and Disease-linked Mutations in p97

The Golgi apparatus mediates protein and lipid transport and is a central cell organelle. The Golgi apparatus and endoplasmic reticulum experience substantial change by regulated disassembly and reassembly during mitosis where the fragmented vesicles and tubules gets reassembled in daughter cells[70,71]. These are heavily controlled by phosphorylation and ubiquitination of the proteins involved. Several proteins play an important role in the Golgi membrane dynamics, among which two well established ATPases known are the *N*-ethylmaleimide-sensitive fusion protein (NSF) and p97/valosin-containing protein (VCP)[72]. The NSF mediated pathways have been well characterized to understand its interaction to adapters α/γ-SNAP[73,74]. p97-mediated Golgi membrane fusion pathways are interesting because of their role in the UPS but are not yet structurally characterized completely[70,75]. There are two major adapters which guide the

p97 path, p47 and p37[76-78]. The p97-p47 pathway is ubiquitin dependent, initiated by the mono-ubiquitination of Syn-5 in the early mitotic steps of disassembly, thus recruiting the p47 to interact by its ubiquitin associated (UBA) domain (Figure 1-6). On the contrary, p37 lacks UBA, and does not require ubiquitin for functioning. Both the p47 and p37 mediated pathways require a co-adapter Valosin-containing protein p97–p47 complex-interacting protein, p135 (VCIP135), a de-ubiquitinase (DUB) [75,78,79], which removes the ubiquitin from Syn-5 and is implicated in the later steps of interphase and post-mitotic Golgi cisternae regrowth[78-80].

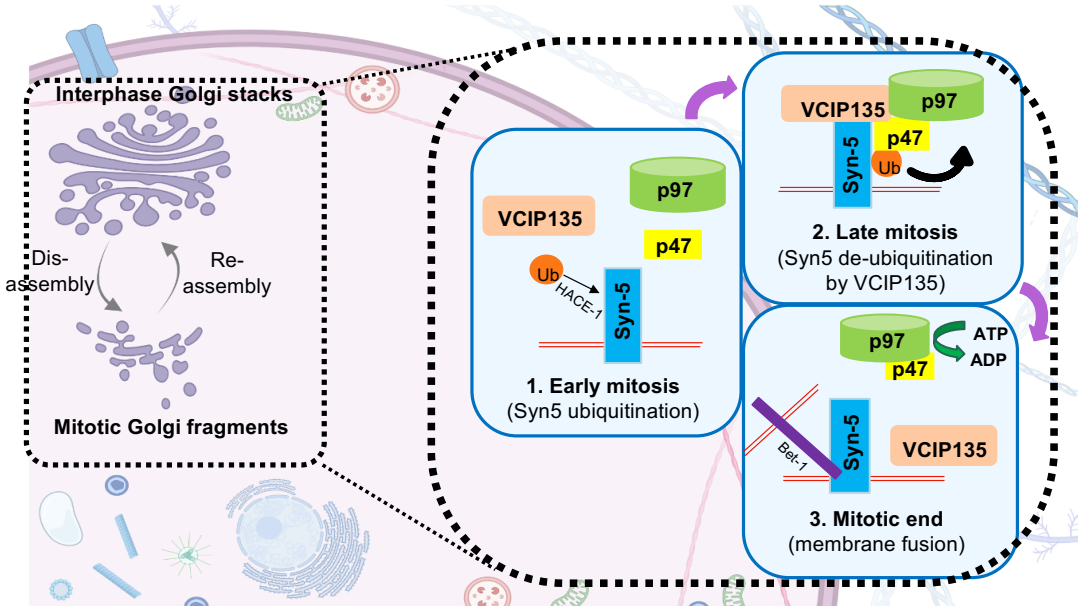


Figure 1-6. Proposed mechanism how p47 mediates p97 based Golgi membrane fusion in a ubiquitin (Ub) dependent cycle[77].

Several studies have indicated that single amino acid mutations in p97 results in rare neurodegenerative diseases like multisystem proteinopathy type I (MSP1), amyotrophic lateral sclerosis (ALS) and inclusion body myopathy with Paget's disease of the bone and frontotemporal dementia (IBMPFD)[81]. The disorder of MSP is genetically inherited and heterozygous, mostly caused by a single amino-acid mutation of the

p97/valosin-containing protein (VCP) gene[82-86]. The patients suffer from respiratory, skeletal, cardiac muscle atrophy and weakness, Paget's disease of bones and in more than 50% cases involves motor neuron damage leading to frontotemporal dementia. The disease may further aggravate to inclusion body myopathy with early-onset Paget's disease of bones and frontotemporal dementia (IBMPFD). The symptoms of Paget's disease of bones involves fragile bones and deformations due to higher activity of osteoclasts. It is rarely accompanied by defective cardiac, visual, auditory, or sensory systems. The pathways which lead to p97 dysfunction are very controversial. However, several MSP mutants are associated with defects in autophagy, endosomal trafficking, ERAD and aggregate removal pathways, lysosomal activity and NF-kB signalling[87-92]. The MSP mutations are clustered mostly in the N domain or the N-D1 linker region of p97 and are associated with elevated ATPase p97 activity under basal conditions in *in-vitro* experiments[93-97]. The functional significance is not clearly structurally characterized. The most probable explanation suggests that the domain communication between the N and D1 ATPase is dysregulated[86,98], along with unstable ADP binding[86,97,99-102], and does not result in an effective gain of p97 function. Most MSP mutants structurally display an N-up configuration, regardless of the nucleotide binding state, and hence it is speculated to increase access of adapters to p97-NTD, resulting in altered binding to various cofactors and adapters. The mixed population of MSP mutant-p97 and wild type p97 in the diseased cells disturb the basic functions of p97. Additionally cellular assays have indicated p97's affinity to bind adapters gets differentially affected in case of various MSP mutants[95,103]. Hence to probe the direct effects of the MSP mutations on p97 functions and adapter dynamics, it is essential to

explore the high-resolution structures of p97-adapter complexes to compare healthy and diseased states. Among almost 41 mutations identified[104], the $p97^{R155H}$ is the predominant MSP1 mutant and the p47 is one of the earliest studied co-factors of p97, where the partial crystal structure[105] of $p97^{ND1}$-p47 lays the background to further probe the molecular details of the full-length structures. This understanding can be further extended to study other MSP1 mutants.

In Chapter 3, I have worked on a collaboration project with Prof. Chou's group (Caltech) to perform the structural studies using cryo-EM and unravel mechanistic details of p47 binding to p97. The AAA+ ATPase p97 (Cdc48/VCP) is associated with the regulation of multiple cellular activities leading to protein homeostasis. p97 derives energy from ATP hydrolysis to bring about a conformational change of the p97 hexamer which enables downstream substrate processing. The substrate recognition and ubiquitin-dependent homeostasis pathways are under the control of various co-factors or adaptor proteins that form crucial p97-subcomplexes. This Mg^{+2} dependent ATPase has an N-terminal domain, followed by a conserved D1 and catalytic D2 ATPase domain and an unstructured C-terminal tail. The adaptor proteins of the p97 interact with either the N-domain or the C-terminal tail, resulting in a modulation of the ATPase activity of p97, thereby affecting its cellular function. The mutation of R155H on the N- domain is one of the highest mutated sites on p97. p47 an essential cofactor to the p97-mediated membrane reassembly of the Golgi apparatus and nuclear envelope at the end stage of mitosis and has a preference to bind the R155H mutant over the wild-type p97[94]. The p47 cofactor binding also alters the ATPase activities of both the p97 and $p97^{R155H}$. The mechanistic detail of the disease mutation of $p97^{R155H}$ on the p47 binding has not been fully

investigated yet. To deeply understand the molecular details, I assembled the proteins *in vitro* and determined the structures to answer the disease-relevance of the complex.

In Chapter 4, I have further investigated the role of the D1-D2 linker of p97 and relevance of the conserved residue L464 in the domain-domain communication by performing cryo-EM analysis of the full length $p97^{L464P}$ proteins. The L464P mutation dramatically affects the p97 ATPase activity in presence or absence of cofactors like p47 and our results shed light on the D1-D2 linker's relevance in the domain crosstalk.

1.6.3 Role of p97 in Cancer

Cellular processes are largely dependent on the coordination of proteins. Many diverse pathways are involved in proteostasis, including protein synthesis and trafficking. The p97 mediated ubiquitin-proteasome system (UPS) is the major contributor in controlling protein degradation by various methods, like programmed cell death, stress response, protein quality control or clearing the unfolded and damaged proteins to maintain healthy individuals[106]. p97 serves as a molecular chaperone to recognize polyubiquinated proteins, and pools them out from the unmodified proteins to enable targeted degradation by the 26S proteosome[107]. p97 acts as an unfoldase and segregase and play an important role in cell cycle control and DNA damage pathways. The p97 mediated ubiquitin dependent degradation pathways are affected in cancer resulting in higher cell proliferation and lower rates of cell death[108]. Due to increased cell proliferation[108], cancer cells encounter intrinsic stress factors (like expression of oncogenes, DNA replication being upregulated and oxidative stress) and extrinsic stress factors (like environmental change resulting in lack of nutrients and various chemotherapeutic agents), which result in the cells seeking strategies to overcome for survival. Thus, in most cases of cancer, the

UPR is upregulated and to restore the cell homeostasis, cancer cells have a reduced sensitivity due to higher environmental stress[109-112], resulting in overexpression of p97, correlating to the aggressiveness of the tumor. Previous research in the various labs began developing potent inhibitors such as NMS-873[113] and CB-5083[114] to treat diseased conditions, and investigating the mechanistic pathway can help in improving existing treatments.

In the final chapter, I have concluded the results of the cryo-EM structures solved from all the above research and discussed the future directions.

1.7 Recent Advances and Existing Challenges in Cryo-EM

Over the past years there is a "resolution revolution" of cryo-EM afforded by the development of better detectors, cameras, faster GPUs, and image processing algorithms, resulting in exponential increase in structures deposited to the electron microscopy data bank (EMDB). Numerous developments enabled cryo-EM to enter this era for structure determination. The TEM have high performance FEG, stages which have minimum drift during data-collection, and complete automation in data collection over long stretches of time. The development of DED has been a great benefit than the previously used detection technologies like photographic films or CCD's. The DED's can directly convert the detected electrons into electrical signal, significantly increasing DQE with overall increase in the SNR in the images[115]. Additionally, DEDs are popular for their high "read-out-speed" and that allows us to record movies instead of single frames, with improved alignment of movies to correct effects of "beam-induced motion" and thus increases final resolution manifold[40]. The introduction of "Volta phase plates" which improves the phase contrast[116], enabled structure determination of very small proteins

possible, like the 64 kDa hemoglobin solved at 3.2 Å[117], which are typically difficult targets for cryo-EM. Advances in developing newer instruments for sample application to cryo-EM grids like the Spotiton[118], Chameleon (SPT Labtech), VitroJet[119], and cryoWriter[120], which require negligible volumes of sample and have controlled settings, have paved the way for easy, reproducible and precise semi-automated cryo-EM sample vitrification. Besides these hardware developments, the software's in data processing have been continuously streamlined and become user-friendly. For example, the comprehensive software of cryoSPARC[52,53] can be efficiently learnt by a beginner. The field also benefits from the existing crystallization software for *de-novo* model building like COOT[10,11,121] and Phenix[9].

The application, throughput and resolution of cryo-EM has undergone remarkable growth over the past decade; however, several bottle necks exist, which limit the true potential of this groundbreaking method. The structure determination is heavily dependent on the protein biochemistry, and preparation of an ideal cryo-EM grid, having a high concentration of proteins distributed in all orientations in a thin layer of ice is often difficult. The sample preparation encounters hurdles like preferred orientation bias of notorious protein samples like p97 ATPase, or their preference to stick to the carbon, decreasing overall concentration in the holes. To mitigate the issues of homogeneous sample distribution on grids, or ice thickness several grid treatments can be performed as discussed in section 1.3.

To summarize, cryo-EM is a powerful method for structural biologists to understand the molecular details of the intricate cellular processes and the defective biological systems in diseases. If integrated with other techniques like cryo-focused ion

beam (FIB)[122] and cryo-ET[123], or correlative light and electron microscopy (CLEM)[124], it becomes hugely versatile to successfully determine structures of biological specimens over a broad range of sizes from small membrane proteins to larger viruses or multi-protein complexes. The overall time required from sample preparation to structure determination in cryo-EM is less than a week, and the future will be bright with the gradual development of time-resolved cryo-EM[125] capable of capturing snapshots of short-lived states that enables modelling the non-equilibrium conformational changes over time.

CHAPTER 2

PURIFICATION AND STRUCTURAL STUDIES OF THE NEURONAL APOPTOTIC COMPLEX OF THE proNGF-p75^{NTR}-SORTILIN USING CRYO-EM

2.1 Introduction

Development and maintenance of the nervous system are largely affected by the neurotrophin signaling. Neurodegenerative diseases like dementia, Alzheimer's and Parkinson's diseases are attributed to improper signal transduction between the neurotrophin receptors and their ligands. This commonly leads to gradual degeneration of neurons, mostly resulting from improper signal transduction or abnormal synaptic plasticity[65-67]. The growth factors in neurotrophin (NT) family consist of nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), NT-3, and NT-4. They play a critical role in governing neuronal survival or induction of apoptosis by partnering to two categories of receptors: the high affinity tropomyosin-related kinase (Trk) tyrosine kinase receptors and low affinity p75 neurotrophin receptor (p75^{NTR}), respectively. Both of these have been studied to demonstrate distinct differences in their structures and functions[126]. Among all the neurotrophin receptors, p75^{NTR} plays a flexible but pivotal role in determining a neuron's fate[127].

Neurotrophic genes commonly code for the neurotrophin precursors (31-35 kDa), which are later processed to their mature forms (13.2-15.9 kDa) by proteases or other prohormone convertases. Various members of proprotein convertase (PC) family of serine proteinases like furin, PACE4, PC2 and PC1/3 facilitate the removal of the pro-domain of neurotrophins by cleaving at a single or a pair of basic residues[128]. In the

nervous system, numerous growth factors that are interlinked with the extracellular matrix follow a similar origin to form their active mature form upon fragmentation from their precursors. The role of pro-domains was implicated in facilitating proper protein folding of the mature forms and directing them to various secretion pathways[129]. Recent studies have hinted about additional roles of the precursor forms as it is found to be in large excess in human brain in the mast cells, prostate gland, skeletal muscles, thyroid gland, hippocampus and the hair follicles[130]. Additionally, proNGF is highly re-expressed under pathological conditions in the injured cells after spinal cord injury, seizures or the neurodegenerative diseases like Alzheimer's disease[131].

$p75^{NTR}$ can bind both the mature and precursor forms of neurotrophins for activation of different signaling pathways, resulting in entirely opposing biological outcomes ranging from cell growth to apoptosis[69]. Heterodimerization of $p75^{NTR}$ with other neurotrophin receptors like TrkA increases its affinity for the mature neurotrophins, NGF and leads to healthy growth signaling, promoting neuron proliferation[132]. On the contrary, when complexed with the non-neurotrophic receptor sortilin, $p75^{NTR}$ has a higher affinity for the precursor of NGF, proNGF, leading to apoptotic signal transduction [133]. This imbalance between the precursor and mature forms of the NGF leading to formation of different protein complexes is one of the leading causes of neurodegenerative diseases[134-137]. Thus, it is crucial to understand the protein complex structure promoting the unfavorable outcome of neuronal death.

$p75^{NTR}$ is a type-I transmembrane receptor belonging to the tumor necrosis factor receptor (TNFR) superfamily, whose members participate in maintaining neuronal homeostasis[138]. The N-terminal extracellular domain (ECD) of $p75^{NTR}$ is comprised of

four negatively charged cysteine-rich domains (CRDs) which bind to ligands, along with several O-linked and N-linked glycosylation sites. This is followed by the stalk-region and chopper domain that connects to the intracellular (ICD) $p75^{NTR}$ death-domain[139]. The crystal structure of $p75^{NTR}$ ECD with proNGF (PDB:1SG1)[140] shows a symmetric mode of binding of 2:2, where the ECDs of $p75^{NTR}$ braces the proNGF dimer[141] interacting at two sites as described in Figure 2-1.

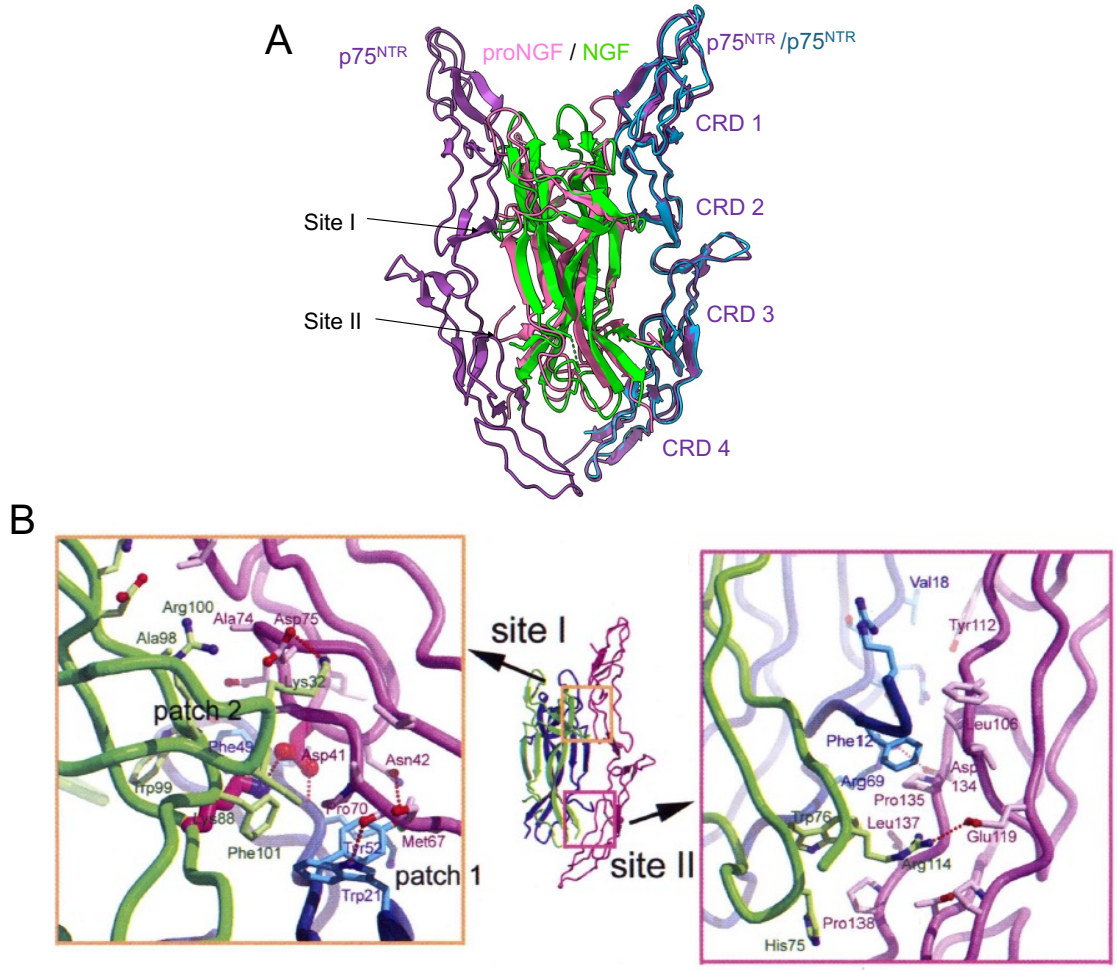


Figure 2-1. Crystal structure of $p75^{NTR}$ ECD bound to proNGF in a symmetric mode. (A) Superimposed crystal structures of $p75^{NTR}$: proNGF and $p75^{NTR}$: NGF. proNGF has a symmetric binding mode to dimeric $p75^{NTR}$ of 2:2 (PDB ID *3IJ2*)[141], though mature NGF has an asymmetric binding ratio of 2:1(PDB ID *1SG1*)[140] ($p75^{NTR}$ monomers are shown in purple/cyan, proNGF in pink and mature NGF in green). Due to high flexibility the NGF pro-domain was not modelled in the crystal structure. The complex is tethered together by two distant binding sites I and II (B) Inset shows the

close-up interactions of mature part of NGF at two sites I and II with the $p75^{NTR}$ cysteine rich domains (CRD) 2 and 3 (backbone is shown as tubes and sidechains in sticks)[140].

Sortilin, another type-I transmembrane receptor, belonging to the VPS10p[142] (vacuolar protein sorting 10 protein domain) family which acts as a cell surface receptor and plays an important role to mediate the apoptotic signaling with $p75^{NTR}$ and proNGF[142-145]. Sortilin possesses a characteristic ten-bladed β-propeller domain that facilitates ligand-binding, with subsequent cysteine-rich domains 10CC-a and 10CC-b, followed by a single transmembrane domain and cytosolic tail[146]. Sortilin uses an environmental pH-dependent monomer-dimer equilibrium to bind and release a diverse set of ligands to propagate its various cellular functions[146,147]. Sortilin is a monomer at neutral pH and dimerizes at acidic pH using the luminal portion of the β-propeller domain in its extracellular portion[148,149] as described in Figure 2-2. The proNGF has higher affinity for binding sortilin than the mature NGF[150], indicating the participation of the pro-domain. The mutual binding affinity of $p75^{NTR}$ for sortilin, is higher in presence of proNGF, thereby leading to the ternary complex formation and the induction of apoptosis[145]. The closely related homolog of sortilin, sortilin related CNS-expressed receptor 2 (SorCS2) forms a homo-dimeric complex with mature NGF[151], which can provide a basis to understand the protein interactions stabilizing this apoptotic complex.

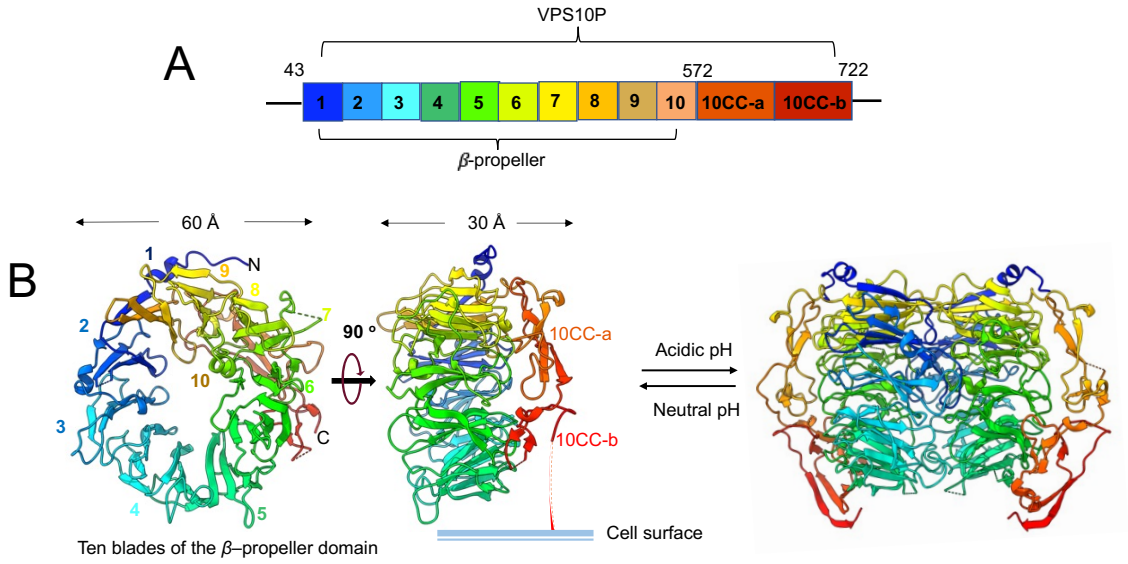


Figure 2-2. Low pH induced dimerization of sortilin. (A) Extracellular portion of mouse sortilin comprised of ten bladed β-propeller domain B) Ribbon representation of a single chain of the mouse sortilin viewed from the top face of the β-propeller (left) and from the side (right) where the domains are numbered and colored according to scheme (A) (PDB ID: *5NMR*)[148]. The side view (right) of the crystal structure at acidic pH of 5.5 depicting the dimerization surface of sortilin using β-propeller and its proposed orientation on the cell surface (PDB: *5ZNN*)[152].

The extracellular domains of $p75^{NTR}$ and sortilin are sufficient to drive the complex formation[153]. The interaction is initiated after activation of $p75^{NTR}$ by α- and γ-secretases controlled regulated intramembrane proteolysis (RIP), which separates the $p75^{NTR}$ ECD from the membrane segment[154-156]. Sortilin plays an important role in the $p75^{NTR}$-RIP and enhances further receptor heterodimerization[150]. The cytoplasmic domain of $p75^{NTR}$ has low affinity for sortilin, while the cytoplasmic domain of sortilin plays an important role in $p75^{NTR}$ cleavage. The $p75^{NTR}$ stalk region is seen to have maximum interaction with sortilin, confirmed from co-immunoprecipitation experiments on various $p75^{NTR}$ truncations[150]. Sortilin originates in a *trans*-Golgi network (TGN) and is responsible for sorting a variety of proteins involved in secretion, internalization, and endosomal pathways[146]. Surprisingly, the sortilin is seen to be present at the cellular

surface when complexed with p75^{NTR} to regulate death signaling, and the connection is not clearly understood why this interaction results in apoptosis.

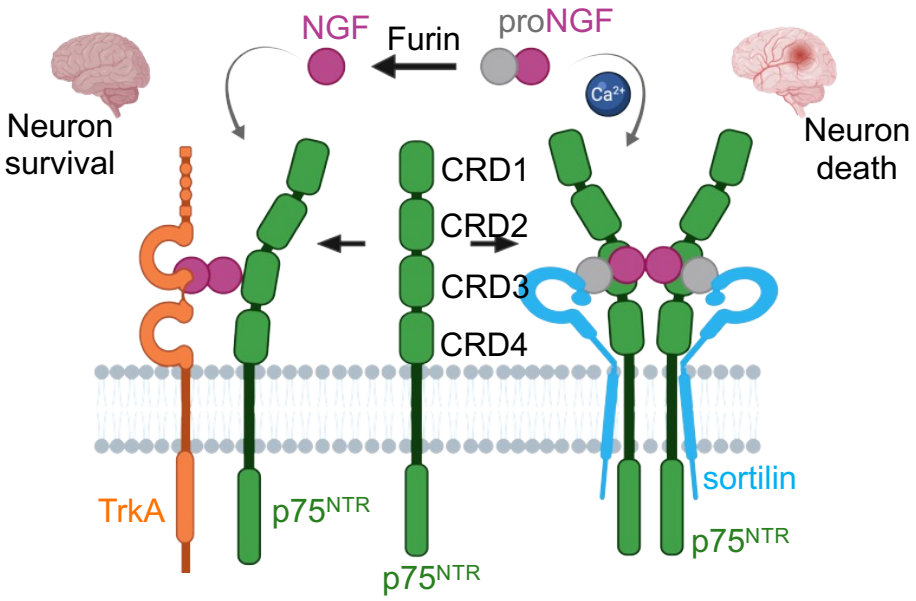


Figure 2-3. Schematic representing the dual opposing roles of p75^{NTR}. In the healthy neuronal survival pathway (left), the p75^{NTR} forms a complex with TrkA in the presence of NGF[157], whereas the neuronal apoptotic pathway assisted by sortilin involves the preferential binding[141] of precursor proNGF to p75^{NTR} (right). The model on the right is hypothetical with the probable stoichiometry of 2:2:2 for proNGF: p75^{NTR}: sortilin based on existing crystal structures of p75^{NTR} ECD with NGF or proNGF[126,140,141,151].

The structure of the neuronal apoptotic complex of proNGF-sortilin-p75^{NTR} has not been solved till date. There is no clear understanding about the interaction residues, and how the established ternary complex propagates downstream signaling. The *in-vitro* experiments revealed that the preferential binding affinity of the precursor proNGF (K_d=0.77 μM) to p75^{NTR} than to mature NGF (K_d=8 μM)[158,159] gets further upregulated (K_d=0.14 μM) in presence of sortilin. Because the proNGF, not NGF, stimulates the apoptotic signaling via p75^{NTR}, the pro-domain is the key to the initiation of the complex formation of proNGF-p75^{NTR}-sortilin. The pro-domain of proNGF is highly

disordered[160,161], making it difficult to be crystallized. Thus, we hypothesize that the pro-domain in the ternary complex may be stabilized by interacting with $p75^{NTR}$ and sortilin (Figure 2-3). To probe the molecular details of the protein-protein interactions, single-particle cryo-EM will be useful for structural studies on this protein complex. To explore the possible models of the interaction of the full-length domains of proNGF-$p75^{NTR}$-sortilin ternary complex it would primarily require identification of the stoichiometry of the interaction obtained from extracellular domains, using size exclusion chromatography and negative-stain EM analysis. NGF dimerization is known to facilitate NGF-$p75^{NTR}$ or NGF-$p75^{NTR}$-TrkA crystal packing by yielding suitable surface for interactions[139]. Furthermore, proNGF also dimerizes[141] as is seen in its crystal structure with dimeric $p75^{NTR}$. Thus, our hypothesis predicts homo-dimerization of proNGF may guide the complex formation, resulting in increased stability of the pro-domain.

Cryo-EM is a highly advanced technique with the recent revolution in resolution due to advanced detectors and image processing algorithms, enabling the structural studies of proteins which are difficult to crystallize[162]. Single particle analysis (SPA) allows a three-dimensional (3D) reconstruction from averaging multiple projections of the target protein that are randomly oriented in vitreous ice [163] .The advantage of the SPA over all other structural methods is that this method does not require crystal sample, which makes it a powerful technique. To date, various studies have elucidated some possible signaling pathways for $p75^{NTR}$; however, the molecular mechanisms of $p75^{NTR}$ receptor activation and downstream signaling are still a mystery. The project aims to study the interactions between the proNGF-$p75^{NTR}$-sortilin ternary complex and gain insights into neuronal apoptosis.

2.2 Materials and Methods

2.2.1 Bac-to-Bac® Baculovirus Expression Vector System *(BEVS*)

This method helps to rapidly generate recombinant baculovirus for high throughput expression of proteins in insect cell lines from the fall armyworm like *Spodoptera frugiperda (Sf9, Sf21)* cells and the cabbage looper *Trichoplusia ni* (Hi-5) cells. The insect cell system is more suitable than the prokaryotic system since it incorporates post-translational modifications required for proper folding of our target glycoproteins. *BEVS* allows the choice of the pFastBac donor plasmid where our gene of interest was cloned under the control of the baculovirus-specific polyhedrin promoter. The recombinant bacmid with a kanamycin resistance gene is thereby generated by transposition of the pFastBac in an *E. coli* strain DH10Bac^{TM}, which has the baculovirus shuttle vector (Bacmid) and a helper plasmid. The mini-attn 7 site from the pFastBac donor plasmid helps to insert the foreign gene at the transposon Tn7 site, which is the N-terminus to the lacZα peptide sequence. Thus, successful recombination disrupts the lacZα reading frame and helps in the blue-white selection, as blue colonies grown in the presence of Bluo-gal and IPTG, indicate the negative results. In my experiments, the pFastBac1 plasmids with our gene of interests were transformed into DH10Bac^{TM} competent cells and plated on LB (Luria broth)-Agar plates with 50 µg/ml kanamycin, 7 µg/ml gentamicin, 10 µg/ml tetracycline, 100–300 µg/ml Bluo-gal, and 40 µg/ml IPTG, which were incubated for 24-48 hours at 37 °C. The recombinant Bacmid was thereafter isolated from the white colonies by mini-cultures and used to transfect the healthy insect cells (*Sf9)* in the purchased media of Insectragro® Sf-900 II serum free medium (Corning 13-410-CV), using Cellfectin II- lipid-based reagent. A 1 ml volume of *Sf9* insect cells

seeded at ~10^6/ml density was used to generate a P1 generation virus in seven days at 27 °C, which was further used to generate a P2 generation virus from 4ml of *Sf9* cells at ~10^6/ml in the next five days. Healthy *Sf9* at ~$3*10^6$/ml were treated with the P2 virus to express our target proteins. This method is widely used over others as it eliminates multiple rounds of plaque purification and is highly site-specific[164].

Primer name	**Sequence 5'-3'**	**Used for**
M13 Fwd	CCCAGTCACGACGTTGTAAAACG	This study
M13 Rev	AGCGGATAACAATTTCACACAGG	This study
Sortilin Gibson Fwd	GCGCGGATCCATGCGGCTGGACG CGCCG	This study
Sortilin Gibson Rev	CTTTGAATTCTTACACGTGATGGT GGTGATGGTGATGGTG	This study
Sortilin pFB1 gibson Fwd	TCACGTGTAAGAATTCAAAGGCC TACGTCGACGAG	This study
Sortilin pFB1 gibson Rev	CCAGCCGCATGGATCCGCGCGCC CCGATGGT	This study
proNGF site 2 KRtoAA Fwd	GCCCTCCGTTCACCCCGTGTG	This study
proNGF site 2 KRtoAA Rev	CGCAAACAGCCTGGGGTCCAC	This study
proNGF site 1 RRtoAA Fwd	CACTGCCCTTGCCGCGGCCCGCA GCG	This study
proNGF site 1 RRtoAA Rev	TCAAGGGAATGCTGAAGTTTAG	This study
proNGF site 3 R312A_Fwd	CAGGACTCACGCCAGCAAGCGGT CATCATC	This study
proNGF site 3 R312A_Rev	TTGAAGGGGGCAGCACCA	This study
proNGF site 3 R321A_Fwd	CGCCAGCAAGGCCTCATCATCCC ATCC	This study
proNGF site 3 R321A_Rev	TGAGTCCTGTTGAAGGGG	This study

Table 2-1. Primers used in this study.

2.2.2 Cell Culture

The cell line of *Sf9* insect cells was purchased from Thermo Fisher Scientific (SKU #11496-015) and passaged every 48 hours in commercial media of Insectragro® Sf-900 II serum-free medium (Corning 13-410-CV), to maintain viable cells for protein expression.

2.2.3 Cloning, Over-expression and Purification of Recombinant $p75^{NTR}$-ecto in *Sf9*

A single-chain construct with an N-terminal tandem hexa-histidine (His_6) tag and a maltose-binding protein (MBP) tag was prepared. The $p75^{NTR}$-RFP plasmid was purchased from Addgene (Watertown, MA; Cat No. 24092), and the ecto-domain containing the residues 29-250 was cloned into a *pFastBac1*™ vector between the two restriction sites, *EcoRI* and *BamHI*. Sequencing and restriction double digestion was used to validate the cloning results. Bacmid colonies were analyzed by colony polymerase chain reaction (PCR) using M13 forward and reverse primers (Table 2-1) to verify the recombinant baculovirus genome. Protein overexpression was performed in *Sf9* insect cells using the baculovirus expression system. The protein was harvested from the virus-infected *Sf9* insect cells, 72 hours post-infection, followed by extraction from the insect cell membranes by douncing with lysis buffer of 20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 1 mM EDTA, and a cOmplete protease inhibitor tablet (Roche #11836170001). Furthermore, the supernatant from centrifugation was loaded to a 5 mL MBP-Trap HP column (GE Healthcare, SKU-28-9187-80) equilibrated with 5 column volumes (CV) of binding buffer (20 mM Tris-HCl (pH 7.4), 200 mM NaCl and 1 mM EDTA) and was eluted using 10 mM maltose in binding buffer. The elution fractions were purified by size-exclusion chromatography (SEC) using a Superose 6 10/30 GL

column (GE ÄKTA Healthcare) equilibrated with the gel-filtration buffer (20 mM Tris-HCl (pH 7.4) and 150 mM NaCl). The proteins were confirmed from various biochemical analysis as SDS-PAGE and Western immunoblotting using monoclonal anti-His antibody (Sigma Aldrich #A5588) as discussed in Figure 2-4.

2.2.4 Cloning, Over-expression and Purification of Recombinant sortilin- ecto in *Sf9*

Our construct was designed for the extracellular domain of sortilin (sortilin-ecto; residues 33-725) and amplified from a full-length plasmid sortilin (DNASU plasmid repository, Biodesign, ASU) with a C-terminal tandem hexa-histidine tag (His_6) and a MBP tag. It was cloned into baculovirus transfer vector *pFastBac1*™ (Thermo fisher Scientific, Waltham, MA) between the *EcoRI* and *BamHI* sites using Gibson assembly (primer designs shown in Table 2-1). The correct orientation and sequence of the insert was verified by sequencing (Genewiz, South Plainfield, NJ) and restriction double digestion. Bacmid colonies were analyzed by colony PCR using M13 forward and reverse primers (sequences shown in Table 2-1) to verify the recombinant baculovirus genome. The protein overexpression was performed in *Sf9* insect cells using the baculovirus expression system. After optimizing the over-expression of the protein in *Sf9* cells at 27°C for 48-96 hours in 250 mL flasks shaken at 125 rpm, the protein was harvested from the P2 virus-infected *Sf9* insect cells, 72-hours post-infection. Recombinant sortilin-ecto was extracted from the insect cell membranes by douncing with lysis buffer of 25 mM HEPES (pH 7.2), 500 mM NaCl, 2mM $CaCl_2$ and EDTA-free cOmplete protease inhibitor tablet (Roche #11836170001). Furthermore, the supernatant from centrifugation was loaded to a 5ml MBP-Trap HP column (GE Healthcare, SKU-28-9187-80), equilibrated with 5 column volumes (CV) of binding buffer (25 mM

HEPES (pH 7.2), 200 mM NaCl and 2 mM $CaCl_2$) and was eluted using 10 mM maltose in binding buffer. The elution fractions were purified by SEC using a Superose 6 10/300 GL increase column (GE ÄKTA Healthcare) equilibrated with the gel-filtration buffer (25 mM HEPES (pH 7.2), 150 mM NaCl and 2 mM $CaCl_2$). The proteins were confirmed from various biochemical analysis such as SDS-PAGE and Western immunoblotting using monoclonal anti-His antibody (Sigma Aldrich #A5588) as discussed in Figure 2-5.

2.2.5 Cloning, Over-expression, and Purification of Recombinant proNGF in *E. coli*

Our construct has a N-terminal hexahistidine (His_6) tag followed by proNGF residues (-103-120), synthesized from Integrated DNA Technologies (IDT) and cloned into the pET-24a (+) (Novagen) vector between the two restriction sites, namely *EcoRI* and *HindIII*, and confirmed by sequencing and restriction double digestion. This protein was overexpressed in *E. coli* (BL21DE3*) cells using 1mM IPTG induction in the mid-log culture phase. The cell lysis was conducted by sonification using the lysis buffer of 20 mM HEPES (pH 7.4), 150 mL NaCl, 1mM PMSF, 1mM DTT and cOmplete protease inhibitor tablet (Roche #11836170001). The cell lysates were centrifuged at 40,000 xg at 4 °C for 40 minutes. Supernatant from centrifugation was loaded to a chelating 5mL HisTrap HP column (GE Health Care) and was washed with two low-imidazole buffers wash buffer 1 (20 mM HEPES (pH 7.4), 150 mM NaCl and 10 mM imidazole), and wash buffer 2 (20 mM HEPES (pH 7.4), 150 mM NaCl and 25 mM imidazole) to remove non-specifically bound proteins. The His-tagged proNGF was further eluted out using an elution buffer (20 mM HEPES (pH 7.4), 150 mM NaCl and 250 mM imidazole), and loaded to a size exclusion Superose 6 10/30 GL column (GE Healthcare) equilibrated

with a gel-filtration buffer (20 mM HEPES (pH 7.4) and 150 mM NaCl, 1mM DTT). The purified proNGF was further analyzed by SDS-PAGE as discussed in Figure 2-6.

The whole cell pellets were further evaluated and I successfully extracted proNGF proteins refolded from inclusion body pellets[165]. The pellets were re-suspended in lysis buffer (1 mM EDTA and 50 mM phosphate buffer at pH 7.0) and crushed using a sonicator (7.5 mins, 1s on 3s off pulse; on ice). The cell homogenate was then centrifuged at 8,000 rpm and 4 °C for 15 minutes. The collected inclusion bodies were re-suspended in 100 mL of detergent buffer (0.5% Triton X-100, 2 M urea, 1 mM EDTA and 50 mM sodium phosphate buffer at pH 7.0) and stirred with the magnetic agitator for 4 hours. This was followed by centrifugation at 8000 rpm for 15 minutes, and the proNGF-inclusion body pellets were obtained. A 1 g pellet of the washed proNGF- inclusion body was solubilized into 10 mL denaturing buffer (8 M urea, 10 mM DTT, 1 mM EDTA and 50 mM phosphate buffer at pH 7.0) and stirred at room temperature for 6 hours. The solution containing the denatured protein was then centrifuged at 10,000 rpm and 4 °C for 30 minutes, to remove the insoluble particles. Dialysis was used to exchange to refolding buffer (3 M urea, 10 mM EDTA, 0.15 M NaCl and 50 mM phosphate buffer at pH 7.0). Further stepwise dialysis (decreased salt gradually from 3M, 2M, 1M, 0.5 M, with change of buffer every 2.5 hours) was used to reduce final urea concentration to 0.5 M, as protein precipitated below that concentration. The sample was assessed for proper protein folding using circular dichroism (CD) spectroscopy and verified with 15% SDS-PAGE. Far-UV CD spectra were recorded on an JASCO corp, J810 spectropolarimeter from 190 to 260 nm. For all measurements, protein samples were buffer exchanged to 50 mM Na-phosphate (pH 7.0), 100 mM NaCl, 0.5 M urea, at 20 °C. The spectra were

buffer corrected, and the results concluded that proNGF was deposited in the pellets as confirmed from a 15% SDS-PAGE gel electrophoresis in Figure 2-6.

2.2.6 Cloning, Over-expression and Purification of Recombinant proNGF in *Sf9*

The above proNGF construct with an N-terminal hexa-histidine (His_6) tag was subcloned into the baculovirus transfer vector pFastBac1™ (Thermo Fisher Scientific, Waltham, MA) between the *EcoRI* and *BamHI* sites. This was done since the protein expressed in *E. coli* lacked glycosylation and the required post-translational modifications, as well as the refolded proteins using strong denaturant like urea will likely be unsuitable for future cellular functional studies. The correct orientation and sequence of the insert was verified by sequencing (Genewiz, South Plainfield, NJ) and restriction double digestion. Bacmid colonies were analyzed by colony PCR using M13 forward and reverse primers (sequences shown in Table 2-1) to verify the recombinant baculovirus genome. The protein overexpression was performed in *Sf9* insect cells using the baculovirus expression system and was optimized at 27 °C for 48-96 hours in 250 mL flasks shaken at 125 rpm. The protein was harvested from the P2 virus infected *Sf9* insect cells, 72-96 hours post-infection. The protein was extracted from the insect cell membranes by douncing with lysis buffer of 10 mM HEPES (pH 7.2), 150 mM NaCl, 1 mM DTT, 0.02 % NaN_3 and EDTA-free cOmplete protease inhibitor tablet (Roche #11836170001). The cell lysates were centrifuged at 40,000 xg at 4 °C for 40 minutes. Supernatant from centrifugation was loaded to a chelating 5mL-HisTrap HP column (GE Healthcare) and was further purified on Superose 6 10/30 GL column (GE Healthcare) using similar conditions as stated in section 2.2.5. The purified proNGF was further analyzed by SDS-PAGE,

Western immunoblotting using monoclonal anti-His antibody (Sigma Aldrich #A5588) and mass spectrometry-based proteomics as discussed in Figure 2-7B.

2.2.7 Mutagenesis Experiments to Generate the Stable Triple-mutant proNGF

The proNGF molecule is processed by cellular proteases like furin at three dibasic sites located in the pro-domain and results in processed mature NGF[166]. Site 3 adjacent to the N-terminus of mature NGF is the major site responsible for processing of proNGF to mature NGF. However previous studies[166] have indicated that single and double mutants are partially cleaved, and triple mutation renders maximum resistance to degradation to NGF. To generate this stable triple-mutant construct, the Q5 site directed mutagenesis was performed stepwise according to manufacturer's instructions, mutating one site at a time with a set of primers as described in Table 2-1. This enabled to generate multiple constructs including site-3 mutant, several double mutants and the triple-mutant (pro123) construct. Each step was confirmed using Genewiz sequencing. Further the pro123 protein was expressed and purified from insect cells (*Sf9*) using identical conditions as discussed in sections 2.2.5 and 2.2.6, and the results are discussed in Figure 2-7A, C.

2.2.8 Expression and Purification of the proNGF-p75^{NTR}-sortilin Complex by Co-transfection in *Sf9*

To increase incubation time of the three baculoviruses responsible for generating the proteins, the three P2-viruses of p75^{NTR}-ecto, sortilin-ecto and wild-type proNGF were co-transfected to healthy *Sf9* insect cells. The cells were shaken at 125 rpm, at 27 °C for 72 hours post transfection followed by harvesting. The protein complex was extracted from the insect cell membranes by douncing with lysis buffer of 25 mM HEPES (pH 7.2), 500 mM NaCl, 2mM $CaCl_2$ and EDTA-free cOmplete protease inhibitor tablet

(Roche #11836170001). Since each of the $p75^{NTR}$-ecto and sortilin-ecto constructs had a MBP tag, the supernatant from centrifugation was loaded to a 5ml MBP-Trap affinity column (GE Life Sciences, Piscataway, NJ, USA), equilibrated with binding buffer (25 mM HEPES (pH 7.2), 200 mM NaCl and 2 mM $CaCl_2$) and was eluted using 10 mM maltose in binding buffer. The elution fractions were purified by SEC using a Superose 6 10/300 GL increase column (GE ÄKTA Healthcare) equilibrated with the buffer (25 mM HEPES (pH 7.2), 150 mM NaCl and 2 mM $CaCl_2$). As a control, the experiment was repeated without adding the P2-virus of proNGF to the *Sf9* cell culture. The biochemical analysis with SDS-PAGE and Western immunoblotting using monoclonal anti-His antibody (Sigma Aldrich #A5588) was done and the results are discussed in Figure 2-8.

2.2.9 Polyacrylamide Gel Electrophoresis (PAGE)

SDS-PAGE was performed by preparing 4% stacking gels and 10% or 15% resolving gels. The protein samples were mixed with equal volume of 62.5 mM Tris-HCl (pH 6.8), 5 mM DTT, 25% glycerol, 0.1% bromophenol blue and 0.2% SDS loading buffer. The running buffer used was 25 mM Tris (pH 8.3), 250 mM glycine, 0.1% SDS. The protein bands were stained by 0.1% Coomassie blue R-250.

2.2.10 Negative-stain Electron Microscopy

Protein samples were stained with 0.75% uranyl formate on an EM grid using previous method[167]. 0.01 mg/mL of the protein sample was used to applied onto a continuous carbon-coated copper grid. The specimen was imaged using a Philips CM12 transmission electron microscope (TEM) (80 keV) at Eyring Materials Center (EMC), Arizona State University (ASU) with a side-mounted CCD camera (Model 791, Gatan, Pleasanton, CA)

with 1,024x1,024 pixels. The image pixel size at the specimen level was found to be 8 Å /pixel.

2.2.11 Cryo-EM Data Collection:

A holey-carbon C-Flat grid (2/1 4C; Protochips, Morrisville, NC) was glow-discharged for 15 seconds using a Pelco easiGlow glow-discharge system (Ted Pella, Redding, CA). 6 µl of purified protein complex sample at a concentration of 0.3 mg/mL was applied onto the pretreated grid, and the excess solution was blotted away by a filter paper (Whatman 55/20 nm, 595 Filter paper, Ted Pella) using a Thermo Fisher/FEI Vitrobot Mark IV automated freeze plunger (Thermo Fisher/FEI, Hillsborough, OR) for 6 seconds in a chamber with a humidity of 100%. The grid was then quickly plunge frozen into liquid ethane and transferred to grid storage. Particle homogeneity and ice thickness of the grid specimen were screened using a FEI Tecnai TF20 TEM. Grids with thin ice and a homogenous protein dispersion were used for the subsequent cryo-EM data collection.

For the purified sortilin-p75^{NTR}-proNGF protein complex, the grid specimen was imaged at the EMC at ASU using a Thermo Fisher/FEI Titan Krios TEM (Thermo Fisher/FEI, Hillsborough, OR) at an accelerating voltage of 300 keV. Cryo-EM movies were recorded using a Gatan K2 Summit direct electron detector (DED) camera (Gatan, Pleasanton, CA). Defocus range was set to -0.8 to -2.5 µm. Nominal magnification was 48,077X, resulting in a physical pixel size of 1.04 Å/pixel at the specimen level. The movie data was recorded at a counted rate of 2 e^-/sub-pixel/sec and a sub-frame rate of 200 msec in super-resolution mode. Total exposure was 6 seconds, accumulating to a dosage of 67.64 e^-/A^2. The beam-image shift was applied to accelerate data acquisition[168]. Data collection was automated using the customized SerialEM macros (version 3.9)[169].

2.2.12 Image Processing:

The image processing was generally conducted using cryoSPARC software (version 2.15)[52]. 1490 movies were unpacked and imported. The frame registration and averaging were performed using patch motion correction. The defocus parameters were estimated using patch CTF estimation. The particle selection was performed using Topaz program (version 0.2.3)[170]. The two-dimensional (2D) unsupervised classification was performed to generate the averages from the selected particle pool. The classes with discernible features were selected for *ab initio* volume generation using stochastic gradient descent (SGD) method [52]. 135011 particle images were used to reconstruct a 3D volume at 6.20 Å resolution. The processing schematics is shown in Figure 2-9.

2.2.13 Model Building:

The templates used for atomic modelling or molecular docking were the previous coordinates of sortilin (PDB code: *5NMR*)[148] and coordinates for $p75^{NTR}$ (PDB code:*1SG1*[140], *3IJ2*)[141]. The templates were first docked into individual cryo-EM densities using 'Fit in the Map' function in UCSF Chimera (version 1.14)[171]. The cryo-EM density maps and atomic models were prepared using UCSF Chimera and UCSF ChimeraX (version 0.91)[172]. Due to insufficient resolution, the atomic sidechains could not be docked in the cryo-EM density with precision.

