technische universität
dortmund

mpi

Structural studies on human RZZ and Cln3p

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

des Fachbereichs Chemie und Chemische Biologie der Technischen Universität Dortmund

angefertigt am

Max-Planck-Institut für Molekulare Physiologie

vorgelegt von

Shyamal Narayan Mosalaganti

geboren am 20.09.1985 in Jamshedpur, Indien

Dortmund, Februar 2014

Declaration/Erklärung

The work described in this dissertation was carried out between September 2009 and February 2014 under the guidance of Prof. Dr. Stefan Raunser at the Max-Planck-Institute for Molecular Physiology, Dortmund.

I hereby declare that I carried out the work independently and did not use any aid, other than the ones mentioned.

Die vorliegende Arbeit wurde in der Zeit von September 2009 bis Februar 2014 am Max-Planck-Institut für Molekulare Physiologie in Dortmund unter der Anleitung von Prof. Dr. Stefan Raunser durchgeführt.

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständing und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Shyamal Narayan Mosalaganti, Dortmund

.

1. Gutachter/1st Examiner: Prof. Dr. Roger S. Goody

2. Gutachter/2nd Examiner: Prof. Dr. Roland Winter

To my mother

Table of contents

List of tables

Zusammenfassung

Die vorliegende Dissertation ist in zwei Thematiken unterteilt:

1.Molekulare Mechanismen der Segregation von Chromosomen

Die Segregation der Chromosomen ist ein zentraler Prozess während der Zellteilung. Dieser wird durch hierarchische Proteinkomplexe reguliert, die Kinetochore. Sie kontrollieren die synchrone Teilung der Schwesterchromatiden in höheren Eukaryoten in einem Prozess, der als *spindle assembly checkpoint* (SAC) bezeichnet wird. Der 800 kDa große ROD-Zw10-Zwilch (RZZ) Proteinkomplex besitzt eine Schlüsselfunktion in der Formierung des SAC sowie der Anheftung