2.2.14 Label-free Proteomics:

These experiments were performed at Prof. Tsui-Fen Chou's lab (Caltech) where the protein samples for LC-MS were prepared using the Thermo EasyPep Mini MS Sample Prep Kit (Thermo Scientific, A4006) according to the manufacturer's instructions using 500 ng of protein samples. The dried peptides were then dissolved in 0.1% formic acid

(Thermo Scientific, 5178) solution, and the peptide concentration was confirmed using the Pierce Quantitative Fluorometric Peptide Assay (Thermo Scientific, 23290). LC-MS/MS experiments were done on an EASY-nLC 1000 (Thermo Fisher Scientific, Waltham, MA, USA) connected to an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) following previously reported procedure[173]. Proteome Discoverer 2.4 (Thermo Scientific) software was used for the proteomics analysis using the Uniprot human database. From the spectrum of ~2500 peptides the signature peptides for the respective proteins were identified.

2.3 Results and Discussion:

2.3.1 The Extracellular Human p75^{NTR} Purifies as a Monomer at Physiological pH

The extracellular construct of p75^{NTR} (residues 29-250) was designed with a C-terminal His-MBP tag and was overexpressed in insect cells (*Sf9*) as shown in Figure 2-4A. It was purified by MBP-Trap affinity chromatography followed by SEC as a monomer of 75 kDa at pH 7.2, as seen in Figure 2-4B-C. p75^{NTR} being a founding member of the TNFR family, is reported to co-exist as trimers and monomers at the cell-surface in prenatal mouse brains, like the other members of the TNFR family[174]. Pre-clustering[175] and homodimerization of p75^{NTR} is proposed to increase the concentration of the receptors at the cell surface. Full-length p75^{NTR} forms homodimers through covalent bonds in the transmembrane residue Cys^{257}, as well as the pre-ligand assembly domain (PLAD)[176] distributed over the CRDs. This oligomerization does not affect the ligand-dependent or independent p75^{NTR} activation as confirmed from the growth cone retraction assay. Though oligomers have been reported for p75^{NTR} purification earlier[177-179], our

extracellular construct purified as a monomer, with no unfavorable aggregation as seen in Figure 2-4C.

The insect cell expressed $p75^{NTR}$-ecto proteins are non-glycosylated and the molecular weights of the purified proteins were confirmed using SDS-PAGE and Western immunoblotting analysis using anti-His antibody. The extracellular construct of $p75^{NTR}$ is 32 kDa in weight, and because of the additional 42.5 kDa MBP tag, it purified at 75 kDa as shown in the SDS-PAGE and Western immunoblotting results in Figure 2-4D. $p75^{NTR}$ is reported to have one N-linked glycosylation site[180] in the extracellular residue Asn^{32}, and several putative O-Linked glycosylation sites in the CRDs, however our results did not show any glycosylation pattern by shift in molecular weight. The role of $p75^{NTR}$ glycosylation in inducing ligand binding is highly debated, since the glycans do not participate in neurotrophin binding in the existing crystal structures of $p75^{NTR}$-NGF[140], $p75^{NTR}$-proNGF[141] or $p75^{NTR}$- NT3[180]. Thus, the $p75^{NTR}$-ECD purified was used for further used for binding experiments with proNGF and sortilin for structural studies.

Further, we analyzed the protein homogeneity using negative-stain EM and identified mono-disperse species. The ECD of $p75^{NTR}$ is characterized by four tandem CRDs, which have twelve pairs of disulfide bonds evenly distributed and limits flexibility by a ladder-like arrangement. This is similar to other members of the TNFR family, solved by X-ray crystallization like TNFR1[181-183], TNFR2[184] and OX40[185], all of which share a similar repeating stretch of four CRDs in the extracellular portion. These cylindrical features of the $p75^{NTR}$ -ECD were observed in our negative-stain EM results and the subsequent 2D class averages, as seen in Figure 2-4E. The extracellular structure of $p75^{NTR}$ undergoes minimal change upon ligand binding[186], and hence our results also

resemble the p75^{NTR}-ECD structure as observed in the p75^{NTR}- NGF^{140}, p75^{NTR}-proNGF^{141} and p75^{NTR}-NT3^{181} crystal structures.

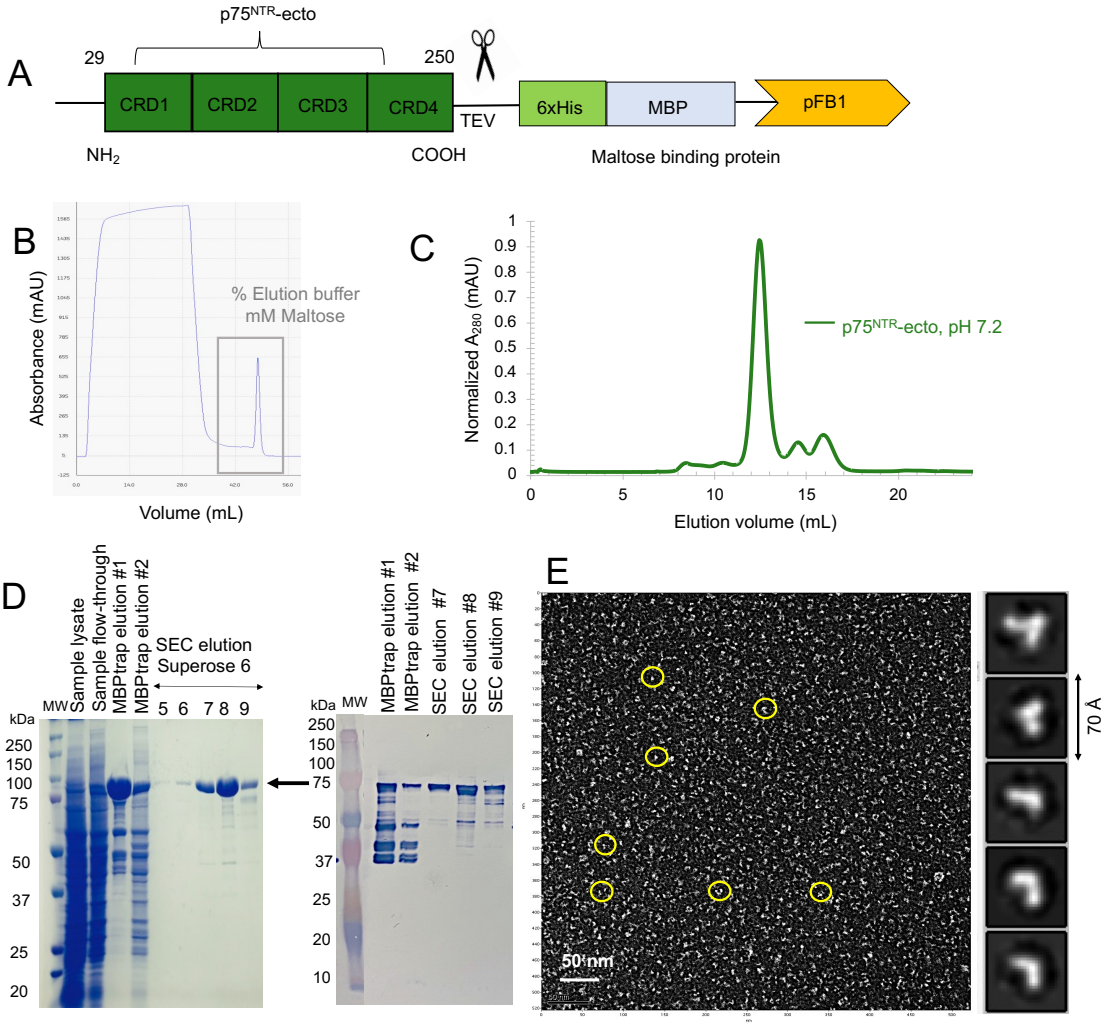


Figure 2-4. Purification of recombinant extracellular p75^{NTR} proteins expressed in insect cells (*Sf9*). (A) Domain architecture of extracellular p75^{NTR} used in our studies, including the four cysteine rich domains. Our construct had a C-terminal His-MBP tag which could be cleaved using TEV protease treatment. (B)MBP-Trap affinity chromatography showing an elution peak of the MBP-tagged protein p75^{NTR} using 10 mM maltose in elution buffer (C) Normalized chromatographic profile on a Superose 6 10/30 GL size-exclusion column at pH 7.2 from the MBP-Trap p75^{NTR} elutants; Green curve represents p75^{NTR}-ecto run. (D)10 % SDS-PAGE gel, stained with Coomassie blue, corresponding to p75^{NTR}purification, indicating an expected molecular weight of 75 kDa as indicated by the black arrow. The molecular marker is labelled on the left. The right panel shows the Western immunoblotting results of the p75^{NTR} using anti-His antibody. (E)A representative negative-stain EM micrograph of SEC elution peak sample, Scale bar indicates 50 nm. Protein particles are homogeneously distributed and shown in white and

a few circled yellow, in contrast to the black stain background, indicating a diameter of approximately 7 nm, which has close resemblance to the atomic model of $p75^{NTR}$-ECD (PDB *1SG1*)[141]. Right panel shows the representative two-dimensional (2D) class averages generated using RELION (version 3.1)[51] using *k*=50 classes,15600 particle images using an extraction box size of 70 Å.

2.3.2 The Extracellular Human sortilin is a Monomer at Physiological pH

The extracellular construct of sortilin (residues 36-725) was overexpressed in insect cells (*Sf9*) and purified using SEC as seen in Figure 2-5A. Our construct bears an MBP-tag and eluted as a predominantly as a monomer of 120 kDa in size exclusion chromatography at pH 7.2 (Figure 2-5B, C). The sortilin extracellular residues corresponds to 67 kDa, and an extra 42.5 kDa increase in molecular weight is observed due to the MBP tag. There is a leading shoulder as seen in the SEC profile in Figure 2-5C which represents the scarce amount of sortilin dimers present in equilibrium with the monomers. Our results are consistent with the previous studies that sortilin is monomeric at neutral pH[149]. The biochemical analysis using SDS-PAGE gel electrophoresis and Western immunoblotting performed with anti-His antibody, confirmed the pure proteins as seen in Figure 2-5D. We also evaluated the protein quality using negative-stain EM, and our results indicated homogeneous particle distribution, with no aggregation as seen in Figure 2-5E. The 2D class averages were generated by particle picking from hundreds of negative-stain EM micrographs using RELION (version 3.1)[51], and they depict the features of the ten bladed β-propeller domain as seen in mouse sortilin (PDB *5NMR*)[148] crystal structure, with the top face resembling a circular donut as observed clearly in the 2D class averages in Figure 2-5E (right panel).

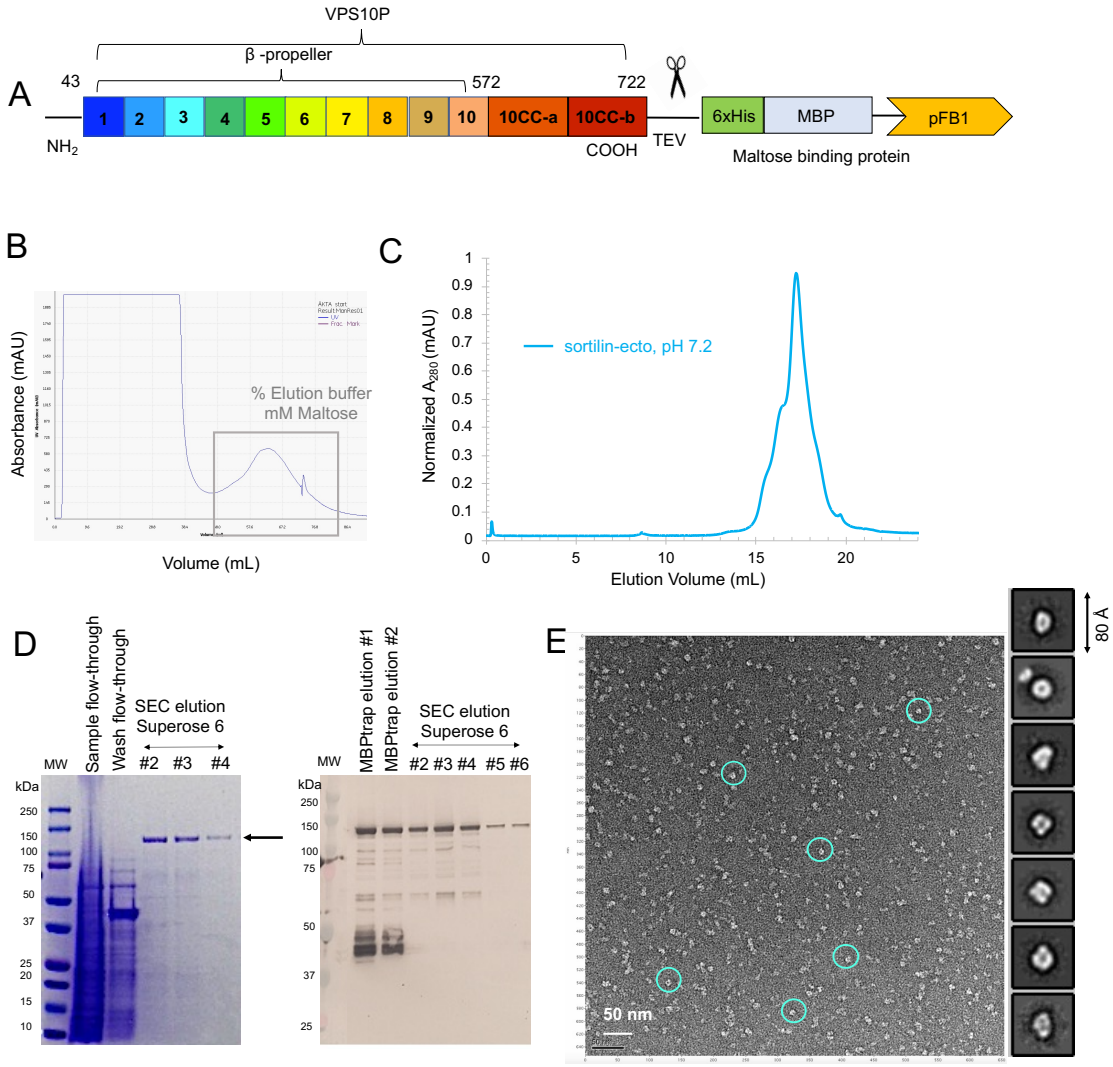


Figure 2-5. Purification of recombinant extracellular sortilin proteins expressed in insect cells (*Sf9*). (A) Domain architecture of extracellular sortilin used in our studies. Our construct had a C-terminal His-MBP tag which could be removed using TEV protease treatment. (B) MBP-Trap affinity chromatography showing an elution peak of the MBP-tagged protein sortilin using 10 mM maltose in elution buffer (C) Normalized chromatographic profiles on a Superose 6 10/30 GL size-exclusion column at pH 7.2 from the MBP-Trap elutants; blue curve for sortilin-ecto purified as a monomer. (D) 10 % SDS-PAGE gel, stained with Coomassie blue, corresponding to sortilin purification (peak in C). The molecular marker is labelled on the left MW. The SEC peak indicates an expected molecular weight of 120 kDa as shown by the black arrow. The right panel shows the Western immunoblotting results of the sortilin using anti-His antibody. (E) A representative micrograph of SEC peak sample screening by negative-stain EM. Scale bar indicates 50 nm. Proteins were homogeneously distributed and shown in white, in contrast to the black stain background. Features of the characteristic donut shaped ten bladed β-propeller domain, with a close resemblance to the atomic model of mouse

sortilin-ECD (PDB *5NMR*)[148] were seen as indicated by blue circles. Inset shows the representative two-dimensional (2D) class averages generated using RELION (version 3.1)[51] using *k*=50 classes, from 18500 particle images from negative-stain EM micrographs with an extraction box size of 80 Å.

The luminal section of the β-propeller domain is responsible for facilitating sortilin to bind a diverse set of ligands in the monomeric state. Though about six possible N-linked glycosylation sites are reported for sortilin-ECD purifications[187], our results do not indicate any increase in molecular weight due to addition of glycans, and it remains unclear if the glycans help sortilin-ECD in binding $p75^{NTR}$ or proNGF.

2.3.3 Purification of proNGF

proNGF is susceptible to post-translational cleavage by proteases like furin, and in order to perform downstream experiments to test the neuronal cell apoptosis, it is important to preserve the intact pro-domain. We initially purified the wild-type (WT) proNGF constructs from *E. coli* to rule out the effects of post-translational modifications (Figure 2-6A), but could recover only small quantities of soluble proteins even after scaling up the culture volume. This proNGF construct bears a His-tag and was purified by Ni-NTA affinity chromatography and size exclusion chromatography as shown in Figure 2-6B, C. The protein purified as a monomer of 27 kDa as confirmed from SDS-PAGE gel electrophoresis in the Figure 2-6D. We tested the cell pellets to locate that most of the proteins were deposited in the pellets. This is due to the three possible disulfide linkages in the mature NGF domain, that targets the proteins to aggregate in pellets[188]. For rescuing the inclusion body proteins from the proNGF pellets, strong urea-based denaturation was performed[165], and the protein was refolded using stepwise dialysis. The refolded protein matched the expected monomeric molecular weight of 27 kDa as seen

from the gel electrophoresis results in Figure 2-6E. The protein was correctly folded as confirmed from the circular dichroism (CD) spectroscopy experiments, a dip at 225 nm was observed representing the signature β- sheets of the proNGF as seen in Figure 2-6F.

Future directions of our studies involve treating neuronal stem cell lines like DOAY and U87 with the pure proNGF proteins to trigger formation of apoptotic complexes in the cells and purify them using co-immunoprecipitation. The harsh denaturation conditions encountered by the inclusion body proteins in the process of getting refolded for WT proNGF (*E. coli*) rendered them unsuitable for such studies. We further generated the baculovirus for expressing the WT proNGF proteins in eukaryotic systems like the insect cells; construct shown in Figure 2-7A. We successfully purified WT proNGF from *Sf9*, as confirmed from the 15% SDS-PAGE, subsequent Western immunoblotting with anti-His antibody results as seen in Figure 2-7B, and mass spectrometry-based proteomics analysis. However, we were unable to detect the pure proteins at their monomeric molecular weights from SEC or biochemical analysis. Our SDS-PAGE and subsequent Western immunoblotting depicted a multi-band pattern as seen in Figure 2-7B, which presented upshifted bands than the proNGF monomeric molecular weight of 27 kDa. This is most likely due to different degrees of glycosylation, facilitated by the baculovirus system, resulting in slower migration of some bands in the gel than others. Using tandem MS analysis, two N-linked and two O-linked glycosylation sites were identified in the pro-domain[189] of NGF. The presence of glycans on our purified WT proNGF must be confirmed by treatment with enzymes like PNGaseF or sialidase which can remove N-linked or O-linked glycans respectively.

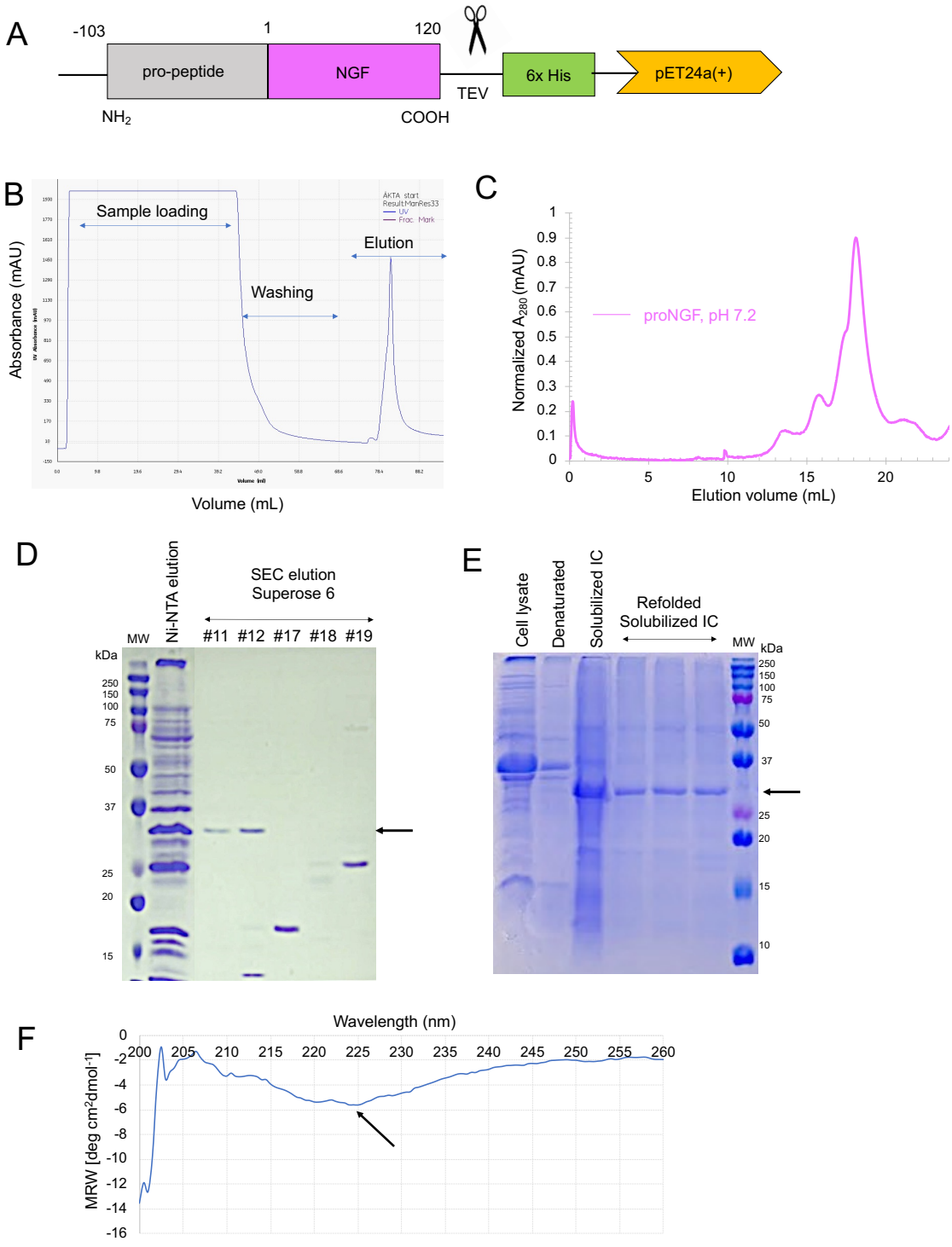


Figure 2-6. Expression and purification of recombinant wild-type proNGF. (A) ProNGF construct domain architecture for expression in *E. coli* was designed with pET-24a (+) backbone (B) Ni-His Trap affinity chromatography showing an elution peak of

the His-tagged protein proNGF using 250 mM imidazole in elution buffer. (C) Size exclusion chromatographic profiles on a Superose 6 10/30 GL size-exclusion column of the elution fractions from Ni-NTA purification. (D) Western immunoblotting results using anti-His antibody after transfer from a 15% SDS-PAGE gel, corresponding to proNGF purification. Molecular weight standards are indicated at the left side. SEC peak fractions of proNGF indicating a molecular weight of ~27 kDa as indicated by the black arow (E) 15% SDS-PAGE, stained with Coomassie blue showing the purification of proNGF (black arrow) by refolding using stepwise dialysis from the inclusion body pellets in *E. coli*. (F) Circular dichroism (CD) spectroscopy of proNGF purified in (E) form inclusion body pellets, recorded using a JASCO J810 spectropolarimeter from 190 to 260 nm where we observe the signature β-sheets of the proNGF signal as a dip at 225 nm [188,190](black arrow).

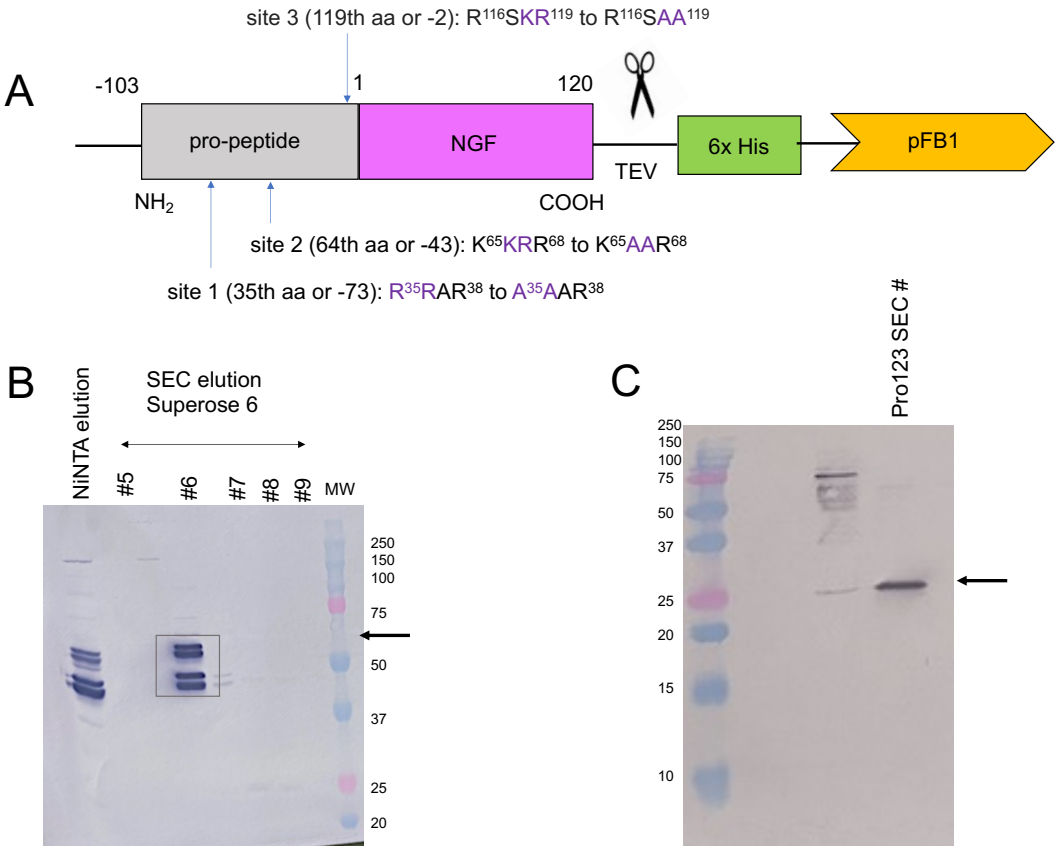


Figure 2-7. Generation of the proNGF triple-mutant construct (pro123) which is resistant to protease degradation[166]. (A)The proNGF construct was re-designed by introducing three mutations to remove the dibasic sites in the pro-domain and render stability against cellular proteases[166]. Q5 site directed mutagenesis was performed using primers in Table 2-1 (B) Western immunoblotting results using anti-His antibody after transfer from a 15% SDS-PAGE, corresponding to wild-type proNGF purification in insect cells (*Sf9*) indicated by the black arrow, the molecular weight is upshifted than the

expected size of 27 kDa, showing a multi-band pattern resulting from possible glycosylation. (C) Western immunoblotting results using anti-His antibody after transfer from a 15% SDS-PAGE, corresponding to the pro123 purification in insect cells (*Sf9*), the molecular weight is at the expected size of 27 kDa (black arrow).

The insect cell system comprises of many inherent cellular proteases which assist proper protein folding, and hence the WT proNGF may be unfavorably cleaved at the dibasic sites to generate the mature NGF. In our attempts to scale up highly pure proNGF, we mutated the cleavage-prone residues[166] in the pro-domain (Figure 2-7A). Though single and double mutants of proNGF have increased resistance to degradation than the WT proNGF, they are only partially stable, and the triple-mutant proNGF (pro123) is the most stable form to a set of diverse proteases[166]. The pro123 could be purified as a monomer of 27 kDa from the insect cells as indicated from Western immunoblotting results with anti-His antibody in the Figure 2-7C and mass spectrometry-based proteomics analysis. Affinity based purifications with His-tag may co-elute non-specific proteins which contaminates our target proteins. Hence, the future studies involve re-cloning the pro123 with FLAG-tag for highly specific FLAG-based affinity chromatography.

2.3.4 The proNGF-p75^{NTR}-sortilin Ternary Complex Formation by Co-transfection

To promote stronger interactions among the protein binding partners, we added the P2-viruses of the extracellular protein constructs of p75^{NTR} and sortilin, as well as the ligand of proNGF and co-expressed the protein complex in insect cells (*Sf9*) for optimal time (72 hours). The formed complex was purified from crude cell lysate by MBP-Trap affinity chromatography followed by SEC. Size exclusion chromatography profiles of the proNGF-p75^{NTR}-sortilin complex shows a trend of four peaks, corresponding to

molecular weights of 440 kDa, 120 kDa, 75 kDa and 60 kDa, indicating four different species as seen in Figure 2-8A. The leading peak fraction at ~440 kDa corresponds to the formed ternary complex, following the stoichiometry of 2:2:2 for proNGF: $p75^{NTR}$: sortilin. The second SEC peak corresponds to 120 kDa, which is most likely the excess unbound sortilin-ecto monomer, third peak at 75 kDa representative of excess $p75^{NTR}$-ecto monomer. The fourth peak is negligible and represents the excess proNGF ligand.

We performed biochemical analysis on the leading SEC peak 1, which has the assembled ternary proNGF-$p75^{NTR}$-sortilin complex, and each of the three proteins proNGF, $p75^{NTR}$ and sortilin were confirmed from 10% SDS-PAGE and Western immunoblotting with anti-His antibody, as shown in Figure 2-8B. The samples were further digested by trypsin for mass spectrometry-based proteomics analysis to confirm the constituents in the leading SEC peak 1. The nerve growth factor (NGF), nerve-growth factor receptor, i.e., $p75^{NTR}$ (NGFR) and sortilin (Sort1) family were all detected in the proteomics results of peak 1 as seen in the results in Figure 2-8C. These data closely resemble the previously reported complex formed by *in-vitro* assembly which eluted at ~500 kDa in size-exclusion chromatography[141], indicating similar stoichiometry. We performed biophysical analyses using negative-stain EM, and stable complexes with a diameter spanning 180 Å were identified from the SEC fractions of peak 1, as seen in Figure 2-8D. The negative staining also confirmed the sample homogeneity in the assembled peak 1 fractions, showing the ternary complex can be transiently expressed and purified form the baculovirus system. Overall, the quantity of the assembled complex (peak 1 in the orange curve) expressed is higher than the remaining peaks (peak 2-4 in the orange curve), confirming that the complex formation imparts added stability, and only

sparse quantities of individual excess proteins are eluted. The subsequent 2D classification elucidates the structure of the formed ternary complex where the ten-bladed β-propeller domain of extracellular domains of sortilin are seen to form a dimeric-sandwich arrangement (Figure 2-8E). The low-resolution 3D volume (Figure 2-8F) generated from the representative 2D class averages, shows that the atomic coordinates of previously determined mouse s-sortilin (PDB *5NMR*)[148] can be well fitted.

The cryo-EM analysis of the purified complex peak (Figure 2-9) allowed to reconstruct a 3D cryo-EM density map at 6.2 Å, where the domains of extracellular sortilin dimers are better resolved and atomic models corresponding to sortilin (PDB *5NMR*)[148] or $p75^{NTR}$ (PDB *1SG1* or *3IJ2*)[140,141] crystal structures could be partially fitted as seen in Figure 2-9D. The interaction surfaces of $p75^{NTR}$ or proNGF cannot be modelled with certainty, due to insufficient resolution and map features. The datasets also represented high sample heterogeneity on the cryo-EM grids with multiple conformations captured in the 2D class averages, unfortunately implying that the ternary complex is unstable. The sample may denature during the vitrification and hence future studies will need to stabilize the purified complex using mild crosslinking[16]. Previous studies suggest that the $p75^{NTR}$ dimerization is not affected by the presence or absence of the pro-domains of NGF (PDB *1SG1* or *3IJ2*)[140, 141]. However, the NGF pro-domain is necessary to initiate interaction with sortilin. It is most favorable if the homodimerization happens in a 2:2 arrangement for the proNGF: $p75^{NTR}$. We cannot infer yet if both the pro-domains of NGF will bind to each sortilin, or both will bind to one sortilin. The previous structures of neurotensin or progranulin bound to sortilin displays that the β-propeller domain of sortilin binds ligands[191]. Thus, the dimeric sandwich arrangement using

sortilin-ecto presents a favorable binding platform for the heterocomplex. Sortilin is monomeric at physiological pH and our purification of sortilin-ecto in the absence of ligands also yielded monomers. Sortilin dimerizes in acidic environment[149], however our proNGF-p75^{NTR}-sortilin heterocomplex structures formed at neutral pH of 7.2 suggest that the dimerization of sortilin provides a binding interface and drives the ternary complex formation under physiological conditions.

The previous crystal structures of NGF:p75^{NTR} or NGF:p75^{NTR}:TrkA report that the dimerization interface is provided by the dimeric NGF[126]. For the closely related homolog of sortilin, SorCS2, a crystal structure is available SorCS2-NGF, where the NGF dimer is bound to cross-braced SorCS2 dimers[151]. The proNGF is stable as a dimer in solution, thus the pro-domain does not hinder dimerization[141]. Hence, it is possible that the proNGF dimer additionally stabilizes this heterocomplex formation, but further studies will be required to confirm this. The p75^{NTR} is capable to bind sortilin in absence of a ligand, thus it's crucial to understand the changes in the ligand binding domains of p75^{NTR} or sortilin upon the ternary complex formation, to understand the apoptotic signaling. As a control experiment, the co-transfection was repeated with identical conditions without the introduction of the baculovirus for proNGF to the insect cells. The purification yielded a lower molecular weight species in SEC (~100 kDa), as indicated by the blue overlay curve in Figure 2-8A. The peak fractions were assessed with SDS-PAGE, Western immunoblotting, and negative-stain EM analysis to find that no appreciable new species was formed. It resulted in a simple mixture of p75^{NTR}-ecto and sortilin-ecto purified in the monomeric states. These results suggest that enough proNGF

is indispensable to guide the ternary complex formation proNGF: p75^{NTR}: sortilin in insect cells.

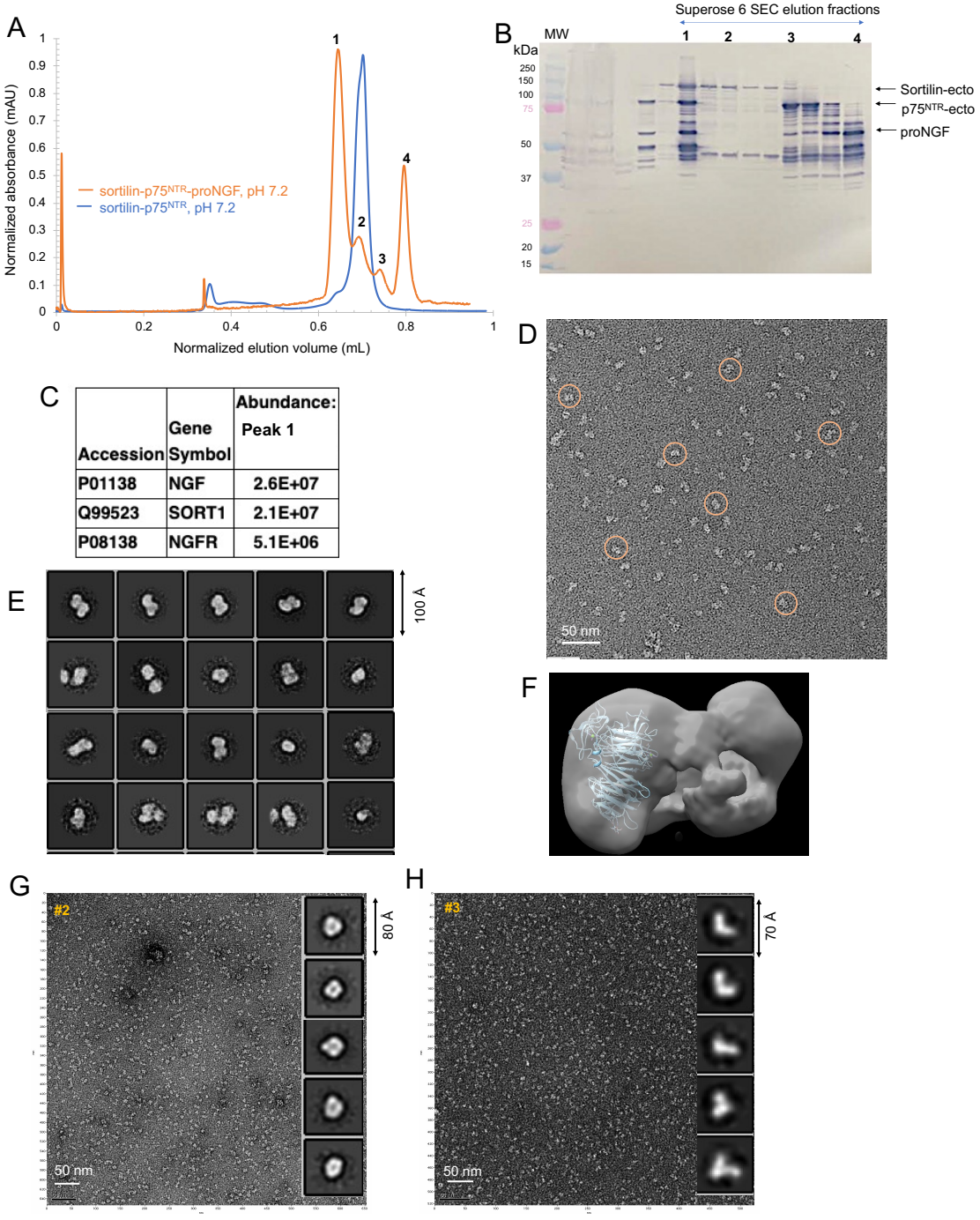


Figure 2-8. Biochemical and negative-stain EM analysis of proNGF-p75^{NTR}-sortilin complex formed by co-transfection in insect cells (*Sf9*). (A) Size exclusion

chromatography profiles of the co-infected protein complex proNGF-p75^{NTR}-sortilin overlayed on the control experiment lacking proNGF. The Western immunoblotting results using anti-His antibody indicate the presence of sortilin (120 kDa), p75^{NTR} (75 kDa) and proNGF (60 kDa) in the leading SEC peak 1 (indicated by the black arrows) (C) The mass spectrometry analysis confirms the presence of sortilin, proNGF (NGF family) and p75^{NTR} (NGFR family) in the leading peak 1. (D) A representative negative-stain EM micrograph from the SEC elution peak fractions 1 shows formation of the proNGF-p75^{NTR}-sortilin ternary complex, as indicated by orange circles (E) Representative two-dimensional (2D) classes performed using RELION (version 3.1)[51], using k = 50 classes, and 20,010 particle images showing dimers of sortilin in the complex. (F) The three-dimensional (3D) volume at 15 Å generated in RELION from the 2D class averages, with the fitted atomic coordinates of sortilin monomer (PDB *5NMR*)[148] in light blue. (G) A representative negative-stain EM micrograph from the SEC elution peak fractions 2 shows the unbound monomeric sortilin-ecto in peak 2, inset shows representative two-dimensional (2D) classes performed using RELION (version 3.1)[51], using k = 50 classes with 21034 particle images. (H) A representative negative-stain EM micrograph from the SEC elution peak fractions 3 shows the unbound p75^{NTR}-ecto in peak 3, inset shows representative two-dimensional (2D) classes performed using RELION (version 3.1)[51], using k = 50 classes with 18760 particle images.

The baculovirus system can impart all the necessary post-translational modifications necessary for proper protein folding of the target glycoproteins[192], hence the formed complex represents the native conditions. The p75^{NTR}-ecto and sortilin-ecto were non-glycosylated in the formed complex, as their molecular weights correlated to their monomeric weights in SDS-PAGE and Western immunoblotting results. In previous studies, the non-glycosylated or glycosylated p75^{NTR} behaved similarly to form the 2:2 heterocomplex with proNGF[141]. Thus, the role of glycans on p75^{NTR} or sortilin in imparting stability to this apoptotic ternary complex remains unclear.

proNGF on the other hand seems to be heavily glycosylated, as the molecular weight is upshifted to ~60 kDa, from the expected monomeric weight of 27 kDa of our construct. The proNGF purified from CHO cells was characterized to have two N-linked and two O-linked glycosylation[189], which may assist protein folding, cellular functions or assist the pro-domain interaction with sortilin[145]. The speculated glycosylation in the

baculovirus expressed proNGF in our assembly must be further verified for N-linked or O-linked glycosylation sites using enzymes like PNGase-F and sialidase which can remove the glycans.

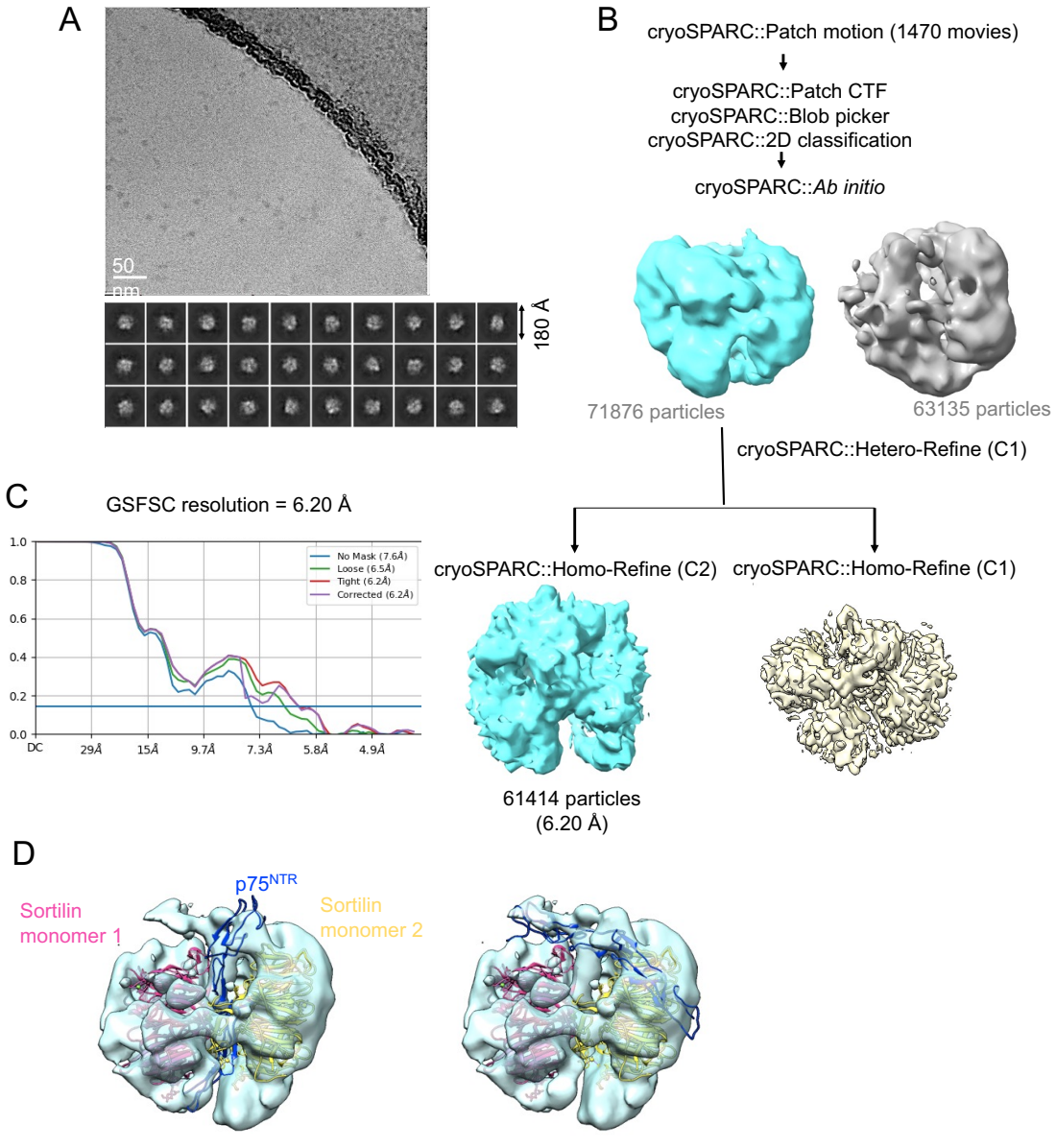


Figure 2-9. Cryo-EM structural analysis of the proNGF-p75^{NTR}-sortilin complex. (A) Electron micrograph of cryogenic preparation of the proNGF-p75^{NTR}-sortilin complex. Black contrast represents protein and white is background. Scale bar indicates 50 nm. Insets below show representative 2D class averages. Box side length is 180 Å. (B) Image processing flowchart for single-particle cryo-EM reconstruction using

cryoSPARC software (version 2.15)[52] (C) Fourier-shell correlation (FSC) plot of the 3D reconstructions. (D) The atomic coordinates of two sortilin monomer (PDB *5NMR*)[148] shown in pink and yellow and $p75^{NTR}$ monomer (PDB *1SG1*) shown in blue can be fitted in cryo-EM density.

Calcium ions increase the interactions of proNGF with sortilin almost 2.3 times as seen in cell-based assays, when the complexes were co-precipitated[141]. Supplementing the cell cultures with calcium ions in the physiological range of 1-2 mM were beneficial to induce interactions, which were otherwise absent in calcium depleted conditions[141]. Calcium ligation helps to stabilize some cadherins which have a well-known acidic pocket[193]. However, no such pocket is well-characterized in proNGF, though there are potential pockets detectable in the pro-region for proNGF[194], and exposed acidic loops in sortilin. In our studies, all the purification buffers were supplemented with 2mM calcium ions, and our data validated that the complex proNGF: $p75^{NTR}$: sortilin was stably formed. How calcium ions help the assembly is not clearly understood, and one possible reason is the stabilization of the formed complex during endosomal internalization as there are high influxes of calcium ions during compartmentalization. Further mutagenesis and functional studies will need to confirm how the calcium ions assist proNGF binding to sortilin.

2.4 Conclusions

There underlies a close interconnection between increased levels proNGF, $p75^{NTR}$, sortilin, and other extracellular proteases during aging or diseased neuron in Alzheimer's brain. Characterization of the expression patterns and receptor specificities would shed light on the major differences in growth factor activity under healthy and

pathophysiological conditions. The acquired structural information will thus help better understand apoptotic Alzheimer's disease pathway and enable potential drug designing.

The purpose of this study was to explore the binding interfaces that facilitates proNGF to form a complex with $p75^{NTR}$ and sortilin. We successfully concluded that the individual protein extracellular constructs of $p75^{NTR}$-ecto, sortilin-ecto and proNGF could be overexpressed and purified from the *Sf9* insect cells system. Further, co-transfection experiments in *Sf9* cells showed that longer incubation time allowed the proNGF ligands to interact with the neuronal receptors at a favorable stoichiometry and form the proNGF-$p75^{NTR}$-sortilin heterocomplex which could be purified for the first time.

The insect cells facilitated rich glycosylation and proper post-translational modification of these target glycoproteins, which rendered them suitable for structural studies. Our data suggests that the proNGF-$p75^{NTR}$-sortilin heterocomplex is stabilized by cross-braced dimers of sortilin. However, the stability of the complex was poor and sample heterogeneity prevented preparation of optimal cryo-EM sample and precise atomic modelling using the reconstructed 3D volume. For future cryo-EM studies, the expression and purification of the complex will be further improved, to retrieve high concentration of the complex. Obtaining a model of the proNGF-$p75^{NTR}$-sortilin complex will help to understand the interaction and binding characteristics of this proteins.

2.5 Acknowledgements

We thank Dr. Dewight Williams for assisting in cryo-electron microscopy data-collection and acknowledge the use of the Titan Krios cryo-TEM, in the Eyring Materials Center (EMC), at Arizona State University (ASU) supported from grant number NSF MRI 1531991. We acknowledge the support and resources from the Life Sciences Electron

Microscopy Lab for the use of the electron microscope Philips CM12, at ASU and thank Dr. David Lowry for assistance. We acknowledge the support and resources of the DNASU plasmid repository, part of the Biosciences Core Facilities at Biodesign Institute, ASU since we purchased a plasmid for the studies.

CHAPTER 3

STRUCTURAL AND FUNCTIONAL ANALYSIS OF DISEASE-LINKED p97 ATPASE MUTANT COMPLEXES

This work has been published in Nandi, P.[#]; Li, S.[#]; Columbres, R.C.A.; Wang, F.; Williams, D.R.; Poh, Y.-P.; Chou, T.-F.[*]; Chiu, P.-L.[*] Structural and Functional Analysis of Disease-Linked p97 ATPase Mutant Complexes. *Int. J. Mol. Sci.* **2021**, *22*, 8079 and this chapter is adapted from the manuscript.

3.1 Introduction

3.1.1 p97 in Dementia-related Diseases

Human p97/VCP (valosin-containing protein) belongs to the type II AAA+ (ATPase Associated with various cellular Activities) protein family, which has two AAA+ ATPase domains[195,196]. The orthologs are known as Cdc48 in *Saccharomyces cerevisiae* and transitional endoplasmic reticulum (TER) ATPase in archaebacteria and eukaryotes. p97 acts as a molecular hub, interacting with various cofactors to perform a wide variety of cellular functions, including autophagy, cell-cycle regulation, ubiquitin-dependent proteostasis and reassembly of Golgi and nuclear membranes[76,197-200]. p97 is abundant in the cytosol, comprising 1% of the cytosolic proteins[201,202], and carries out ATP hydrolysis to gain energy to fuel the conformational change underlying its activities[203-206]. Abnormal functions or uncontrolled regulations of p97 can cause serious diseases. Single amino acid mutations in human p97 have been linked to diseases of aging and neurodegeneration, including IBMPFD (inclusion body myopathy with Paget's disease of bone and frontotemporal dementia) and familial amyotrophic lateral sclerosis (ALS)[82,104,207,208].

3.1.2 p97 Domain Architecture

p97 comprises an N-terminal domain (NTD), two ATPase domains, D1 and D2, containing conserved Walker A and B motifs, and an unstructured C-terminal tail[196] (Figure 3-1). The NTD and C-terminal tail are the key domains to regulating ATPase function[96,203,209]. Six p97 monomers form a functional complex, featuring a double-ring structure based on the association of D1 and D2. The D1 domain is responsible for p97

oligomerization regardless of their nucleotide-binding states. Upon ATP binding to the D1 domain [97,210,211], the D2 domain is the primary mediator of ATP hydrolysis[97,105,212-215]. The two stacked rings create a central channel in the p97 hexamer that binds substrates, facilitating the segregation of substrates destined for proteolysis from their usual location in multi-subunit complexes, organellar membranes and chromatins[107,216]. Previous cryogenic electron microscopic (cryo-EM) structures of the AAA+ ATPases revealed that the substrate processing by the Cdc48 ATPase results in a broken ring and an asymmetric structure of the six p97 monomers. These structures have been used to suggest a hand-over-hand model for the substrate's induced motion[217-219]. The substrate movement along the central channel is dependent on the D1 nucleotide state and is mainly driven by the ATP hydrolysis on the D2 domain[101,220]. This movement would call for cooperation between domains within the AAA+ ATPase complex and cofactor bindings[203,221,222]. However, it is not certain whether the hand-over-hand model applies to other p97 functions, such as membrane fusion.

The p97 NTD mediates the binding of various cofactors and ubiquitylated substrates and modulates ATPase activity[205,223]. Although the p97 NTD does not directly relate to nucleotide binding, it has two distinct configurations that are affected by the D1 nucleotide-binding states: the 'up' (ATP-bound D1) configuration, rising above the plane of the D1 ring, and the 'down' (ADP-bound D1) configuration, coplanar with the D1 ring[214,215,224]. However, previous EM structural analysis revealed variant NTD positions, rather than the only two distinct states[81,96]. One hypothesis suggests that the NTD configuration affects linker conformation, initiating inter-subunit or intra-domain communication, thus influencing ATPase activity[225-227]. This is supported by the position

of IBMPFD-associated disease mutants that result in elevated ATPase activities, mostly from the D2 ATPase[96,228]. The IBMPFD mutations cluster on the NTD or the N-D1 linker of the p97[85,214]. Because the mutation sites are far away from the D2 nucleotide-binding site, it has been suggested that the NTD mutation changes the interactions between p97 inter-domains, leading to functional alterations[101]. However, the connection between the pathological effects of the IBMPFD mutations and elevated p97 ATPase activity remains unclear.

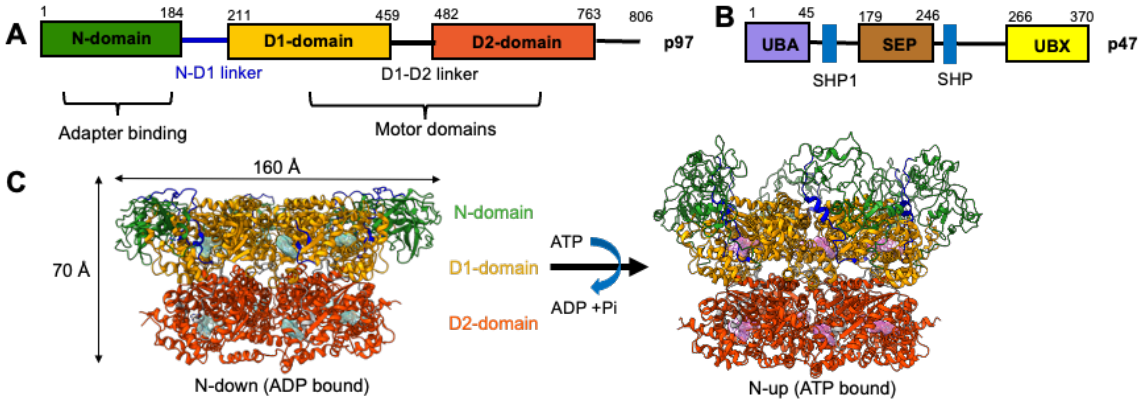


Figure 3-1. Domain architecture of p97 and p47. (A) Domain organization of p97 and (B) its co-factor p47. The cofactor p47 is known to interact via the UBX domain (in yellow) with p97 N-domains. (C) The N-domains of p97 can adopt an down-configuration (ADP-bound) or up-configuration (ATP-bound) based on its nucleotide–binding state and the molecular motions of p97 may influence p47 binding. ADP molecules shown in light blue density and ATP molecules shown in pink density in the D1 and D2 nucleotide binding pockets of p97.

In addition to mutations, the binding of p97 cofactors on the NTDs also impacts the ATPase activities. p47 is one of the p97 cofactors involved in p97-mediated membrane fusion[76]. It contains an N-terminal UBA (ubiquitin-associated) domain, followed by a SEP (shp1, eyc, and p47) and UBX (ubiquitin-regulatory X) domain, which interacts with the p97 NTD (Figure 3-1A and Figure 3-1B)[105,226,229]. The interaction reduces the D1 ATPase activities of both wild-type and p97 mutants[94]. Also,

p47 enhances the D2 ATPase activity of wild type p97 ($p97^{WT}$), whereas it reduces the activities of the p97 disease mutants[94]. Thus, p47 has different effects on the two ATPase domains of both $p97^{WT}$ and its mutants, resulting in that the mutant p97-p47 complex blocks the ATP hydrolysis in one single sigmoidal phase, rather than a biphasic response of the wild-type p97-p47 complex[94].

3.1.3 Existing p97-p47 Structures

The binding mode of the p47 UBX domain to the p97 NTD is conserved among the UBX or UBX-like (UBX-L) containing cofactors[230]. Previous structures of partial or complete complexes reveal this interaction as seen in $p97^{N\text{-}D1}$-$p47^{UBX}$, $p97^{N}$-$FAF1^{UBX}$, $p97^{N}$-$OTU1^{UBXL}$, and p97-ASPL[105,231-235]. The high-resolution crystal structure of the $p97^{N\text{-}D1}$-$p47^{UBX}$ truncated complex showed that a conserved β sheet with an S3/S4 loop in the UBX domain binds to a hydrophobic pocket formed by the two p97 N-subdomains (comprised of an N_a (four-stranded β barrel) and an N_b (ψ barrel) domain)[105]. Recently, the bipartite and tripartite binding modes have been proposed for the interactions between multiple p47 and wild type p97 NTDs[235]. However, how this binding in the combination of disease NTD mutations modulates the D1 and D2 ATPase activities remain uncertain. Most of the current data for p97 mutants were based on the structures of truncated domains[99]. Because the structural information of the full-length diseased complexes is still lacking, more efforts on studying their functions and structures are needed to understand the mechanistic details.

Here we performed functional and structural analyses on the disease-linked complex, $p97^{R155H}$-p47, to understand how the cofactor binding and NTD mutation leads

to functional alterations. We chose $p97^{R155H}$ as our target, since the NTD mutation of arginine 155 to histidine (R155H) is the most commonly found in the IBMPFD phenotype[86]. We have previously identified that p97 disease mutants are defective in the cofactor-regulated ATPase hydrolysis cycle[94]. Also, the increased p47 expression can correct the autophagy defect caused by the p97 disease mutant[94]. To understand how p47 impacts the function of the $p97^{R155H}$ disease mutant, we characterized the interactions between the $p97^{R155H}$ and p47 using single-particle cryo-EM image analysis. The structural results revealed two major forms in the molecular population: the $p97^{R155H}$ dodecamer and the $p97^{R155H}$ hexamer-p47 complex. The nucleotide-binding pockets are empty in the $p97^{R155H}$ dodecamer. The presence of the nucleotides, such as ADP or ATPγS, destabilizes the dodecameric formation. The NTDs showed all up configurations when bound to ATPγS or in the absence of nucleotides. However, in the presence of ADP, the NTDs of the mutant allow variable intermediate positions, which may affect the accessibility of the p47 cofactor. Our structures also showed that the arginine fingers exhibit different conformations than those of the wild type when $p97^{R155H}$ binds to p47. We propose a possible model that impacts functional alteration of the $p97^{R155H}$ upon p47 binding, providing an insight into the pathological function of the IBMPFD-associated p97 mutants.

3.2 Materials and Methods

3.2.1 Overexpression of the Wild-type p97, $p97^{R155H}$ Disease-mutant, and p47 Proteins

Plasmid constructs used for generating p97 and its mutants and p47 proteins are listed in Table 3-1. Overexpression and purification of the p97 and its mutants and p47 followed previous methods[236].

Plasmid Number	Plasmid name	Vector	Source and Reference
TCB-328	Rat p47 pET15_T	pET15b_TEV linker	Chou et al. 2014
TCB-197	Human p97 pET15_T	pET15b_TEV linker	Chou et al. 2014
TCB-210	Human R155H p97 pET15_T	pET15b_TEV linker	Chou et al. 2014
TCB-423	Human R155H, K251R p97 pET15_T	pET15b_TEV linker	This study
TCB-436	Human R155H, K524R p97 pET15_T	pET15b_TEV linker	This study
TCB-460	Human R155H, R359A p97pET15T_T	pET15b_TEV linker	This study
TCB-461	Human R155H, R635A p97pET15T_T	pET15b_TEV linker	This study

Table 3-1. List of plasmids used for overexpression of p97 and p47 proteins[236]

3.2.2 ATPase Activity Measurements

Detection of ATPase activity using Biomol Green reagent (Enzo Life Sciences, Farmingdale, NY) was performed as previously described with slight modifications[94]. To compare the ATPase activity of wild type and mutant p97 proteins, each purified protein was diluted to a final monomer concentration of 25 nM in 50 µL ATPase assay buffer (50 mM Tris (pH 7.4), 20 mM $MgCl_2$, 1 mM EDTA, 0.5 mM TCEP, 0.01% Triton X-100 and 80 nM BSA). After adding 200 µM ATP, the reaction was carried out at room temperature for optimal reaction times. 50 µL Biomol Green reagent (Enzo Life Sciences Inc., Farmingdale, NY) was added to stop the reaction and the absorbance at 635 nm was

measured using a BioTek Synergy Neo 2 plate reader (BioTek, Winooski, VT). The eleven-dose titrations with p47 cofactor were performed by adding the varying amount of p47 protein in ATPase buffer with p97. The results were calculated from five replicates using GraphPad Prism 7.0.

3.2.3 p47 Binding Affinity Measurements

The binding affinity (K_d values) between p47 and $p97^{WT}$ or p97 mutants was determined by measuring the temperature-related intensity change (TRIC) signals using a Dianthus NT.23 instrument (Nano-Temper Technologies, München, Germany). p47 was labeled with a RED-NHS dye using a Monolith Protein Labeling Kit RED-NHS second-generation (Nano-Temper Technologies, CAT# LO-L011). The full-length wild type or mutant p97 was titrated against 10 nM of p47-NHS in two-fold steps from 2.8 μM to 1.37 nM in 20 μL working buffer (20 mM HEPES (pH 7.4), 150 mM KCl, 1 mM $MgCl_2$, 5% (w/v) glycerol and 0.0025% (v/v) Tween 20). Assays were performed in a Dianthus 384-well plate and reproduced in three independent experiments. Data from the three independent measurements were fitted using non-linear regression analysis in Prism 7.0.

3.2.4 Assembling p97-p47 Complexes

To assemble p47 with $p97^{WT}$ or p97 mutants, 1.67 μM p97 hexamer was mixed with 80 μM p47 for 10 minutes and fractionated in 20 mM HEPES (pH 7.4), 150 mM KCl, 1 mM $MgCl_2$ and 5% glycerol. Three different assemblies were prepared in the absence of nucleotides and the presence of 100 μM ADP or ATPγS in the SEC elution buffer. The mixture of the p97-p47 complex was loaded onto a Superdex 200 10/300 GL (GE Healthcare, Chicago, IL) for size-exclusion chromatography (SEC). When the samples

were used for cryo-EM analysis, 0.25% (v/v) glutaraldehyde (Electron Microscopy Sciences, Hatfield, PA) was applied for on-column cross-linking by following the previous method[16]. Peak fractions were characterized using SDS-PAGE, Western blotting and negative-stain electron microscopy (EM).

3.2.5 Negative-stain Electron Microscopy for Single-particle Analysis

The negatively stained specimens were prepared using the previous method[15]. 0.01 mg/mL of the protein sample was used to applied onto a continuous carbon-coated copper grid. The specimen was imaged using a Philips CM12 transmission electron microscope (TEM) (80 keV) with a side-mounted CCD camera (Model 791, Gatan, Pleasanton, CA) or a FEI Tecnai TF20 TEM with a CCD camera.

For imaging the Walker A mutants, the specimens were imaged under a FEI Tecnai TF20 TEM, recording at a pixel size of 1.04 Å/pixel at the specimen level. 153 and 144 electron images were collected for $p97^{R155H\text{-}K251R}$ and $p97^{R155H\text{-}K524R}$, respectively, and imported into Relion (version 3.1-beta-commit-ca101f)[51] for image processing. 23,342 and 23,020 particles of $p97^{R155H\text{-}K251R}$ and $p97^{R155H\text{-}K524R}$ were manually selected from the electron images, respectively, and the two-dimensional (2D) class averages with an assigned k of 50 were calculated. For the $p97^{R155H\text{-}K251R}$ dataset, the numbers of side views for dodecamer and hexamer were 2,173 and 313, respectively. For the $p97^{R155H\text{-}K524R}$ dataset, the numbers are 2,007 and 300, respectively.

3.2.6 Cryo-EM Data Collection

A holey-carbon C-flat grid (2/1 4C; Protochips, Morrisville, NC) was glow-discharged for 15 seconds using a Pelco easiGlow glow-discharge system (Ted Pella, Redding, CA).

6 μL protein sample was applied to the pretreated grid, and the excess solution was blotted away using a filter paper (retention 20 μm) (Product #:47000-100, Ted Pella, Redding, CA). The grid specimen was quickly plunge frozen into liquid ethane using a Thermo Fisher/FEI Vitrobot Mark IV automated freeze plunger (Thermo Fisher/FEI, Hillsborough, OR) for 6 seconds in a chamber with a humidity of 100%. Particle homogeneity and ice backgrounds of the grid specimen were screened using a FEI Tecnai TF20 TEM. Grids with thin ice and a homogenous particle dispersion were used for subsequent cryo-EM data collection.

All the cryo-EM data in this report were collected at the Eyring Materials Center (EMC) at Arizona State University (ASU) (Tempe, AZ) using a Thermo Fisher/FEI Titan Krios TEM (Thermo Fisher/FEI, Hillsborough, OR) at an accelerating voltage of 300 keV. Cryo-EM movies were recorded using a Gatan K2 Summit direct electron detector (DED) camera (Gatan, Pleasanton, CA). Defocus range was set to -0.8 to -2.5 μm. Nominal magnification was 48,077X, resulting in a physical pixel size of 1.04 Å/pixel at the specimen level. The movie data was recorded at a counted rate of 2 e^-/sub-pixel/sec and a sub-frame rate of 200 msec in super-resolution mode. Total exposure was 6 seconds, accumulating to a dosage of 44.4 e^-/A^2. The beam-image shift was applied to accelerate data acquisition[168]. Data collection was automated using the customized SerialEM macros (version 3.9)[169].

3.2.7 Image Processing

For the $p97^{R155H}$-p47 assembly, 4,223 collected movies were unpacked and gain-normalized using Relion (version 3.1-beta-commit-ca101f)[51]. Image frames of a movie

were translationally registered and averaged with a dose-weighting scheme. The final frame average was Fourier-cropped at the spatial frequency of 1.04 $Å^{-1}$. The defocus and astigmatism of the images were estimated using the CTFFIND4 program (version 4.1.13)[47]. A few particles were manually selected from the images, and their average was then served as a searching template for the subsequent automated particle picking in Relion[51]. 368,575 selected particle images were automatically selected, and the data curation was completed using iterative 2D unsupervised classification. The classes with discernible features were selected for *ab initio* volume generation using cryoSPARC software (version 2.15)[52]. The two three-dimensional (3D) densities of a $p97^{R155H}$ dodecamer and a $p97^{R155H}$-p47 assembly were generated.

For particle images of the $p97^{R155H}$ dodecamer, one additional round of 3D classification was performed to remove poorly aligned particle images. The selected 3D class average was then refined against 64,252 experimental particle images with enforcing a D6 symmetry. The density was further improved using CTF refinement (fit of the defocus and astigmatism per particle and estimation of beam tilt)[237] and Bayesian polishing[238] procedures to a final resolution of 3.34 Å. The final map was sharpened using a *b*-factor of -78.6 $Å^2$.

For the $p97^{R155H}$-p47 complex particle images, the consensus 3D volume of the $p97^{R155H}$-p47 complex was calculated at 4.30 Å resolution with a sharpened *b*-factor of -125.7 $Å^2$. The C6 symmetry was applied to further improve the densities of the D1 and D2 domains. The generated map reached a final resolution of 3.98 Å and was sharpened using a *b*-factor of -139.3 $Å^2$. The final resolutions of all the density maps were

determined using the golden FSC criteria at 0.143 cutoff[239], and the local resolution estimations were performed using the implementation in Relion software. The processing schematic is shown in Figure 3-3. On the other hand, further 3D classification procedures with k = 6 or larger or multiple hierarchical layers were attempted (Figure 3-3B). However, no major structural differences between the generated class averages were discernible, and no improvements were shown in the density quality in local regions of NTDs and p47 (Figure 3-3B).

For the $p97^{R155H}|_{ADP}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 datasets, the image processing was generally conducted using cryoSPARC software (version 2.15)[52]. 3,512 and 2,796 movies of the $p97^{R155H}|_{ADP}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 were unpacked, gain normalized and imported into cryoSPARC, respectively. Frame registration and averaging were performed using patch motion correction. Defocus and astigmatism parameters were estimated using patch CTF estimation. Particle locations were automatically selected using the Topaz program (version 0.2.3)[170]. For the $p97^{R155H}|_{ADP}$-p47 complex, 1,124,232 particle images were selected and curated using iterative 2D unsupervised image classification. The classes with discernible features were selected for *ab initio* volume generation. The two densities of a $p97^{R155H}$ dodecamer and a $p97^{R155H}|_{ADP}$-p47 assembly were generated. 2,219 and 160,050 particle images were used to calculate the final volumes of the $p97^{R155H}$ dodecamer and the $p97^{R155H}|_{ADP}$-p47 complex, respectively. The two volumes were refined against their experimental particle images and reached resolutions of 6.10 Å ($p97^{R155H}$ dodecamer) and 4.50 Å ($p97^{R155H}|_{ADP}$-p47). Like the $p97^{R155H}$-p47 dataset, further 3D classification did not yield discernible features in the local regions for NTDs and p47. Processing schematic and local resolution estimation are

shown in Figure 3-10. For the $p97^{R155H}|_{ATP\gamma S}$-p47 dataset, we reconstruct the volume followed the same procedure as above. Final 63,353 particle images were used to reconstruct a 3D volume at 4.23 Å resolution, sharpened using a *b*-factor of -77.6 $Å^2$. Processing schematic and local resolution estimation are shown in Figure 3-11.

Structural heterogeneity of the three $p97^{R155H}$-p47 datasets was further analyzed using a deep coordinate neural network by the cryoDRGN program (version 0.3)[240]. Before latent encoding, the particle images were Fourier-cropped to a box size corresponding to pixel sizes of 3.90 and 3.71 Å/pixel for $p97^{R155H}$-p47 and $p97^{R155H}|_{ADP}$-p47 dataset, respectively. For each group, an eight-dimensional latent variable model was trained for 25 epochs. The encoder and decoder architectures were three layers with 1,024 nodes. The latent encodings were visualized using a principal component factor plot or UMAP (uniform manifold approximation and projection) representation. After training, *k*-means clustering with $k = 20$ was conducted on the latent encodings, and reconstructions were calculated at the cluster centers using the decoder network.

3.2.8 Model Building

The previous p97 and p47 coordinates (PDB code: *5FTK* and *1S3S*) were used as the templates for atomic modeling or molecular docking[105,224]. The templates were first docked into individual cryo-EM densities using the 'Fit in the Map' function in UCSF Chimera (version 1.14)[171]. The fitted model was manually rebuilt using Coot (version 0.9.1)[10,11] and then refined against the cryo-EM densities using the 'phenix.real_space_refine' program in Phenix software package (version 1.18.2-3874)[9]. Hydrogen atoms were added using the 'phenix.reduce' program for the model refinement

and removed after the refinement. Secondary structure restraints were applied during the model refinement. The refinement and validation statistics were listed in Table 1 (Appendix section). The figures for the cryo-EM density maps and atomic models were prepared using UCSF Chimera and ChimeraX (version 0.91)[172].

3.3 Results and Discussion

3.3.1 $p97^{R155H}$-p47 Assembly Leads to Two Distinct Molecular Populations

To analyze the activity of mutated p97, we first assembled an *in vitro* complex of full-length $p97^{R155H}$ with p47 in the absence of nucleotides (molar ratio of $p97^{R155H}$ hexamer: p47 monomer 1:10). A peak representing the $p97^{R155H}$-p47 complex, corresponding to a molecular weight of about 670 kDa, was detected in the size-exclusion chromatographic (SEC) profile (Figure 3-2B). The peak profile showed a leading shoulder, implying that multiple molecular species were present in the peak fraction. We further characterized the peak fractions using SDS-PAGE, which confirmed the presence of both $p97^{R155H}$ and p47 proteins in the main peak fraction and only the $p97^{R155H}$ proteins in the leading shoulder (Figure 3-2C). This observation corroborates the previous report on the IBMPFD mutants, which showed that a full-length $p97^{R155H}$ dodecamer was present in the leading fraction [19]. In contrast, the $p97^{WT}$-p47 assembly did not show an asymmetric leading peak profile (Figure 3-2B). Thus, the $p97^{R155H}$ mutant is found in additional forms, forming complexes beyond the hexameric arrangement of the wild-type protein. Although the higher-order form has been reported for the wild type[96], inhibitor-bound[241,242] or engineered p97[243], we did not find the same pattern in the SEC of the wild type, possibly that the high-order complex was in a less proportion in the wild type that could not be detected. In the SDS-PAGE analysis, these higher-order complexes do not

associate with p47, even in the presence of excess p47 (Figure 3-2B). This suggests that the higher-order arrangement of $p97^{R155H}$ may hinder the interaction with the p47 cofactor.

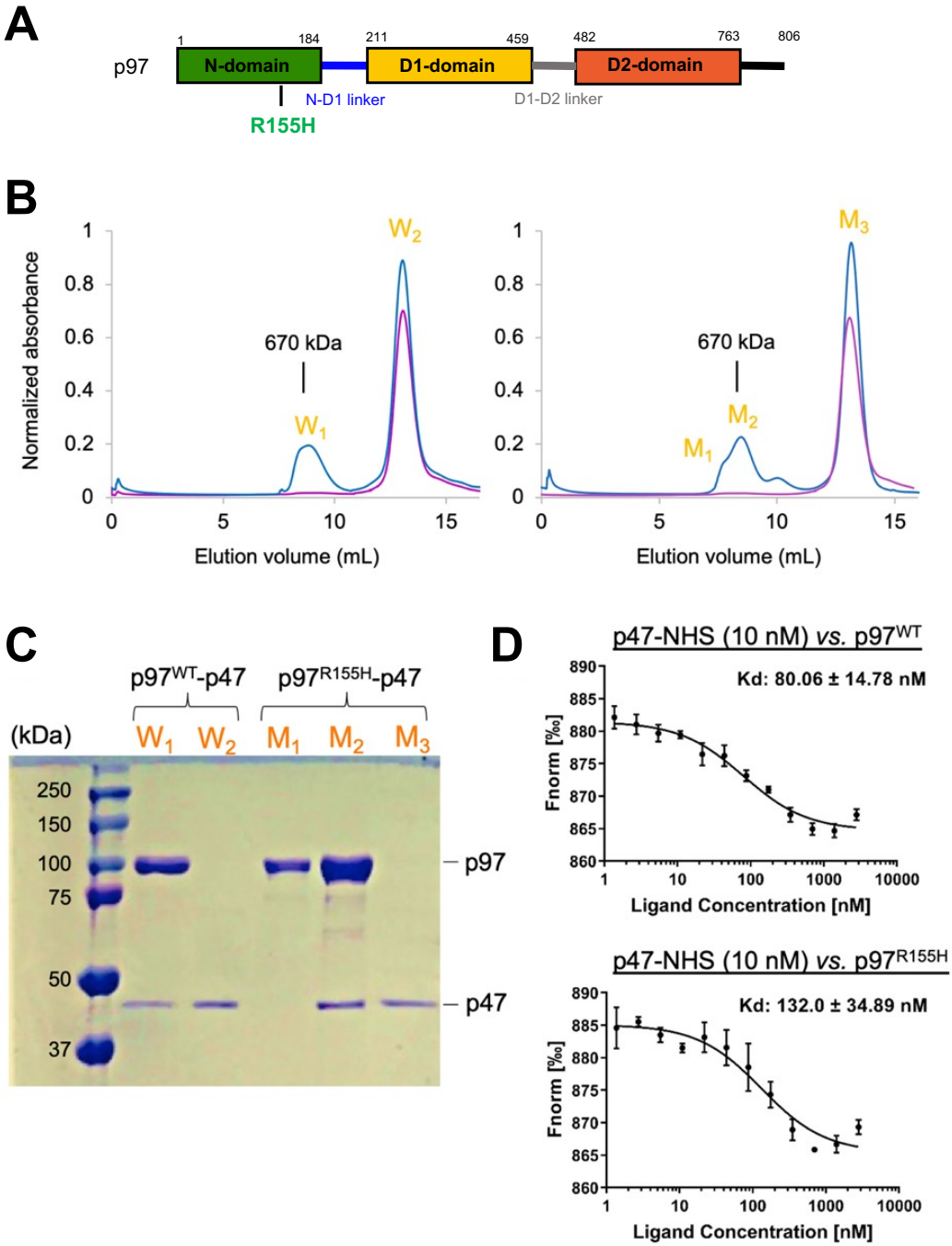


Figure 3-2. Characterization of the p97^{R155H}-p47 and p97^{WT}-p47 complex formation. (A). The disease-linked mutation of R155H is labeled on the NTD of p97. (B) Size-exclusion chromatographic (SEC) profiles of the p97^{WT}-p47 (left) and p97^{R155H}-p47 (right) assemblies. Blue and purple curves are for the p97-p47 assembly and p47 protein alone. Peak fractions are labeled. (C) SDS-PAGE analysis of the SEC eluted peaks of the p97-p47 assemblies. (D) Determination of dissociation constants, K_d's, for the p97^{WT}-p47 and p97^{R155H}-p47 complexes. Unlabeled p97 was titrated against 10 nM of RED-NHS-labeled p47 in two-fold steps from 2.80 μM to 1.37 nM. The temperature-related intensity change (TRIC) signals were recorded and plotted against p47 concentrations.

If the high order p97^{R155H} does not interact with p47 and forms in a larger proportion than the wild type, the p47 binding affinity to the p97^{R155H} on average is lower than that of the wild type p97^{WT}. To test this idea, we further employed fluorescent labeling and measured the temperature-related intensity change (TRIC) signal to determine the affinities of the p47 binding on p97^{WT} and p97^{R155H} and their dissociation constants, K_d. In the absence of nucleotides, the resulting K_d averages of p97^{WT}-p47 and p97^{R155H}-p47 are 80 nM and 132 nM, respectively, suggesting a slightly higher binding affinity for the wild type (Figure 3-2D and Table 3-2).

K_d (μM)	WT	R155H	R155H-R359A	R155H-R635A
1	0.133	0.186	0.197	0.205
2	0.083	0.096	0.220	0.237
3	0.074	0.138	0.173	0.186
Average ± SD	0.097±0.032	0.140±0.045	0.197±0.024	0.209±0.026

Table 3-2. Kinetic studies of the interaction between p97 and p47 proteins.

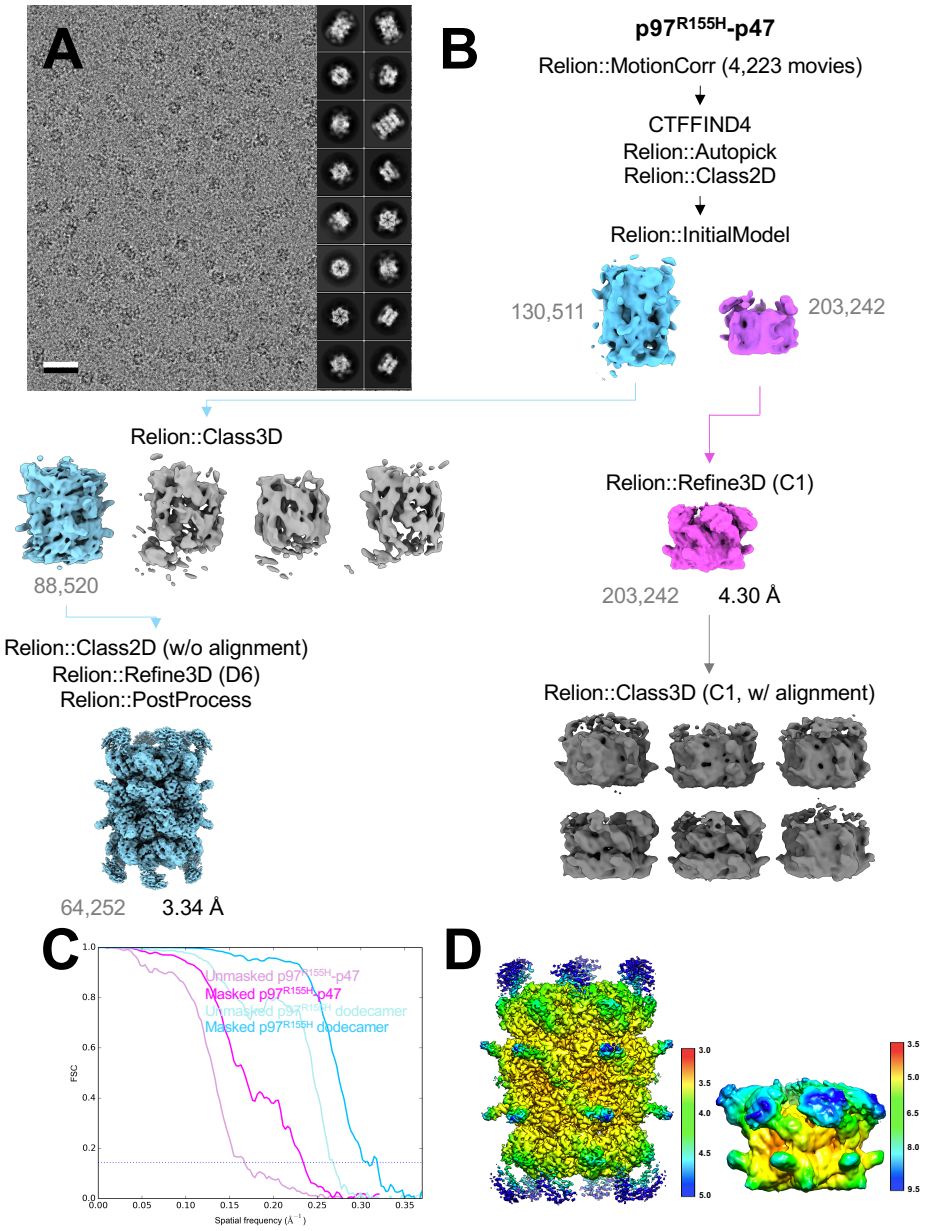


Figure 3-3. Cryo-EM analysis of the full-length p97^{R155H}-p47 assembly in the absence of nucleotides. (A) Electron micrograph of cryogenic p97^{R155H}-p47 complexes. Black contrast represents protein, and white is background. Scale bar indicates 50 nm. The inset to the side shows representative 2D class averages. Box side length is 374 Å. (B) Image processing flowchart for single-particle cryo-EM reconstruction. (C) Fourier-shell correlation (FSC) plot of the 3D reconstructions. (D) Local resolution estimation of the reconstructions.

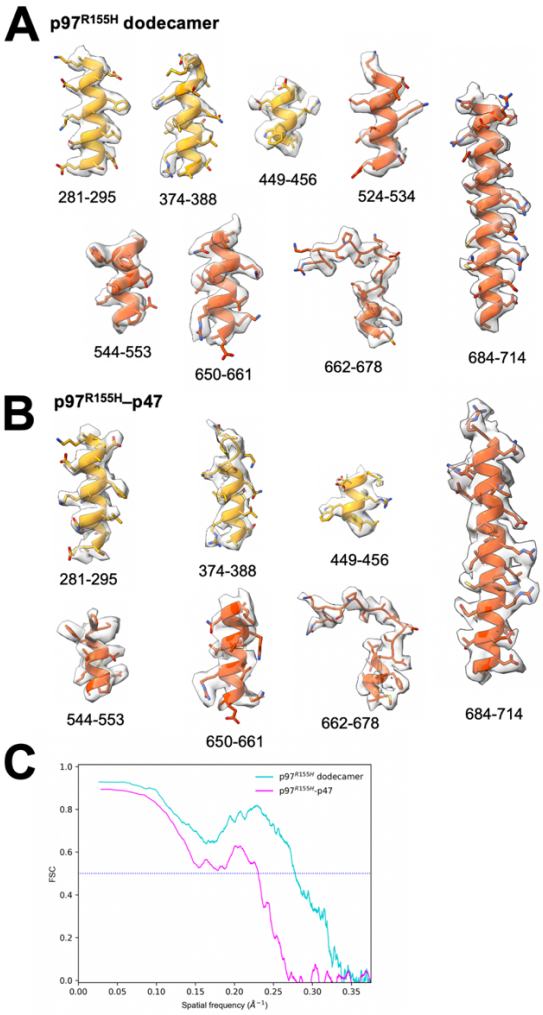


Figure 3-4. Modelling of atomic coordinates in the cryo-EM densities of the p97^{R155H} dodecamer and p97^{R155H}-p47 complexes in the absence of nucleotide. (A) Model fitting of the p97^{R155H} dodecamer. Orange and orange red are D1 and D2 domains, respectively. (B) Model fitting of the p97^{R155H}-p47 assembly. (C) Fourier-shell correlation (FSC) plot. Light blue and pink curves are p97^{R155H} dodecamer and p97^{R155H}-p47 assembly. Blue dash indicates where FSC values are 0.5.

3.3.2 Up NTDs of the p97^{R155H} mutant and p47 are structurally disordered

To visualize the interaction between the p97^{R155H} mutant and the p47 cofactor, we plunge-froze the sample of the p97^{R155H}-p47 assembly for single-particle cryo-EM analysis. Notably, two different particle populations were discernible in the two-dimensional (2D) class averages (Figure 3-3A). One showed a typical side view of the

p97 hexameric form, and the other showed two stacked p97 hexameric double rings (Figure 3-3A). *Ab initio* map generation produced two different three-dimensional (3D) densities presenting one with general p97 hexameric features and the other containing a dimer of hexamers (Figure 3-3B). We reconstructed the 3D density maps of the $p97^{R155H}$ dodecamer complex at 3.34 Å resolution and the $p97^{R155H}$-p47 complex at 4.30 Å (Figures 3-5, 3-3B and 3-3C). The ratio of the single-particle images of the $p97^{R155H}$-p47 complex to the $p97^{R155H}$ dodecamer is 1.56:1.

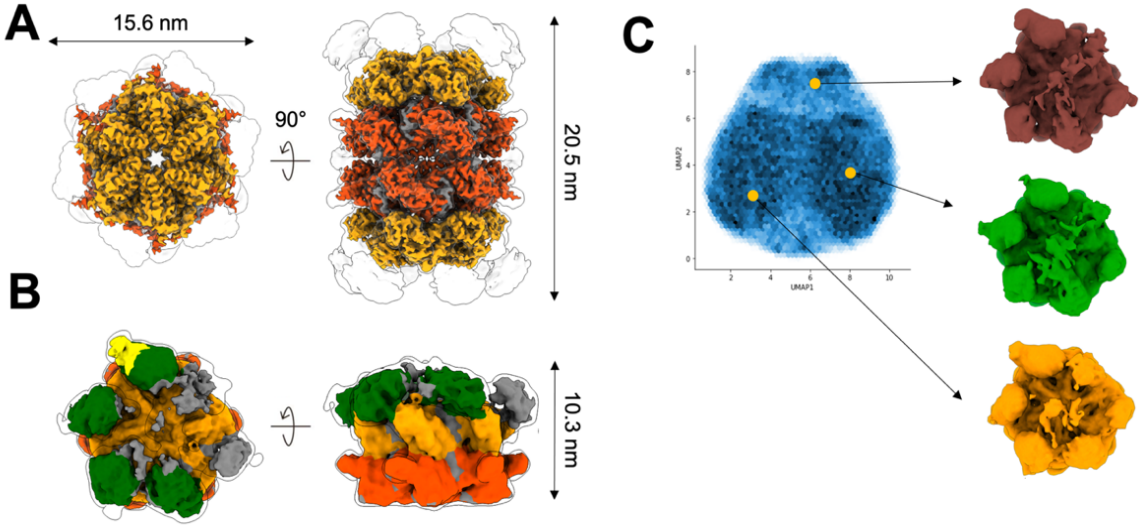


Figure 3-5. Cryo-EM reconstructions of the $p97^{R155H}$ dodecamer and the $p97^{R155H}$-p47 complex. Cryo-EM densities of (A) $p97^{R155H}$ dodecamer (5.6**σ**) and (B) $p97^{R155H}$-p47 complex (2.3**σ**). Green, orange, orange-red and yellow highlight the NTD, D1 and D2 of the $p97^{R155H}$ and $p47^{UBX}$, respectively. Grey color indicates regions where densities could not be assigned. Envelopes are the cryo-EM maps at a lower contour ($p97^{R155H}$ dodecamer: 2.4**σ**; $p97^{R155H}$-p47: 1.6**σ**). (C) UMAP (uniform manifold approximation and projection) representation of a generative neural network analysis of the $p97^{R155H}$-p47 cryo-EM density. Blue intensities represent the distribution of particle images in the embedded latent space. Three different 3D averages (brown, green and light orange) were calculated from the chosen clusters (orange circles).

Local resolution estimation showed that the density of the $p97^{R155H}$ dodecamer or the $p97^{R155H}$-p47 complex is well resolved in D1 and D2 domains but not the NTDs (Figure 3-4D). For $p97^{R155H}$ dodecamer, we did not find densities that could be assigned to p47. Based on specific map features of the $p97^{R155H}$-p47 complex, we were able to identify the NTD and $p47^{UBX}$ domains at a lower contour level (2.7σ) by docking available atomic models (PDB *1S3S*) into the density map (Figure 3-5B)[105]. Four NTD densities could be assigned and shown in up configurations, and they are rendered into slightly different orientations in the azimuthal direction along the C6 pseudo-symmetrical axis (Figure 3-5B). Only one $p47^{UBX}$ domain was found to fit into the density attached to the up NTD (Figure 3-5B). We did not identify any high-order form of the p47 bound to p97, and it is consistent with the previous finding that the p97 NTDs interfere with p47 oligomerization[235,244]. Also, the flattened densities of the NTD of p97 and p47 are possible due to their dynamic and partially disordered nature, leading to the weak scattering signals and limiting resolution of these components (Figures 3-5B and 4-3D). We also performed image classification in 3D, but the result did not yield high-resolution densities for modeling the SEP or UBA domain of the p47 (Figure 3-3D). This observation is also corroborated by a neural network data analysis[240], showing a high variation in the NTD and p47 densities (Figure 3-5C). The densities of D1 and D2 rings were invariable across the reconstructions. Unlike the $p97^{WT}$-p47 complex[235], the mutant NTD may be too disordered or mobile to present a structural feature with the p47 factor to be resolved by cryo-EM.

In this complex assembling, we did not supply any nucleotides in the purification steps, and nucleotide was not detectable in the D1 and D2 nucleotide-binding sites of the

$p97^{R155H}$ dodecamer or the $p97^{R155H}$-p47 cryo-EM density maps. This observation differs from the previous study where ADP is always observed in the wild type p97 D1 ATPase, even when the nucleotide was not supplied in the sample buffers[113,212,214,236,245]. It is the first time that the full-length $p97^{R155H}$ disease mutant structure is revealed by single-particle cryo-EM. One possibility could be that the nucleotide binding for the $p97^{R155H}$ mutant is weak in the extremely low concentrations of nucleotides, so that the densities of the nucleotides were insufficient to be detected in the averages of the cryo-EM images. Since the previous analyses were conducted using $p97^{WT}$ or truncated p97 mutants, it is likely that the R155H mutation or the domain-domain interactions in the full-length $p97^{R155H}$ reduces the affinity of the D1 domain for ADP[94,246]. Also, because the ADP off-rate of the $p97^{R155H}$ D1 ATPase is two-fold faster than that of the $p97^{WT}$ D1 ATPase and the nucleotide binding in the D2 domain depends on the D1 nucleotide state, it indicates a high likelihood of the ADP binding instability in D1 ATPase of the full-length $p97^{R155H}$[246]. Apart from this, the previous SPR experiments showed that the R155H mutation reduces the ADP-binding affinity to p97[236,247]. Thus, the $p97^{R155H}$ mutant is likely to be structurally stable when the nucleotide-binding pockets are empty.

3.3.3 $p97^{R155H}$ Dodecamer is Stabilized by the Two Oppositely Stacked D2 Rings

Our SEC profile and SDS-PAGE analysis showed that the dimeric $p97^{R155H}$ hexamers do not bind p47 (Figures 3-2B and 3-2C). To elucidate how the two hexamers organize into a higher-order complex, we built an atomic model along with the cryo-EM density map of the $p97^{R155H}$ dodecamer (Figure 3-4). The two $p97^{R155H}$ hexamers are oppositely packed against their D2 rings with a D6 symmetrical arrangement (Figures 3-5A and 3-

6A). We did not find any density assigned for the p47 cofactor, as corroborated with our biochemical characterization and the K_d determination of the p97-p47 binding (Figures 3-2B, 3-2C, 3-2D and 3-5A). Thus, this suggests that the p47 cofactor does not access the $p97^{R155H}$ dodecamer.

The C-terminal tail of the p97 (residues Q764 to G806) that interacts with the VCP cofactors was previously reported as flexible and has not previously been structurally resolved for a full-length $p97^{R155H}$ mutant[96,248,249]. The crystal structure of the C-terminus up to P774 of an engineered p97 variant has also been determined, but at a weak signal level[243]. Since in our $p97^{R155H}$ dodecamer, the C-terminal tail is sandwiched by the two hexameric D2 rings, its mobility is limited and allows visualization of high-resolution details (Figure 3-6A). Our cryo-EM density map of the $p97^{R155H}$ dodecamer was able to show an apparent density of the partial C-terminal tail (residues Q764 to F773) (Figure 3-6A). The C-terminal tail points away from the pore center, interacting with another C-terminal tail of the opposite neighbor to stabilize the packing of the two $p97^{R155H}$ hexamers (Figure 3-6B). Packing occurs via largely polar-polar interactions between the D2 α helices and a hydrogen-bonding network of the C-terminal residues (D749, E756, Q760, Q764 and R766) (Figure 3-6B). **π**-stacking forces between the aromatic side chains of the C-terminal tails and the D2 α helices (F674, F682, F768, F771 and F773) also contribute to the packing of the two $p97^{R155H}$ hexamers.

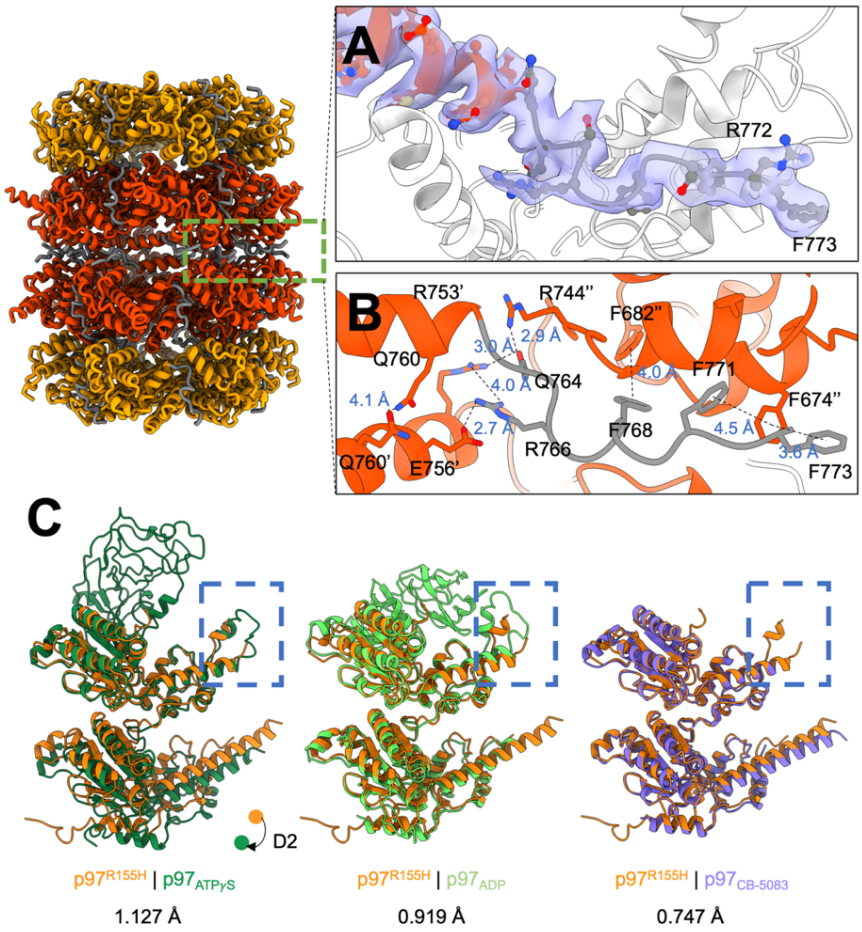


Figure 3-6. Cryo-EM structure of the p97^{R155H} dodecamer. Orange, orange-red, and grey indicate the D1, D2 and linker domains of the p97^{R155H}. (A) Cryo-EM density of the partial C-terminal tail. Purple surface shows the cryo-EM density (4.6$\boldsymbol{\sigma}$). The side chains are shown in stick representation. (B) Interaction network of the binding interface of the two p97^{R155H} hexamers. The two hexamers interact via hydrogen bonds and π-π aromatic packing. (C) Superposition of the p97^{R155H} with p97 structures in different nucleotide states. Dark green, light green, and purple are p97$_{ATP\gamma S}$, p97$_{ADP}$, and p97$_{CB\text{-}5083}$, respectively. RMSDs are 1.127 Å, 0.919 Å, and 0.747 Å for p97$_{ATP\gamma S}$, p97$_{ADP}$, and p97$_{CB\text{-}5083}$, respectively. Blue rectangles highlight conformational changes in the D1 helix-turn-helix motif. The direction of the D2 movement upon ATPγS binding is shown in dots and an arrow.

3.3.4 Nucleotide Binding Destabilizes the p97^{R155H} Dodecameric Formation

We did not find nucleotide binding in the p97^{R155H} dodecamer structure, which occupies about 39.1% of the population (Figures 3-2B, 3-2C and 3-3). In contrast, p97^{WT} is always associated with ADP in the D1 ATPase and has a trace amount of dodecameric form[96]. This could be possible that the nucleotide binding affects p97 oligomerization. We would

like to test whether the nucleotide in the solution affects the dodecameric formation. We then mutated $p97^{R155H}$ on the Walker A motif of the D1 (K251R) or D2 (K524R) ATPase to disrupt the nucleotide binding. The Walker A mutants do not bind any nucleotides nor perform the function. These double mutants were then imaged using negative-stain electron microscopy (EM) for single-particle image analysis. 2D class averages of these mutants showed the ratios of the dodecamer to hexamer are 6.94:1 and 6.69:1 for $p97^{R155H\text{-}K251R}$ and $p97^{R155H\text{-}K524R}$, respectively (Figure 3-7). Compared to the above cryo-EM analysis of the $p97^{R155H}$-p47 assembling, the ratios of the $p97^{R155H}$ dodecamer in D1 and D2 Walker A mutants increase 10.8 and 10.4 times, respectively. This suggests that the $p97^{R155H}$ dodecamer is stable when it does not bind nucleotides. Moreover, the double mutants lack ATPase activity, suggesting that the dodecameric $p97^{R155H}$ is likely to represent an inactive state.

3.3.5 $p97^{R155H}$ Dodecamer is Likely to be an Inactive Form

To compare the monomeric structure of the $p97^{R155H}$ dodecamer with others, we superimposed our structure with $p97^{WT}$ in ADP or ATPγS bound state and a CB-5083-bound p97 (PDB codes: *5FTK* (ADP), *5FTN* (ATPγS), and *6MCK* (CB-5083)) (Figure 3-6C)[224,241]. The superposition of $p97^{R155H}$ and $p97|_{ATP\gamma S}$ revealed a downward rigid-body movement of the D2 domain, which may be induced by the torsional change of the D1-D2 linker (Figure 3-6C). Another significant structural change occurs in the C-terminal tail conformation. The C-terminal tail of the $p97|_{ATP\gamma S}$ points to the pore center, which is in the opposite direction against that of the $p97^{R155H}$ dodecamer (Figure 3-6C). R766 on the C-terminal tail has been found to interact with the γ-phosphate of the D2-bound

nucleotide directly[243]. Because the C-terminal tail of the $p97^{R155H}$ dodecamer extends outwards from the pore center and away from the D2 nucleotide-binding site, it is impossible to stabilize the nucleotide binding.

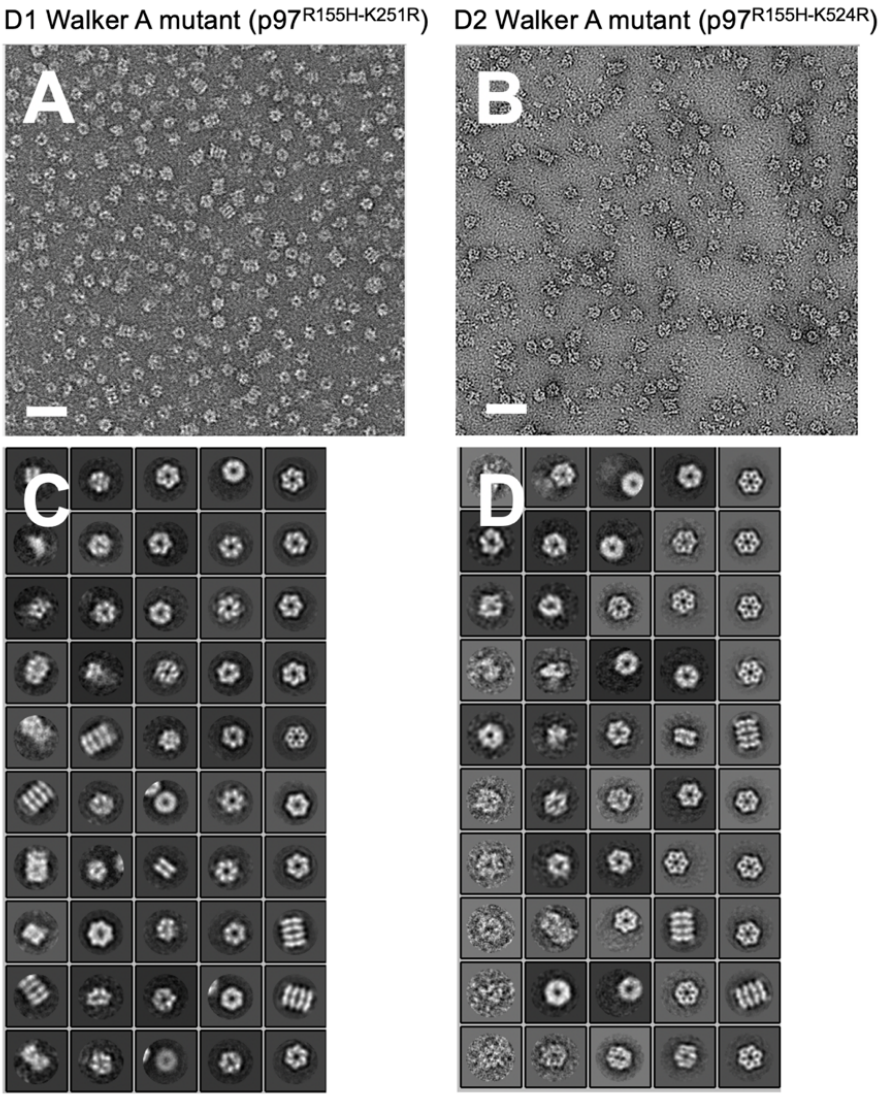


Figure 3-7. Single-particle EM image analysis on the negatively stained Walker A mutants in the context of $p97^{R155H}$. Electron images of the negatively stained (A) $p97^{R155H-K251R}$ and (B) $p97^{R155H-K524R}$ double mutants are shown. Scale bars indicate 50 nm. Two-dimensional class averages of selected particle images were calculated in (C) and (D), respectively. Box side lengths are 333 Å.

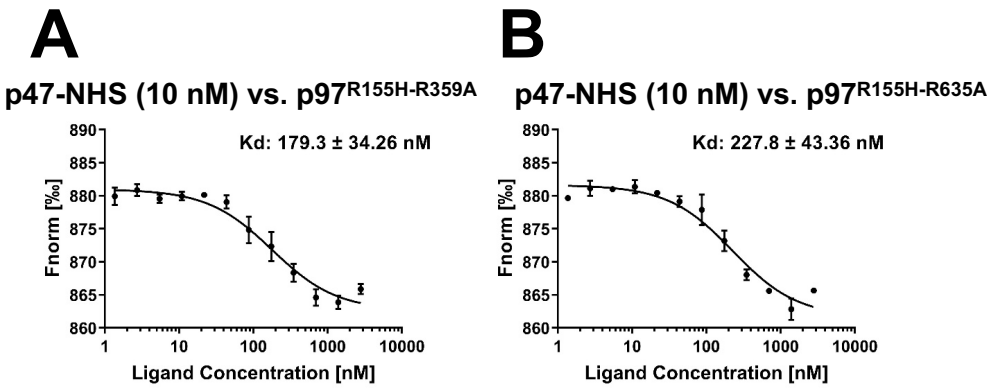


Figure 3-8. Determination of the binding constant, K_d, for the interaction between p97 arginine finger mutants and p47 proteins. Temperature-related intensity change (TRIC) signals on a DianthusNT.23 PicoDuo instrument (Nano-Temper Technologies) was used to determine the equilibrium binding constant (K_d) for p47 in solution with four p97 proteins: (A) $p97^{R155H\text{-}R359A}$ and (B) $p97^{R155H\text{-}R635A}$. Unlabeled p97 protein was titrated against 10 nM of RED-NHS-labeled p47 in two-fold steps from 2,800 nM to 1.37 nM.

The D1 helix-turn-helix (HTH) motif (410-445) of the $p97^{R155H}$ dodecamer differs from that of the $p97|_{ADP}$, which has the HTH motif interacting with the NTD and stabilize it in the down configuration (Figure 3-6C). However, the HTH motif in the $p97^{R155H}$ seems not interacting with the NTD, allowing NTD to be less spatially restricted in the up configuration. Although the NTD densities were resolved in low resolutions partly due to its mobility, the NTD densities in low contours (2.4$\boldsymbol{\sigma}$) could still be identified in up configuration (Figure 3-3D).

CB-5083 is a highly selective D2 ATP-competitive inhibitor on p97 ATPase, and the packing of the dimeric hexamers was also found in the crystal structure of the NTD-truncated p97 with CB-5083[86]. Based on these superpositions, the structure of the $p97^{R155H}$ dodecamer more closely resembles $p97|_{CB\text{-}5083}$ (RMSD 0.747 Å) than the $p97|_{ADP}$ (RMSD 0.919 Å) or $p97|_{ATP\gamma S}$ (RMSD 1.127 Å) (Figure 3-6C). The crystal

structure is consistent with our dodecamer structure, suggesting that p97^{R155H} dodecamer could be functionally equivalent to the inhibited p97 with CB-5083 (Figure 3-6C)[86].

3.3.6 Nucleotide Binding Influences the p47 Binding onto p97^{R155H}

The nucleotide-bound state of the p97^{WT} D1 domain affects the D2 ATPase activity and influences the binding of the p47 cofactor[196,246]. Also, p47 has a differential influence on D1 and D2 ATPase functions of wild-type and p97 mutants[94]. To uncover how nucleotide binding modulates p47 binding to p97^{R155H} mutant, we assembled the complexes in the presence of ADP or ATPγS and biochemically analyzed the resulting complexes. ATPγS was used to mimic the ATP-bound site since it cannot be hydrolyzed when bound to the nucleotide-binding site. The SEC profiles of the two assemblies in the presence of ADP (p97$^{R155H}|_{ADP}$-p47) and ATPγS (p97$^{R155H}|_{ATP\gamma S}$-p47) were similar, and the peaks suggested a uniform distribution for both complexes (Figure 3-9A). SDS-PAGE gel analyses verified the peak fractions containing p97^{R155H} and p47 (Figure 3-9B).

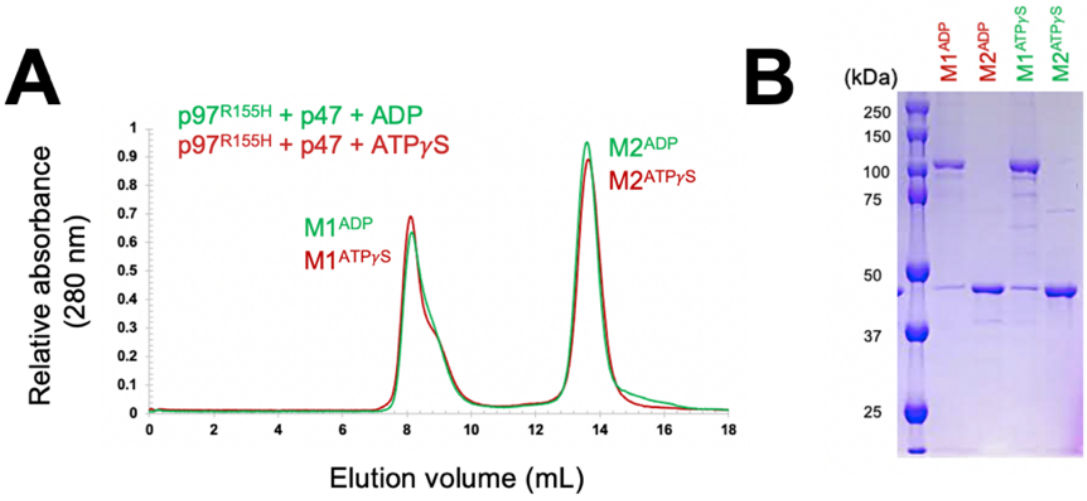


Figure 3-9. Biochemical characterization of the p97$^{R155H}|_{ADP}$-p47 and p97$^{R155H}|_{ATP\gamma S}$-p47. (A) Size-exclusion chromatographic (SEC) profiles. Green and red curves represent

$p97^{R155H}|_{ADP}$ and $p97^{R155H}|_{ATP\gamma S}$-p47, respectively. (B) SDS-PAGE analysis of the SEC eluted peaks of the complex assemblies.

We then collected the peak fractions and froze the samples for cryo-EM structural analysis (Figure 3-10 and 3-11). Although no leading fraction was observed in the SEC profile of the $p97^{R155H}|_{ADP}$-p47, a small fraction of the $p97^{R155H}$ dodecamers were identified in the 2D image class averages (Figure 3-10). After 3D classification, the number of particle images representing the $p97^{R155H}$ dodecamer was only 2,219 (1.37%) among an overall 162,269 particle images. Like the 3D reconstruction of the $p97^{R155H}$ without nucleotides, the dodecamer does not show any sign of bound nucleotide or p47 cofactor. However, the $p97^{R155H}|_{ATP\gamma S}$-p47 image dataset did not have the $p97^{R155H}$ dodecameric population, suggesting that $p97^{R155H}$ dodecamer does not form in the presence of ATPγS (Figure 3-11). Thus, these suggest that the nucleotide presence, particularly ATPγS, favors the dissociation of the $p97^{R155H}$ dodecamer into hexamers. Furthermore, because the ATPase activity of the $p97^{R155H}$ is higher than $p97^{WT}$ [94-96,197,228,236] but the $p97^{R155H}$ dodecamer does not interact with nucleotides, it is very likely that the increased activity of the $p97^{R155H}$ is contributed from the functional hexamers, rather than dodecamers.

We calculated the 3D reconstructions of the $p97^{R155H}|_{ADP}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 complexes at 4.50 Å and 4.23 Å resolution, respectively (Figures 3-10 and 3-11). Cryo-EM densities of the nucleotide can be identified in both the D1 and D2 nucleotide-binding pockets of the $p97^{R155H}|_{ATP\gamma S}$-p47 density (Figure 3-12). Unlike the $p97^{R155H}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 densities, the $p97^{R155H}|_{ADP}$-p47 density map showed a heterogeneous distribution of NTD orientations (Figure 3-13A). An NTD density with

bound $p47^{UBX}$ domain was identified (1.0$\boldsymbol{\sigma}$), located at the highest height among all the other NTDs (Figure 3-13B). The tilting angles and heights of the NTDs relative to the ring plane varied sequentially (Figure 3-13B). This observation of the high structural variabilities is consistent with the deep coordinate neural network analysis (Figure 3-14).

We calculated the 3D reconstructions of the $p97^{R155H}|_{ADP}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 complexes at 4.50 Å and 4.23 Å resolution, respectively (Figures 3-10 and 3-11). Cryo-EM densities of the nucleotide can be identified in both the D1 and D2 nucleotide-binding pockets of the $p97^{R155H}|_{ATP\gamma S}$-p47 density (Figure 3-12). Unlike the $p97^{R155H}$-p47 and $p97^{R155H}|_{ATP\gamma S}$-p47 densities, the $p97^{R155H}|_{ADP}$-p47 density map showed a heterogeneous distribution of NTD orientations (Figure 3-13A). An NTD density with bound $p47^{UBX}$ domain was identified (1.0$\boldsymbol{\sigma}$), located at the highest height among all the other NTDs (Figure 3-13B). The tilting angles and heights of the NTDs relative to the ring plane varied sequentially (Figure 3-13B). This observation of the high structural variabilities is consistent with the deep coordinate neural network analysis (Figure 3-14).

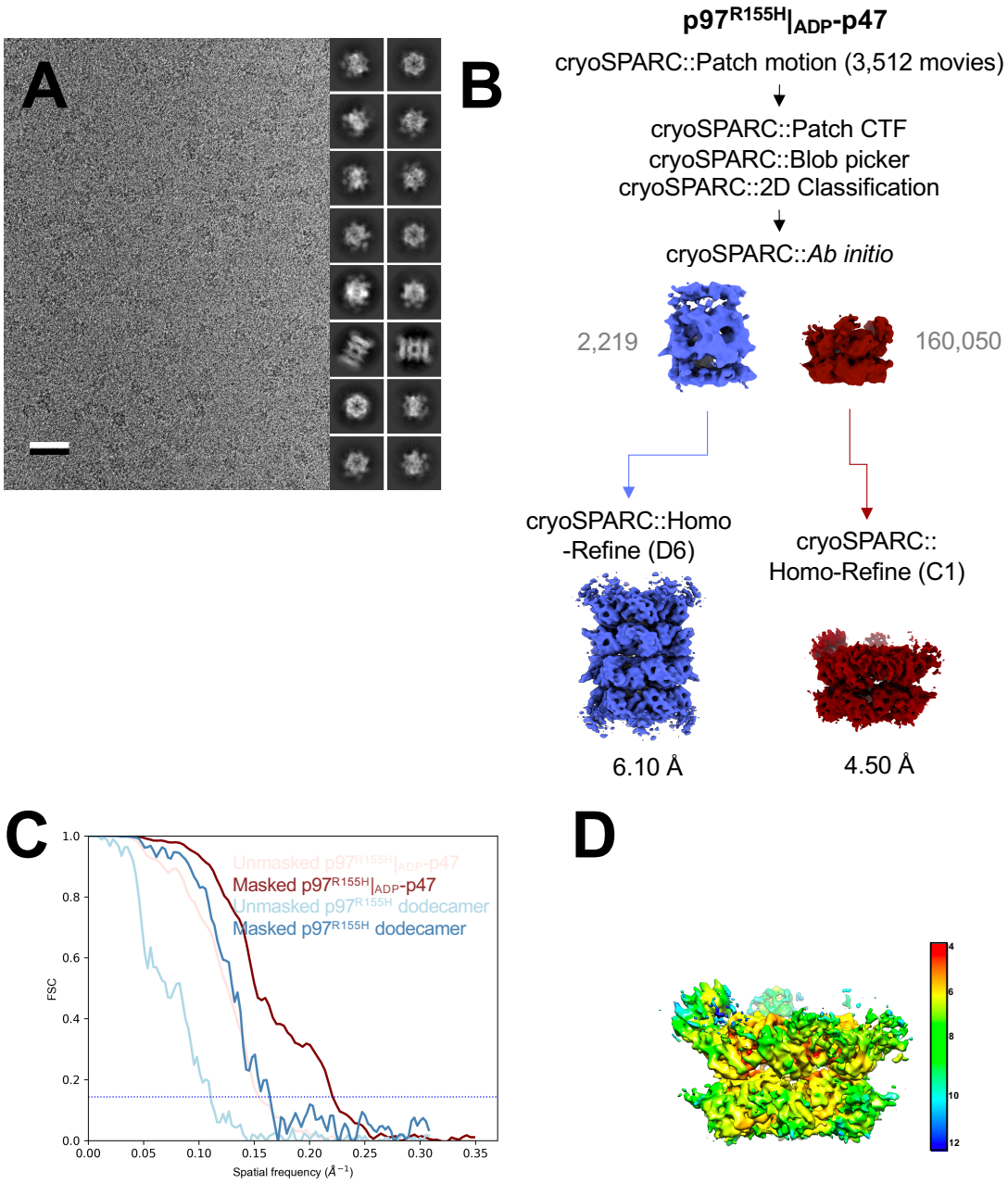


Figure 3-10. Cryo-EM structural analysis of the full-length p97^{R155H}|$_{ADP}$-p47 assembly. (A) Electron micrograph of cryogenic preparation of the p97^{R155H}|$_{ADP}$-p47 complex. Black contrast represents protein and white is background. Scale bar indicates 50 nm. Insets to the side show representative 2D class averages. Box side length is 382 Å. (B) Image processing flowchart for single-particle cryo-EM reconstruction. (C) Fourier-shell correlation (FSC) plot of the 3D reconstructions. (D) Local resolution estimation of the reconstructions.

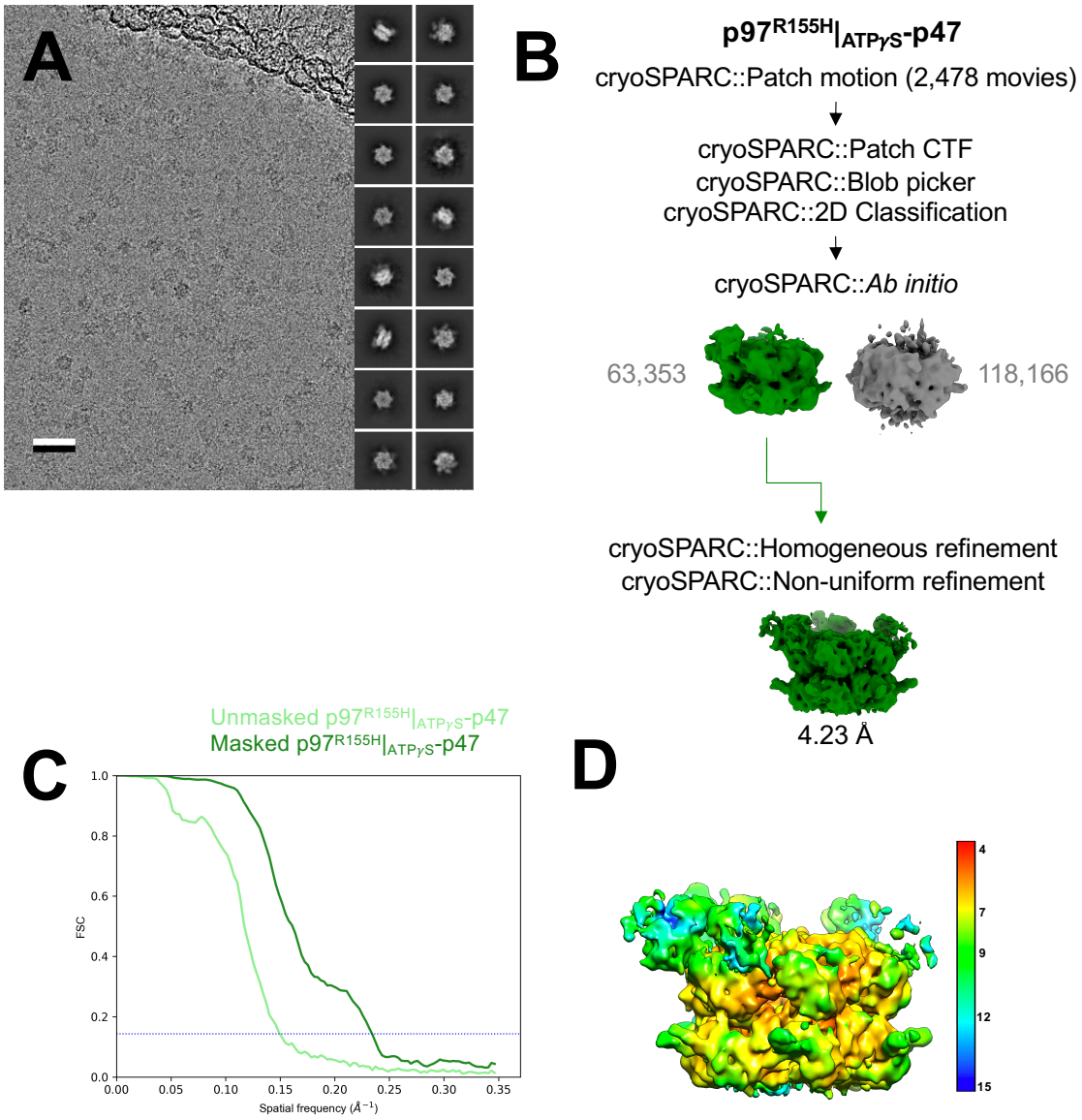


Figure 3-11. Cryo-EM structural analysis of the full-length p97^{R155H}|$_{ATP\gamma S}$-p47 assembly (A) Electron micrograph of the cryogenic p97^{R155H}|$_{ATP\gamma S}$-p47 complex. Black contrast represents proteins and white is background. Scale bar indicates 50 nm. Inset shows representative 2D class averages. Box side length is 382 Å. (B) Image processing flowchart for single-particle cryo-EM reconstruction. (C) Fourier-shell correlation (FSC) plot of the 3D reconstructions. (D) Local resolution estimation of the reconstructions.

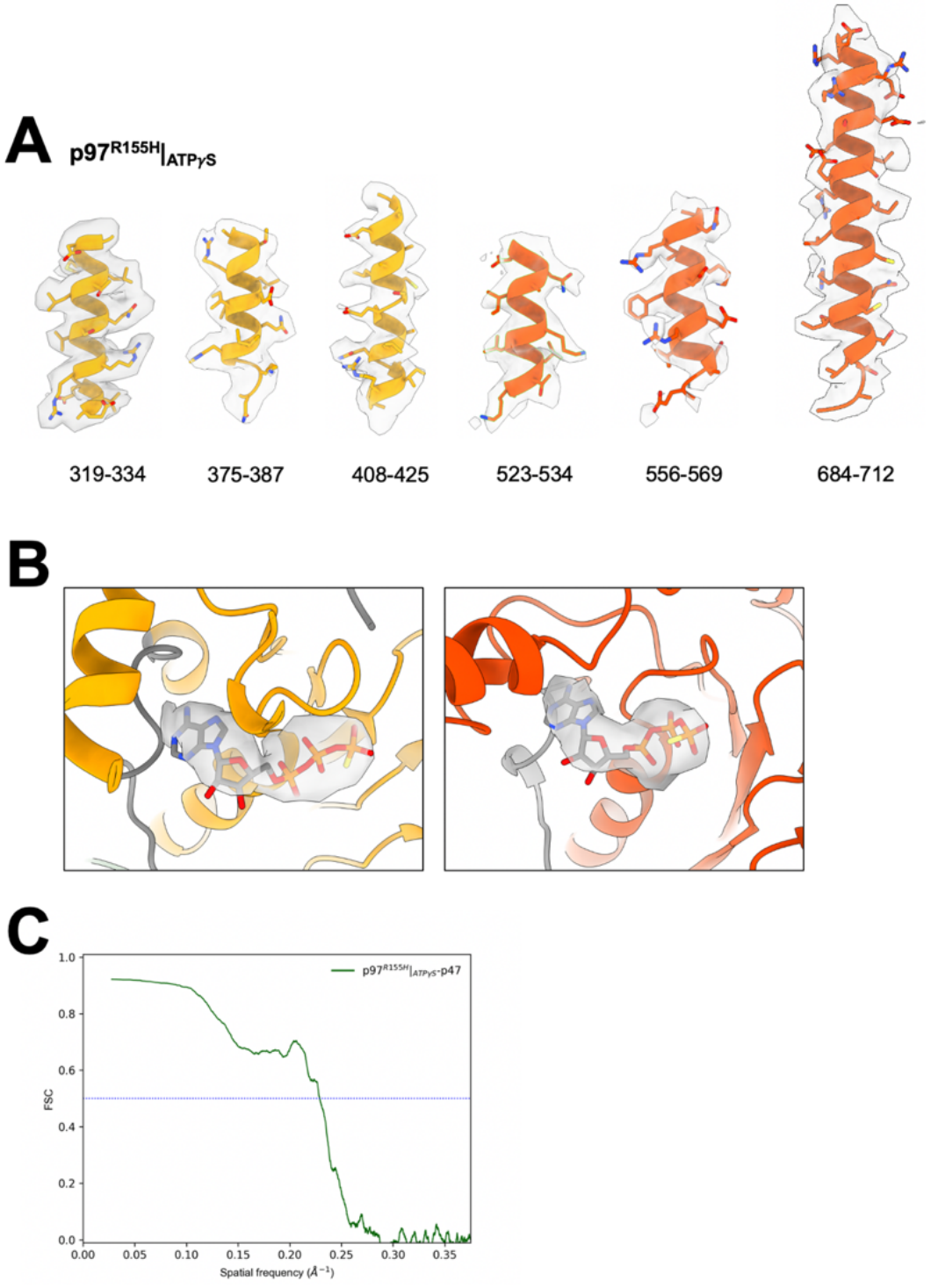


Figure 3-12. Modelling of atomic coordinates into the cryo-EM densities of p97^{R155H}|$_{ATP\gamma S}$-p47 complexes. (A) Model fitting of the p97^{R155H}|$_{ATP\gamma S}$-p47 assembly. Orange and orange red are D1 and D2 domains, respectively. (B) Model fitting of

ATPγS nucleotides in D1 (left) and D2 (right) binding sites. (C) Fourier-shell correlation (FSC) plot. Green curve is $p97^{R155H}|_{ATP\gamma S}$-p47 assembly. Blue dash indicates FSC values of 0.5.

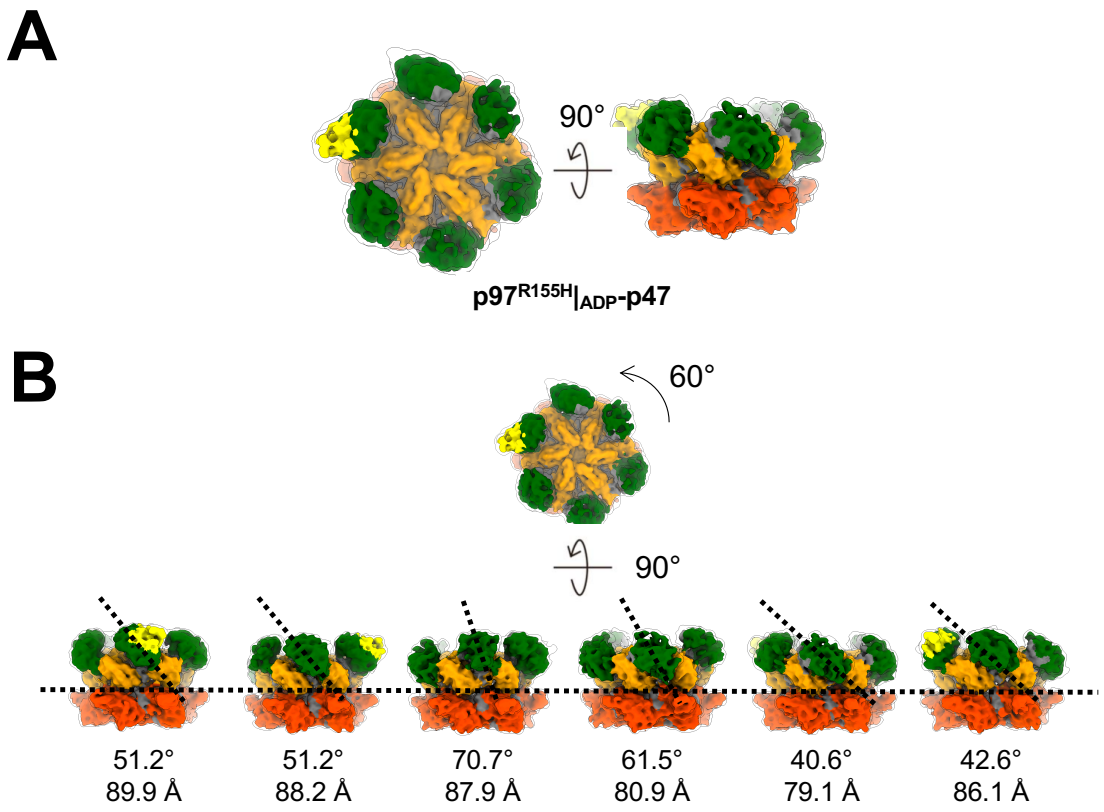


Figure 3-13. Cryo-EM structure of the $p97^{R155H}|_{ADP}$-p47 complex. (A) Cryo-EM densities of $p97^{R155H}|_{ADP}$-p47. Green, orange, orange red and yellow indicate the NTD, D1 and D2 domains of the $p97^{R155H}$ and $p47^{UBX}$, respectively. Grey indicates the densities for the linkers between domains of $p97^{R155H}$. Envelopes are the cryo-EM maps at a lower contour (1.0**σ**). (B) NTD conformations. The tilting angles of the NTDs are labeled, and the measured heights are the distance between the NTD center and the bottom of the D2 ring plane.

For the $p97^{R155H}|_{ATP\gamma S}$-p47 complex, four NTDs and two $p47^{UBX}$ structures can be assigned to the cryo-EM density, and the unassigned two NTD densities were fragmented (Figure 3-15A). The four up NTD are consistent with the crystal structure of the truncated N-D1 domain of the $p97^{R155H}|_{ATP\gamma S}$ [214]. Like the $p97^{R155H}|_{ADP}$-p47 complex, the

densities of NTDs and p47 were highly variable, but those for D1 and D2 domains were almost invariant, as shown in the local resolution estimation (Figure 3-11D).

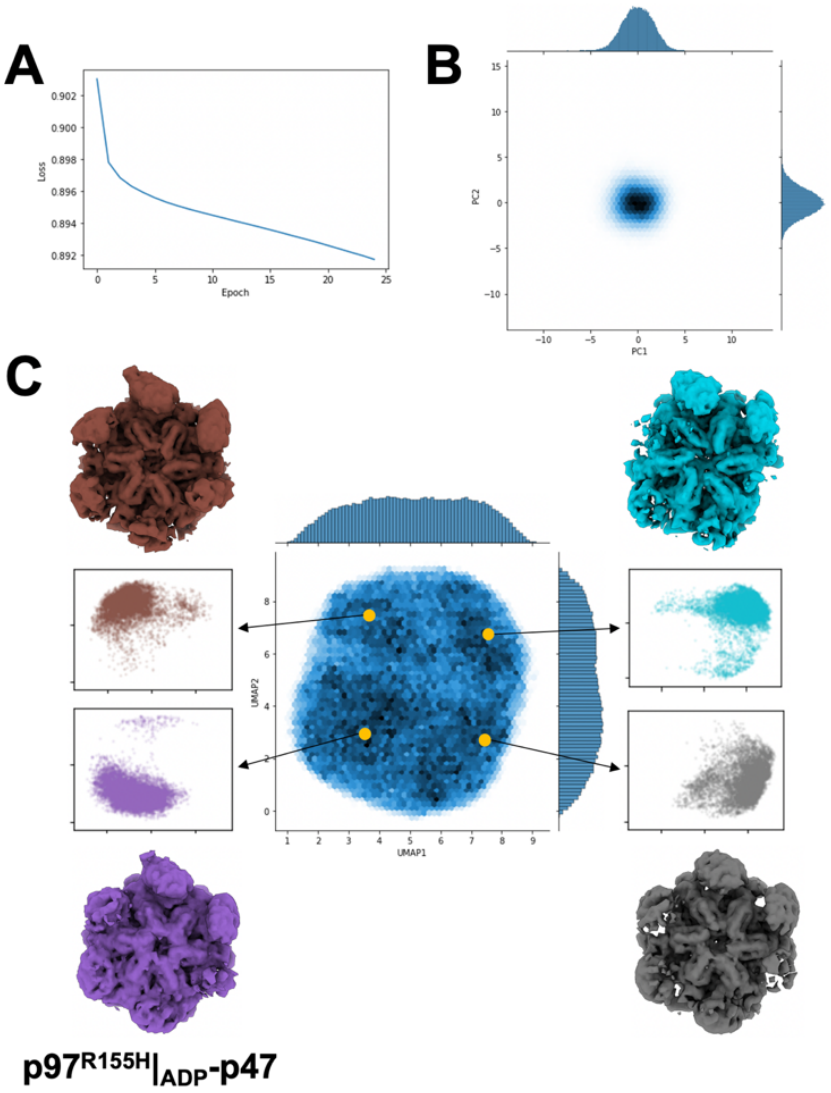


Figure 3-14. Deep coordinate neural network analysis of the p97^{R155H}|$_{ADP}$-p47 single-particle cryo-EM images. (A) Plot of the loss function versus learning epochs. (B) Plot of the two major principal components in the hyperspace. (C) UMAP representation and averaged reconstructions of the four selected clusters.

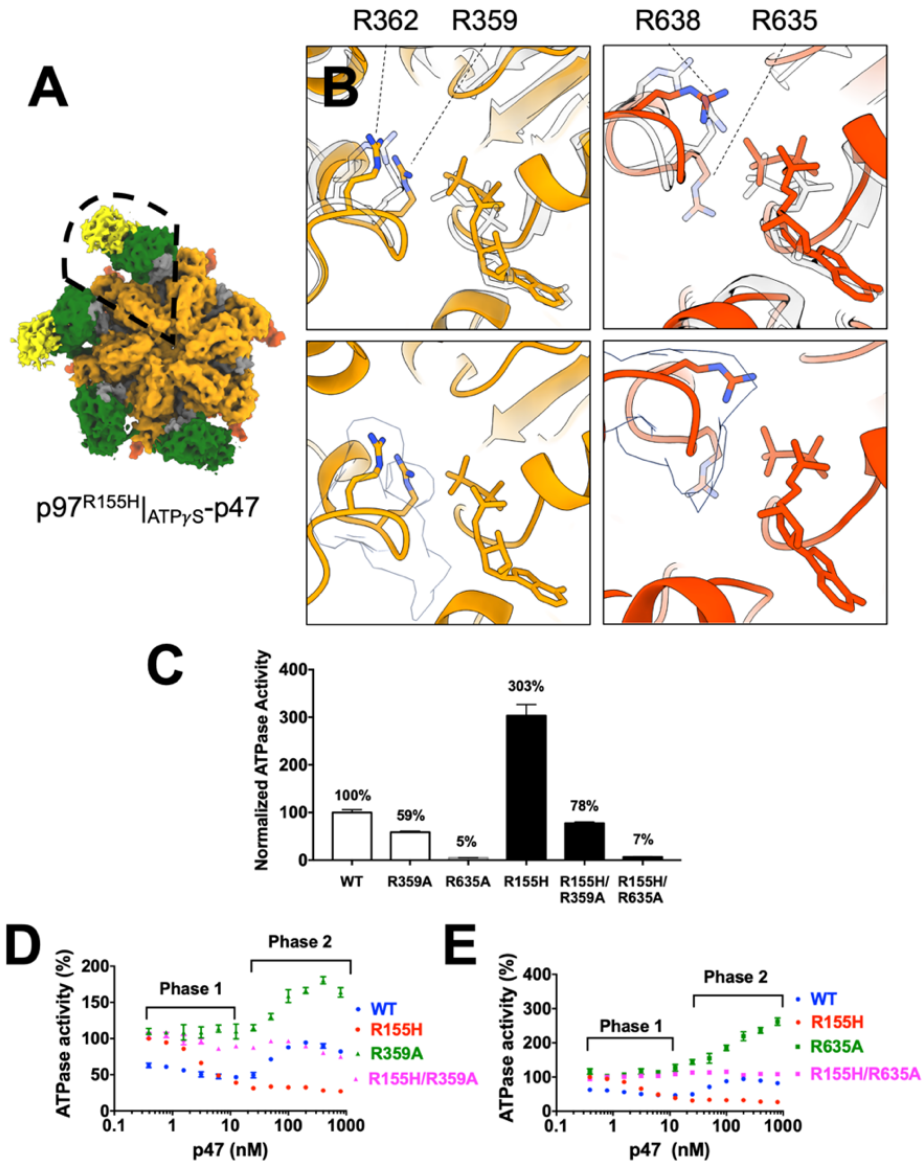


Figure 3-15. Cryo-EM structural analysis of the p97^{R155H}|$_{ATP\gamma S}$-p47 complex and functional measurements of the p97 arginine finger mutants. (A) Cryo-EM density of the p97^{R155H}|$_{ATP\gamma S}$-p47complex. Enclosed is the p97^{R155H}|$_{ATP\gamma S}$-p47 monomer used for structural comparison with p97^{WT}|$_{ATP\gamma S}$. (B) Superposition of the structures of p97^{R155H}|$_{ATP\gamma S}$-p47 and p97^{WT}|$_{ATP\gamma S}$ (PDB code: *5FTN*)[224]. Bound ATPγS and arginine fingers are shown in stick representation. Upper panels are the superpositions of the two structures and white are the p97^{WT}|$_{ATP\gamma S}$ structure. Lower panels are the map-model fitting of the two arginine fingers in D1 and D2 nucleotide-binding sites. White surfaces are cryo-EM densities. (C) Full-length (FL) p97 ATPase activities of wild type and R155H mutants. Activity measurements were normalized relative to the p97^{WT} activity and measured in the presence of 200 μM ATP (n = 4). (D) Normalized ATPase activities of p97 ATPase: WT (blue circles), R155H (red circles), R359A (green triangles) and

R155H-R359A (magenta triangles). The R359A mutation is in the D1 arginine finger. p97 ATPase activity was normalized relative to its basal activity in the absence of p47. The addition of p47 increased from 0 to 800 nM. ATPase activity measurements were performed in the presence of 200 μM ATP. (E) Experiments as in (B), but the mutated residue was located at the R635A, the D2 arginine finger. p97 ATPase activity in R635A and R155H-R635A mutants is shown in green squares and magenta squares, respectively. All error bars indicate ±SD (n = 4).

3.3.7 p47 Binding Impacts p97^{R155H} Function Via an Allosteric Effect on D2 ATPase

Previous functional analyses demonstrated that cofactor binding or NTD mutation alters p97 ATPase function, mainly the D2 ATPase[94,228]. However, the NTD is far away from the nucleotide-binding pocket and how do NTDs allosterically regulate p97 ATPase functions? Because D1 and D2 ATPase activities are correlated, one possibility is that the NTD regulates ATPase functions via inter-domain or interprotomer communications between these functional modules[228,250]. Because the cryo-EM density of p97$^{R155H}|_{ATP\gamma S}$-p47 allows us to visualize the side chains in its D1 and D2 domains, we compared it with the model with p97$^{WT}|_{ATP\gamma S}$ structure (PDB code: *5FTN*)[224]. We identified two p47$_{UBX}$ bound onto the p97^{R155H} NTDs, implying majority of the p97^{R155H} particles bind two p47 molecules. The superposition of the structures of the p47-bound p97$^{R155H}|_{ATP\gamma S}$ and p97$^{WT}|_{ATP\gamma S}$ alone showed that the D2 domain slightly moves and most of the critical residues in the nucleotide-binding pockets do not exhibit major conformational changes (RMSD: 1.015 Å) (Figure 3-15B). However, the D2 R635, the D2 arginine finger, of the p97$^{R155H}|_{ATP\gamma S}$ changes its conformation upon p47 binding, although the conformations of the D1 arginine fingers seem not changed (Figure 3-15B). It is also known that the arginine fingers extended from the neighbor protomer can affect the ATP hydrolysis efficiency[197,251,252].

We would then like to test whether the D2 ATPase activity change is affected by NTD mutation or p47 binding via the arginine fingers. We introduced two mutations of the arginine fingers on R359A and R635A and measured the mutant ATPase activities (Figure 3-15C). The $p97^{R359A}$ and $p97^{R635A}$ mutants have reduced activities compared to wild type (59% and 5%, respectively), and the mutation of the D2 R635 nearly abolished the p97 ATPase activity, consistent with the previous findings[253]. We next prepared the two double mutants, $p97^{R155H\text{-}R359A}$ and $p97^{R155H\text{-}R635A}$, and the results showed that both double mutants reduced $p97^{R155H}$ activities to 26% and 2.3%, respectively. Because the $p97^{R155H}$ has a higher ATPase activity than $p97^{WT}$ [94-96,197,228,236] (Figure 3-15C), the results showed that the increased $p97^{R155H}$ ATPase activity due to its R155H mutation was diminished by the D1 and D2 arginine mutations. Both D1 and D2 arginine fingers appear to affect the communications within p97 subunits for the R155H functional enhancement. Thus, the gained ATPase activity from the R155H mutation of the NTD is very likely to associate with functional arginine fingers.

Next, we would like to know whether p47 impacts $p97^{R155H}$ function via arginine fingers in the same manner as R155H mutation. As seen previously, upon binding to p47, $p97^{WT}$ exhibits a biphasic response, but $p97^{R155H}$ does not exhibit the rebound Phase 2 (Figure 3-15D)[94]. However, the D1 arginine finger mutants, $p97^{R359A}$, lacked Phase 1 inhibition but showed significant activation in the later phase (1.9-fold), similar to the D1 Walker B mutant (E305Q), which binds to nucleotide but does not perform the catalysis (Figure 3-15D)[94]. Similar to $p97^{R359A}$, the D2 arginine finger mutants, $p97^{R635A}$, showed a significant Phase 2 activation (3-fold), but different from the D2 Walker B mutant (E578Q) (Figure 3-15E)[94]. These data indicate that upon p47 binding, the Phase 1

inhibition of the $p97^{WT}$ associates with both D1 and D2 arginine fingers, but the Phase 2 activation is mainly contributed by the D2 ATPase activity that associates with the D2 arginine finger (Figures 3-15D and 3-15E).

The responses of the double mutants showed both impacts from R155H mutation and p47 binding. The p47 inhibitory regulation on the $p97^{R155H}$ was abolished for the two double mutants, the $p97^{R155H\text{-}R359A}$ and $p97^{R155H\text{-}R635A}$ mutants (Figures 3-15C and 3-15D). Especially for the $p97^{R155H\text{-}R635A}$ mutant for the D2 arginine finger, the curve is flat, implying that the p47 binding has no impact on the ATP hydrolysis. Thus, this may indicate a functional connection between p47 binding and the arginine fingers, R359 and R635, of the $p97^{R155H}$ mutant. The structural superpositions showed that the change of the ATPase function may result from the slight conformational changes of the arginine fingers (Figure 3-15A). Thus, these findings suggest that the arginine fingers are critical for the p47-induced communications between the NTD and ATPase domain. In addition to the functional modulation of p47 via arginine fingers, we would like to test whether the p47 binding is affected when the arginine finger is mutated. We fractionated a mixture of $p97^{R155H\text{-}R359A}$ or $p97^{R155H\text{-}R635A}$ (1.67 μM in hexameric form) and p47 (80 μM in monomeric form) proteins in a gel-filtration column, using $p97^{R155H}$ as a positive control (Figure 3-16). We found that the double mutant $p97^{R155H\text{-}R359A}$ and $p97^{R155H\text{-}R635A}$ retained the ability to bind p47, eluting in the same fraction as the $p97^{R155H}$-p47 complex (Figure 3-16). We also determined the binding affinities, K_d, of the p47 to $p97^{R155H}$, $p97^{R155H\text{-}R359A}$ and $p97^{R155H\text{-}R635A}$ as 132, 179 and 228 nM, respectively (Figure 3-16 and Table 3-2). Thus, these results showed that the gained ATPase activities of the $p97^{R155H\text{-}R635A}$ from $p97^{R155H}$ impacted by p47 may be partly due to weak p47 binding affinity to the $p97^{R155H\text{-}}$

^{R635A} double mutant (Figure 3-2D, 3-15E and 3-16B). In addition, the impaired arginine fingers reduce the p47 binding affinity to p97^{R155H}, implying that the domain-domain communications within p97 ATPase affect how p97 responds to cofactor binding.

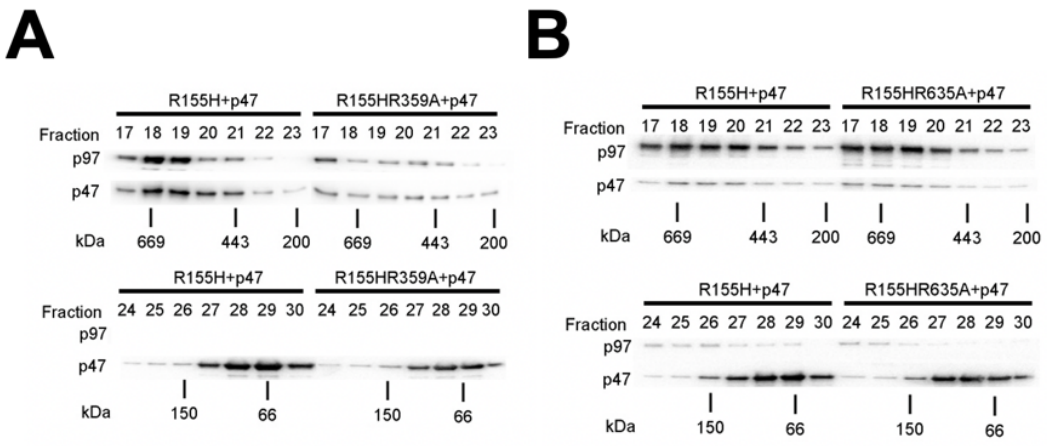


Figure 3-16. Western blotting analyses of the SEC-eluted fractions of the p97-p47 complexes. The SEC of p97-p47 complex was performed using a Superdex 20010/300 GL column and subsequent Western blotting was performed for the SEC fractions from (A) 17 to 23 and (B) 24 to 30.

3.3.8 Conformational Change of the D2 Arginine Fingers Play an Important Role

Single amino acid mutations in p97 have long been linked to diseases including IBMPFD and ALS[254], and these disease mutants alter the p97 ATPase activity and cofactor binding on the NTDs[94]. Because the NTD is far away from the D2 ATPase, where ATP hydrolysis mostly occurs, it is unclear how the NTD allosterically affects the p97 ATPase function. The molecular mechanism underlying the disease is still unresolved today. Our goal aims to understand (1) how NTD conformational change, such as up and down configurations, connects to ATPase function, (2) the structural change of p97 ATPase caused by R155H mutation, and (3) how the mutant responds to p47 binding in the context of a complete full-length complex. Here we conducted biochemical and structural

analyses to study potential pathological changes in the $p97^{R155H}$ mutant, the most prevalent disease-linked mutation[86].To our knowledge, our report reveals the first full-length $p97^{R155H}$ mutant structures in dodecameric and hexameric forms, as well as the $p97^{R155H}$-p47 complex in different nucleotide states. Previous studies have placed less emphasis on the $p97^{R155H}$ dodecamer. Our report found that the $p97^{R155H}$ dodecamer seems unstable in the presence of nucleotides and insensitive to bind p47. Although we do not know if the $p97^{R155H}$ dodecamer plays a physiological role *in vivo* or leads to IBMPFD disease, our cryo-EM image analysis showed that this high order form can be stabilized in an aqueous solution and occupies about 40%. Also, we found a possible connection for the p47 and NTD regulation on ATPase function, likely arginine fingers, and the miscommunication between these domains alters the normal ATPase activity. This may interfere with the ability of the p97 to carry out the p47-related function such as Golgi-membrane reassembly or autophagy[217]. These new findings thus reveal how cellular functions may be so profoundly impacted in the $p97^{R155H}$ mutant.

3.3.9 Relevance of the Dodecamer

The formation of a $p97^{WT}$ dodecamer has been previously reported[231,243,255]. However, our SEC analysis did not show the dodecameric fraction in the wild-type assembly, and the proportion of the $p97^{WT}$ dodecamer may be too low to be detected in our SEC analysis (Figure 3-2B). Although it is not known whether the formation of the $p97^{R155H}$ dodecamer occurs *in vivo* or whether it plays any role in a cell, the dodecamer at least can be stably formed in the solution and likely in the cell as well. Also, the structural comparison of the CB-5083-bound structure [241] and its empty nucleotide-binding pockets

strongly suggest the dodecameric form as an inactive state. The inactivation of the p97 dodecamer could be caused by its high-order organization, which may hinder the mobility of the monomers, interfere with the hexameric ring to break, prevent hand-over-hand movements and possibly, in turn, down-regulate normal p97 ATPase function during substrate processing[218,219]. Only when the nucleotides are present, the $p97^{R155H}$ dodecamer will dissociate into two functional hexamers with an elevated ATPase activity. However, we have no information about why the R155H mutation favors the dodecameric formation. One possible explanation would be that the dodecameric formation induced by R155H mutation may keep p97 inactive while under the stress of the limited ATP concentrations in the cell[241]. The definition of its functional role *in vivo* requires further investigations using cell biological tools.

Our data showed that the p47 binding or NTD mutation affects the communications between the functional modules of the full-length $p97^{R155H}$ ATPase in either an intra-domain or an inter-domain manner or both. Also, the affinity of p47 to $p97^{R155H}$ is affected by the D1 or D2 ATPase activities. Also, we showed that either D1 or D2 arginine finger mutation can affect the p47 binding affinity (Table 3-1). These views might not be observed using the truncated or incomplete p97 complexes to show the effects caused by domain communications. The partial p97 ATPase may also lead to inconsistent observations on the stoichiometry of the p97-p47 bindings. For example, the crystal structure of the $p97^{N\text{-}D1}$-$p47^{UBX}$ showed that two p47 UBX domains bind to two adjacent p97 monomers, while a third is poorly resolved located away from those two bound NTDs[105]. However, this view is not consistent with our observation from single-particle cryo-EM. Therefore, a complete full-length complex will be required to provide a

view of the interactions between overall functional modules that are close to their native states.

Previous structural studies revealed the interactions between $p97^{WT}$ and p47 proteins[105,235,244,256]. NMR (nuclear magnetic resonance) spectroscopic analysis of the $p97^{N\text{-}D1\text{-}L}$-p47 (residues 1-480 of p97) indicated a possible tripartite p47 binding mode[235]. This interaction involves the UBX and the C-terminal SHP of p47 binding to one p97 NTD, while a third p47 N-terminal SHP domain binds the adjacent NTD of the p97. The up NTD configurations allow organizing p47 in this arrangement, rather than down NTDs. On the other hand, the $p97^{R155H}$ mutant, especially when bound to ATP, should interact with p47 cofactors followed by the proposed tripartite binding mode, since the NTDs are all in up configuration[235,246]. However, our $p97^{R155H}$-p47 complex structures indicate a disordered or flexible structure of NTD or $p47^{SHP}$ domain, implying that p97 NTD mutant may either interact with p47 more transiently or require a concerted motion on NTDs to adapt the proposed binding mode[226,246]. In this sense, the oscillating up NTDs of the $p97^{R155H}|_{ADP}$ could easily access the p47 cofactors than the $p97^{WT}|_{ADP}$, which has all down NTDs [37]. This corroborates with the previous finding that the binding of p47 on the $p97^{WT}$ ATPase in the presence of ADP is much weaker than those in the presence of ATP or the absence of nucleotides[246]. However, the p47 binding for the $p97^{R155H}$ mutant is independent of the nucleotide states[246].

Recent cryo-EM structures of the AAA+ ATPase with p47 cofactor or its homolog showed an asymmetric arrangement of the hexameric structure[218,221]. These structures have shown the substrate binding in the central channel of the ATPase double

ring[218,219,257]. However, we did not observe the same view in the structures of our complexes or p97^{R155H} ATPase alone. It is possible that the p47 binding alone or the R155H NTD mutation cannot induce the domain movements and the presence of the substrate seems to be required. Ufd1/Npl4 (UN) is one of the p97 cofactors that increase p97 unfoldase activity[258,259]. The p97 disease mutant exhibited an elevated unfoldase activity when binding to the UN cofactor and the substrate[81]. Although p47 seems not to be involved in the unfoldase process, it may regulate p97 ATPase in a similar manner for Golgi membrane reassembly or autophagy, which may not require substrate in the pathway. The p97-p47-mediated membrane fusion requires the participation of VCIP135 (valosin-containing protein p97/p47 complex-interacting protein p135) and syntaxin-5 (Syn5)[79], which help in the coordination of ubiquitin transfer and ATP hydrolysis. These cellular functions may utilize an action mode of p97 different from the ERAD (endoplasmic-reticulum-associated protein degradation) process or require more proteins to initiate the conformational changes to perform the function.

3.3.10 Proposed Mechanism

Here we propose a possible diseased mechanistic model for p97^{R155H} regulation (Figure 3-17). In the absence of nucleotides, the p97^{R155H} mutant prefers to form a higher-order p97^{R155H} dodecamer, immobilizing C-terminal tails and prohibiting interaction of p47 cofactor with the NTD. When the nucleotide is present and binds to the enzyme mutant, the p97^{R155H} dodecamer dissociates into two functional hexamers, allowing cofactor binding and performing ATP hydrolysis. The regulation of the p47 cofactor binding and functional alteration by R155H mutation may induce a conformational change of the

arginine fingers and modulate the ATPase function. Our analyses suggest the importance of using a full-length, rather than truncated, p97 for functional or structural characterization to gain a complete view of the domain-domain communications within the p97 hexamer. To build on our finding regarding dysregulation of interprotomer communication in $p97^{R155H}$ hexamer, we expect the incorporation of additional physiological substrates for p97-p47 assembly for the next critical step to understand the p97 pathological mechanism.

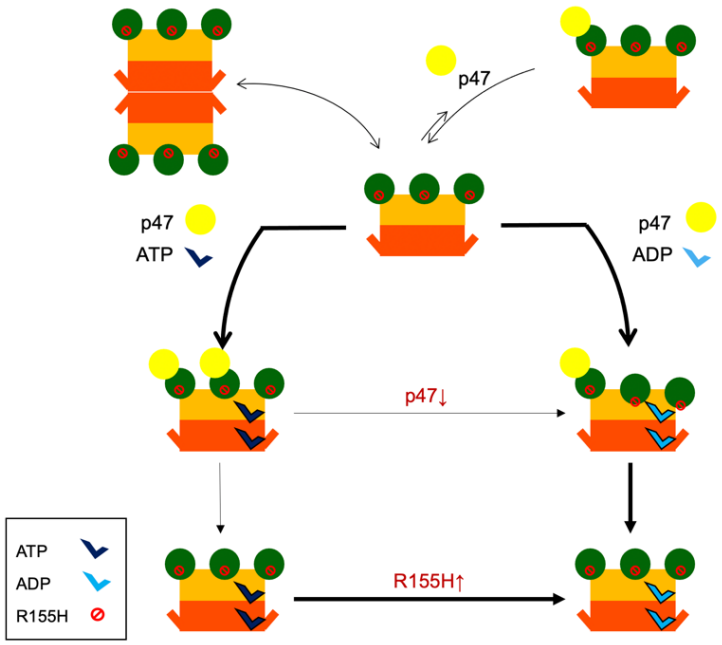


Figure 3-17. Proposed model for p47 binding to the disease-linked $p97^{R155H}$ mutant. The NTD, D1 and D2 domains of the $p97^{R155H}$ are colored green, orange and orange red. In the absence of nucleotides, $p97^{R155H}$ can form a dodecamer and become inactive. p47 does not access the $p97^{R155H}$ dodecamer in the absence of nucleotides. The $p97^{R155H}$ hexamer does not stably bind to p47 in the absence of nucleotides. Once $p97^{R155H}$ binds nucleotides, it stably interacts with p47. The p47 binding blocks the $p97^{R155H}$ ATPase activity. Although the NTDs are all above the D1 ring plane in the presence of ADP, they adopt slightly different tilting angles and heights. However, in the presence of ATP, the NTDs are all in the same up configuration. The p47 interacts with the up-NTD, but not the down-NTD. Thicker lines indicate favor direction for reaction.

3.4 Conclusions

IBMPFD/ALS is a genetic disorder caused by a single amino acid mutation on the p97 ATPase, promoting ATPase activity and cofactor dysregulation. The disease mechanism underlying p97 ATPase malfunction remains unclear. To understand how the mutation alters the ATPase regulation, we assembled a full-length $p97^{R155H}$ with its p47 cofactor and first visualized their structures using single-particle cryo-EM. More than one-third of the population was the dodecameric form. Nucleotide presence dissociates the dodecamer into two hexamers for its highly elevated function. The N-domains of the $p97^{R155H}$ mutant all show up configurations in ADP- or ATPγS-bound states. Our functional and structural analyses showed that the p47 binding is likely to impact the $p97^{R155H}$ ATPase activities via changing the conformations of arginine fingers. These functional and structural analyses underline the ATPase dysregulation with the miscommunication between the functional modules of the $p97^{R155H}$.

3.5 Data Availability Statement

Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-24302 ($p97^{R155H}|_{ATP\gamma S}$-p47), EMD-24305 ($p97^{R155H}$-p47), EMD-23191 ($p97^{R155H}$ dodecamer), EMD-24304 ($p97^{R155H}|_{ADP}$-p47) and EMD-23192 ($p97^{R155H}$ dodecamer II). Model coordinates were deposited in the Protein Data Bank (PDB) under accession numbers 7R7S ($p97^{R155H}|_{ATP\gamma S}$-p47), 7R7U ($p97^{R155H}$-p47), 7L5W ($p97^{R155H}$ dodecamer) and 7R7T ($p97^{R155H}|_{ADP}$-p47). All the data are available from the corresponding author upon request.

3.6 Acknowledgments

We thank Ellen Zhong (Massachusetts Institute of Technology) for the brief discussion on the data analysis using cryoDRGN neural networks. We thank Dr. Xiaoyi Zhang, Gui Lin, and Daniel Wong for preliminary ATPase activity determination for p97 mutants. We thank Dr. Dewight Williams for assisting in cryo-EM data-collection and acknowledge the use of the Titan Krios TEM, in the Eyring Materials Center (EMC), at Arizona State University (ASU) supported from grant number NSF MRI 1531991. We thank the NVIDIA GPU Grant Program to Prof. Po-Lin Chiu for GPU device support. The project was partially supported by a grant from the National Institute of Neurological Disorders and Stroke (NINDS) (R01NS102279) to Prof. Tsui-Fen Chou (Caltech) and the ASU startup fund to Prof. Po-Lin Chiu. We acknowledge the support and resources from the Life Sciences Electron Microscopy Lab for the use of the electron microscope Philips CM12, at ASU.

CHAPTER 4

CONSERVED L464 IN p97 D1–D2 LINKER IS CRITICAL FOR p97 COFACTOR REGULATED ATPASE ACTIVITY

This work has been published in Zhang, X.[#]; Gui, L.[#]; Li, S.[#]; **Nandi, P.[#]**; Columbres, R.C.A.; Wong, D.E.; Moen, D.R.; Lin, H.J.; Chiu, P.-L.[*]; and Chou, T.-F.[*] Conserved L464 in p97 D1-D2 linker is critical for p97 cofactor regulated ATPase activity. *Biochem J.* **2021**, *478 (17)*: 3185–3204 and this chapter is adapted from the manuscript[260].

4.1 Introduction

p97, also known as VCP (valosin-containing protein), TER (translational endoplasmic reticulum ATPase, in mammals), VAT (valosin-containing protein-like ATPase of *Thermoplasma acidophilum*, in archaea) and Cdc48p (in yeast), is a ubiquitous type II ATPase protein. It is highly conserved and plays essential roles in multiple cellular processes, including protein quality and cell cycle control[261-265], ubiquitin-proteasome system (UPS)[204,266], transcriptional activation[267], DNA-damage repair[268] and autophagy[89,269-272]. The involvement of p97 in all major proteolysis pathways makes it a central player in cellular homeostasis. Given its prominent role in eukaryotic cells, p97 malfunction is associated with human disease[106,206,272,273]. Single amino acid mutations in p97 (e.g., R155H, Figure 4-1A) cause multisystem proteinopathy type 1 (MSP1), also known as inclusion body myopathy with early-onset Paget disease of the bone and frontotemporal dementia (IBMPFD). Some p97 mutations caused amyotrophic lateral sclerosis (ALS)[274]. Variants in p97 have been identified in 2% of familial IBMPFD cases[83,104,275].

p97 is a 540 kDa homo-hexamer (6 × 89 kDa) in which each monomer contains an N-terminal domain (NTD, residues 1–184) and two tandem AAA+ ATPase domains, D1 (residues 211–459) and D2 (residues 482–762) domains, and a C-terminal tail (residues 763 to 806) (Figure 4-1A). The D1 and D2 domains contain highly conserved Walker A and B motifs and a second region of homology (SRH). During an ATPase cycle, the D2 domain undergoes significant conformational changes, whereas ATP hydrolysis induces only minor changes in D1. The D1 domain is required to form stable hexamers, regardless of the nucleotide-binding state[196,276]. Cooperativity between the D1

and D2 domains is essential for competent ATPase function[236]. However, the mechanisms of motion transmission and crosstalk between the two domains remain to be elucidated.

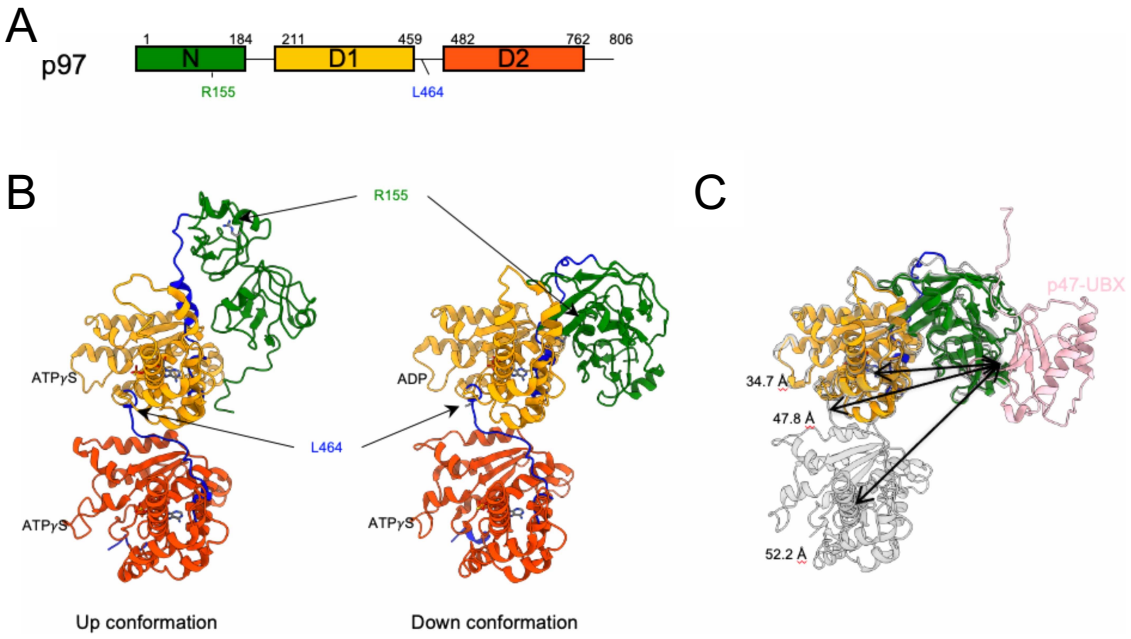


Figure 4-1. p97 domain organization (A) p97 AAA+ ATPase domains. p97 consists of an N terminal region (green; residues 1–184) containing cofactor binding sites and the disease-related, conserved arginine residue (R155), followed by the D1 (orange; residues 211–459) and D2 (orange red; residues 482–762) ATPase domains. These two domains are joined by a linker (blue; residues 460–481), containing a highly conserved leucine residue (L464). (B) Structures of the p97 monomer with the N-domain in conformations. At the left is the up-conformation (PDB code: *5FTN*), and at the right is the down-conformation (PDB code: *5FTM*)[224] (C) Superposition of the crystal structure of the $p97^{ND1}$–$p47^{UBX}$ (PDB code: *1S3S*)[105] with the structure of the full-length p97 in the down conformation (grey; PDB code: *5FTM*). The structure of the p47-UBX is shown in pink. The distances of p47-UBX from the D1 domain, the D1–D2 linker, and the D2 domain are 34.7 Å, 47.8 Å, and 52.2 Å, respectively.

Both D1 and D2 domains are organized into a double-ring structure. Cryo-EM structures of full-length p97 showed that the NTD adopts positions up-conformation, down-conformation, and mobile[81,224,277]. The NTD binds to various p97 cofactors, which coordinate p97 involvement in major cellular pathways. The largest group of known p97

cofactors are proteins that contain a ubiquitin-regulatory X (UBX) domain (13 members)[230]. The group is further subdivided based on the presence of an N-terminal ubiquitin-associated (UBA) domain, versus UBX-only proteins. Examples are p37 (UBX-only) and p47 (UBA–UBX)[205,230]. The UBA domain binds the ubiquitinated substrates, and the UBX domain mediates interactions with p97. p47 was the first discovered p97 UBX cofactor and is required for p97-mediated membrane fusion[76,278]. p37 is highly homologous to p47 in protein sequences but is involved in distinct pathways required for Golgi and endoplasmic reticulum (ER) biogenesis[78]. One of well-characterized p97 cofactors is a heterodimer of Npl4 and Ufd1 (NU). It binds to p97 at distinct sites to form a complex with the stoichiometry of one NU heterodimer per p97 hexamer and is involved in p97-dependent processes[279-282]. Npl4 is an effective target of the potential anti-cancer drug, disulfiram[258,283].

The N–D1 linker (residues 185–210) and the D1–D2 linker (residues 460–481) connect NTD–D1 and D1– D2 domains, respectively (Figure 4-1A). This structural arrangement is also found in another AAA ATPase, N-ethylmaleimide-sensitive factor (NSF), the closest homolog of p97. In NSF, the N–D1 linker is involved in motion transmission from the D1 domain to the N-domain, as well as auto-inhibition of NSF activity[284]. Previous work on p97 suggests that flexibility in the D1–D2 linker region is required for ATP hydrolysis dependent motion transmission between D1 and D2. Mutations in this region lead to the accumulation of an endoplasmic reticulum-associated degradation (ERAD) substrate, T cell receptor alpha chain fused to GFP (TCR-α-GFP)[227,285]. p97 proteins containing only the NTD and D1 domains without the D1–D2 linker domain (ND1) exhibit only 1.4% of the ATPase activity of the full-length wild-

type p97 ($p97^{WT}$) protein. In contrast, NTD and D1 domains containing the D1–D2 linker (ND1L) exhibit 79% of $p97^{WT}$ activity[93]. Thus, the linker region is crucial for D1 ATPase competence, but it is unclear how the D1–D2 linker contributes to D1 ATPase activity.

The D1–D2 linker region of p97 is important for motion transmission and D1 activity[227,253,285]. Tang and Xia reported that the linker activates D1 by breaking the six-fold symmetry of the D1 ring and that half of the D1–D2 linker region is highly conserved from humans to fungi[286] . Our work showed that disease-causing mutations in p97 are associated with improper cofactor-dependent regulation of ATPase activity, demonstrating the importance of cofactor regulation to p97 function[94]. In addition, MSP1 disease mutations caused differential binding of adaptor proteins (p37 and p47) to the NTD of p97 in cells[246,287].

The present study connects these areas of p97 mechanics by evaluating D1–D2 linker region participation in cofactor-regulated ATPase activity and in regulating the N-domain conformation via solving full-length $p97^{L464P}$ structures using single-particle cryo-EM. We studied the effect of three p97 cofactors on D1–D2 linker mutants constructed from full-length p97 and on a truncated p97 protein containing the N and D1 domains and the D1–D2 linker (ND1L). Our results suggest that the D1–D2 linker region of WT p97 has essential roles in cofactor-regulated ATPase function and its conformation changes will relay to the N-domains, D1 and D2 ATP binding sites. This provides further insight into why p97 mutants cause the accumulation of the ERAD substrate TCR-α-GFP[285], and highlights the importance of linker region in cofactors-regulated p97 ATPase function.

4.2 Materials and Methods

4.2.1 Ubiquitin Fusion Degradation Pathway Reporter Assay

The HeLa cells that stably expressed Ub^{G76V}-GFP are described previously[288,289]. Cells were treated with TrypLE Select (Thermo Fisher) and counted the day before transfection. Cells were then plated, 3000 cells per well, in 30 ml of 10% FBS clear growth medium. Cells were observed to be 80–90% confluent the day of transfection. Transfection complexes for each sample were prepared by mixing 2 mg of DNA, 3 ml of BioT reagent (Bioland Sci), and 100 ml of Serum-free DMEM, and incubating for 25 min at 25°C. An amount of 10 ml of the DNA-BioT complexes were then added directly to each well-containing cells and mixed on a rocker. Plates were incubated at 37°C in a CO_2 incubator for 24 h, then read using ImageXpress (Molecular Device).

4.2.2 Preparing the p97 ATPase Proteins

Plasmids used for generating p97 and ND1L proteins can be found in following Table 4-1. Proteins were purified as previously described[236,290].

Plasmid Number	Plasmid name	Vector	Source and Reference
TCB-197	Human p97	pET15b_TEV linker	Chou, 2014[236]
TCB-215	L464A Human p97	pET15b_TEV linker	This study
TCB-217	L464E Human p97	pET15b_TEV linker	This study
TCB-216	L464P Human p97	pET15b_TEV linker	This study
TCB-201	Human ND1L p97 (1-	pET24b	Chou, 2014[236]

	480) pET24b		
TCB-224	L464E Human ND1L p97 (1-480) pET24b	pET24b	This study
TCB-223	L464P Human ND1L p97 (1-480) pET24b	pET24b	This study
TCB-125	Ufd1	pET-Duet	Hanzelmann 2011[213]
TCB-126	Npl4	pCOLA-Duet	Hanzelmann 2011[213]
TCB-328	Rat p47	pET15b_TEV linker	Chou, 2014[236]
TCB-327	Human p37	pET15b_TEV linker	Zhang, 2015[94]
TCB-213	E305Q human p97	pET15b_TEV linker	Chou, 2014[236]
TCB-221	E578Q human p97	pET15b_TEV linker	Chou, 2014[236]
TCB-49	WT p97-myc His	pcDNA4.1TO	Dalal, 2004[291]
TCB-77	QQ p97-myc His	pcDNA4.1TO	This study
TCB-81	WT p97-myc His	pcDNA4.1TO	This study
TCB-456	L464P-p97-myc His	pcDNA4.1TO	This study

Table 4-1. Plasmids used in this study.

4.2.3 Determining IC_{50} Values of p97 Inhibitors in ATPase Assays

The detailed method was described previously[93]. Inhibition of human p97 (25 nM monomer) was carried out in assay buffer (50 mM Tris (pH 7.4), 20 mM $MgCl_2$, 1 mM EDTA and 0.5 mM TCEP) containing 0.01% Triton X-100 and 200 mM ATP. The 8-dose titration was performed at 40, 13.3, 4.4, 1.48, 0.49, 0.16, 0.05, and 0 mM. ATPase activity was determined through the addition of Biomol Green Reagent (Enzo Life Sciences). NMS-873 was purchased from Xcess Biosciences Inc.

4.2.4 Gel Filtration to Confirm the Stability of the Mutant p97 and p47 Complex

Gel filtration was carried out with a Superdex 200 10/300 GL (GE Healthcare). The column was calibrated with molecular weight (MW) standards kit (Sigma). The standards are Blue Dextran [MW is 2000 kDa; used to determine the void volume (V_o) of the column], β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), and carbonic anhydrase (29 kDa). The linear molecular weight calibration plotted V_e/V_o versus Log MW (see standard curve), where V_e is the elution volume of the MW standard and V_o is the void volume. To determine the apparent molecular weight of WT-ND1L, L464E-ND1L, and L464P-ND1L, we injected 100 ml of 20 mM proteins, determined their elution volumes (V_e), and then calculated apparent molecular weights using the equation obtained from the standard curve. The calculated MW was divided by the monomer MW to calculate the oligomeric state.

4.2.5 Cryo-EM Data Collection and Image Analysis

For $p97^{L464P}$ mutant, a holey-carbon C-flat grid (2/1 4C; Protochips, Morrisville, NC) was glow-discharged for 15 s using a Pelco easiGlow glow-discharge system (Ted Pella, Redding, CA). An amount of 6 µl protein sample at a concentration of 0.2 mg/ml was applied onto the pretreated grid, and the excess solution was blotted away by a filter paper (Whatman 55/20 nm) using a Thermo Fisher/FEI Vitrobot Mark IV automated freeze plunger (Thermo Fisher/FEI, Hillsborough, OR) for 6 s in a chamber with a humidity of 100%. The grid was then quickly plunged frozen into liquid ethane and transferred to a grid storage. Particle homogeneity and ice thickness of the grid specimen

were screened using a FEI Tecnai TF20 TEM. Grids with a thin ice and a homogeneous protein dispersion were used for the subsequent cryo-EM data collection.

The grid specimen was imaged using a Thermo Fisher/FEI Titan Krios TEM at an accelerating voltage of 300 keV at the Eyring Materials Center (EMC) at Arizona State University (ASU), Tempe, AZ. Cryo-EM movies were recorded using a Gatan K2 Summit direct electron detector (DED) camera (Gatan, Pleasanton, CA). Defocus range was set to −0.8 to −2.5 mm. Nominal magnification was 48 077×, resulting in a physical pixel size of 1.04 Å/pixel at the specimen level. The movie data was recorded at a counted rate of 8.096 e^-/ pixel/sec and a sub-frame rate of 200 msec in counting mode. Total exposure was 8 s, accumulating to a dosage of 64.8 $e^-/Å^2$. Beam-image shift was applied to accelerate the data acquisition[292]. The data collection was automated using the customized SerialEM macros (version 3.9)[169].

The image processing for the $p97^{L464P}$ dataset was generally conducted using cryoSPARC software (version 3.1)[52]. 4490 movies were unpacked and imported into cryoSPARC program. The frame registration and averaging were performed using patch motion correction. The defocus parameters were estimated using patch CTF estimation. The particle selection was performed using Topaz program (version 0.2.3)[170]. The two-dimensional (2D) unsupervised classification was performed to generate the averages from the selected particle pool. The classes with discernible features were selected for *ab initio* volume generation using stochastic gradient descent (SGD) method[52]. Further particle image curation was performed using 2D image classification, *ab initio* 3D generation supplied with a higher k value and heterogeneous refinement. The final 3D density was then refined against the experimental images using homogeneous refinement

and non-uniform refinement[53]. The generated maps reached final resolutions of 4.86 Å. The final resolution of the density map was determined using the golden FSC criteria at 0.143 cutoff[293]. To visualize the details of the D1 and D2 domains, the densities of the D1 and D2 domains were symmetrically averaged supplied with a local mask only containing D1 and D2 domains. The resolution of the symmetrically averaged density reaches 4.12 Å. The processing schematics and data statistics are shown in Figure 4-9 and Table 2 (Appendix section).

4.2.6 Model Building

The templates used for atomic modeling were the previous coordinates of p97 (PDB code: *5FTK*)[224]. The templates were first docked into individual cryo-EM densities using 'Fit in the Map' function in UCSF Chimera (version 1.14)[171]. The fitted model was manually rebuilt using Coot (version 0.9.4)[11] and then refined against the cryo-EM densities using 'phenix.real_space_refine' program in Phenix software package (version 1.18.2-3874)[294]. Hydrogen atoms were added using 'phenix.reduce' for the model refinement and removed after the refinement. Secondary structure restraints were applied during the model refinement. The refinement and validation statistics was listed in Table 2 (Appendix section). The figures for the cryo-EM density maps and atomic models were prepared using UCSF Chimera or ChimeraX (version 1.0)[295].

4.2.7 Biomol Green ATPase Assay

Detection of ATPase activity using BIOMOL Green reagent (Enzo Life Sciences) was performed as previously described[94]. Purified p97 or ND1L proteins (25 ml of 50 mM;

final concentration = 25 nM monomer concentration) were used with 400 nM cofactor proteins. For generating Michaelis–Menten (MM) curves, at least eight concentrations of ATP were utilized in at least six replicate reactions, and all calculations were made using GraphPad Prism 7.0 by fitting data to Equation 1.

$$n ¼ V_{max} [ATP]= (KM þ [ATP]) \qquad (1)$$

4.2.8 Molecular Graphics

The figure of the p97 structures was prepared using UCSF ChimeraX[295] with the atomic coordinates from the PDB codes of 5TFN (up)[105], 5TFM (down)[224], and 1S3S ($p97^{N\text{-}D1}$–$p47^{UBX}$)[105].

4.2.9 p47 Binding Affinity Measurements

The binding affinity (K_d values) between p47 and $p97^{WT}$ or p97 mutants was determined by measuring the temperature-related intensity change (TRIC) signals using a Dianthus NT.23 instrument (Nano-Temper Technologies, München, Germany). p47 was labeled with a RED-NHS dye using the Monolith Protein Labeling Kit RED-NHS second-generation (Nano-Temper Technologies, CAT# LO-L011). Full-length wild type or mutant p97 was titrated against 10 nM of p47-NHS in two-fold steps from 2.8 mM to 1.37 nM in 20 ml working buffer (20 mM HEPES (pH 7.4), 150 mM KCl, 1 mM $MgCl_2$, 5% (w/v) glycerol, and 0.0025% (v/v) Tween 20). Assays were performed in a Dianthus 384-well plate in three independent experiments. Data from three independent measurements were fitted using non-linear regression analysis in Prism 7.0.

4.3 Results and Discussion

4.3.1 L464 Mutations Affect the Degradation of the Ubiquitin Proteosome Substrate Ub^{G76V}-GFP

The N–D1 linker and the D1–D2 linker (residues 460–481) connect NTD–D1 and D1–D2 domains (Figure 4-1A). The p97 crystal structure features an NTD linked to two stacked AAA domains (Figure 4-1B). The NTD binding site for p47-UBX (pink; Figure 4-1C) resides 37.4 Å and 51.6 Å away from the D1 and D2 nucleotide-binding sites, respectively. Therefore, cofactor regulation of ATPase kinetics must involve interactions relayed through the NTD, D1, and D2 domains, as well as the linker regions. We are intrigued by the interesting model and the result suggested that the D1–D2 linker region of p97 is important for motion transmission and to activate D1 activity[227,253,285]. Li and coworkers demonstrated that p97 mutants L464A, L464E and L464P cause substantial accumulation of TCR-α-GFP, indicating a dramatic loss of ERAD activity[285]. To determine whether the L464 amino acid is critical for the degradation of another p97 dependent proteasome substrate, we expressed L464A, L464E or L464P p97 plasmids in HeLa cells that stably expressed Ub^{G76V}-GFP[288]. We observed an accumulation of Ub^{G76V}-GFP in the L464A, L464E and L464P mutants. We used a p97 ATPase dead mutant as a positive control, which was created by mutating the conserved glutamate residues of the Walker B motifs to glutamines in both the D1 and D2 domains and resulting in single mutants D1-E305Q and D2-E578Q. The mutants were then combined as a double mutant D1-E305Q/D2-E578Q (QQ). This led to the double mutant that allows to bind ATP but is defective in hydrolysis[211]. Strong accumulation of Ub^{G76V}-GFP

was also observed in QQ, the ATPase dead mutant. (Figure 4-2). We confirmed that the conserved L464 is a critical amino acid for p97 to maintain its role in mediating degradation of p97-dependent ubiquitin-proteasome substrates Ub^{G76V}-GFP[289]. These results suggest that the structural arrangement at position 464 is crucial for proper p97 function, we then evaluated the biochemical and structural effect of L464 by study the effect on cofactor regulated ATPase activity of these mutants. We selected three previously studied cofactors, p47, p37 and Npl4–Ufd1[94,296].

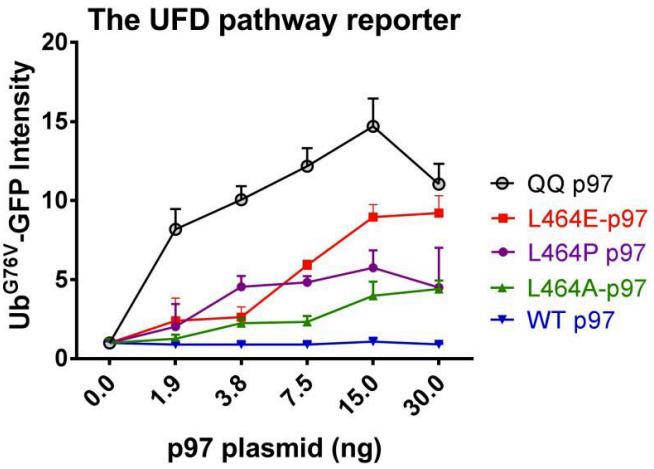


Figure 4-2. UFD pathway reporter assay titration of p97 variants- Increasing amounts of plasmid (0-30 ng) containing full-length variants E305Q/578Q (QQ), L464A, L464E, L464P and WT p97 were transfected into cells stably expressing the UFD substrate UbG76V to measure accumulation of ubiquitin fusion degradation (UFD) pathway products.

4.3.2 D1-D2 Linker Mutants are Defective in p47 Regulated p97 ATPase Activity

We and others have previously used cofactor regulated ATPase activity as a function readout of cofactor binding to N-domain and communicate to D1 and D2 activity activities[94,296]. To evaluate whether the conserved L464 is critical for the cofactor-

regulated p97 ATPase activity, we purified full-length L464A, L464E and L464P p97 mutants. Consistent with previous reports[285], we found that the basal ATPase activity of mutants L464A and L464E was similar to that of $p97^{WT}$ (Figure 4-3A). However, the proline mutant L464P had a two-fold reduction in basal ATPase activity (49% of $p97^{WT}$), suggesting that the proline side chain possibly changes the D1–D2 linker conformation thereby reducing the basal ATP hydrolysis of $p97^{253}$. We then performed cofactor titration experiments on these three L464 mutants to determine the effects of these mutations on cofactor-regulated ATPase activity. Regulation of $p97^{WT}$ activity by p47 is biphasic, with strong inhibition at lower concentrations of p47 (Phase 1), and less inhibition at higher stoichiometries of the cofactor (Phase 2, or the 'rebound' phase)[94]. With the three L464 mutants, however, Phase 1 inhibition at low concentrations of p47 was abolished (Figure 4-3B), suggesting that this conserved residue is required for proper cofactor-regulation of the D1 and D2 domains. Moreover, during Phase 2, a >2.2-fold increase in ATPase activity was observed for $p97^{L464A}$ and $p97^{L464E}$ and 3.8-fold increase was for $p97^{L464P}$ — far beyond the usual $p97^{WT}$ level (0.8-fold). It appears that the conserved L464 residue within the D1–D2 linker plays an important role in p47-regulated p97 ATPase activity. This p47 concentration-dependent effect is similar to that of the D1-E305Q mutant, in which the D1 ATPase is unable to perform ATP hydrolysis. Thus, the linker mutants seem to block communication between D1 and D2, and the L464 linker residue is critical for mediating D2 domain cross inhibition by D1 during p47 binding.

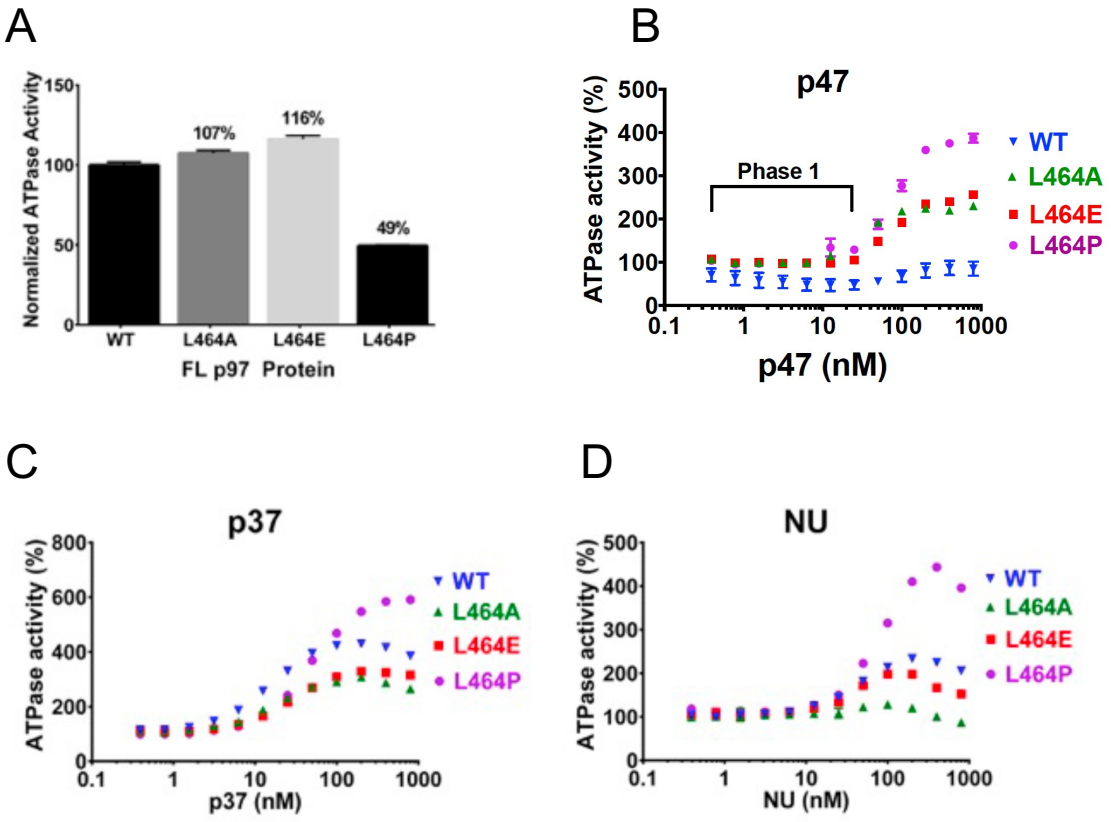


Figure 4-3. ATPase activity of full-length (FL) D1–D2 linker mutants and concentration-dependent effects of cofactors(A) ATPase activity of FL p97 proteins normalized to WT activity and measured with 200 mM ATP (n = 8). (B–D) ATPase activity of WT-p97 (blue) and D1-linker mutants L464E (red), L464A (green), and L464P (purple) at increasing concentrations (0–800 nM) of cofactors (B) p47 (C), p37, and (D) Npl4–Ufd1 (NU). Experiments were performed using p97 proteins at 4.17 nM hexamer and 200 mM ATP, and the activity of each p97 protein was normalized to its basal activity in the absence of cofactors (n = 6). All error bars indicate ± SD.

4.3.3 L464 Mutants Display p37 and NU Activation Profiles Like Those for $p97^{WT}$

Instead of biphasic regulation of p47, we have shown that p37, a p47 homolog, activates the $p97^{WT}$ function (4.2-fold)[78,94]. In the presence of p37, the D1–D2 linker mutants followed an activation profile like profiles for $p97^{WT}$. L464A and L464E had slightly decreased activation by p37 (3.1- and 3.3-fold increases in ATPase activity, respectively),

whereas the increase in ATPase activity for L464P reached a maximum of 5.8-fold at 800 nM p37 (Figure 4-3C). We further investigated the binding of the Npl4–Ufd1 (NU) with p97 ATPase. Electron microscopic (EM) study demonstrated that the NU interacts with p97–NTD via the Npl4-UBX-L domain, which induces a conformational change for ATP hydrolysis[279,281,282,297]. Our titration results identified the NU as an activator of $p97^{WT}$ ATP hydrolysis, starting at 20 nM and reaching a maximal 2.2-fold increase in activity at 200 nM, consistent with previously published studies (Figure 4-3D)[296,298]. Using the Walker B mutants of p97 (D1-E305Q and D2-E578Q) (Figure 4-6B), we showed that D1-E305Q and $p97^{WT}$ have similar titration results. The result suggests that the NU activates D2 ATPase, because the D1 dead mutant (E305Q-p97) can hydrolyze ATP by the D2 ATPase, but not the D1 ATPase. The NU-mediated increases in ATPase activity (Figure 4-3D) for linker mutants L464A and L464E were slightly less than the increases for $p97^{WT}$. The activation of the $p97^{L464P}$ mutant was 4.4-fold, about twice that of the $p97^{WT}$, and like the effect of p37. This supports that cofactor regulation by p37, and NU is affected by D1–D2 linker mutations to some extent.

4.3.4 Steady-state Kinetic Analysis of the D1-D2 Linker Mutants and p97-cofactor Complexes

To characterize the changes observed during the cofactor titrations, we measured the apparent steady-state kinetic constants for ATP hydrolysis for each enzyme-cofactor complex. The $p97^{L464P}$ mutant cofactor interactions produced the most significant changes in Michaelis–Menten (MM) curves (Figure 4-4A). The apparent kinetic constants are summarized in Table 4-3 and their ratio to without cofactor are shown in Table 4-5. The

apparent maximum rate of ATP turnover (k_{cat}) for the $p97^{L464P}$ mutant was 3.3-fold lower than for $p97^{WT}$ (1.5 versus 4.7 min^{-1}) (Figure 4-4B). However, the apparent Michaelis constant (K_M) for $p97^{L464P}$ did not differ substantially from the $p97^{WT}$ (Figure 4-4C). Thus, the apparent overall catalytic efficiency (k_{cat}/K_M) of the $p97^{L464P}$ mutant was 3.8-fold lower than that of $p97^{WT}$, primarily due to the decreased rate of ATP hydrolysis (Figure 4-4D). $p97^{L464A}$ and $p97^{L464E}$ mutants had apparent k_{cat}, apparent K_M, and apparent k_{cat}/K_M values within 2-fold of $p97^{WT}$ values (Figure 4-4B-D).

p47 decreased the apparent ATP turnover number slightly for $p97^{WT}$ but increased the apparent k_{cat} for all three L464 mutants (Figure 4-4B), most pronounced for $p97^{L464P}$ (3.3-fold; Figure 4-4B) and comparable to levels for $p97^{WT}$ alone (5.0 min^{-1} vs. 4.7 min^{-1}, respectively). The apparent K_M values of $p97^{WT}$ and all three mutants, L464A, L464E and L464P, were all decreased by p47 compared with its basal ATPase activity without p47, supporting an interpretation that L464 mutations primarily affect p47-mediated changes in hydrolysis (apparent k_{cat}, in Figure 4-4B) more than that of ATP binding (apparent K_M, in Figure 4-4C). Compared with WT p97, the apparent K_M value is generally higher for the Linker mutants, thus the linker does affect ATP binding in the absent of p47. The presence of p47 rescued the defect of L464P in ATP hydrolysis (apparent k_{cat}) and made WT p97 and L464P similar. We previously demonstrated that p47 inhibits D2 ATPase hydrolysis, which requires active D1 hydrolysis[94]. Here, our results suggest that during p47 binding, L464 is required for the cross-inhibition of D1 on D2-driven ATP hydrolysis. Unlike p47, p37 and NU complexes produced similar steady-state kinetic profiles within 2-fold of $p97^{WT}$ values across the three mutants. This suggests that the distinct conformational changes that occur during p37 and NU binding were not

significantly impaired by these L464 mutations. Thus, p97 control by p37 and NU differs from p47-mediated regulation.

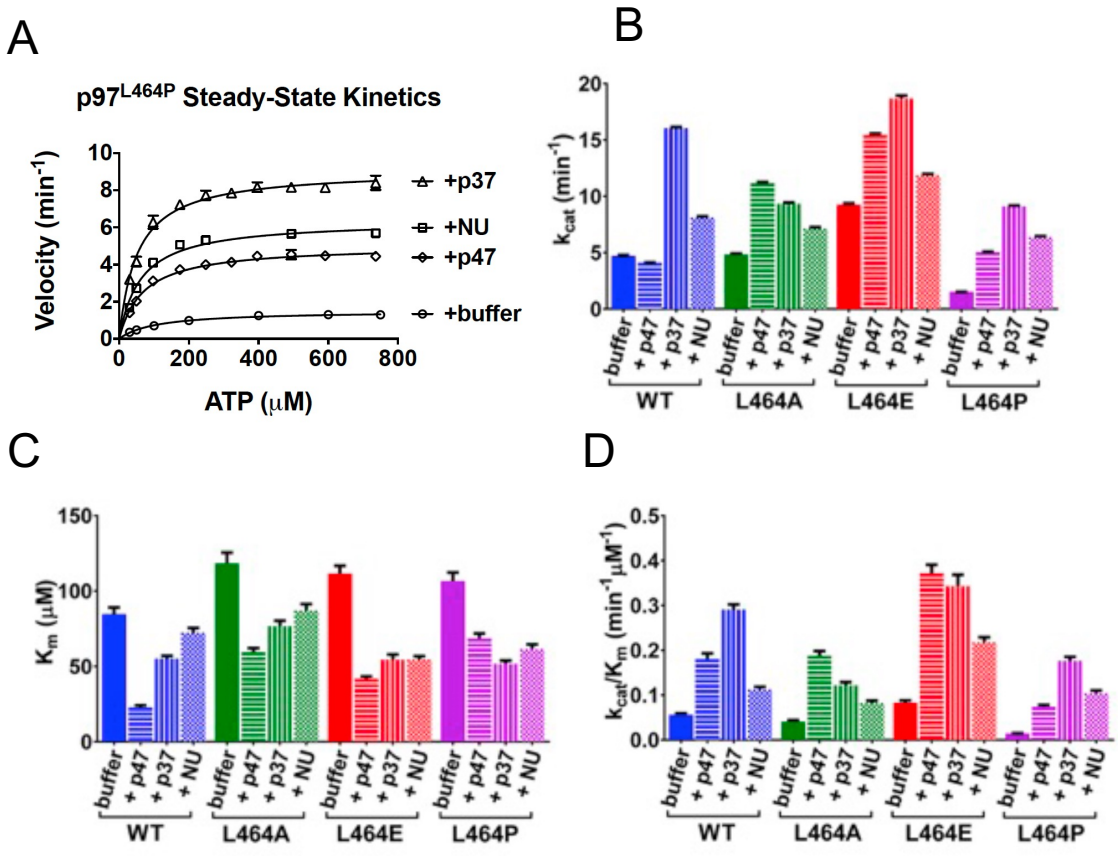


Figure 4-4. Steady-state kinetic analysis of full-length D1–D2 linker mutants and cofactor-enzyme complexes(A) Michaelis–Menten curves of ATP hydrolysis by L464P-p97 with and without cofactors (n = 10). (B–D) Steady-state kinetic analyses of WT, L464A, L464E, and L464P-p97 catalyzed ATP hydrolysis performed alone (solid bar) and in the presence of 400 nM p47, p37, or Npl4–Ufd1 (NU) (n = 8). Dependence of apparent k_{cat} (B), apparent K_M (C), and apparent k_{cat}/K_M (D) values on cofactors p47 (horizontal stripes), p37 (vertical stripes), and NU (checkered) are shown for WT-p97 (blue), L464A (green), L464E (red), and L464P (purple). Experiments were performed using p97 proteins at 4.17 nM hexamer and 30–750 mM ATP. All error bars indicate ± SD.

Taken together, these findings indicate that the L464, located between the D1 and D2 domains (Figure 4-1C), does not affect the apparent Michaelis constant (K_M) (Figure

4-4C) but is instead critical for the hydrolysis steps of both basal and cofactor-regulated ATPase cycles of full-length p97. The $p97^{L464P}$ mutant had a 3.1-fold reduction in the k_{cat} for basal ATPase activity, which was rescued by all three cofactors, shown by increases in apparent k_{cat} of 3.3-fold (with p47), 6.0-fold (with p37), and 4.2-fold (with NU).

4.3.5 Elucidating the Effect of Linker Mutants on D1 ATPase Activity

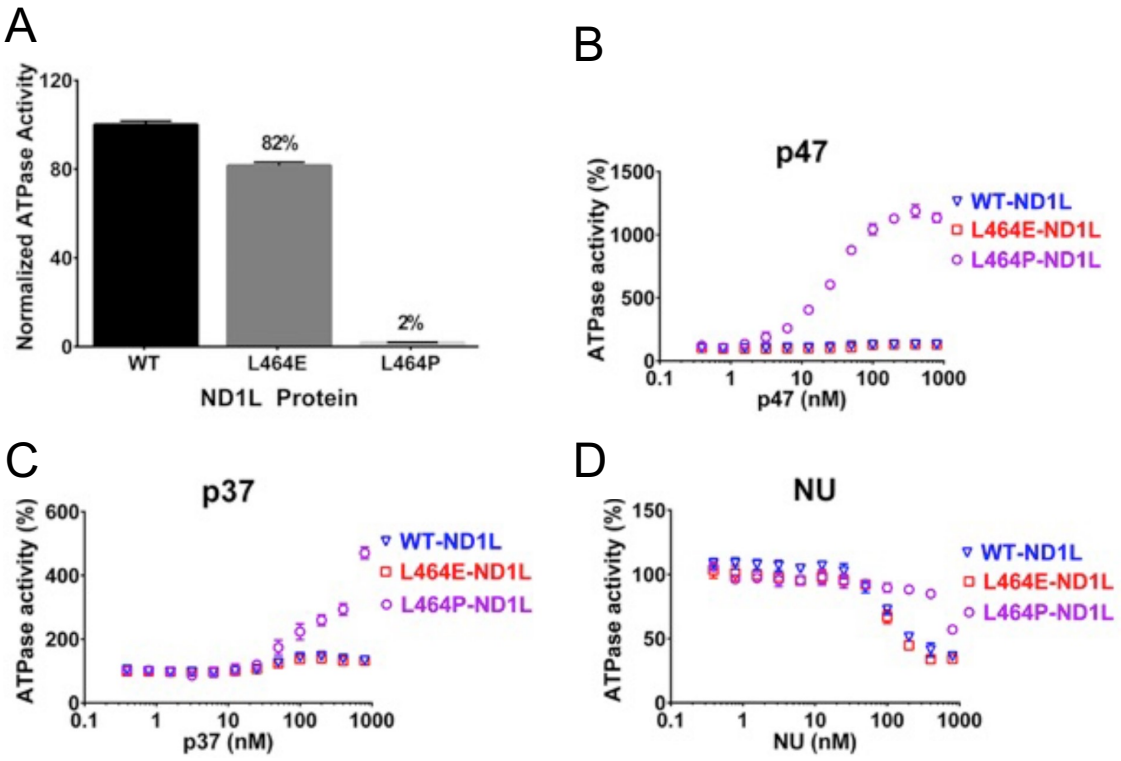


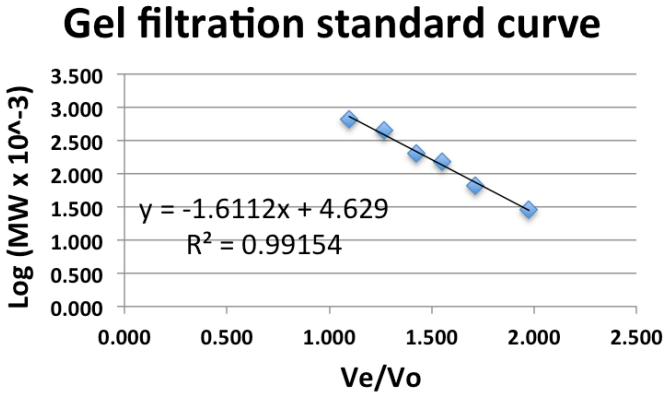
Figure 4-5. ATPase activity of WT–ND1L and ND1L-linker mutants and concentration-dependent effects of cofactors (A) ATPase activity of ND1L proteins normalized to WT-ND1L (n = 8). (B–D) ATPase activity of WT–ND1L (blue open triangles) and ND1L-linker mutants L464E-ND1L (red open squares) and L464P-ND1L (purple open circles) at increasing concentrations (0–800 nM) of cofactors p47 (B) (n = 6), p37 (C) (n = 12), and Npl4–Ufd1 (NU; D) (WT and L464E n = 9; L464P n = 4). Experiments were performed using ND1L proteins at 4.17 nM hexamer for WT and L464E, 20.85 nM hexamer for L464P-ND1L, and 200 mM ATP. All error bars indicate ± SD.

Recently, NMR (nuclear magnetic resonance) spectroscopic analysis revealed that the WT–ND1L (residues 1–480 of p97) binding to monomeric p47 in solution with a K_d value of 230 nM determined by ITC[235]. The ND1L p97 protein is a robust, catalytically competent ATPase, like the D1 ATPase in full-length p97[236]. Notably, ND1 proteins lacking the D1–D2 linker are enzymatically dead, indicating that the linker region is vital to D1 activity[236]. Therefore, ND1L can be used to scrutinize the D1-specific changes in ATPase kinetics for the linker mutants, with and without cofactor binding. Like the full-length version, L464E-ND1L exhibited basal ATPase activity (81.6%) comparable to activity for WT–ND1L. In contrast, L464P-ND1L retained only 1.7% of the activity of WT–ND1L (Figure 4-5A), a level like that of ND1 (1.8% of WT–ND1L activity)[236]. The result suggests that introducing proline 464 into ND1L significantly impairs the alignment of residues into the proper conformation so diminish the ability of D1 to hydrolyzing ATP as if the linker is completely removed. (Figure 4-5A).

Size-exclusion chromatographic (SEC) profiles confirmed that the loss of activity in L464P-ND1L was not due to differences in the formation of proper oligomers, as compared with WT–ND1L. All three proteins had the same elution volume, and the oligomeric state of each was calculated to be 7.6 (Table 4-2). Based on published ND1 structures, we think ND1L is a hexamer; the calculated 7.6 oligomeric state for all the ND1L proteins suggested their oligomeric state is the same. One crystal structure of p97 D2 domain showed a heptamer but the rest of p97 structures are all hexamers[299]. As with the full-length protein, dose-dependent cofactor titrations found differences in cofactor-regulated ATPase activity in wild-type ND1L versus ND1L mutants. In the presence of p47, the titration curves of WT–ND1L and L464E-ND1L remained flat, with activities

close to baseline levels (Figure 4-5B). However, L464P-ND1L had a dramatic 12-fold increase in ATPase activity at 800 nM p47, suggesting that the p47 binding to NTD causes a conformational change that alleviates the L464P defects in D1 hydrolysis.We observed minor increases in the ATPase activity (1.3-fold) of the WT–ND1L with p37 (Figure 4-5C), and a 3.7-fold increase for full-length $p97^{WT}$ (Figure 4-3C), suggesting that p37 activation occurs primarily at D2 ATPase, consistent with previous data[94] . L464P-ND1L ATPase activity rose 4.7-fold with 800 nM p37. Heterodimer NU is an activator of full-length p97 (Figure 4-3D). Yet, we found that NU inhibited hydrolysis in three ND1L proteins (Figure 4-5D). For L464P-ND1L, we were unable to detect activity above background levels when using 25 nM monomer, due to the low basal activity of L464P-ND1L and the inhibitory effect of NU. We therefore increased the L464P-ND1L monomer concentration to 125 nM (5-fold) for NU titration. Still, we did not detect activity changes over 0.4 to 600 nM NU. At 800 nM (the equivalent of 160 nM NU per 25 nM L464P-ND1L monomer), we observed nearly 2-fold inhibition. WT-ND1L and L464E-ND1L variants followed similar titration curves, with inhibition starting at 60 nM, and ~2-fold decreases in ATPase activity at 200– 800 nM NU (Figure 4-5D).

Molecular Weight Standard	Molecular Weight (kDa)	Elution Volume (mL)	Ve/Vo	Log MW
Carbonic Anhydrase	29	16.52	1.974	1.462
Bovine Serum Albumin	66	14.32	1.711	1.820
Alcohol Dehydrogenase	150	12.97	1.550	2.176
β-Amylase	200	11.94	1.427	2.301
Thyroglobulin	443	10.62	1.269	2.646
Apoferritin	669	9.18	1.097	2.825
Blue Dextran	2000	8.37	1.000	3.301



Proteins	Ve (mL)	Ve/Vo	Log (calculated MW)	Calculated MW (kDa)	Monomer MW (kDa)	Oligomer*
WT-ND1L	10.42	1.24	2.62	419	55	7.6
L464E-ND1L	10.42	1.24	2.62	419	55	7.6
L464P-ND1L	10.42	1.24	2.62	419	55	7.6

* The calculated MW was divided by the monomer MW to calculate the oligomer number.

Table 4-2. Gel filtration to determine the oligomeric states of WT-ND1L, L464E-ND1L, and L464P-ND1L.

To examine NU-mediated D1 and D2 ATPase activities in full-length p97, we used Walker B mutants, D1E305Q and D2-E578Q, which allow ATP and ADP binding but block activation of hydrolysis. The NU titration curve for $p97^{WT}$ was like the curve for D1-305Q p97, whereas the D2-E578Q curve was relatively flat, with a small decrease (Figure 4-6). NU appears to inhibit D1 activity slightly while activating D2 activity in $p97^{WT}$. Comparison of titrations for ND1L (Figure 4-5D) and D2-E578Q (Figure 4-6) suggest that NU inhibits D1 activity more when the D2 domain is absent.

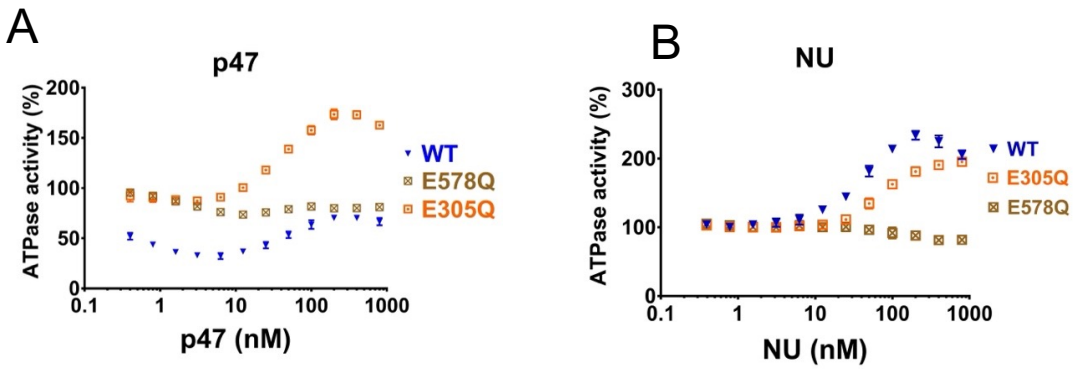


Figure 4-6. Concentration-dependent effects of cofactors on Walker B mutants: ATPase activities of WT, E305Q, and E578Q (4.17nM hexamer) were measured at increasing concentrations of p47 (A) and NU (B) at 200μM ATP. The activity of each p97WT protein was normalized to its basal activity in the absence of cofactor.

4.3.6 Steady State Kinetic Analysis of ND1L-domain Mutants

The apparent steady-state kinetic constants of the ND1L proteins (Tables 4-3,4-4, 4-5) with p47, p37 and NU were determined at 400 nM, the concentration that produced the maximum changes observed for wild-type ND1L. Examples of MM curves for L464P-ND1L are shown in Figure 4-7A. For WT-ND1L, only NU produced a >2-fold change in ATP turnover, reducing apparent k_{cat} values from 3.7 to 1.4 min^{-1} (Figure 4-7B and Table 4-5). All three cofactors produced >2-fold decreases in apparent K_M values, indicating that they enhanced ATP binding specifically at the D1 domain (Figure 4-7C). It appears that p37 and p47 enhance ATP binding to D1, thereby improving the apparent catalytic efficiency of D1 (k_{cat}/K_M). Intriguingly, NU increased the turnover number in full-length p97 but decreased the turnover number in ND1L, suggesting that NU specifically activates D2 catalysis.

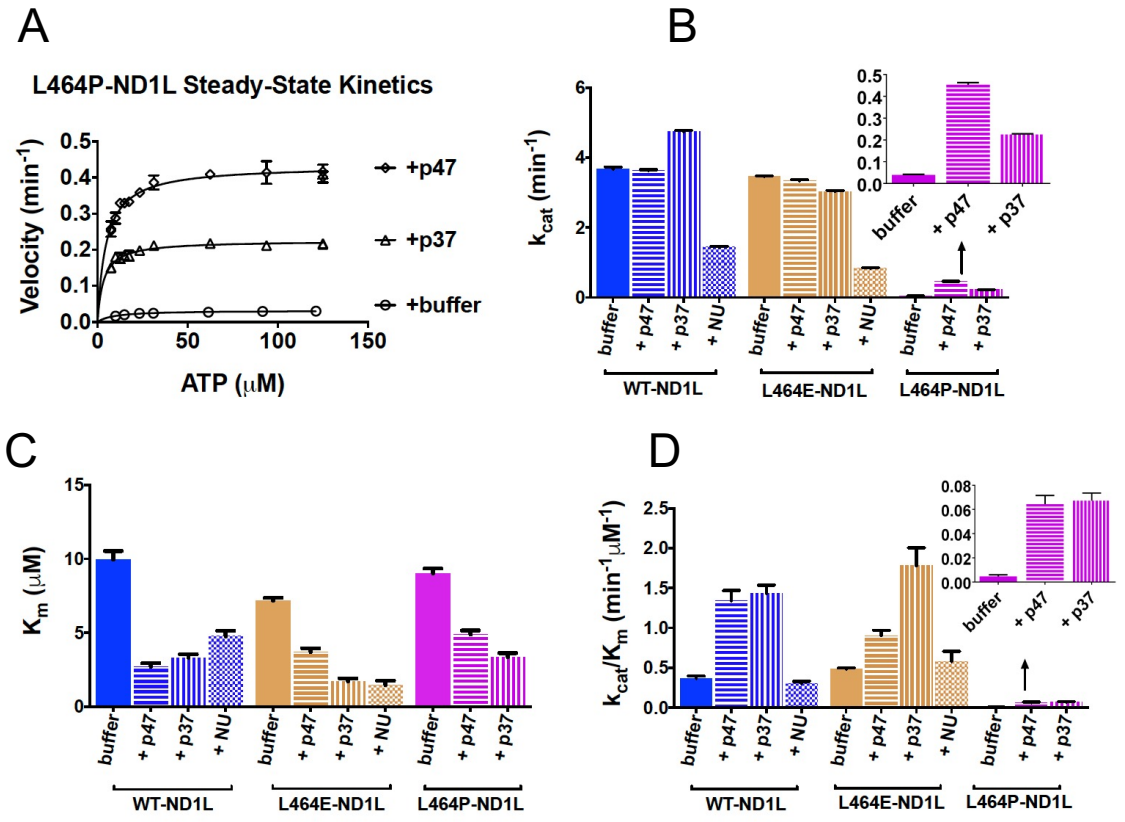


Figure 4-7. Steady-state kinetic analysis of WT-ND1L, ND1L-linker mutants, and cofactor-enzyme complexes (A) Michaelis–Menten curves of ATP hydrolysis by L464P-ND1L in buffer (circles), p47 (diamonds), and p37 (triangles). (B–D) Steady-state kinetic analyses of WT, L464E, and L464P-ND1L ATP hydrolysis performed alone (solid bar) and in the presence of 400 nM p47, p37, or NU. Dependence of apparent k_{cat} (B), apparent K_M (C), and apparent k_{cat}/K_M (D) values on cofactors p47 (horizontal stripes), p37 (vertical stripes), and NU (checkered) are shown for WT-ND1L (blue), L464E-ND1L (red), and L464P-ND1L (purple). Experiments were performed using ND1L proteins at 4.17 nM hexamer and 10–121 mM ATP. All error bars indicate ± SD and n = 9.

L464E-ND1L and WT-ND1L displayed similar kinetic constants with all three cofactors (Figure 4-7B–D), indicating that the hydrophobic nature of L464 is not a strict requirement for either basal ATPase activity or cofactor regulation at isolated D1 domains. In contrast, the L464P-ND1L dramatically reduced the apparent ATP turnover

rate (k_{cat}) of D1 by 116-fold, from 3.7 to 0.032 min^{-1}, with no significant change in the K_M value. This observation suggests that the position 464 has structural requirements that are vital for D1 ATPase hydrolysis but not for ATP binding, consistent with previous comparisons of nucleotide-binding for ND1 versus ND1L[236].

In the presence of p47, L464P-ND1L apparent k_{cat} values (Figure 4-7B) increased 14-fold, from 0.032 to 0.434 min^{-1}, without significant changes in apparent K_M (Figure 4-7C). It appears that binding to p47 primarily increases the catalytic ability of the L464P-ND1L protein, possibly through stabilization of the linker region that becomes distorted by the proline side chain. With p37, L464P-ND1L displayed a 7-fold increase in k_{cat} (Figure 4-7B), and the apparent K_M value decreased 2.7-fold (Figure 4-7C). This implies that p37, like p47, may also help stabilize the linker region. Cofactor complex NU inhibited L464P-ND1L to undetectable levels of ATPase activity across the measured ATP concentrations, like NU effects on WT-ND1L. Thus, we cannot report kinetic values for NU-regulated L464P-ND1L.

To summarize our results on full-length $p97^{WT}$ compared with ND1L: (1) p47 does not affect the activity of ND1L but does regulate $p97^{WT}$ activity in a biphasic manner (Figure 4-3B,4-5B). (2) p47 decreases the apparent K_M of both ND1L and $p97^{WT}$ (Figures 4-4C, 4-7C). (3) NU decreases ND1L activity while activating $p97^{WT}$ (Figures 4-3D, 4-5D). (4) p37 caused a 1.5-fold increase in activity for ND1L (Figure 4-5C) and a 4-fold increase for full-length p97 (Figure 4-3C), which supports an earlier finding that p37 mainly activates D2[94]. Overall, these results suggest that the mechanisms by which cofactors regulate D1 and full-length p97 (D1 and D2) differ, with distinct mechanisms for each cofactor.

Protein	Kinetic Constant		buffer	w/ p47	w/ p37	w/ NU
WT	k_{cat} (min^{-1})	Mean	4.71	4.09	16.0	8.13
		SE	0.06	0.04	0.1	0.09
	K_m (μM)	Mean	84.7	22.8	55.2	72.4
		SE	4.4	1.4	1.8	3.0
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.056	0.180	0.291	0.112
		SE	0.004	0.013	0.011	0.006
L464A	k_{cat} (min^{-1})	Mean	4.83	11.2	9.34	7.14
		SE	0.08	0.1	0.11	0.11
	K_m (μM)	Mean	118	59.4	76.6	87.0
		SE	7.0	2.7	3.6	4.3
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.041	0.188	0.122	0.082
		SE	0.003	0.010	0.007	0.005
L464E	k_{cat} (min^{-1})	Mean	9.25	15.4	18.7	11.9
		SE	0.12	0.1	0.2	0.1
	K_m (μM)	Mean	111	41.5	54.5	54.5
		SE	5.2	1.7	3.3	2.2
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.083	0.372	0.343	0.217
		SE	0.005	0.019	0.025	0.011
L464P	k_{cat} (min^{-1})	Mean	1.51	5.03	9.10	6.39
		SE	0.02	0.06	0.08	0.08
	K_m (μM)	Mean	106	68.5	51.7	61.5
		SE	5.7	3.3	2.2	3.1
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.014	0.073	0.176	0.104
		SE	0.001	0.004	0.009	0.006

Table 4-3. Steady-state kinetic constants for full-length p97 proteins with and without cofactors.

Protein	Kinetic Constant		buffer	w/ p47	w/ p37	w/ NU
WT-ND1L	k_{cat} (min^{-1})	Mean	3.67	3.63	4.75	1.45
		SE	0.06	0.03	0.04	0.02
	K_m (μM)	Mean	9.95	2.70	3.31	4.77
		SE	0.58	0.24	0.22	0.36
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.369	1.34	1.43	0.303
		SE	0.027	0.13	0.11	0.026
L464E-ND1L	k_{cat} (min^{-1})	Mean	3.46	3.34	3.04	0.84
		SE	0.02	0.03	0.02	0.01
	K_m (μM)	Mean	7.17	3.70	1.70	1.45
		SE	0.19	0.25	0.20	0.30
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.483	0.903	1.78	0.580
		SE	0.016	0.068	0.22	0.127
L464P-ND1L	k_{cat} (min^{-1})	Mean	0.032	0.434	0.225	N/A
		SE	0.002	0.005	0.003	
	K_m (μM)	Mean	9.02	4.87	3.35	N/A
		SE	0.32	0.28	0.28	
	k_{cat} / K_m (min^{-1}μM^{-1})	Mean	0.004	0.065	0.067	N/A
		SE	0.001	0.007	0.006	

Table 4-4. Steady-state kinetic constants for ND1L proteins with and without cofactors.

Enzyme	Constant	p47	p37	NU
WT	k_{cat}	0.87	3.4	1.7
	K_m	0.27	0.65	0.85
	k_{cat} / K_m	3.2	5.2	2.0
L464A	k_{cat}	2.3	1.9	1.5
	K_m	0.50	0.65	0.73
	k_{cat} / K_m	4.6	3.0	2.0
L464E	k_{cat}	1.7	2.0	1.3
	K_m	0.37	0.49	0.49
	k_{cat} / K_m	4.5	4.1	2.6
L464P	k_{cat}	3.3	6.0	4.2
	K_m	0.64	0.48	0.58
	k_{cat} / K_m	5.2	12	7.3
WT-ND1L	k_{cat}	0.99	1.3	0.39
	K_m	0.27	0.33	0.48
	k_{cat} / K_m	3.6	3.9	0.81
L464E-ND1L	k_{cat}	0.97	0.88	0.24
	K_m	0.52	0.24	0.20
	k_{cat} / K_m	1.9	3.7	1.2
L464P-ND1L	k_{cat}	13.6	7.0	N/A
	K_m	0.54	0.37	N/A
	k_{cat} / K_m	18	19	N/A

Table 4-5. The ratio of cofactor over buffer of each enzyme.

4.3.7 L464P Reduced p47 Binding to Full-length p97 but not to the ND1L Truncated Mutant

To evaluate whether changing the conserved L464 to E or P will affect p47 binding, we used temperature related intensity change (TRIC) on a DianthusNT.23 PicoDuo instrument to determine the equilibrium binding constant (K_d) for p47 in solution with six p97 proteins (Figure 4-8). The K_d values are 80, 182, 376 nM for $p97^{WT}$, $p97^{L464E}$, and $p97^{L464P}$, respectively. These results suggested a 4.7-fold reduction in p47 binding affinity caused by proline for full-length p97. The K_d values are 343, 264, 271 nM for WT-ND1L,

L464E-ND1L, and L464P-ND1L, respectively. These results suggested a 4.3-fold reduction in p47 binding affinity for ND1L as compared with full length p97 and the binding affinity of L464P-ND1L is like WT-ND1L.

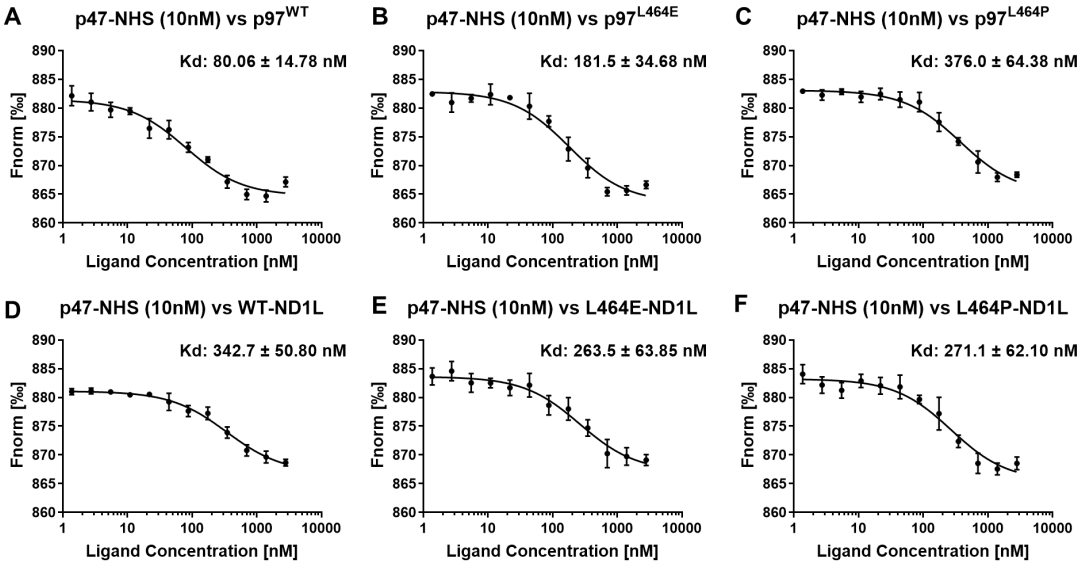


Figure 4-8. Determination of the binding constant, K_d, for the p97 and p47 interaction. Temperature-related intensity change (TRIC) signals on a DianthusNT.23 PicoDuo instrument (Nano-Temper Technologies) was used to determine the equilibrium binding constant (K_d) for p47 in solution with six p97 proteins: (A) $p97^{WT}$, (B) $p97^{L464E}$, (C) $p97^{L464P}$, (D). WT-ND1L, (E). L464E-ND1L, and (F) L464P-ND1L. Unlabeled p97 protein was titrated against 10 nM of RED-NHS-labeled p47 in two-fold steps from 2,800 nM to 1.37 nM. All error bars indicate ± SD (n=3)

4.3.8 Structure of the Full Length $p97^{L464P}$ Reveals Conformational Changes Caused by Changing the Conserved L464 at the D1-D2 Linker

We showed that L464 is the key residue on the D1–D2 linker for modulating p97 ATPase function, we want to determine whether the L464 mutation alters the D1–D2 linker conformation, thereby affecting the ATPase activity. We utilized single-particle cryo-EM imaging to determine the three-dimensional (3D) structure of the $p97^{L464P}$ (Figures 4-9, 4-11). The cryo-EM reconstruction of the $p97^{L464P}$ showed the same hexameric organization as the wild-type p97, suggesting that the L464P mutation does not affect the p97

oligomerization (Figures 4-9A, 4-12A). The cryo-EM reconstruction showed four clear NTD densities in the down conformation, with two NTDs in lower signal levels (Figure 4-10A). The density maps allowed us to dock the p97 coordinates (Figure 4-12B). The resolved NTDs are all in the down configuration, implying less mobility than the others (Figure 4-10A). The unresolved mobile NTD densities may be due to their high flexibility, resulting in density flattening. Unlike the homogeneous NTD configuration in the wild type[224], the highly flexible NTDs of the $p97^{L464P}$ could be due to the linker mutation and the altered communication between domains, resulting in undefined positions of the NTDs in cryo-EM density maps. Unlike the NTD densities, the D1 and D2 densities were well resolved in our cryo-EM reconstructions. To visualize the details of the structural changes between domains, we then attempted to average six D1 and D2 densities symmetrically (Figure 4-11D). We further superimposed our structure with $p97^{WT}$ (PDB code: *5FTK*) (RMSD 1.162 Å) (Figure 4-10B). The D1–D2 linker changes its structure upon L464P mutation, which causes a slight movement of the helix-turn-helix (HTH) motif of K543–A569 (Figure 4-10C, D). Also, the loop densities (K584–D598) for the D2 ring in the L464P mutant were missing, suggesting a flexible nature and size variation of the D2 ring. These local structural changes could be attributed to the change of the loop torsion contributed from the geometric constraints of proline residue (Figure 4-10C, D). The structural changes of the HTH motif and D2 ring could affect the concerted movement between protomers and thereby the ATPase activities.

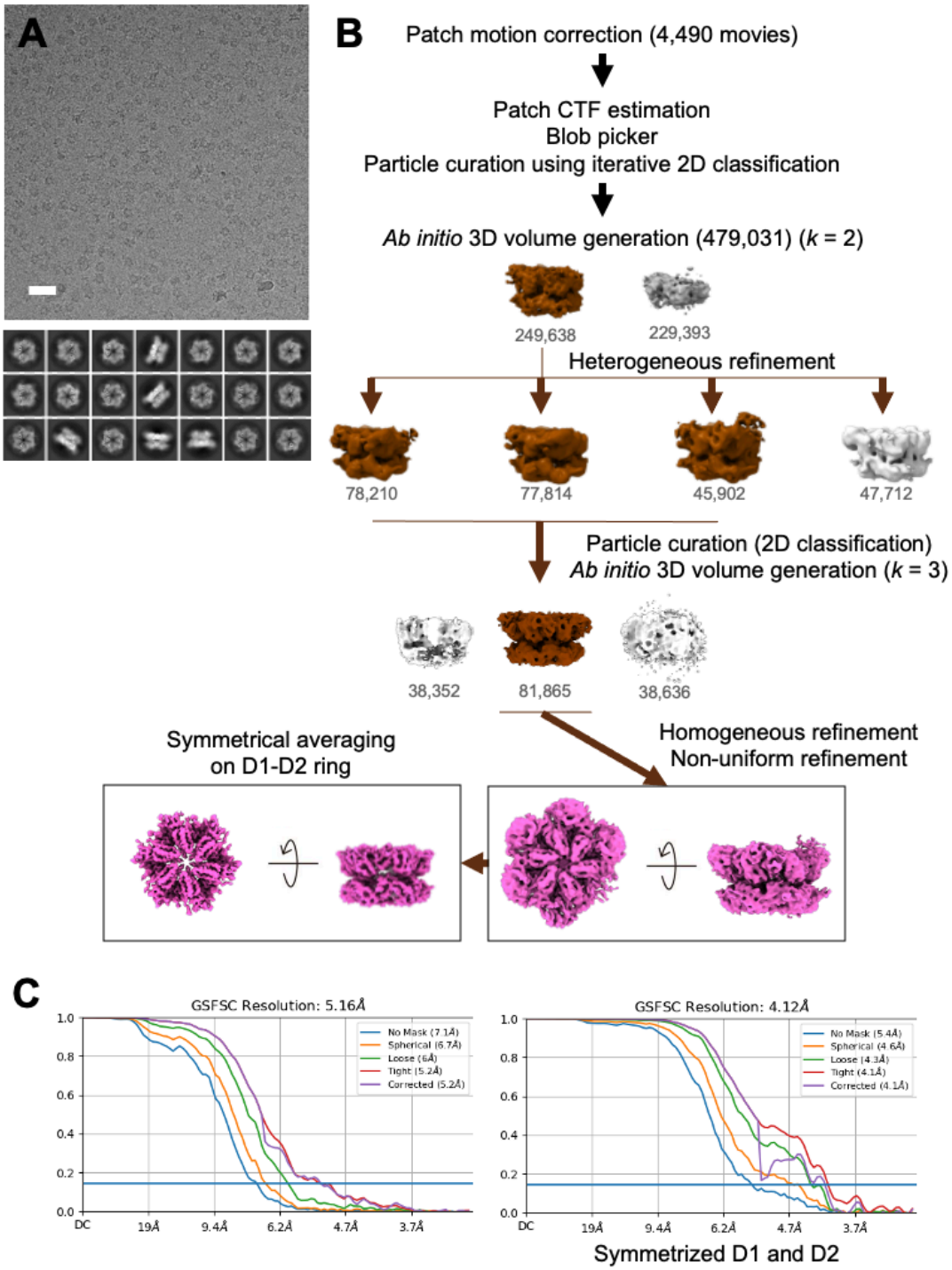


Figure 4-9. Cryo-EM image analysis of the p97^{L464P} mutant. (A) Electron micrograph of cryogenic p97^{L464P} proteins. Scale bar indicates 30 nm. Lower panel shows representative two-dimensional class averages of the p97^{L464P} projections. Box lengths are 204 Å. (B) Flowchart of the cryo-EM data processing. (C) Fourier-shell correlation

(FSC) plots of the final three-dimensional reconstructions. Left and right panels are for unsymmetrized and symmetrized averaging, respectively.

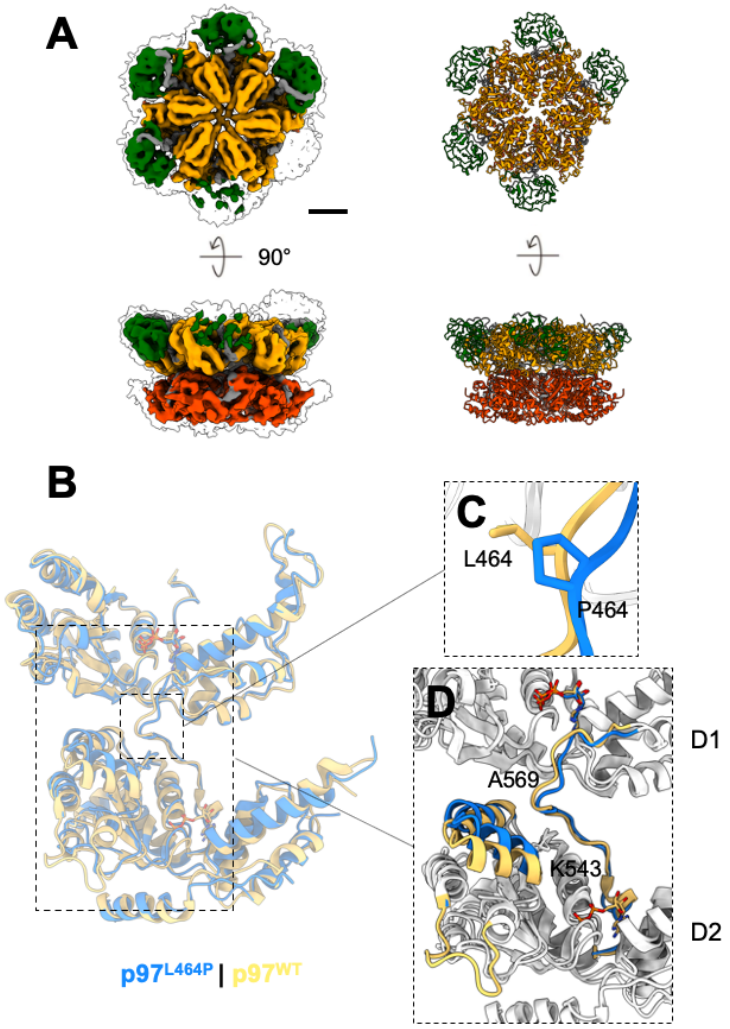


Figure 4-10. Cryo-EM structural analysis of the p97^{L464P} mutant.
(A) Cryo-EM density map and model of human p97^{L464P}. NTD, D1 and D2 domains are dark green, orange and orange red, respectively. Linkers and C-terminal tail are colored in grey. Outer white surface and colored surface densities are present in signal levels of 2.7σ and 7.6σ, respectively. Scale bar indicates 25 Å. (B) Superposition of structures of the p97^{L464P} mutant and ADP-bound p97WT (PDB code: *5FTK*). The RMSD of the superimposed structures is 1.162 Å. p97^{L464P} and p97WT are in blue and golden yellow, respectively. (C) Enlarged view of the mutated residues from leucine 464 (golden yellow) to proline (blue). (D) Enlarged view of the superimposed D1–D2 linkers and surrounding areas. D1–D2 linker, helix-turn-helix (HTH) motif (K543–A569) and flexible central D2 loop (586–598) are colored. Bound nucleotides are shown in sticks.

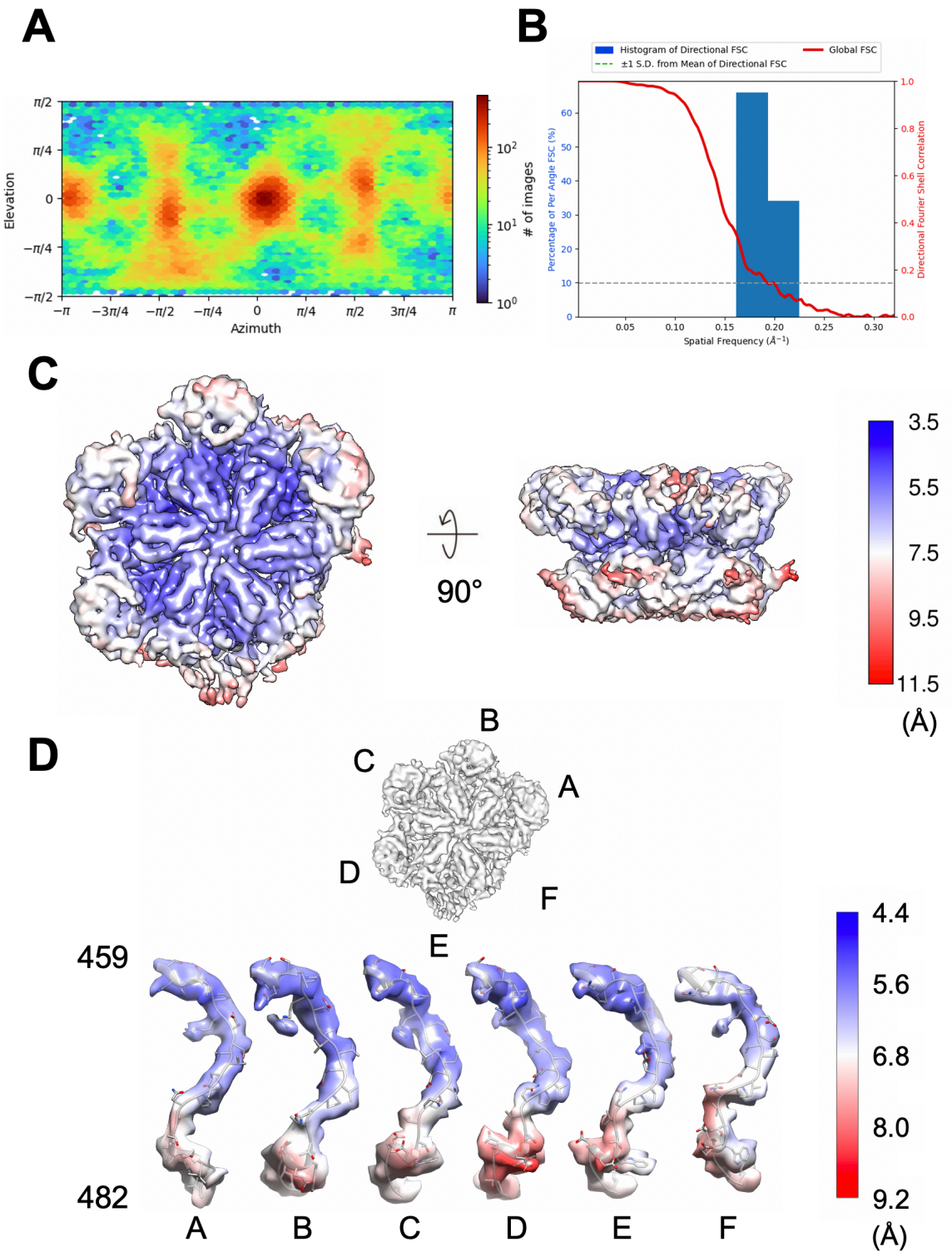


Figure 4-11. Analysis of the cryo-EM reconstruction of human p97^{L464P} mutant. (A) Projection direction distribution of the 3D reconstruction. (B) Directional Fourier-shell correlation (FSC) plot of the 3D reconstruction. (C) Local resolution estimation of the overall cryo-EM map (6.7σ). (D) Local resolution estimations of the densities of

individual linkers (3.2σ). Linkers of individual chains are labeled. Fitted models are shown in sticks.

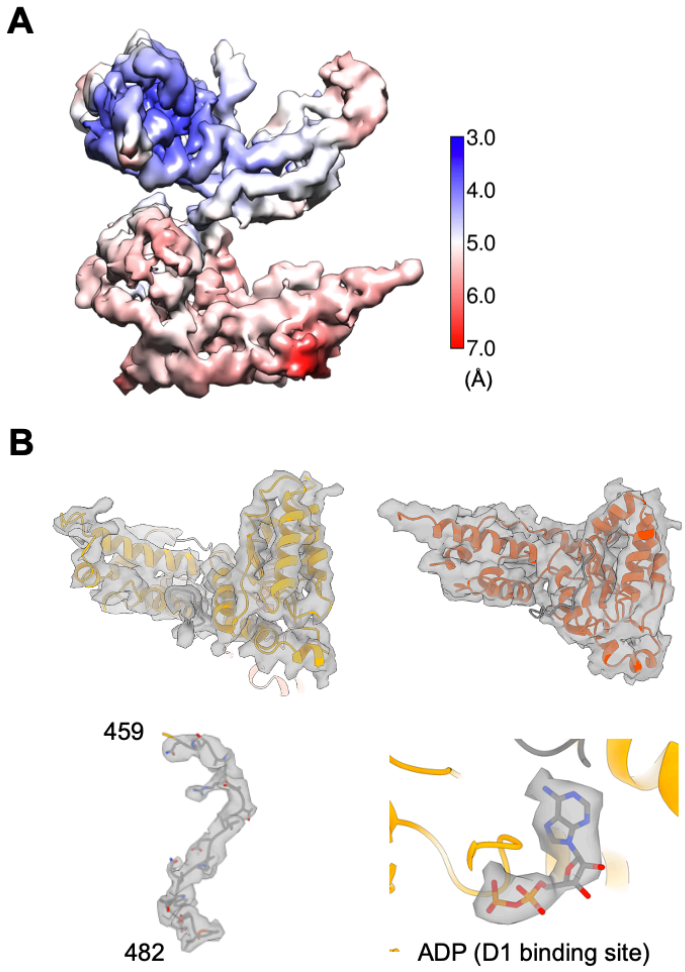


Figure 4-12. Cryo-EM structure of the p97^{L464P} D1-D2 domain. (A) Local resolution estimation of the symmetrized D1-D2 density. (B) Model fitting of the D1 (orange; 7.5σ) and D2 (orange red; 5.8σ) domains, linker domain (grey sticks; 5.9σ) and D1-bound ADP nucleotide (sticks; 7.3σ). Grey surfaces represent cryo-EM densities.

4.3.9 L464 Mutation Decreases the Potency of p97 Inhibitors

Because p97 plays a vital role in protein quality control, much has been done in recent years to target p97 with potent and specific inhibitors for potential cancer therapy[106,300-302]. These compounds are divided into two major classes: ATP-competitive inhibitors[290,303] and noncompetitive inhibitors[113,173,304]. We previously evaluated the

domain and cofactor-complex selectivity of these compounds, finding that communication between the D1 and D2 domains changes the ability of each inhibitor to affect ATPase activity[236,298,305]. For example, mutations within the ND1 domain can significantly increase the IC_{50} of the D2-specific inhibitor ML240[236]. In addition, the allosteric p97 inhibitor NMS-873 binds in a cavity formed by the D1 and D2 regions of the hexamer in close proximity to L464 and $p97^{L464E}$ is 2-fold less sensitive to NMS-873[113]. Furthermore, we established that CB-5083-resistant ovarian cancer cells carried E470K and E470D missense mutants that lead to increased p97 ATPase activity[306]. Therefore, we want to test if L464 mutants might affect the potency of the ATP-competitive p97 inhibitors.

To examine inhibitor potency in relation to L464 mutations, we determined IC_{50} values for four ATP-competitive inhibitors we previously identified — DBeQ, ML240, ML241, CB-5083)[290,303] and the noncompetitive inhibitor (NMS-873)[113] — against each of the L464 mutants and $p97^{WT}$ (Figure 4-13). Structures of p97 inhibitors are shown in Figure 4-13A. The allosteric inhibitor NMS-873, which binds close to the D1–D2 linker, had a 2.6-fold increase in the IC_{50} value for the L464E mutant (Figure 4-13B). The result corroborates previously published data[113]. The L464A and L464P mutants did not significantly affect the IC50 values for NMS-873, which suggested that NMS-873 could still bind to these two mutants.

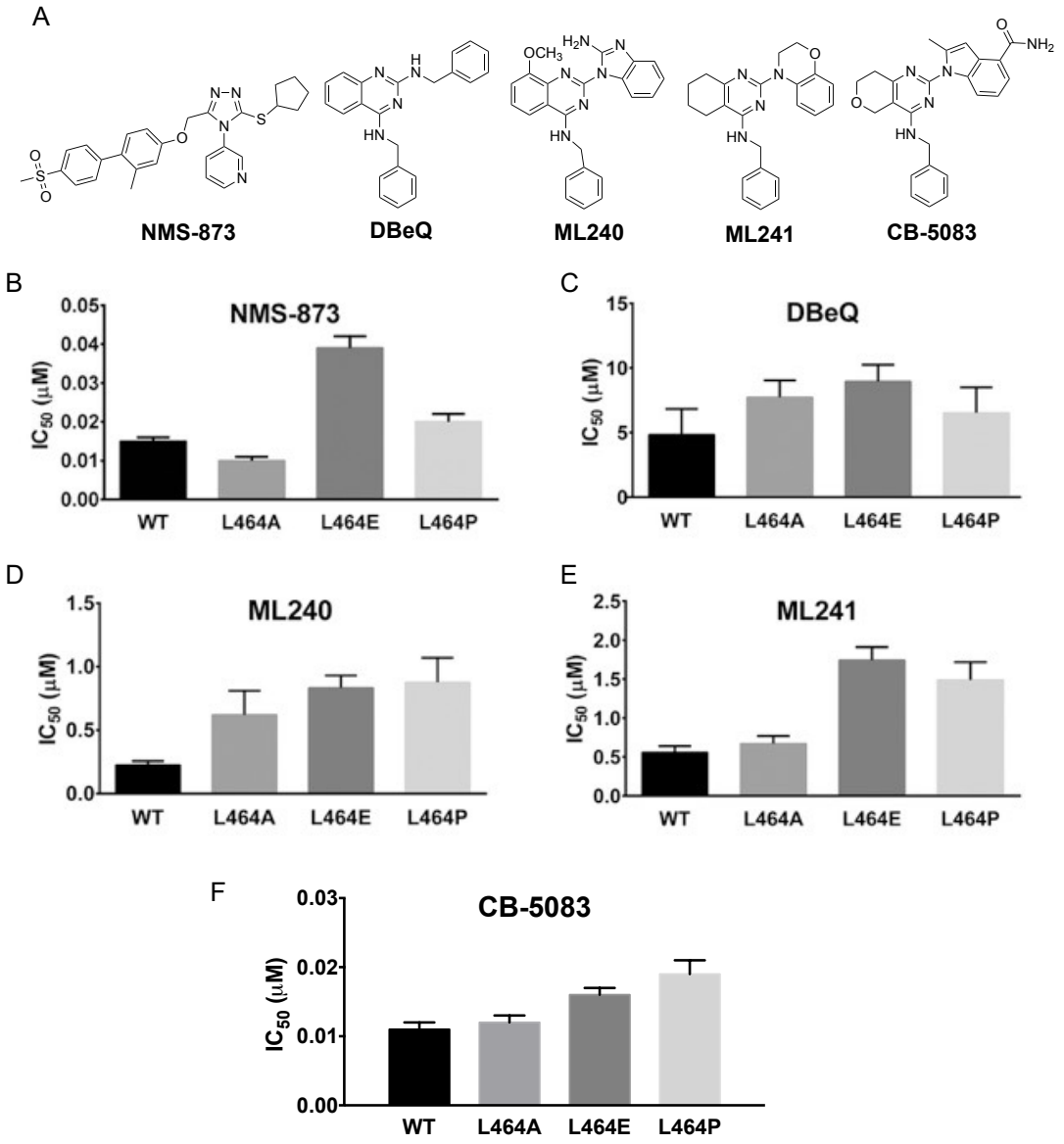


Figure 4-13. Effect of D1–D2 linker mutations on potency of p97 inhibitors. (A) Structures of p97 inhibitors. IC_{50} values for p97 inhibitors (B) NMS-873, (C) DBeQ, (D) ML240, (E) ML241, and (F) CB-5083 for full-length WT-p97 and L464 mutants at 200 mM ATP. All error bars indicate ± SD, and n = 3.

The ATP-competitive inhibitor DBeQ, which inhibits both D1 and D2 ATPase activities, was not affected by L464 mutations, showing similar potencies towards the three linker mutants and $p97^{WT}$ (Figure 4-13C). On the other hand, the D2-specific inhibitors, ML240, ML241, and CB-5083, had 2 to 3-fold increases in IC50 values for the

L464 mutants (Figure 4-13D, E). Interestingly, the IC_{50} increased is more for ML240 than that of CB-5083, which is consistent with the finding for E470K and E470D mutants[306]. Because the L464 resides in the D1– D2 linker region and because the mutations reduced the effects of the D2-binding inhibitors, the experiment further supports the conclusion that altering L464 can impair either the D2 conformation or communication between the D1 and D2 domains, to hinder the binding of D2-specific inhibitors.

4.3.10 Critical Role of the L464 in the D1-D2 Linker

As an AAA+ ATPase protein, p97 acts as the primary molecular machine in several cellular processes vital to protein homeostasis and cell survival. Over the past several decades, structural analyses and biochemical assays on p97 have identified many essential and conserved regions. Mutations in these regions result in cell-based defects and/or ATPase activity defects. Several p97 cofactors are thought to direct p97 to various cellular functions via discrete and/or overlapping binding sites in the NTD or the C-terminal tails[205,217,307]. Cofactor binding results in a conformational change of p97, leading to modulation of the enzyme kinetics of ATP hydrolysis[222]. Previous studies suggest that p97 disease mutants cause imbalanced cofactor binding such as Npl4–Ufd1 but not p47 impairs binding of UBE4B to p97[103] and also impairs UBXD1 binding to p97[87]. A better understanding of these cofactor-regulated changes is imperative for increased knowledge of disease pathogenesis and for providing valuable insight into possible therapeutic approaches.

We explored the role of cofactor-mediated p97 regulation in IBMPFD, a rare disease characterized by large cellular inclusion bodies, indicative of possible defects in the p97 protein degradation pathways[94]. Several disease-causing variants have been

biochemically characterized and found to raise basal levels of ATPase activity. Therefore, ATPase activity alone is not the sole indicator of an efficient, functional p97 protein. We found that three disease-causing p97 mutations display altered cofactor regulation. Specifically, the D2 domain of IBMPFD mutants could not be activated by either p37 or p47. This suggests that these defects may contribute to the pathogenesis of the disease. We also identified amino acids 69 to 92 of p47 as responsible for the observed difference for p37 and p47 in regulating p97 ATPase activity[94]. The binding modes of the two cofactors are highly similar and it is possible that upon binding to p97 the amino acids 69 to 92 of p47 induce conformational changes to case inhibition of both D1 and D2 domain at lower concentration of p47. Without this region p37 cannot cause such inhibition. Recently, the role of two related p97 cofactors, p47 and p37, in protein phosphatase 1 (PP1) maturation revealed an interesting ubiquitin-independent unfolding function to regulated PP1 complex disassembly and assembly[308] .

Here we have demonstrated that the Npl4–Ufd1 heterodimer is an activator of the D2 domain and an inhibitor of the D1 domain of the $p97^{WT}$ (Figures 4-3D, 4-5D). This supports the validity of studying cofactor-regulated p97 activity. The Npl4–Ufd1 heterodimer binds via the ubiquitin-like (UBX-L) domain of Npl4, similar to the p47-UBX domain. The binding of the two cofactors is mutually exclusive, suggesting that they either bind at the same site or prevent the other from binding[309]. The increase in ATPase activity observed for NU may also be involved in coordinating and accelerating its function in the ERAD pathway. Tang and Xia showed that the N-terminal half of the D1–D2 linker is essential for the activity of the D1 domain. By comparing the X-ray crystal structures of the isolated D1 domain with and without the D1–D2 linker, they

showed that the linker region is required to break the symmetric arrangement of the D1 ring and enable ATP hydrolysis[196,286]. Bastola and coauthors reported a potentially direct effect of the E470 mutations on the interactions between the ligand and target[306]. Here, we pinpoint the critical role of the conserved L464 residue in the D1–D2 linker region of p97 by studying the enzymatic activity of point mutants of L464 in both full-length p97 and the isolated D1 domain.

We showed that the L464P mutation reduced by 51% the activity of full-length $p97^{WT}$ (in which both the D1 and D2 domains are active ATPases; Figures 4-3A, 4-14A). Consistent with the published data[285] on $p97^{L464E}$, both $p97^{L464A}$ (with a smaller side chain) and $p97^{L464E}$ (with a negatively charged side chain) retained 100% of $p97^{WT}$ activity. However, the L464P mutation disrupted the activity of $p97^{WT}$, possibly due to the proline side chain caused conformational change (Figure 4-10). To evaluate the effect of L464P specifically on D1 activity, we made an L464P-ND1L protein (which has an active D1 domain with the linker) and showed that L464P diminished ND1L activity to the levels found for deletion of the entire linker (Figures 4-5A, 4-14B). The results suggest that the structural flexibility of L464 in the D1–D2 linker is important for the basal activity of $p97^{WT}$. The severe effect of the proline mutation on the isolated D1 domain (L464P-ND1L) indicates that the proper structure of the D1–D2 linker is vital to D1 function. Due to the proximity of L464 to the nucleotide binding pocket of the D1 domain (Figure 4-1B), it is possible that its flexibility is required for the overall movement that occurs during ATP hydrolysis, but not ATP binding. This interpretation is supported by the fact that the apparent KM value of L464P-ND1L does not differ from

the apparent K_M value of WT-ND1L (Figure 4-7C), but the apparent k_{cat} value of L464P-ND1L is 115-fold lower than the WT-ND1L value (Table 4-4).

Although the basal ATPase activities of $p97^{L464A}$ and $p97^{L464E}$ are similar to that of the $p97^{WT}$, the mutations caused severe defects in the clearance of the ERAD substrate TCR-α-GFP[285] and the UFD pathway reporter (UbG76V-GFP) (Figure 4-2). Here we showed that the cellular defects observed with the L464 mutants may be due to abnormal cofactor-regulated p97 enzyme function. The L464 mutants and the full-length $p97^{WT}$ displayed similar basal steady-state kinetics, but the mutants had a complete loss of p47-regulated inhibition (Figures 4-3, 4-4). Instead, p47 had only single-phase activation with the mutants, which appeared to enhance the phase 2 rebound normally observed with the $p97^{WT}$. L464P-p97 had much greater activation, compared with the more flexible alanine and glutamate mutants, for the three cofactors tested (Figure 4-3B–D). Furthermore, $p97^{L464A}$ and $p97^{L464E}$ reduced activation by p37 (Figure 4-3C) and NU (Figure 4-3D), whereas L464P enhanced activation of full-length p97 by both p37 and NU. These three-point mutations in the D1–D2 linker region clearly altered cofactor-regulated p97 ATPase activities.

Previous structural studies on p97 and yeast homolog Cdc48 have provided valuable insight into this versatile ATPase[105,218,235,256]. $p97^{WT}$ is highly dynamic upon cofactor binding, and some mutation sites were found to stabilize the complex; for example, $p97^{A232E,\ E578Q}$ double mutations were introduced[283]. NMR study suggested that an intrinsically disordered region regulates the interaction between p47 and the truncated p97–ND1L[235]. Shp1, a yeast p47 homolog, was used to pull down Cdc48 in the presence of nonhydrolyzable ATP analog with an endogenous polypeptide. The cryo-EM analysis

of this complex implied a hand-over-hand mechanism of substrate translocation of the AAA+ ATPase[218]. Using Cdc48^{E588Q}, a D2 ATPase-dead mutant complex with Npl4–Ufd1 can pull a ubiquitinylated substrate through the central pore[219] .

Blythe et al. showed that in the presence of ATP, p97^{WT} has 50% up and down NTD and the NTDs of the p97^{A232E} are all in up position[258]. In the presence of ADP, all resolved p97^{WT} NTD showed down positions and p97^{A232E} show more than 50% mobile NTD and 20% visible down position and less than 10% up position[81]. In our p97^{L464P} structure, we did not supply any nucleotides and we observed four clear NTD densities in the down conformation and two NTDs in lower signal levels (Figure 4-10A). Our results indicate that the resolved NTDs in the p97^{L464P} resemble the NTDs in the ADP-bound p97^{WT}.

Recently, Pan and coauthors solved several cryo-EM structures of p97^{A232E,E578Q} with Npl4–Ufd1; they used the mutant A232E to lock N-domain in a more up-conformation and E578Q to block D2 ATP hydrolysis to enable a stable complex with Npl4/Ufd1 and polyubiquitinated substrate, and they observed conformational change at L464 amino acid by comparing the ATPγS-bound non-translocating structure with the ATPγS-bound plus substrate (PDB ID: *7JY5*)[283,310]. By comparing two ATPγS-bound p97 structures, with and without substrate, they demonstrated L464A has reduced unfoldase activity using the polyubiquitinated substrate[310]. *In vitro* study using artificial mutant such as E578Q to abolish ATP hydrolysis is one of the methods to enable visualization of the dynamic p97 protein. Similarly, even though ATPase activity alone is not the sole indicator of p97 function, it does provide direct evidence of how a single amino acid change can change it basal and cofactor-regulated ATPase activities. We think the

difference in cofactor-regulated ATPase activity can potentially indicate difference in regulating p97's endogenous substrates associated with a specific cofactor, but this remains to be determined.

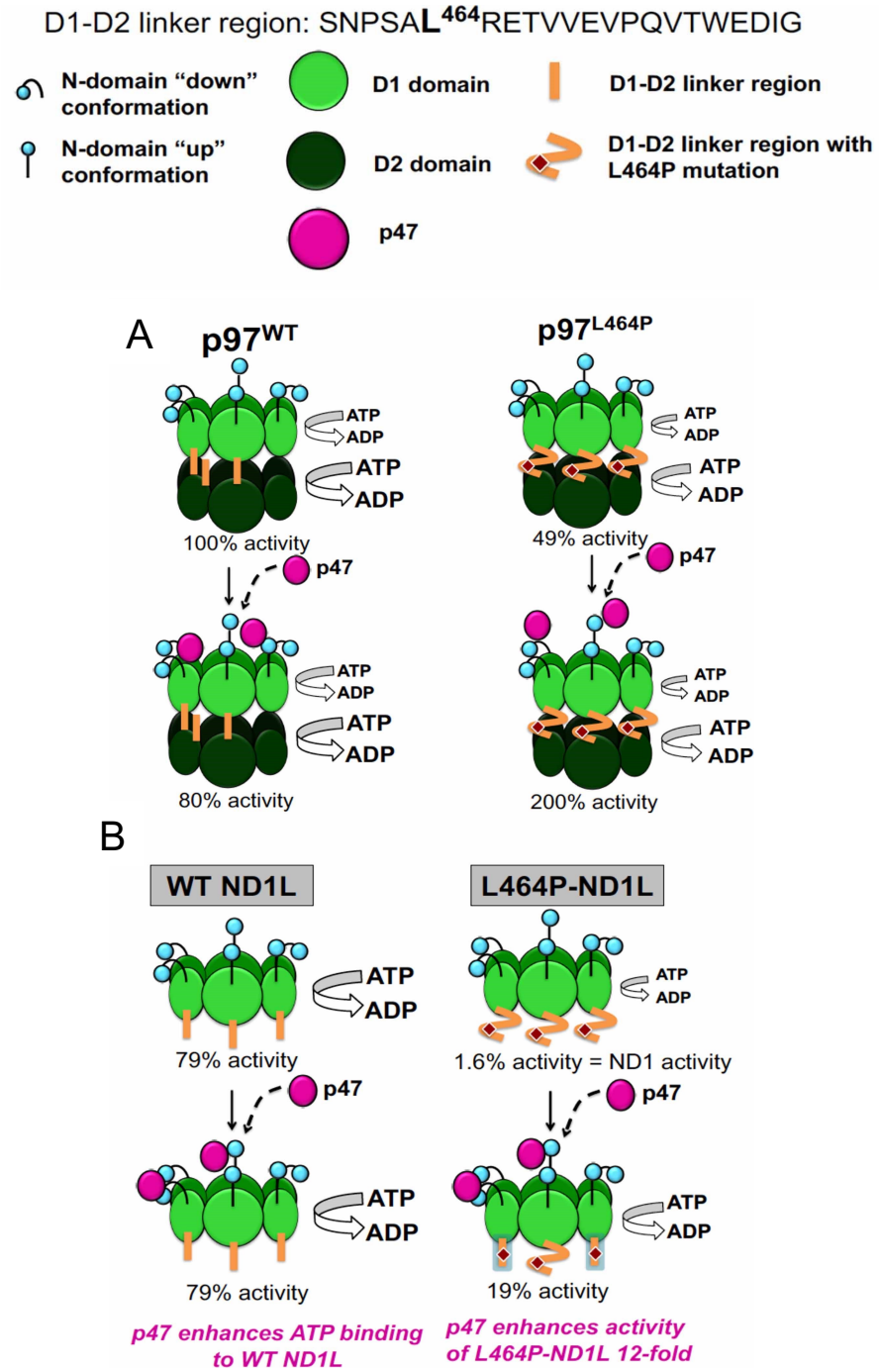


Figure 4-14. Model for the role of the D1–D2 linker region in regulating p97 ATPase activity (A) Mutation of L464 to Pro reduces basal ATPase activity of full-length p97 by 50% and binding with excess p47 cofactor inhibits $p97^{WT}$ activity to be 80% of WT but active $p97^{L464P}$ to be 200% of WT. (B) L464P mutation in the ND1L protein produces a severe defect, equal to that of complete truncation of the D1–D2 linker region (1.6% of $p97^{WT}$) and binding to excess p47 cofactor does not affect ND1L activity but enhances binding to ATP, whereas binding to L464P-ND1L affects the conformation of the D1–D2 linker of L464P and enhance activity to 19% of $p97^{WT}$.

In recent years, an upsurge in p97 inhibitor development has yielded promising starting points for future therapies. Several structure-activity relationship studies have been aimed towards tailoring inhibitors to specific p97 functions[173,296,298,304,305]. We found that L464 is critical for determining the functionality of p97 inhibitors. Specifically, mutating L464 diminished by three-fold the inhibitory effects of D2 ATP-competitive inhibitors, ML240 and ML241 (Figure 4-13D, E). We reconfirmed findings that an L464 mutant reduced the potency of NMS-873 (Figure 4-13B), an inhibitor that binds in the linker region[113]. From these results, we conclude that L464 is critical for maintaining the correct conformation of the linker region, and that disturbing the conformation of the linker can impact how D2-selective ATP competitive inhibitors bind to the D2 ATPase domain.

Our model suggests that the conserved L464 in the D1–D2 linker (22 amino acids, Figure 4-14) is necessary to maintain communication among the domains of p97, including NTD, D1, and D2, regardless of their distances from one another in space. The L464P mutation changes the conformation of the linker region, lowering the efficiency of ATP hydrolysis by reducing ATPase activities (Figure 4-14A). For the isolated D1 domain with the linker (ND1L), the L464P mutation induces a conformational change that nearly abolishes its ability to hydrolyze ATP (Figure 4-14B). The L464P-ND1L is

enzymatically comparable to the ND1 protein, which lacks the linker region entirely[93]. This substantial loss of activity compared with WT–ND1L demonstrates the importance of the linker region for D1 ATPase function, a finding consistent with our previous results[93]. In addition, we propose that although the L464P mutation alters the conformation of the p97 D1–D2 linker, the presence of p47 restores the conformation to some degree, through stabilization of the linker region of L464P (Figure 4-14B). This interpretation would explain why p47 has no appreciable effect on WT-ND1L ATPase activity but activates the L464P-ND1L mutant by 12-fold (Figures 4-7B, 4-14B).

4.4 Conclusion

This study adds to our growing understanding of p97 and its function. We have determined that the conformation of the linker region is crucial to ATP hydrolysis at the D1 and D2 domains. We identified the importance of these regions in communication within the enzyme during cofactor-regulated ATPase activity and demonstrated that three p97 cofactors regulate the enzyme differently. Even though the total p97 protein level is more than individual p97 co-factors, but published result suggests that in normal condition, most p97 cofactors are not associated with p97 and the association can be enhanced under some conditions. For example, Xue and coauthors used size exclusion chromatography to separate cellular p97 and p97 cofactor complexes in HEK293 coupled with mass spectrometry to infer that the cofactors interaction with p97 are very dynamic; they also found p47 and Npl4–Ufd1 do not cofractionate with p97 in unperturbed cell lysate and upon p97 inhibition by an allosteric inhibitor NMS-873 enhanced formation of the complexes, but most of cofactors are still not associated with p97[311]. Therefore, future studies of the remaining known UBX cofactors are warranted. Additionally,

comprehensive structural data of p97 cofactors are not available. For example, even though p47 has been studied with great interest since its discovery in 1997[76] and the structures of p47 revealed that conformational changes in p97 are induced[76,77,79]. Structural data for p97 bound with the full-length p47 and other p97 cofactors would provide evidence for the conformational changes resulting from cofactor interaction. In addition, it is likely that the p97 function is modulated by its substrates, as well as its cofactors. Future work should examine the role those endogenous substrates play in cofactor-regulated p97 activity. Given our knowledge gained from more thoroughly studying the enzyme cofactor relationship, it is reasonable to infer that each p97 enzyme-cofactor complex is fine-tuned for the microenvironment where it performs its role. Therefore, a better understanding of these complexes will yield valuable insights into the biochemical intricacies of their functions. We believe that our research provides new information useful for understanding how p97 cofactors regulate p97 function and for developing p97-specific therapeutic agents.

To investigate the relationship between the D1–D2 linker and the D1 domain, we produced p97 ND1L mutants and demonstrated that this 22-residue linker region is essential for D1 ATPase activity. The conserved amino acid leucine 464 (L464) is critical for regulating D1 and D2 ATPase activity by p97 cofactors p37, p47, and Npl4– Ufd1 (NU). Changing leucine to alanine, proline, or glutamate increased the maximum rate of ATP turnover (k_{cat}) of p47-regulated ATPase activities for these mutants, but not for WT. p37 and p47 increased the k_{cat} of the proline substituted linker, suggesting that they induced linker conformations facilitating ATP hydrolysis. NU inhibited D1 ATPase

activities of WT and mutant ND1L proteins but activated D2 ATPase activity of full-length p97. To further understand the mutant mechanism, we used single-particle cryo-EM to visualize the full-length $p97^{L464P}$ and revealed the conformational change of the D1–D2 linker, resulting in a movement of the helix-turn-helix motif (543–569). Taken together with the biochemical and structural results we conclude that the linker helps maintain D1 in a competent conformation and relays the communication to/from the N-domain to the D1 and D2 ATPase domains, which are ~50 Å away.

4.5 Data Availability Statement

Cryo-EM density maps were deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers of EMD-23775 and EMD-23776 (symmetrized) and the Protein Data Bank (PDB) under accession numbers of 7MDM and 7MDO (symmetrized). All data are available from the corresponding author upon request.

4.6 Acknowledgments

We thank Dr. Dewight Williams for assisting in cryo-EM data-collection and acknowledge the use of the Titan Krios TEM, in the Eyring Materials Center (EMC), at Arizona State University (ASU) supported from grant number NSF MRI 1531991. We thank the NVIDIA GPU Grant Program to Prof. Po-Lin Chiu for GPU device support. The project was partially supported by a grant to Prof. Tsui-Fen Chou (Caltech) and the ASU startup fund to Prof. Po-Lin Chiu. We acknowledge the support and resources from the Life Sciences Electron Microscopy Lab for the use of the electron microscope Philips CM12, at ASU.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Cryo-EM enabled the studies of proteins in multiple conformations and helped in the study of challenging protein complexes which are tough to crystallize, and hence served as a catalyst to provide rich two-dimensional (2D) and three-dimensional (3D) structural information.

In Chapter 2, I performed the structural studies to identify the protein complex which triggers neuronal apoptosis proNGF-p75^{NTR}-sortilin and could successfully purify the extracellular-domain constructs of the neurotrophin receptor p75, it's co-receptor sortilin and the ligand proNGF from insect cells (*Sf9*) and confirmed the functional proteins from various biochemical and biophysical techniques. The proNGF-p75^{NTR}-sortilin complex forms transiently after co-transfecting the insect cells, and more efforts are needed to generate the 3D atomic model of the apoptotic assembly. I found the complex formation is aided in the presence of calcium ions (at physiological concentrations like 2 mM) and dimerization interface of sortilin initiates the assembling. The wild type proNGF is prone to degradation to its mature form NGF by cellular proteases, like furin and thus triple mutation can render stability to pro-domain for further studies. Future studies will involve the purification of a stable and highly pure FLAG-tagged triple-mutant proNGF (purified using Flag-affinity purification) and the protein can be supplied to neuronal stem cells like DOAY and U87 to study cell viability and biological activity. The exogeneous ligand proNGF will be supplied to cells to pull-down natively formed complex using co-immunoprecipitation to pursue structural studies. Optimizing proNGF-p75^{NTR}-sortilin complex formation will involve exploring the

dimerization interface of sortilin and proNGF individually, trying sequential binary binding experiments (adding proNGF to the sortilin-$p75^{NTR}$ assembly or sortilin to $p75^{NTR}$-proNGF assembly) and needs stabilization using mild crosslinking techniques. The formed complex can be characterized by temperature dependent intensity change (TRIC) or surface plasmon resonance (SPR) experiments and followed by subsequent cryo-EM data collection.

In Chapter 3, I performed structural studies to investigate the disease mechanism linked to the p97 R155H mutation. Using single-particle cryo-EM results, I found that the $p97^{R155H}$ mutant is organized as a dodecamer when it is inactive. This may represent an inactive form. Due to ordering in our structure, we have reconstructed a high-resolution structure that includes the C-terminal tail domain. In the presence of nucleotides, the dodecamer dissociates into functional hexamers. $p47^{UBX}$ cofactor binding to the $p97^{R155H}$ is asymmetric and binds the up N-domain. We determined that D1 and D2 arginine fingers are critical for increased $p97^{R155H}$ ATPase activity. Based on the structure, we argue that D1 and D2 interactions are altered in the $p97^{R155H}$ resulting in interdomain miscommunication. To complement our structural studies, we have also conducted a series of biochemical analyses that reveal how interactions change in the context of the p97 mutant. These results provide insight into the disease mechanism of the R155H mutation, and its role in pathological function. Our analysis will support next steps that include identifying and characterizing drugs that may tackle disorders resulting from p97 dysfunction[94].

The dodecameric complex has been reported in other studies when the wild type p97 was inhibited by the potent ATP-competitive drug CB-5083[241], in studies using p97

affinity grids as well, however little is known about its *in-vivo* cellular implications. Future studies will investigate the relevance of the dodecamer formation from functional p97 hexamers in cellular studies, by screening other common MSP1 mutants, where the residues in the N-terminal domain or N-D1 linker region are mutated to mimic the MSP1 diseased p97.

In our preliminary results, we observed the dodecamer formation in several other MSP1 mutants as seen in the Figure 5-1. Since the MSP1 mutations are densely located on the N-terminal domain as well as the N-D1 linker region of p97, we designed several mutants as well as a few MSP1 mutants known for the p97 D1 and D2 domains as shown in Figure 5-1A. Negative-stain EM shows a high proportion of dodecamers formed for three mutants $p97^{R89G}$, $p97^{I369F}$, $p97^{K663R}$ as shown in Figure 5-1B. We further pursued cryo-EM data collection of these mutants and were able to reconstruct high resolution details of the higher order dodecameric structure as shown in Figure 5-1C.

Further, to understand the relevance of p97-p47 complex in the action of substrate processing it would be necessary to optimize the assembling of the $p97^{WT}$-VCIP135-$Syn5^{Ub}$-p47 and perform cryo-EM analysis. We observed that the complex formation is unstable in the absence of ubiquinated Syn-5. For continuing these future studies, we purified the enzymes E1, E2, E3 which assists the in-vitro ubiquitination of Syn-5 in a cascaded reaction, and purified VCIP135 which acts as a deubiquitinase, to help stabilize the assembly.

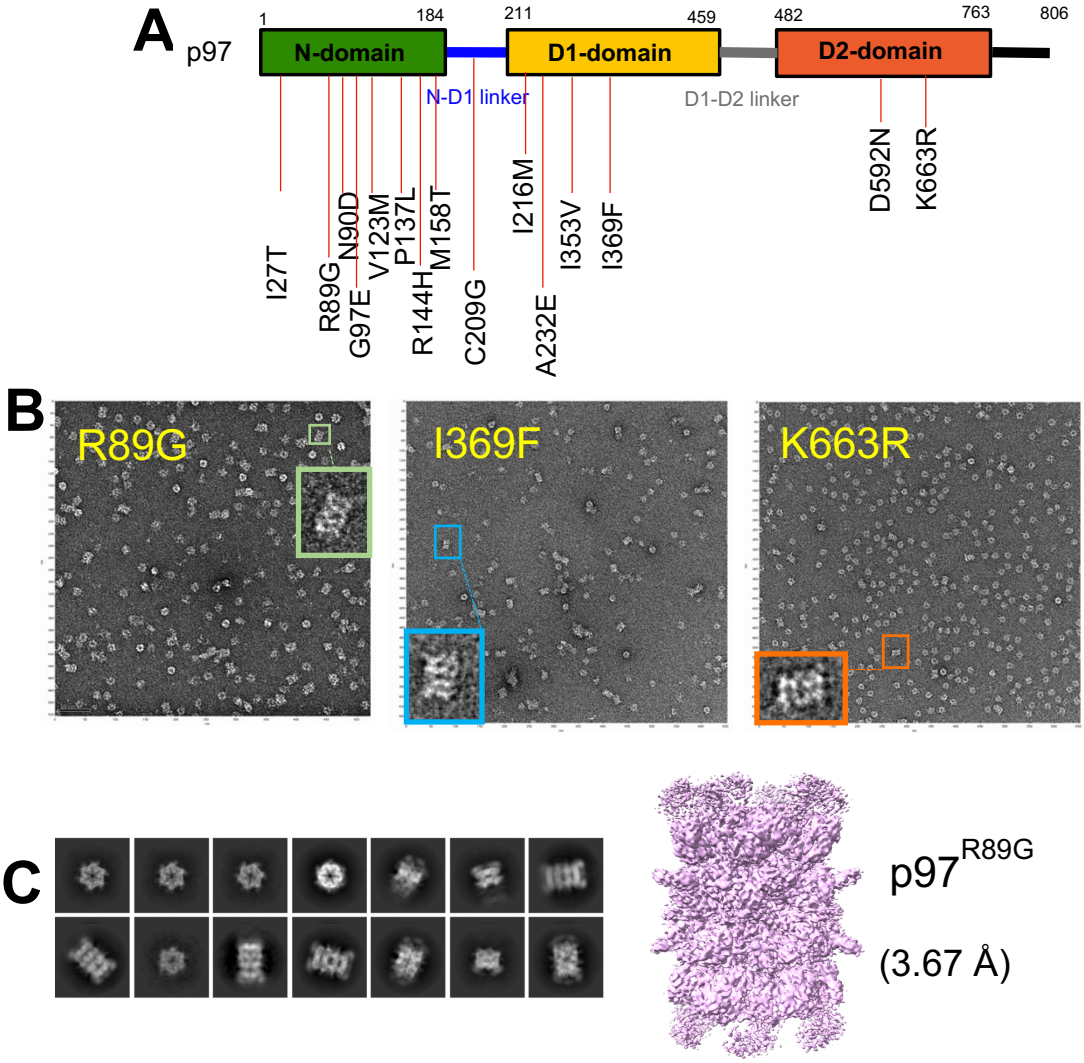


Figure 5-1. Study of dodecamer formation in various p97 MSP1 mutants (A) MSP1 mutations are primarily clustered in the p97 N domain or N-D1 linker[86]. The panel describes some p97 MSP1 mutants expressed for our structural studies. (B) Negative-stain electron micrographs showing a high proportion of dodecameric population. Inset shows the dodecamer in each of the three mutants $p97^{R89G}$, $p97^{I369F}$, $p97^{K663R}$ located in the N, D1 and D2 domains. (C) The cryo-EM dataset of $p97^{R89G}$ shows high resolution dodecamers in the 2D class averages, with the reconstructed cryo-EM density map shown at 3.67 Å (right).

In Chapter 4, I described that the residue L464 in the p97 D1-D2 linker region plays an important role in controlling the basal activity of the D1 ATPase domain and in the cofactor-regulated ATPase activity. We solved the $p97^{L464P}$ full length cryo-EM

structure and found that changing the residue leucine 464 to proline caused conformational changes. This might affect communication of cofactor binding on p97 from the N-domain to the D1 and D2 domains due to torsional restraint. The cryo-EM reconstruction showed clear four N-terminal domain (NTD) densities in down conformation (chains A-D) and two NTDs in lower signal levels (chains E and F), possibly due to uncontrolled flexibility. Chains A-D in the NTD-down conformation were also less mobile. The full length $p97^{L464P}$ is hexameric as the wild type p97 and the L464P mutation does not affect the p97 oligomerization. The L464P mutation structurally changes the D1-D2 linker by causing a slight movement of the helix-turn-helix (HTH) motif of K543-A569 and the loop densities (K584-D598) for the D2 ring were absent. Previous studies showed that there is an interaction between the D1 and D2 ATPase domains of p97/VCP[236]. We also showed that p97/VCP cofactors control its enzymatic activity and p97/VCP disease mutants behave abnormally, due to lack of appropriate control by these cofactors. In this study, we discovered an important functional property of the conserved L464 residue in the D1-D2 linker region of p97 ATPase, by demonstrating its role in domain crosstalk of p97 upon cofactor binding from the N-domain to the D1 and D2 ATPase domains. We found that changing the conserved linker residue 464 from leucine to alanine (L464A), glutamate (L464E), or proline (L464P) abolished p47's inhibitory effects at low p47 concentrations, and unexpectedly caused p97 activation at higher p47 concentrations. Our findings provide strong biochemical evidence for L464 mutants being a cause of cellular defects, as reported[285]. To understand the effect of the D1-D2 linker on the D1 ATPase domain specifically, we made linker mutations in a p97 ND1L protein. We showed that a single amino acid

change (L464P) abolished D1 ATPase activity - to the same level seen for the ND1 domain without the 22 amino acid D1-D2 linker. Thus, the flexible linker region is critical for the basal ATPase activity of the D1 domain for p97. Interestingly, binding of cofactors like p37 and p47 to the N domain of L464P-ND1L restored ATPase activity by increasing k_{cat} values by seven and 14-fold, respectively. We found that the Npl4-Ufd1 cofactor complex activated full-length p97 but inhibited ND1L, in both WT-ND1L as well as ND1L linker mutants. We thus concluded that the D1-D2 linker region not only assists D1 in maintaining a competent conformation, but also transmits binding effects of p97 cofactors from the N-domain to control the ATPase activities of the D1 and D2 domains. This study improves our understanding of p97 and its function, and the importance of the D1-D2 linker region in interdomain communication.

Cryo-EM has emerged as a cutting-edge technique to study macromolecular protein complexes thanks to the recent advances of the direct electron detectors (DED) and the improved data processing algorithms. The generated knowledge helped us to understand the unknown symmetries in protein complexes or in labelling components of the complex[312,313]. It enabled the studies of very flexible domains in the p97 complexes studied in Chapters 3 and 4 and allowed to capture various functional states. Our studies thus contributed to develop a deeper knowledge of the intricate molecular mechanisms affected in disease pathways and provide insight into the design of therapeutics.

REFERENCES

1 Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683-694, doi:10.1038/nature06404 (2007).

2 Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* **77**, 443-477, doi:10.1146/annurev.biochem.77.060407.135530 (2008).

3 Hope, H. Crystallography of biological macromolecules at ultra-low temperature. *Annu Rev Biophys Biophys Chem* **19**, 107-126, doi:10.1146/annurev.bb.19.060190.000543 (1990).

4 Terwilliger, T. C., Stuart, D. & Yokoyama, S. Lessons from structural genomics. *Annu Rev Biophys* **38**, 371-383, doi:10.1146/annurev.biophys.050708.133740 (2009).

5 Ilari, A. & Savino, C. Protein structure determination by x-ray crystallography. *Methods Mol Biol* **452**, 63-87, doi:10.1007/978-1-60327-159-2_3 (2008).

6 Dubochet, J. Cryo-EM--the first thirty years. *J Microsc* **245**, 221-224, doi:10.1111/j.1365-2818.2011.03569.x (2012).

7 McDowall, A. W. *et al.* Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. *J Microsc* **131**, 1-9, doi:10.1111/j.1365-2818.1983.tb04225.x (1983).

8 Chlanda, P. & Sachse, M. Cryo-electron microscopy of vitreous sections. *Methods Mol Biol* **1117**, 193-214, doi:10.1007/978-1-62703-776-1_10 (2014).

9 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

10 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

11 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

12 Quentin, D. & Raunser, S. Electron cryomicroscopy as a powerful tool in biomedical research. *J Mol Med (Berl)* **96**, 483-493, doi:10.1007/s00109-018-1640-y (2018).

13 Martynowycz, M. W., Khan, F., Hattne, J., Abramson, J. & Gonen, T. MicroED structure of lipid-embedded mammalian mitochondrial voltage-dependent anion channel. *Proc Natl Acad Sci U S A* **117**, 32380-32385, doi:10.1073/pnas.2020010117 (2020).

14 Harris, J. R. Negative staining of thinly spread biological samples. *Methods Mol Biol* **369**, 107-142, doi:10.1007/978-1-59745-294-6_7 (2007).

15 Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* **6**, 23-34, doi:10.1251/bpo70 (2004).

16 Shukla, A. K. *et al.* Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* **512**, 218-222, doi:10.1038/nature13430 (2014).

17 Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol* **481**, 109-126, doi:10.1016/S0076-6879(10)81005-5 (2010).

18 Kang, J. Y. *et al.* RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell* **69**, 802-815.e805, doi:10.1016/j.molcel.2018.01.018 (2018).

19 Zhang, C. *et al.* Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell* **175**, 212-223.e217, doi:10.1016/j.cell.2018.09.001 (2018).

20 Kasinath, V. *et al.* Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science* **359**, 940-944, doi:10.1126/science.aar5700 (2018).

21 Gestaut, D. *et al.* The Chaperonin TRiC/CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellular Proteostasis. *Cell* **177**, 751-765.e715, doi:10.1016/j.cell.2019.03.012 (2019).

22 Chiu, P. L., Kelly, D. F. & Walz, T. The use of trehalose in the preparation of specimens for molecular electron microscopy. *Micron* **42**, 762-772, doi:10.1016/j.micron.2011.06.005 (2011).

23 De Carlo S, Adrian, M., Kälin, P., Mayer, J. M. & Dubochet, J. Unexpected property of trehalose as observed by cryo-electron microscopy. *J Microsc* **196**, 40-45, doi:10.1046/j.1365-2818.1999.00591.x (1999).

24 Crowe, J. H., Crowe, L. M. & Chapman, D. Preservation of membranes in anhydrobiotic organisms: the role of trehalose. *Science* **223**, 701-703, doi:10.1126/science.223.4637.701 (1984).

25 Crowe, J. H., Hoekstra, F. A., Nguyen, K. H. & Crowe, L. M. Is vitrification involved in depression of the phase transition temperature in dry phospholipids? *Biochim Biophys Acta* **1280**, 187-196, doi:10.1016/0005-2736(95)00287-1 (1996).

26 Lee, C. W., Waugh, J. S. & Griffin, R. G. Solid-state NMR study of trehalose/1,2-dipalmitoyl-sn-phosphatidylcholine interactions. *Biochemistry* **25**, 3737-3742, doi:10.1021/bi00361a001 (1986).

27 Jain, N. K. & Roy, I. Trehalose and protein stability. *Curr Protoc Protein Sci* **Chapter 4**, Unit 4.9, doi:10.1002/0471140864.ps0409s59 (2010).

28 Kilburn, D. *et al.* Organization and mobility of water in amorphous and crystalline trehalose. *Nat Mater* **5**, 632-635, doi:10.1038/nmat1681 (2006).

29 Harris, J. R. & Scheffler, D. Routine preparation of air-dried negatively stained and unstained specimens on holey carbon support films: a review of applications. *Micron* **33**, 461-480, doi:10.1016/s0968-4328(01)00039-7 (2002).

30 Yimo Han , X. F., Haozhe Wang , Fang Zhao , Christopher G. Tully , Jing Kong , Nan Yao , Nieng Yan. (bioRxiv 827089; doi: https://doi.org/10.1101/827089, 2019).

31 Naydenova, K., Peet, M. J. & Russo, C. J. Multifunctional graphene supports for electron cryomicroscopy. *Proc Natl Acad Sci U S A* **116**, 11718-11724, doi:10.1073/pnas.1904766116 (2019).

32 Palovcak, E. *et al.* A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J Struct Biol* **204**, 80-84, doi:10.1016/j.jsb.2018.07.007 (2018).

33 Wang, F. *et al.* General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. *Proc Natl Acad Sci U S A* **117**, 24269-24273, doi:10.1073/pnas.2009707117 (2020).

34 Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. The complete structure of the human TFIIH core complex. *Elife* **8**, doi:10.7554/eLife.44771 (2019).

35 Fan, X. *et al.* Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat Commun* **10**, 2386, doi:10.1038/s41467-019-10368-w (2019).

36 Yu, G., Li, K., Huang, P., Jiang, X. & Jiang, W. Antibody-Based Affinity Cryoelectron Microscopy at 2.6-Å Resolution. *Structure* **24**, 1984-1990, doi:10.1016/j.str.2016.09.008 (2016).

37 Snijder, J. *et al.* Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. *J Struct Biol* **198**, 38-42, doi:10.1016/j.jsb.2017.02.008 (2017).

38 McMullan, G., Vinothkumar, K. R. & Henderson, R. Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy. *Ultramicroscopy* **158**, 26-32, doi:10.1016/j.ultramic.2015.05.017 (2015).

39 Ruskin, R. S., Yu, Z. & Grigorieff, N. Quantitative characterization of electron detectors for transmission electron microscopy. *J Struct Biol* **184**, 385-393, doi:10.1016/j.jsb.2013.10.016 (2013).

40 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).

41 Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665, doi:10.7554/eLife.03665 (2014).

42 Egelman, E. H. The Current Revolution in Cryo-EM. *Biophys J* **110**, 1008-1012, doi:10.1016/j.bpj.2016.02.001 (2016).

43 Dutta, M. Recent advances in single particle cryo-electron microscopy and cryo-electron tomography to determine the structures of biological macromolecules. *J Indian Inst Sci* **98**, 231-245 doi: 10.0007/s41745-018-0087-z (2018).

44 Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* **177**, 630-637, doi:10.1016/j.jsb.2012.02.003 (2012).

45 Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97, doi:10.1006/jsbi.1999.4174 (1999).

46 Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-347, doi:10.1016/s1047-8477(03)00069-8 (2003).

47 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

48 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).

49 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

50 Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166**, 205-213, doi:10.1016/j.jsb.2009.01.004 (2009).

51 Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).

52 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

53 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221, doi:10.1038/s41592-020-00990-8 (2020).

54 Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237-247, doi:10.1016/j.str.2011.12.007 (2012).

55 Xu, Y., Wu, J., Yin, C. C. & Mao, Y. Unsupervised Cryo-EM Data Clustering through Adaptively Constrained K-Means Algorithm. *PLoS One* **11**, e0167765, doi:10.1371/journal.pone.0167765 (2016).

56 Sigworth, F. J., Doerschuk, P. C., Carazo, J. M. & Scheres, S. H. An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol* **482**, 263-294, doi:10.1016/S0076-6879(10)82011-7 (2010).

57 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* **183**, 377-388, doi:10.1016/j.jsb.2013.07.005 (2013).

58 Scheres, S. H. *et al.* Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* **348**, 139-149, doi:10.1016/j.jmb.2005.02.031 (2005).

59 Doerr, A. Single-particle cryo-electron microscopy. *Nat Methods* **13**, 23, doi:10.1038/nmeth.3700 (2016).

60 Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**, 406-418, doi:10.1016/j.jmb.2011.11.010 (2012).

61 Joachim Frank, B. S., Helen Dowse. (Ultramicroscopy,Volume 6, Issue 1, 1981, Pages 343-357, ISSN 0304-3991, 1981).

62 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, doi:10.7554/eLife.36861 (2018).

63 Jaitly, N., Brubaker, M. A., Rubinstein, J. L. & Lilien, R. H. A Bayesian method for 3D macromolecular structure inference using class average images from single particle electron microscopy. *Bioinformatics* **26**, 2406-2415, doi:10.1093/bioinformatics/btq456 (2010).

64 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745, doi:10.1016/j.jmb.2003.07.013 (2003).

65 Chao, M. V., Rajagopal, R. & Lee, F. S. Neurotrophin signalling in health and disease. *Clin Sci (Lond)* **110**, 167-173, doi:10.1042/CS20050163 (2006).

66 Roux, P. P. & Barker, P. A. Neurotrophin signaling through the p75 neurotrophin receptor. *Prog Neurobiol* **67**, 203-233, doi:10.1016/s0301-0082(02)00016-3 (2002).

67 Kaplan, D. R. & Miller, F. D. Neurotrophin signal transduction in the nervous system. *Curr Opin Neurobiol* **10**, 381-391, doi:10.1016/s0959-4388(00)00092-1 (2000).

68 Huang, E. J. & Reichardt, L. F. Neurotrophins: roles in neuronal development and function. *Annu Rev Neurosci* **24**, 677-736, doi:10.1146/annurev.neuro.24.1.677 (2001).

69 Meeker, R. & Williams, K. Dynamic nature of the p75 neurotrophin receptor in response to injury and disease. *J Neuroimmune Pharmacol* **9**, 615-628, doi:10.1007/s11481-014-9566-9 (2014).

70 Tang, D. & Wang, Y. Cell cycle regulation of Golgi membrane dynamics. *Trends Cell Biol* **23**, 296-304, doi:10.1016/j.tcb.2013.01.008 (2013).

71 Wang, Y. & Seemann, J. Golgi biogenesis. *Cold Spring Harb Perspect Biol* **3**, a005330, doi:10.1101/cshperspect.a005330 (2011).

72 Rabouille, C. *et al.* Syntaxin 5 is a common component of the NSF- and p97-mediated reassembly pathways of Golgi cisternae from mitotic Golgi fragments in vitro. *Cell* **92**, 603-610, doi:10.1016/s0092-8674(00)81128-9 (1998).

73 Rabouille, C., Levine, T. P., Peters, J. M. & Warren, G. An NSF-like ATPase, p97, and NSF mediate cisternal regrowth from mitotic Golgi fragments. *Cell* **82**, 905-914, doi:10.1016/0092-8674(95)90270-8 (1995).

74 Shorter, J. & Warren, G. Golgi architecture and inheritance. *Annu Rev Cell Dev Biol* **18**, 379-420, doi:10.1146/annurev.cellbio.18.030602.133733 (2002).

75 Wang, Y., Satoh, A., Warren, G. & Meyer, H. H. VCIP135 acts as a deubiquitinating enzyme during p97-p47-mediated reassembly of mitotic Golgi fragments. *J Cell Biol* **164**, 973-978, doi:10.1083/jcb.200401010 (2004).

76 Kondo, H. *et al.* p47 is a cofactor for p97-mediated membrane fusion. *Nature* **388**, 75-78, doi:10.1038/40411 (1997).

77 Meyer, H. H., Kondo, H. & Warren, G. The p47 co-factor regulates the ATPase activity of the membrane fusion protein, p97. *FEBS Lett* **437**, 255-257 (1998).

78 Uchiyama, K. *et al.* p37 is a p97 adaptor required for Golgi and ER biogenesis in interphase and at the end of mitosis. *Dev Cell* **11**, 803-816, doi:10.1016/j.devcel.2006.10.016 (2006).

79 Uchiyama, K. *et al.* VCIP135, a novel essential factor for p97/p47-mediated membrane fusion, is required for Golgi and ER assembly in vivo. *J Cell Biol* **159**, 855-866, doi:10.1083/jcb.200208112 (2002).

80 Huang, S., Tang, D. & Wang, Y. Monoubiquitination of Syntaxin 5 Regulates Golgi Membrane Dynamics during the Cell Cycle. *Dev Cell* **38**, 73-85, doi:10.1016/j.devcel.2016.06.001 (2016).

81 Blythe, E. E., Gates, S. N., Deshaies, R. J. & Martin, A. Multisystem Proteinopathy Mutations in VCP/p97 Increase NPLOC4·UFD1L Binding and Substrate Processing. *Structure* **27**, 1820-1829.e1824, doi:10.1016/j.str.2019.09.011 (2019).

82 Abramzon, Y. *et al.* Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* **33**, 2231.e2231-2231.e2236, doi:10.1016/j.neurobiolaging.2012.04.005 (2012).

83 Guinto, J. B., Ritson, G. P., Taylor, J. P. & Forman, M. S. Valosin-containing protein and the pathogenesis of frontotemporal dementia associated with inclusion

body myopathy. *Acta Neuropathol* **114**, 55-61, doi:10.1007/s00401-007-0224-7 (2007).

84 Miller, J. W. *et al.* Mutation analysis of VCP in British familial and sporadic amyotrophic lateral sclerosis patients. *Neurobiol Aging* **33**, 2721.e2721-2722, doi:10.1016/j.neurobiolaging.2012.06.003 (2012).

85 Saracino, D. *et al.* Novel VCP mutations expand the mutational spectrum of frontotemporal dementia. *Neurobiol Aging* **72**, 187.e111-187.e114, doi:10.1016/j.neurobiolaging.2018.06.037 (2018).

86 Tang, W. K. & Xia, D. Mutations in the human AAA+ chaperone p97 and related diseases. *Front Mol Biosci* **3**, 79, doi:10.3389/fmolb.2016.00079 (2016).

87 Ritz, D. *et al.* Endolysosomal sorting of ubiquitylated caveolin-1 is regulated by VCP and UBXD1 and impaired by VCP disease mutations. *Nat Cell Biol* **13**, 1116-1123, doi:10.1038/ncb2301 (2011).

88 Nalbandian, A. *et al.* The multiple faces of valosin-containing protein-associated diseases: inclusion body myopathy with Paget's disease of bone, frontotemporal dementia, and amyotrophic lateral sclerosis. *J Mol Neurosci* **45**, 522-531, doi:10.1007/s12031-011-9627-y (2011).

89 Tresse, E. *et al.* VCP/p97 is essential for maturation of ubiquitin-containing autophagosomes and this function is impaired by mutations that cause IBMPFD. *Autophagy* **6**, 217-227, doi:10.4161/auto.6.2.11014 (2010).

90 Ramanathan, H. N. & Ye, Y. The p97 ATPase associates with EEA1 to regulate the size of early endosomes. *Cell Res* **22**, 346-359, doi:10.1038/cr.2011.80 (2012).

91 Ching, J. K. *et al.* mTOR dysfunction contributes to vacuolar pathology and weakness in valosin-containing protein associated inclusion body myopathy. *Hum Mol Genet* **22**, 1167-1179, doi:10.1093/hmg/dds524 (2013).

92 Papadopoulos, C. *et al.* VCP/p97 cooperates with YOD1, UBXD1 and PLAA to drive clearance of ruptured lysosomes by autophagy. *EMBO J* **36**, 135-150, doi:10.15252/embj.201695148 (2017).

93 Chou, T. F. *et al.* Specific inhibition of p97/VCP ATPase and kinetic analysis demonstrate interaction between D1 and D2 ATPase domains. *J Mol Biol* **426**, 2886-2899, doi:10.1016/j.jmb.2014.05.022 (2014).

94 Zhang, X. *et al.* Altered cofactor regulation with disease-associated p97/VCP mutations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1705-1714, doi:10.1073/pnas.1418820112 (2015).

95 Manno, A., Noguchi, M., Fukushi, J., Motohashi, Y. & Kakizuka, A. Enhanced ATPase activities as a primary defect of mutant valosin-containing proteins that cause inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia. *Genes Cells* **15**, 911-922, doi:10.1111/j.1365-2443.2010.01428.x (2010).

96 Niwa, H. *et al.* The role of the N-domain in the ATPase activity of the mammalian AAA ATPase p97/VCP. *J Biol Chem* **287**, 8561-8570, doi:10.1074/jbc.M111.302778 (2012).

97 Tang, W. K. & Xia, D. Altered intersubunit communication is the molecular basis for functional defects of pathogenic p97 mutants. *J Biol Chem* **288**, 36624-36635, doi:10.1074/jbc.M113.488924 (2013).

98 Ju, J. S. & Weihl, C. C. Inclusion body myopathy, Paget's disease of the bone and fronto-temporal dementia: a disorder of autophagy. *Hum Mol Genet* **19**, R38-45, doi:10.1093/hmg/ddq157 (2010).

99 Schuetz, A. K. & Kay, L. E. A Dynamic molecular basis for malfunction in disease mutants of p97/VCP. *Elife* **5**, doi:10.7554/eLife.20143 (2016).

100 Schütz, A. K., Rennella, E. & Kay, L. E. Exploiting conformational plasticity in the AAA+ protein VCP/p97 to modify function. *Proc Natl Acad Sci U S A* **114**, E6822-E6829, doi:10.1073/pnas.1707974114 (2017).

101 Huang, R., Ripstein, Z. A., Rubinstein, J. L. & Kay, L. E. Cooperative subunit dynamics modulate p97 function. *Proc Natl Acad Sci U S A* **116**, 158-167, doi:10.1073/pnas.1815495116 (2019).

102 Tang, W. K., Zhang, T., Ye, Y. & Xia, D. Structural basis for nucleotide-modulated p97 association with the ER membrane. *Cell Discov* **3**, 17045, doi:10.1038/celldisc.2017.45 (2017).

103 Fernández-Sáiz, V. & Buchberger, A. Imbalances in p97 co-factor interactions in human proteinopathy. *EMBO Rep* **11**, 479-485, doi:10.1038/embor.2010.49 (2010).

104 Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**, 857-864, doi:10.1016/j.neuron.2010.11.036 (2010).

105 Dreveny, I. *et al.* Structural basis of the interaction between the AAA ATPase p97/VCP and its adaptor protein p47. *EMBO J* **23**, 1030-1039, doi:10.1038/sj.emboj.7600139 (2004).

106 Deshaies, R. J. Proteotoxic crisis, the ubiquitin-proteasome system, and cancer therapy. *BMC Biol* **12**, 94, doi:10.1186/s12915-014-0094-0 (2014).

107 van den Boom, J. & Meyer, H. VCP/p97-Mediated Unfolding as a Principle in Protein Homeostasis and Signaling. *Mol Cell* **69**, 182-194, doi:10.1016/j.molcel.2017.10.028 (2018).

108 Dou, Q. P., Smith, D. M., Daniel, K. G. & Kazi, A. Interruption of tumor cell cycle progression through proteasome inhibition: implications for cancer therapy. *Prog Cell Cycle Res* **5**, 441-446 (2003).

109 Yamamoto, S. *et al.* Expression level of valosin-containing protein (p97) is associated with prognosis of esophageal carcinoma. *Clin Cancer Res* **10**, 5558-5565, doi:10.1158/1078-0432.CCR-0723-03 (2004).

110 Hetz, C., Axten, J. M. & Patterson, J. B. Pharmacological targeting of the unfolded protein response for disease intervention. *Nat Chem Biol* **15**, 764-775, doi:10.1038/s41589-019-0326-2 (2019).

111 Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081-1086, doi:10.1126/science.1209038 (2011).

112 Vembar, S. S. & Brodsky, J. L. One step at a time: endoplasmic reticulum-associated degradation. *Nat Rev Mol Cell Biol* **9**, 944-957, doi:10.1038/nrm2546 (2008).

113 Magnaghi, P. *et al.* Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nat Chem Biol* **9**, 548-556, doi:10.1038/nchembio.1313 (2013).

114 Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an approach to treat cancer through disruption of protein homeostasis. *Cancer Cell* **28**, 653-665, doi:10.1016/j.ccell.2015.10.002 (2015).

115 McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156-163, doi:10.1016/j.ultramic.2014.08.002 (2014).

116 Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc Natl Acad Sci U S A* **111**, 15635-15640, doi:10.1073/pnas.1418377111 (2014).

117 Khoshouei M, R. M., Baumeister W, Danev R. (Nat Commun. 2017 Jun 30;8:16099. doi: 10.1038/ncomms16099. PMID: 28665412; PMCID: PMC5497076.).

118 Dandey, V. P. *et al.* Spotiton: New features and applications. *J Struct Biol* **202**, 161-169, doi:10.1016/j.jsb.2018.01.002 (2018).

119 Ravelli, R. B. G. *et al.* Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat Commun* **11**, 2563, doi:10.1038/s41467-020-16392-5 (2020).

120 Arnold, S. A. *et al.* Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J Struct Biol* **197**, 220-226, doi:10.1016/j.jsb.2016.11.002 (2017).

121 Emsley, P. & Debreczeni, J. E. The use of molecular graphics in structure-based drug design. *Methods Mol Biol* **841**, 143-159, doi:10.1007/978-1-61779-520-6_6 (2012).

122 Mahamid, J. *et al.* A focused ion beam milling and lift-out approach for site-specific preparation of frozen-hydrated lamellas from multicellular organisms. *J Struct Biol* **192**, 262-269, doi:10.1016/j.jsb.2015.07.012 (2015).

123 Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J Struct Biol* **199**, 187-195, doi:10.1016/j.jsb.2017.07.007 (2017).

124 de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. Correlated light and electron microscopy: ultrastructure lights up! *Nat Methods* **12**, 503-513, doi:10.1038/nmeth.3400 (2015).

125 Fu, Z. *et al.* Key Intermediates in Ribosome Recycling Visualized by Time-Resolved Cryoelectron Microscopy. *Structure* **24**, 2092-2101, doi:10.1016/j.str.2016.09.014 (2016).

126 Wehrman, T. *et al.* Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. *Neuron* **53**, 25-38, doi:10.1016/j.neuron.2006.09.034 (2007).

127 Meeker, R. B. & Williams, K. S. The p75 neurotrophin receptor: at the crossroad of neural repair and death. *Neural Regen Res* **10**, 721-725, doi:10.4103/1673-5374.156967 (2015).

128 Lessmann, V., Gottmann, K. & Malcangio, M. Neurotrophin secretion: current facts and future prospects. *Prog Neurobiol* **69**, 341-374, doi:10.1016/s0301-0082(03)00019-4 (2003).

129 Rattenholl, A. *et al.* The pro-sequence facilitates folding of human nerve growth factor from Escherichia coli inclusion bodies. *Eur J Biochem* **268**, 3296-3303, doi:10.1046/j.1432-1327.2001.02232.x (2001).

130 Hasan, W., Pedchenko, T., Krizsan-Agbas, D., Baum, L. & Smith, P. G. Sympathetic neurons synthesize and secrete pro-nerve growth factor protein. *J Neurobiol* **57**, 38-53, doi:10.1002/neu.10250 (2003).

131 Harrington, A. W. *et al.* Secreted proNGF is a pathophysiological death-inducing ligand after adult CNS injury. *Proc Natl Acad Sci U S A* **101**, 6226-6230, doi:10.1073/pnas.0305755101 (2004).

132 Barcelona, P. F. *et al.* p75NTR and Its Ligand ProNGF Activate Paracrine Mechanisms Etiological to the Vascular, Inflammatory, and Neurodegenerative Pathologies of Diabetic Retinopathy. *J Neurosci* **36**, 8826-8841, doi:10.1523/JNEUROSCI.4278-15.2016 (2016).

133 Friedman, W. J. Neurotrophins induce death of hippocampal neurons via the p75 receptor. *J Neurosci* **20**, 6340-6346 (2000).

134 Fahnestock, M. & Shekari, A. ProNGF and Neurodegeneration in Alzheimer's Disease. *Front Neurosci* **13**, 129, doi:10.3389/fnins.2019.00129 (2019).

135 Fahnestock, M., Michalski, B., Xu, B. & Coughlin, M. D. The precursor pro-nerve growth factor is the predominant form of nerve growth factor in brain and is increased in Alzheimer's disease. *Mol Cell Neurosci* **18**, 210-220, doi:10.1006/mcne.2001.1016 (2001).

136 Skeldal, S., Kjaergaard, M. M., Alwasel, S. & Nyengaard, J. R. Establishing a cellular FRET-based fluorescence plate reader assay to monitor proNGF-induced cross-linking of sortilin and the neurotrophin receptor p75(NTR). *Int J Biochem Mol Biol* **6**, 17-25 (2015).

137 Al-Shawi, R. *et al.* ProNGF, sortilin, and age-related neurodegeneration. *Ann N Y Acad Sci* **1119**, 208-215, doi:10.1196/annals.1404.024 (2007).

138 Cabal-Hierro, L. & Lazo, P. S. Signal transduction by tumor necrosis factor receptors. *Cell Signal* **24**, 1297-1305, doi:10.1016/j.cellsig.2012.02.006 (2012).

139 Zeng, F., Lu, J. J., Zhou, X. F. & Wang, Y. J. Roles of p75NTR in the pathogenesis of Alzheimer's disease: a novel therapeutic target. *Biochem Pharmacol* **82**, 1500-1509, doi:10.1016/j.bcp.2011.06.040 (2011).

140 He, X. L. & Garcia, K. C. Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. *Science* **304**, 870-875, doi:10.1126/science.1095190 (2004).

141 Feng, D. *et al.* Molecular and structural insight into proNGF engagement of p75NTR and sortilin. *J Mol Biol* **396**, 967-984, doi:10.1016/j.jmb.2009.12.030 (2010).

142 Nykjaer, A. & Willnow, T. E. Sortilin: a receptor to regulate neuronal viability and function. *Trends Neurosci* **35**, 261-270, doi:10.1016/j.tins.2012.01.003 (2012).

143 Gustafsen, C. *et al.* Sortilin and SorLA display distinct roles in processing and trafficking of amyloid precursor protein. *J. Neurosci.* **33**, 64–71 (2013).

144 Nyborg, A. C., Ladd, T. B., Zwizinski, C. W., Lah, J. J. & Golde, T. E. Sortilin, SorCS1b, and SorLA Vps10p sorting receptors, are novel gamma-secretase substrates. *Mol Neurodegener* **1**, 3, doi:10.1186/1750-1326-1-3 (2006).

145 Nykjaer, A. *et al.* Sortilin is essential for proNGF-induced neuronal cell death. *Nature* **427**, 843-848, doi:10.1038/nature02319 (2004).

146 Quistgaard, E. M. *et al.* Ligands bind to Sortilin in the tunnel of a ten-bladed beta-propeller domain. *Nat Struct Mol Biol* **16**, 96-98, doi:10.1038/nsmb.1543 (2009).

147 Itoh, S., Mizuno, K., Aikawa, M. & Aikawa, E. Dimerization of sortilin regulates its trafficking to extracellular vesicles. *J Biol Chem* **293**, 4532-4544, doi:10.1074/jbc.RA117.000732 (2018).

148 Leloup, N. *et al.* Low pH-induced conformational change and dimerization of sortilin triggers endocytosed ligand release. *Nat Commun* **8**, 1708, doi:10.1038/s41467-017-01485-5 (2017).

149 Januliene, D. *et al.* Acidic Environment Induces Dimerization and Ligand Binding Site Collapse in the Vps10p Domain of Sortilin. *Structure* **25**, 1809-1819.e1803, doi:10.1016/j.str.2017.09.015 (2017).

150 Skeldal, S. *et al.* Mapping of the interaction site between sortilin and the p75 neurotrophin receptor reveals a regulatory role for the sortilin intracellular domain in p75 neurotrophin receptor shedding and apoptosis. *J Biol Chem* **287**, 43798-43809, doi:10.1074/jbc.M112.374710 (2012).

151 Leloup, N., Chataigner, L. M. P. & Janssen, B. J. C. Structural insights into SorCS2-Nerve Growth Factor complex formation. *Nat Commun* **9**, 2979, doi:10.1038/s41467-018-05405-z (2018).

152 Yabe-Wada, T., Matsuba, S., Unno, M. & Onai, N. Crystal structure of the ligand-free form of the Vps10 ectodomain of dimerized Sortilin at acidic pH. *FEBS Lett* **592**, 2647-2657, doi:10.1002/1873-3468.13181 (2018).

153 Kopan, R. & Ilagan, M. X. Gamma-secretase: proteasome of the membrane? *Nat Rev Mol Cell Biol* **5**, 499-504, doi:10.1038/nrm1406 (2004).

154 Skeldal, S., Matusica, D., Nykjaer, A. & Coulson, E. J. Proteolytic processing of the p75 neurotrophin receptor: A prerequisite for signalling?: Neuronal life, growth and death signalling are crucially regulated by intra-membrane proteolysis and trafficking of p75(NTR). *Bioessays* **33**, 614-625, doi:10.1002/bies.201100036 (2011).

155 Shah, S. *et al.* Nicastrin functions as a gamma-secretase-substrate receptor. *Cell* **122**, 435-447, doi:10.1016/j.cell.2005.05.022 (2005).

156 Beel, A. J. & Sanders, C. R. Substrate specificity of gamma-secretase and other intramembrane proteases. *Cell Mol Life Sci* **65**, 1311-1334, doi:10.1007/s00018-008-7462-2 (2008).

157 Hempstead, B. L., Martin-Zanca, D., Kaplan, D. R., Parada, L. F. & Chao, M. V. High-affinity NGF binding requires coexpression of the trk proto-oncogene and the low-affinity NGF receptor. *Nature* **350**, 678-683, doi:10.1038/350678a0 (1991).

158 Beattie, M. S. *et al.* ProNGF induces p75-mediated death of oligodendrocytes following spinal cord injury. *Neuron* **36**, 375-386, doi:10.1016/s0896-6273(02)01005-x (2002).

159 Chao, M. V. & Bothwell, M. Neurotrophins: to cleave or not to cleave. *Neuron* **33**, 9-12, doi:10.1016/s0896-6273(01)00573-6 (2002).

160 Paoletti, F. *et al.* Intrinsic structural disorder of mouse proNGF. *Proteins* **75**, 990-1009, doi:10.1002/prot.22311 (2009).

161 Anastasia, A. *et al.* Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. *Nat Commun* **4**, 2490, doi:10.1038/ncomms3490 (2013).

162 Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim Biophys Acta Gen Subj* **1862**, 324-334, doi:10.1016/j.bbagen.2017.07.020 (2018).

163 Merk, A. *et al.* Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698-1707, doi:10.1016/j.cell.2016.05.040 (2016).

164 Fukumoto, S. *et al.* High-level expression of truncated surface antigen P50 of Babesia gibsoni in insect cells by baculovirus and evaluation of its immunogenicity and antigenicity. *Clin Diagn Lab Immunol* **10**, 596-601, doi:10.1128/cdli.10.4.596-601.2003 (2003).

165 Rattenholl, A. *et al.* Pro-sequence assisted folding and disulfide bond formation of human nerve growth factor. *J Mol Biol* **305**, 523-533, doi:10.1006/jmbi.2000.4295 (2001).

166 Pagadala, P. C., Dvorak, L. A. & Neet, K. E. Construction of a mutated pro-nerve growth factor resistant to degradation and suitable for biophysical and cellular utilization. *Proc Natl Acad Sci U S A* **103**, 17939-17943, doi:10.1073/pnas.0604139103 (2006).

167 Ohi, M., Li. Y., Cheng. Y., Walz. T. (B*iol. Proced. Online* 6**,** 23-34, doi:10.1251/bpo70 . 2004).

168 Cheng, A. *et al.* High resolution single particle cryo-electron microscopy using beam-image shift. *J Struct Biol* **204**, 270-275, doi:10.1016/j.jsb.2018.07.015 (2018).

169 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).

170 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).

171 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

172 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

173 Zhang, G. *et al.* A covalent p97/VCP ATPase inhibitor can overcome resistance to CB-5083 and NMS-873 in colorectal cancer cells. *Eur J Med Chem* **213**, 113148, doi:10.1016/j.ejmech.2020.113148 (2021).

174 Locksley, R. M., Killeen, N. & Lenardo, M. J. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**, 487-501, doi:10.1016/s0092-8674(01)00237-9 (2001).

175 Chan, F. K. *et al.* A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science* **288**, 2351-2354, doi:10.1126/science.288.5475.2351 (2000).

176 Branschädel, M. *et al.* Dual function of cysteine rich domain (CRD) 1 of TNF receptor type 1: conformational stabilization of CRD2 and control of receptor responsiveness. *Cell Signal* **22**, 404-414, doi:10.1016/j.cellsig.2009.10.011 (2010).

177 Agustin Anastasia , P. A. B., Moses V. Chao and Barbara L. Hempstead. (*J Neurosci* **35**, 11911-11920; doi: :10.1523/JNeuroSci.0591-15.2015 (2015).

178 Sykes, A. M. *et al.* The effects of transmembrane sequence and dimerization on cleavage of the p75 neurotrophin receptor by γ-secretase. *J Biol Chem* **287**, 43810-43824, doi:10.1074/jbc.M112.382903 (2012).

179 Grob, P. M., Ross, A. H., Koprowski, H. & Bothwell, M. Characterization of the human melanoma nerve growth factor receptor. *J Biol Chem* **260**, 8044-8049 (1985).

180 Gong, Y., Cao, P., Yu, H. J. & Jiang, T. Crystal structure of the neurotrophin-3 and p75NTR symmetrical complex. *Nature* **454**, 789-793, doi:10.1038/nature07089 (2008).

181 Naismith, J. H., Devine, T. Q., Brandhuber, B. J. & Sprang, S. R. Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor. *J Biol Chem* **270**, 13303-13307, doi:10.1074/jbc.270.22.13303 (1995).

182 Naismith, J. H., Devine, T. Q., Kohno, T. & Sprang, S. R. Structures of the extracellular domain of the type I tumor necrosis factor receptor. *Structure* **4**, 1251-1262, doi:10.1016/s0969-2126(96)00134-7 (1996).

183 Naismith, J. H. & Sprang, S. R. Modularity in the TNF-receptor family. *Trends Biochem Sci* **23**, 74-79, doi:10.1016/s0968-0004(97)01164-x (1998).

184 Mukai, Y. *et al.* Solution of the structure of the TNF-TNFR2 complex. *Sci Signal* **3**, ra83, doi:10.1126/scisignal.2000954 (2010).

185 Compaan, D. M. & Hymowitz, S. G. The crystal structure of the costimulatory OX40-OX40L complex. *Structure* **14**, 1321-1330, doi:10.1016/j.str.2006.06.015 (2006).

186 Dechant, G., Tsoulfas, P., Parada, L. F. & Barde, Y. A. The neurotrophin receptor p75 binds neurotrophin-3 on sympathetic neurons with high affinity and specificity. *J Neurosci* **17**, 5281-5287 (1997).

187 Botta, R. *et al.* Sortilin is a putative postendocytic receptor of thyroglobulin. *Endocrinology* **150**, 509-518, doi:10.1210/en.2008-0953 (2009).

188 Kliemannel, M. *et al.* The mature part of proNGF induces the structure of its pro-peptide. *FEBS Lett* **566**, 207-212, doi:10.1016/j.febslet.2004.04.034 (2004).

189 Esben Trabjerg, F. K., Søren Christensen, Kasper D. Rand,. Conformational characterization of nerve growth factor-β reveals that its regulatory pro-part domain stabilizes three loop regions in its mature part. *J Biol Chem* **292**, 40, 16665-16676, doi: 10.1074/jbc.M117.803320 (2017).

190 Kliemannel, M., Golbik, R., Rudolph, R., Schwarz, E. & Lilie, H. The pro-peptide of proNGF: structure formation and intramolecular association with NGF. *Protein Sci* **16**, 411-419, doi:10.1110/ps.062376207 (2007).

191 Nielsen, M. S., Jacobsen, C., Olivecrona, G., Gliemann, J. & Petersen, C. M. Sortilin/neurotensin receptor-3 binds and mediates degradation of lipoprotein lipase. *J Biol Chem* **274**, 8832-8836, doi:10.1074/jbc.274.13.8832 (1999).

192 Ailor, E. & Betenbaugh, M. J. Modifying secretion and post-translational processing in insect cells. *Curr Opin Biotechnol* **10**, 142-145, doi:10.1016/s0958-1669(99)80024-x (1999).

193 Hudmon, A. & Schulman, H. Neuronal CA2+/calmodulin-dependent protein kinase II: the role of structure and autoregulation in cellular function. *Annu Rev Biochem* **71**, 473-510, doi:10.1146/annurev.biochem.71.110601.135410 (2002).

194 Suter, U., Heymach, J. V. & Shooter, E. M. Two conserved domains in the NGF propeptide are necessary and sufficient for the biosynthesis of correctly processed and biologically active NGF. *EMBO J* **10**, 2395-2400 (1991).

195 Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* **146**, 11-31, doi:10.1016/j.jsb.2003.10.010 (2004).

196 Xia, D., Tang, W. K. & Ye, Y. Structure and function of the AAA+ ATPase p97/Cdc48p. *Gene* **583**, 64-77, doi:10.1016/j.gene.2016.02.042 (2016).

197 Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim Biophys Acta* **1823**, 2-14, doi:10.1016/j.bbamcr.2011.06.014 (2012).

198 Meyer, H. H., Shorter, J. G., Seemann, J., Pappin, D. & Warren, G. A complex of mammalian ufd1 and npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J* **19**, 2181-2192, doi:10.1093/emboj/19.10.2181 (2000).

199 Hetzer, M. *et al.* Distinct AAA-ATPase p97 complexes function in discrete steps of nuclear assembly. *Nat Cell Biol* **3**, 1086-1091, doi:10.1038/ncb1201-1086 (2001).

200 Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol* **9**, 216, doi:10.1186/gb-2008-9-4-216 (2008).

201 Koller, K. J. & Brownstein, M. J. Use of a cDNA clone to identify a supposed precursor protein containing valosin. *Nature* **325**, 542-545, doi:10.1038/325542a0 (1987).

202 Peters, J. M., Walsh, M. J. & Franke, W. W. An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. *EMBO J* **9**, 1757-1767 (1990).

203 Buchberger, A., Schindelin, H. & Hänzelmann, P. Control of p97 function by cofactor binding. *FEBS Lett* **589**, 2578-2589, doi:10.1016/j.febslet.2015.08.028 (2015).

204 Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat Cell Biol* **14**, 117-123, doi:10.1038/ncb2407 (2012).

205 Yeung, H. O. *et al.* Insights into adaptor binding to the AAA protein p97. *Biochem Soc Trans* **36**, 62-67, doi:10.1042/BST0360062 (2008).

206 Stolz, A., Hilt, W., Buchberger, A. & Wolf, D. H. Cdc48: a power machine in protein degradation. *Trends Biochem Sci* **36**, 515-523, doi:10.1016/j.tibs.2011.06.001 (2011).

207 Chapman, E., Fry, A. N. & Kang, M. The complexities of p97 function in health and disease. *Mol Biosyst* **7**, 700-710, doi:10.1039/c0mb00176g (2011).

208 Watts, G. D. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet* **36**, 377-381, doi:10.1038/ng1332 (2004).

209 Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**, 575-597, doi:10.1046/j.1365-2443.2001.00447.x (2001).

210 Nishikori, S., Esaki, M., Yamanaka, K., Sugimoto, S. & Ogura, T. Positive cooperativity of the p97 AAA ATPase is critical for essential functions. *J Biol Chem* **286**, 15815-15820, doi:10.1074/jbc.M110.201400 (2011).

211 Ye, Y., Meyer, H. H. & Rapoport, T. A. Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. *J Cell Biol* **162**, 71-84, doi:10.1083/jcb.200302169 (2003).

212 Davies, J. M., Tsuruta, H., May, A. P. & Weis, W. I. Conformational changes of p97 during nucleotide hydrolysis determined by small-angle X-Ray scattering. *Structure* **13**, 183-195, doi:10.1016/j.str.2004.11.014 (2005).

213 Hänzelmann, P., Buchberger, A. & Schindelin, H. Hierarchical binding of cofactors to the AAA ATPase p97. *Structure* **19**, 833-843, doi:10.1016/j.str.2011.03.018 (2011).

214 Tang, W. K. *et al.* A novel ATP-dependent conformation in p97 N-D1 fragment revealed by crystal structures of disease-related mutants. *EMBO J* **29**, 2217-2229, doi:10.1038/emboj.2010.104 (2010).

215 Zhang, X. *et al.* Structure of the AAA ATPase p97. *Mol Cell* **6**, 1473-1484 (2000).

216 Torrecilla, I., Oehler, J. & Ramadan, K. The role of ubiquitin-dependent segregase p97 (VCP or Cdc48) in chromatin dynamics after DNA double strand breaks. *Philos Trans R Soc Lond B Biol Sci* **372**, doi:10.1098/rstb.2016.0282 (2017).

217 Stach, L. & Freemont, P. S. The AAA+ ATPase p97, a cellular multitool. *Biochem J* **474**, 2953-2976, doi:10.1042/BCJ20160783 (2017).

218 Cooney, I. *et al.* Structure of the Cdc48 segregase in the act of unfolding an authentic substrate. *Science* **365**, 502-505, doi:10.1126/science.aax0486 (2019).

219 Twomey, E. C. *et al.* Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science* **365**, doi:10.1126/science.aax1033 (2019).

220 Bodnar, N. O. & Rapoport, T. A. Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. *Cell* **169**, 722-735.e729, doi:10.1016/j.cell.2017.04.020 (2017).

221 Gates, S. N. & Martin, A. Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Sci* **29**, 407-419, doi:10.1002/pro.3743 (2020).

222 Hänzelmann, P. & Schindelin, H. The Interplay of Cofactor Interactions and Post-translational Modifications in the Regulation of the AAA+ ATPase p97. *Front Mol Biosci* **4**, 21, doi:10.3389/fmolb.2017.00021 (2017).

223 Niwa, H. *et al.* The role of the N-domain in the ATPase activity of the mammalian AAA ATPase p97/VCP. *J. Biol. Chem.* **287**, 8561-8570, doi:10.1074/jbc.M111.302778 (2012).

224 Banerjee, S. *et al.* 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871-875, doi:10.1126/science.aad7974 (2016).

225 Prattes, M. *et al.* A conserved inter-domain communication mechanism regulates the ATPase activity of the AAA-protein Drg1. *Sci Rep* **7**, 44751, doi:10.1038/srep44751 (2017).

226 Beuron, F. *et al.* Conformational changes in the AAA ATPase p97-p47 adaptor complex. *EMBO J* **25**, 1967-1976, doi:10.1038/sj.emboj.7601055 (2006).

227 Huang, C., Li, G. & Lennarz, W. J. Dynamic flexibility of the ATPase p97 is important for its interprotomer motion transmission. *Proc Natl Acad Sci U S A* **109**, 9792-9797, doi:10.1073/pnas.1205853109 (2012).

228 Halawani, D. *et al.* Hereditary inclusion body myopathy-linked p97/VCP mutations in the NH2 domain and the D1 ring modulate p97/VCP ATPase activity and D2 ring conformation. *Mol Cell Biol* **29**, 4484-4494, doi:10.1128/MCB.00252-09 (2009).

229 Buchberger, A., Howard, M. J., Proctor, M. & Bycroft, M. The UBX domain: a widespread ubiquitin-like module. *J Mol Biol* **307**, 17-24, doi:10.1006/jmbi.2000.4462 (2001).

230 Schuberth, C. & Buchberger, A. UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. *Cell Mol Life Sci* **65**, 2360-2371, doi:10.1007/s00018-008-8072-8 (2008).

231 Ewens, C. A. *et al.* The p97-FAF1 protein complex reveals a common mode of p97 adaptor binding. *J Biol Chem* **289**, 12077-12084, doi:10.1074/jbc.M114.559591 (2014).

232 Arumughan, A. *et al.* Quantitative interaction mapping reveals an extended UBX domain in ASPL that disrupts functional p97 hexamers. *Nat Commun* **7**, 13047, doi:10.1038/ncomms13047 (2016).

233 Kim, K. H., Kang, W., Suh, S. W. & Yang, J. K. Crystal structure of FAF1 UBX domain in complex with p97/VCP N domain reveals a conformational change in the conserved FcisP touch-turn motif of UBX domain. *Proteins* **79**, 2583-2587, doi:10.1002/prot.23073 (2011).

234 Kim, S. J. *et al.* Structural basis for ovarian tumor domain-containing protein 1 (OTU1) binding to p97/valosin-containing protein (VCP). *J Biol Chem* **289**, 12264-12274, doi:10.1074/jbc.M113.523936 (2014).

235 Conicella, A. E. *et al.* An intrinsically disordered motif regulates the interaction between the p47 adaptor and the p97 AAA+ ATPase. *Proc Natl Acad Sci U S A* **117**, 26226-26236, doi:10.1073/pnas.2013920117 (2020).

236 Chou, T. F., Bulfer, S. L., Weihl, C. C., Li, K., Lis, L. G., Walters, M. A., Schoenen, F. J., Lin, H. J., Deshaies, R. J., and Arkin, M. R. Specific inhibition of p97/VCP ATPase and kinetic analysis demonstrate interaction between D1 and D2 ATPase domains. *J. Mol. Biol.* **426**, 2886-2899, doi:10.1016/j.jmb.2014.05.022 (2014).

237 Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in. *IUCrJ* **7**, 253-267, doi:10.1107/S2052252520000081 (2020).

238 Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17, doi:10.1107/S205225251801463X (2019).

239 Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).

240 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185, doi:10.1038/s41592-020-01049-4 (2021).

241 Tang, W. K., Odzorig, T., Jin, W. & Xia, D. Structural Basis of p97 Inhibition by the Site-Selective Anticancer Compound CB-5083. *Mol Pharmacol* **95**, 286-293, doi:10.1124/mol.118.114256 (2019).

242 Hoq, M. R. *et al.* Affinity Capture of p97 with Small-Molecule Ligand Bait Reveals a 3.6 Å Double-Hexamer Cryoelectron Microscopy Structure. *ACS Nano* **15**, 8376-8385, doi:10.1021/acsnano.0c10185 (2021).

243 Hänzelmann, P. & Schindelin, H. Structural basis of ATP hydrolysis and intersubunit signaling in the AAA+ ATPase p97. *Structure* **24**, 127-139, doi:10.1016/j.str.2015.10.026 (2016).

244 Yuan, X. *et al.* Structure, dynamics and interactions of p47, a major adaptor of the AAA ATPase, p97. *EMBO J* **23**, 1463-1473, doi:10.1038/sj.emboj.7600152 (2004).

245 Briggs, L. C. *et al.* Analysis of nucleotide binding to P97 reveals the properties of a tandem AAA hexameric ATPase. *J Biol Chem* **283**, 13745-13752, doi:10.1074/jbc.M709632200 (2008).

246 Bulfer, S. L., Chou, T.-F. & Arkin, M. R. p97 Disease Mutations Modulate Nucleotide-Induced Conformation to Alter Protein-Protein Interactions. *ACS Chem. Biol.* **11**, 2112-2116, doi:10.1021/acschembio.6b00350 (2016).

247 Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-951 (1982).

248 Li, G., Zhao, G., Schindelin, H. & Lennarz, W. J. Tyrosine phosphorylation of ATPase p97 regulates its activity during ERAD. *Biochem Biophys Res Commun* **375**, 247-251, doi:10.1016/j.bbrc.2008.08.018 (2008).

249 Ewens, C. A., Kloppsteck, P., Förster, A., Zhang, X. & Freemont, P. S. Structural and functional implications of phosphorylation and acetylation in the regulation of the AAA+ protein p97. *Biochem Cell Biol* **88**, 41-48, doi:10.1139/o09-128 (2010).

250 Trusch, F. *et al.* The N-terminal Region of the Ubiquitin Regulatory X (UBX) Domain-containing Protein 1 (UBXD1) Modulates Interdomain Communication within the Valosin-containing Protein p97. *J Biol Chem* **290**, 29414-29427, doi:10.1074/jbc.M115.680686 (2015).

251 Guo, P. *et al.* Controlling the Revolving and Rotating Motion Direction of Asymmetric Hexameric Nanomotor by Arginine Finger and Channel Chirality. *ACS Nano* **13**, 6207-6223, doi:10.1021/acsnano.8b08849 (2019).

252 Ogura, T., Whiteheart, S. W. & Wilkinson, A. J. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J Struct Biol* **146**, 106-112, doi:10.1016/j.jsb.2003.11.008 (2004).

253 Wang, Q. *et al.* Multifunctional roles of the conserved Arg residues in the second region of homology of p97/valosin-containing protein. *J Biol Chem* **280**, 40515-40523, doi:10.1074/jbc.M509636200 (2005).

254 Bersano, A. *et al.* Inclusion body myopathy and frontotemporal dementia caused by a novel VCP mutation. *Neurobiol Aging* **30**, 752-758, doi:10.1016/j.neurobiolaging.2007.08.009 (2009).

255 Stach, L. *et al.* Crystal structure of the catalytic D2 domain of the AAA+ ATPase p97 reveals a putative helical split-washer-type mechanism for substrate unfolding. *FEBS Lett* **594**, 933-943, doi:10.1002/1873-3468.13667 (2020).

256 Yuan, X. *et al.* Solution structure and interaction surface of the C-terminal domain from p47: a major p97-cofactor involved in SNARE disassembly. *J Mol Biol* **311**, 255-263, doi:10.1006/jmbi.2001.4864 (2001).

257 Ripstein, Z. A., Huang, R., Augustyniak, R., Kay, L. E. & Rubinstein, J. L. Structure of a AAA+ unfoldase in the process of unfolding substrate. *Elife* **6**, doi:10.7554/eLife.25754 (2017).

258 Blythe, E. E., Olson, K. C., Chau, V. & Deshaies, R. J. Ubiquitin- and ATP-dependent unfoldase activity of P97/VCP•NPLOC4•UFD1L is enhanced by a mutation that causes multisystem proteinopathy. *Proc Natl Acad Sci U S A* **114**, E4380-E4388, doi:10.1073/pnas.1706205114 (2017).

259 Sato, Y. *et al.* Structural insights into ubiquitin recognition and Ufd1 interaction of Npl4. *Nat Commun* **10**, 5708, doi:10.1038/s41467-019-13697-y (2019).

260 Zhang, X. *et al.* Conserved L464 in p97 D1-D2 linker is critical for p97 cofactor regulated ATPase activity. *Biochem J* **478**, 3185-3204, doi:10.1042/BCJ20210288 (2021).

261 Boyault, C. *et al.* HDAC6 controls major cell response pathways to cytotoxic accumulation of protein aggregates. *Genes Dev* **21**, 2172-2181, doi:10.1101/gad.436407 (2007).

262 Golbik, R., Lupas, A. N., Koretke, K. K., Baumeister, W. & Peters, J. The Janus face of the archaeal Cdc48/p97 homologue VAT: protein folding versus unfolding. *Biol Chem* **380**, 1049-1062, doi:10.1515/BC.1999.131 (1999).

263 Ju, J. S., Miller, S. E., Hanson, P. I. & Weihl, C. C. Impaired protein aggregate handling and clearance underlie the pathogenesis of p97/VCP-associated disease. *J Biol Chem* **283**, 30289-30299, doi:10.1074/jbc.M805517200 (2008).

264 Wolf, D. H. & Stolz, A. The Cdc48 machine in endoplasmic reticulum associated protein degradation. *Biochim Biophys Acta* **1823**, 117-124, doi:10.1016/j.bbamcr.2011.09.002 (2012).

265 Cao, K., Nakajima, R., Meyer, H. H. & Zheng, Y. The AAA-ATPase Cdc48/p97 regulates spindle disassembly at the end of mitosis. *Cell* **115**, 355-367, doi:10.1016/s0092-8674(03)00815-8 (2003).

266 Richly, H. *et al.* A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* **120**, 73-84, doi:10.1016/j.cell.2004.11.013 (2005).

267 Bonizec, M. *et al.* The ubiquitin-selective chaperone Cdc48/p97 associates with Ubx3 to modulate monoubiquitylation of histone H2B. *Nucleic Acids Res* **42**, 10975-10986, doi:10.1093/nar/gku786 (2014).

268 Dantuma, N. P., Acs, K. & Luijsterburg, M. S. Should I stay or should I go: VCP/p97-mediated chromatin extraction in the DNA damage response. *Exp Cell Res* **329**, 9-17, doi:10.1016/j.yexcr.2014.08.025 (2014).

269 Dargemont, C. & Ossareh-Nazari, B. Cdc48/p97, a key actor in the interplay between autophagy and ubiquitin/proteasome catabolic pathways. *Biochim Biophys Acta* **1823**, 138-144, doi:10.1016/j.bbamcr.2011.07.011 (2012).

270 Bug, M. & Meyer, H. Expanding into new markets--VCP/p97 in endocytosis and autophagy. *J Struct Biol* **179**, 78-82, doi:10.1016/j.jsb.2012.03.003 (2012).

271 Hill, S. M. *et al.* VCP/p97 regulates Beclin-1-dependent autophagy initiation. *Nat Chem Biol* **17**, 448-455, doi:10.1038/s41589-020-00726-x (2021).

272 Fessart, D., Marza, E., Taouji, S., Delom, F. & Chevet, E. P97/CDC-48: proteostasis control in tumor cell biology. *Cancer Lett* **337**, 26-34, doi:10.1016/j.canlet.2013.05.030 (2013).

273 Kakizuka, A. Roles of VCP in human neurodegenerative disorders. *Biochem Soc Trans* **36**, 105-108, doi:10.1042/BST0360105 (2008).

274 Koppers, M. *et al.* VCP mutations in familial and sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* **33**, 837.e837-813, doi:10.1016/j.neurobiolaging.2011.10.006 (2012).

275 Clemen, C. S. *et al.* The heterozygous R155C VCP mutation: Toxic in humans! Harmless in mice? *Biochem Biophys Res Commun* **503**, 2770-2777, doi:10.1016/j.bbrc.2018.08.038 (2018).

276 Wang, Q., Song, C., Yang, X. & Li, C. C. D1 ring is stable and nucleotide-independent, whereas D2 ring undergoes major conformational changes during the ATPase cycle of p97-VCP. *J Biol Chem* **278**, 32784-32793, doi:10.1074/jbc.M303869200 (2003).

277 Banerjee, S. *et al.* 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871-875, doi:10.1126/science.aad7974 (2016).

278 Kaneko, Y. *et al.* p97 and p47 function in membrane tethering in cooperation with FTCD during mitotic Golgi reassembly. *EMBO J* **40**, e105853, doi:10.15252/embj.2020105853 (2021).

279 Bodnar, N. O. *et al.* Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1-Npl4. *Nat Struct Mol Biol* **25**, 616-622, doi:10.1038/s41594-018-0085-x (2018).

280 Hänzelmann, P. & Schindelin, H. Characterization of an Additional Binding Surface on the p97 N-Terminal Domain Involved in Bipartite Cofactor Interactions. *Structure* **24**, 140-147, doi:10.1016/j.str.2015.10.027 (2016).

281 Bebeacua, C. *et al.* Distinct conformations of the protein complex p97-Ufd1-Npl4 revealed by electron cryomicroscopy. *Proc Natl Acad Sci U S A* **109**, 1098-1103, doi:10.1073/pnas.1114341109 (2012).

282 Isaacson, R. L. *et al.* Detailed structural insights into the p97-Npl4-Ufd1 interface. *J Biol Chem* **282**, 21361-21369, doi:10.1074/jbc.M610069200 (2007).

283 Pan, M. *et al.* Seesaw conformations of Npl4 in the human p97 complex and the inhibitory mechanism of a disulfiram derivative. *Nat Commun* **12**, 121, doi:10.1038/s41467-020-20359-x (2021).

284 Liu, C. C., Sun, S. & Sui, S. F. The role of the N-D1 linker of the N-ethylmaleimide-sensitive factor in the SNARE disassembly. *PLoS One* **8**, e64346, doi:10.1371/journal.pone.0064346 (2013).

285 Li, G., Huang, C., Zhao, G. & Lennarz, W. J. Interprotomer motion-transmission mechanism for the hexameric AAA ATPase p97. *Proc Natl Acad Sci U S A* **109**, 3737-3741, doi:10.1073/pnas.1200255109 (2012).

286 Tang, W. K. & Xia, D. Role of the D1-D2 linker of human VCP/p97 in the asymmetry and ATPase activity of the D1-domain. *Sci Rep* **6**, 20037 (2016).

287 Erzurumlu, Y. *et al.* A unique IBMPFD-related P97/VCP mutation with differential binding pattern and subcellular localization. *Int J Biochem Cell Biol* **45**, 773-782, doi:10.1016/j.biocel.2013.01.006 (2013).

288 Dantuma, N. P., Lindsten, K., Glas, R., Jellne, M. & Masucci, M. G. Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nat Biotechnol* **18**, 538-543, doi:10.1038/75406 (2000).

289 Chou, T. F. & Deshaies, R. J. Quantitative cell-based protein degradation assays to identify and classify drugs that target the ubiquitin-proteasome system. *J Biol Chem* **286**, 16546-16554, doi:10.1074/jbc.M110.215319 (2011).

290 Chou, T. F. *et al.* Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *Proc Natl Acad Sci U S A* **108**, 4834-4839, doi:10.1073/pnas.1015312108 (2011).

291 Dalal, S., Rosser, M. F., Cyr, D. M. & Hanson, P. I. Distinct roles for the AAA ATPases NSF and p97 in the secretory pathway. *Mol Biol Cell* **15**, 637-648, doi:10.1091/mbc.e03-02-0097 (2004).

292 Cheng, A. *et al.* High resolution single particle cryo-electron microscopy using beam-image shift. *J. Struct. Biol.* **204**, 270-275, doi:10.1016/j.jsb.2018.07.015 (2018).

293 Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).

294 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

295 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).

296 Her, N. G. *et al.* p97 Composition Changes Caused by Allosteric Inhibition Are Suppressed by an On-Target Mechanism that Increases the Enzyme's ATPase Activity. *Cell Chem Biol* **23**, 517-528, doi:10.1016/j.chembiol.2016.03.012 (2016).

297 Pye, V. E. *et al.* Structural insights into the p97-Ufd1-Npl4 complex. *Proc Natl Acad Sci U S A* **104**, 467-472, doi:10.1073/pnas.0603408104 (2007).

298 Gui, L. *et al.* Evaluating p97 Inhibitor Analogues for Potency against p97-p37 and p97-Npl4-Ufd1 Complexes. *ChemMedChem* **11**, 953-957, doi:10.1002/cmdc.201600036 (2016).

299 Davies, J. M., Brunger, A. T. & Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16**, 715-726, doi:10.1016/j.str.2008.02.010 (2008).

300 Chapman, E., Maksim, N., de la Cruz, F. & La Clair, J. J. Inhibitors of the AAA+ chaperone p97. *Molecules* **20**, 3027-3049, doi:10.3390/molecules20023027 (2015).

301 Huryn, D. M., Kornfilt, D. J. P. & Wipf, P. p97: An Emerging Target for Cancer, Neurodegenerative Diseases, and Viral Infections. *J Med Chem* **63**, 1892-1907, doi:10.1021/acs.jmedchem.9b01318 (2020).

302 Zhang, G., Li, S., Cheng, K. W. & Chou, T. F. AAA ATPases as therapeutic targets: Structure, functions, and small-molecule inhibitors. *Eur J Med Chem* **219**, 113446, doi:10.1016/j.ejmech.2021.113446 (2021).

303 Chou, T. F., Li, K., Frankowski, K. J., Schoenen, F. J. & Deshaies, R. J. Structure-activity relationship study reveals ML240 and ML241 as potent and selective inhibitors of p97 ATPase. *ChemMedChem* **8**, 297-312, doi:10.1002/cmdc.201200520 (2013).

304 Feng Wang. *et al.* Allosteric p97 Inhibitors Can Overcome Resistance to ATP-Competitive p97 Inhibitors for Potential Anticancer Therapy. *ChemMedChem* **15**, 685-694, doi:10.1002/cmdc.201900722 (2020).

305 Fang, C. J. *et al.* Evaluating p97 inhibitor analogues for their domain selectivity and potency against the p97-p47 complex. *ChemMedChem* **10**, 52-56, doi:10.1002/cmdc.201402420 (2015).

306 Bastola, P. *et al.* Specific mutations in the D1-D2 linker region of VCP/p97 enhance ATPase activity and confer resistance to VCP inhibitors. *Cell Death Discov* **3**, 17065, doi:10.1038/cddiscovery.2017.65 (2017).

307 Dreveny, I. *et al.* p97 and close encounters of every kind: a brief review. *Biochem Soc Trans* **32**, 715-720, doi:10.1042/BST0320715 (2004).

308 Weith, M. *et al.* Ubiquitin-Independent Disassembly by a p97 AAA-ATPase Complex Drives PP1 Holoenzyme Formation. *Mol Cell* **72**, 766-777.e766, doi:10.1016/j.molcel.2018.09.020 (2018).

309 Bruderer, R. M., Brasseur, C. & Meyer, H. H. The AAA ATPase p97/VCP interacts with its alternative co-factors, Ufd1-Npl4 and p47, through a common bipartite binding mechanism. *J Biol Chem* **279**, 49609-49616, doi:10.1074/jbc.M408695200 (2004).

310 Pan, M. *et al.* Mechanistic insight into substrate processing and allosteric inhibition of human p97. *Nat Struct Mol Biol* **28**, 614-625, doi:10.1038/s41594-021-00617-2 (2021).

311 Xue, L. *et al.* Valosin-containing protein (VCP)-Adaptor Interactions are Exceptionally Dynamic and Subject to Differential Modulation by a VCP Inhibitor. *Mol Cell Proteomics* **15**, 2970-2986, doi:10.1074/mcp.M116.061036 (2016).

312 Ciferri, C. *et al.* Molecular architecture of human polycomb repressive complex 2. *Elife* **1**, e00005, doi:10.7554/eLife.00005 (2012).

313 Ciferri, C., Lander, G. C. & Nogales, E. Protein domain mapping by internal labeling and single particle electron microscopy. *J Struct Biol* **192**, 159-162, doi:10.1016/j.jsb.2015.09.016 (2015).

APPENDIX A

CRYO-EM DATA STATISTICS OF P97 DATASETS

Table 1. Statistics of Single-particle Cryo-EM Structure Determination on $p97^{R155H}$ -p47 Assembly

Protein	$p97^{R155H}$ – p47				
Data collection					
Electron microscope	Thermo Fisher/FEI Titan Krios TEM				
Accelerating voltage (kV)	300				
Spherical aberration constant (mm)	2.7				
Detector camera	Gatan K2 Summit DED camera				
Defocus (μm)	-0.8 - -2.5				
Nominal magnification	48,077×				
Physical pixel size (Å/pixel)	1.04				
Image dose (e^-/Å^2)	44.3				
Image processing					
Number of movies		4,223		3,512	2,796
Number of particles selected (initial)		368,575		1,124,232	401,303
Group	$p97^{R155H}$ dodecamer (no nucleotides) EMD-23191 PDB code: 7L5W	$p97^{R155H}$-p47 (no nucleotides) EMD-23194 PDB code: 7L5Y	$p97^{R155H}$ dodecamer EMD-23192 PDB code: 7L5X	$p97^{R155H}$-p47 (ADP) EMD-23196 PDB code: 7L5Z	$p97^{R155H}$-p47 (ATPγS) EMD-23198 PDB code: 7L60
Number of particles used for final 3D density (final)	64,252	203,242	2,219	160,050	63,353
Spatial frequency at FSC of 0.143 (Å^{-1})	3.34	4.30	6.10	4.50	4.23
Imposed symmetry	D6	C1	D6	C1	C1
Sharpening *b*-factor (Å^2)	-78.6	-139.3	-183.2	-122.3	-124.8
Modeling					
Initial model used (PDB code)	5FTK	5FTK 1S3S	5FTK	5FTK 1S3S	5FTN 1S3S
Model resolution (Å)	3.5	4.3	7.4	4.2	4.3
Model composition					
Non-hydrogen atoms	51,828	25,290	31,512	26,016	33,239
Protein residues	6,636	3,312	6,384	3,282	4,219
Ligands					
ADP	0	0	0	12	0
ATPγS	0	0	0	0	12
***B* factors (Å^2)**					
Protein	43.9	73.4	433.7	127.1	117.1
Ligands	0	0	0	119.3	63.2
RMS deviations					
Bond length (Å)	0.006	0.004	0.003	0.005	0.004
Bond angle (°)	1.002	0.955	0.657	1.162	0.953
Clash score	11.15	10.62	2.88	8.08	6.53
MolProbity score	2.00	2.10	1.50	2.01	1.89
Rotamer outlier (%)	0.00	0.00	0.00	0.00	0.03
Ramachandran plot (%)					
Disallowed	0.00	0.00	0.00	0.00	0.10
Allowed	6.75	10.07	6.46	10.67	9.26

Table 2. Statistics of Single-particle Cryo-EM Structure Determination on $p97^{L464P}$

Protein	Human $p97^{L464P}$ ATPase mutant	
Data collection		
Electron microscope	Thermo Fisher/FEI Titan Krios TEM	
Accelerating voltage (kV)	300	
Spherical aberration constant (mm)	2.7	
Detector camera	Gatan K2 Summit DED camera	
Defocus (μm)	-0.8 - -3.0	
Nominal magnification	48,077×	
Physical pixel size (Å/pixel)	1.04	
Image dose (e^-/Å^2)	64.8	
Image processing		
Number of movies	4,490	
Number of particles selected (initial)	479,031	
	(PDB code: 7MDM) (EMDB: EMD-23775)	Symmetrized D1 and D2 (PDB code: 7MDO) (EMDB: EMD-23776)
Number of particles used for final 3D density (final)	81,865	
Spatial frequency at FSC of 0.143 (Å^{-1})	5.16	4.12
Imposed symmetry	C1	C6
Sharpening *b*-factor (Å^2)	-279.2	-158.4
Modeling		
Initial model used (PDB code)	5FTK	5FTK
Model composition		
Non-hydrogen atoms	31,986	25,098
Protein residues	31,824	24,936
ADP	162	162
***B* factors (Å^2)**		
Protein	271.1	138.4
Ligands	218.4	114.0
RMS deviations		
Bond length (Å)	0.005	0.005
Bond angle (°)	1.227	1.062
Clash score	9.38	6.95
MolProbity score	1.98	1.82
Rotamer outlier (%)	0.03	0.00
Ramachandran plot (%)		
Disallowed	0.02	0.00
Allowed	8.00	6.67
Favored	91.98	93.33

APPENDIX B

LIST OF PUBLICATIONS

List of Publications During Graduate School

(# denotes co-first authors)

- **Nandi, P.**; Wang, F.; Columbres, R.C.A.; Gang, T.; Li, S.; Poh, Y.-P.; Chou, T.-F.; and Chiu, P.L.[*] Cryo-EM reveals inhibition mechanism of human p97 ATPase by novel small-molecule drug candidates. (*Manuscript in preparation*)

- Zhang, X.[#]; Gui, L.[#]; Li, S.[#]; **Nandi, P.**[#]; Columbres, R.C.A.; Wong, D.E.; Moen, D.R.; Lin, H.J.; Chiu, P.-L.[*]; and Chou, T.-F.[*] Conserved L464 in p97 D1-D2 linker is critical for p97 cofactor regulated ATPase activity. *Biochem J* **478,** 3185–3204 (2021). https://doi.org/10.1042/BCJ20210288

- **Nandi, P.**[#]; Li, S.[#]; Columbres, R.C.A.; Wang, F.; Williams, D.R.; Poh, Y.-P.; Chou, T.-F.[*]; Chiu, P.-L.[*] Structural and functional analysis of disease-linked p97 ATPase mutant complexes. *Int J Mol Sci* **22**, 8079 (2021).doi: 10.3390/ijms22158079.

- Godeshala, S.[#]; Miryala, B.[#]; Dutta, S.[#]; Christensen, M.D.[#]; **Nandi, P.**; Chiu, P.-L.; Rege, K.[*]. A library of aminoglycoside-derived lipopolymer nanoparticles for delivery of small molecules and nucleic acids. *J Mater Chem B* **8**, 8558-8572 (2020). https://doi.org/10.1039/D0TB00924E

- Narayanan, R. P. [#]; Procyk, J. [#]; **Nandi, P.**; Prasad, A.; Xu, Y.; Poppleton, E.; Williams, D.R.; Yan, H.; Chiu, P.-L.[*]; Stephanopoulos, N.[*]; Šulc, P.[*], Experimental characterization and optimization of hybrid protein-DNA nanostructures using coarse-grained simulations. (*under review*)

Conference Proceedings:

- **Nandi, P.**, et al. Mechanism of the alteration in domain-domain communications in human p97/VCP ATPase. *Biophys J* **121**, 298a (2022).

- **Nandi, P.**, et al. Structural and functional analysis of p47 cofactor binding on the p97 disease Mutant. *Microsc & Microana* **27**, 1908–1910 (2021).

- **Nandi, P.**, et al. Structural and functional analysis of p47 cofactor binding on the p97 disease mutant. *Alzheimer's Dement.* **17** (Suppl. 3):e054008 (2021) (doi:10.1002/alz.054008).

- **Nandi, P.**, et al. Structural studies of the neuronal apoptotic complex of the proNGF-$p75^{NTR}$-sortilin. *Biophys J* **120**, 120a (2021).

- **Nandi, P.,** et al. Structural studies using cryo-EM to unravel mechanistic details of p47 binding to p97. *Biophys J* **118**, 501a (2020).

Purbasha Nandi <pnandi@asu.edu>

Case #01587371 - Rights and permission for PhD thesis/dissertation

1 message

customercare@copyright.com <customercare@copyright.com> Mon, Apr 11, 2022 at 2:28 AM
To: "pnandi@asu.edu" <pnandi@asu.edu>

Dear Purbasha Nandi,

Thank you for contacting Copyright Clearance Center's Customer Service. We act on behalf of copyright owners in providing permissions to our customers. Permission availability can vary depending on a rightsholder and also on the permission type.

My name is Anda, and it would be my pleasure to assist you. Based on the information you provided, I am happy to inform you that the article ***'Electron cryomicroscopy as a powerful tool in biomedical research'*** is distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. This means that in order to reuse figure 1 **you are not required to obtain permission to reuse this article.**

I wish you the best of luck and reach back anytime,
Anda

Anda Danciu
Customer Account Specialist
Copyright Clearance Center
222 Rosewood Drive
Danvers, MA 01923
www.copyright.com
Toll Free US +1.855.239.3415
International +1.978-646-2600

Facebook - Twitter - LinkedIn

---------- Original Message ----------
From: Purbasha Nandi [pnandi@asu.edu]
Sent: 4/10/2022 10:17 PM
To: info@copyright.com
Subject: Rights and permission for PhD thesis/dissertation

Hello,

I would like to use figure 1 from the article "Electron cryomicroscopy as a powerful tool in biomedical research" for my dissertation titled " Cryo-EM enables structural studies of proteins relevant to neurodegeneration" to be submitted in April 2022 for my Ph.D. in Chemistry in Arizona State University.

The link is https://pubmed.ncbi.nlm.nih.gov/29730699/

I kindly request you to provide me with permission to do the same.

Many thanks,

Purbasha Nandi

Purbasha Nandi <pnandi@asu.edu>

Chapter permission

Chou, Tsui-Fen <tfchou@caltech.edu> Mon, Apr 4, 2022 at 11:34 AM
To: Purbasha Nandi <pnandi@asu.edu>

Hi Purbasha
You have my permission.
Will try to attend
Best
Tsuifen

Sent from my iPhone

On Apr 4, 2022, at 11:31 AM, Purbasha Nandi <pnandi@asu.edu> wrote:

Dear Tsui-Fen,

Hope you are doing well. My defense is scheduled for next week Tuesday, April 12th at 9:00 am PST and I would like to invite you to the same if you are available. It will be held in a hybrid option, and here is the zoom link: https://asu.zoom.us/j/83399704415

I have Po-Lin's permission but I wanted to request your permission as well if it will be fine on your end if I prepare two chapters based on the p97 manuscripts, and discuss the rest of my research findings on the neuronal complex of proNGF in the other chapters. The current version of my thesis has two chapters adapted and reprinted from the two published papers:

Nandi, P.[#]; Li, S.[#]; Columbres, R.C.A.; Wang, F.; Williams, D.R.; Poh, Y.-P.; Chou, T.-F.[*]; Chiu, P.-L.[*] Structural and functional analysis of disease-linked
p97 ATPase mutant complexes. *Int J Mol Sci* **22**, 8079 (2021).

Zhang, X.[#]; Gui, L.[#]; Li, S.[#]; Nandi, P.[#]; Columbres, R.C.A.; Wong, D.E.; Moen, D.R.; Lin, H.J.; Chiu, P.-L.[*]; and Chou, T.-F.[*] Conserved L464 in p97 D1-D2
the linker is critical for p97 cofactor-regulated ATPase activity. *Biochem J* **478**, 3185–3204 (2021).

This request is non-exclusive, and irrevocable, and it is not intended to interfere with other uses of the same work by you. I would be pleased to include a full citation to your work and other acknowledgment as you might request. Our university also requires this permission to be attached to the thesis appendix (please find the attached document).

I would greatly appreciate your permission and please feel free to let me know at your convenience. If you require any additional information, please let me know.

Best regards,
Purbasha Nandi

BIOGRAPHICAL SKETCH

Purbasha Nandi was born in West Bengal, India on March 9th, 1993. She received her elementary and secondary education at Auxilium Convent School, Bandel, India. For her commendable performance in the Class XII All India Board examination ISC in 2011, she secured INSPIRE-DST Scholarship for being in top 1% (All India-rank: 9), for the next five years 2011-2016 to conduct higher studies and fully funded summer research projects. In 2011, she entered the prestigious Jadavpur University, Kolkata, India to pursue her Bachelor's in Chemistry, with Honors and completed her degree with Merit awards by June 2014. She was also able to conduct summer research during these years where she had the first-hand experience of research and loved it. She further pursued her Master's in Chemistry from the renowned Indian Institute of Technology (IIT), Madras in a two-year program August 2014- June 2016. She received the Master's Best Thesis Award, IIT Madras, India, for the research performed resulting in two peer-reviewed publications. Later in August 2016, she took a daunting decision to travel to the United States to pursue her doctoral studies from Arizona State University (ASU), famous for being top in innovative research in Chemistry and Biochemistry. Her doctoral studies in structural biology using cryo-electron microscopy were rewarding with ~5 peer-reviewed publications and presentations of the research performed at ~10 conferences (poster and platform talks). Along with her research, she also taught undergraduate organic-chemistry and biochemistry courses at School of Molecular Sciences, ASU. She was recognized for her work with the Outstanding Graduate Teaching Assistant (TA) award (2022) and Outstanding Graduate Research Assistant (RA) award (2020) from School of Molecular Sciences, ASU, the Graduate Excellence Award (2022), Graduate College, ASU and Outstanding Mentor Award (2022), Graduate and Professional Student Association (GPSA), ASU. She was a recipient of Graduate Research Support (GRSP) Completion Fellowship (2021), GPSA, ASU. She also enjoyed volunteering for the Biophysical Society Arizona Student Chapter as a General Secretary (2021-2022) and teaching high-school students as part of GPSE, ASU.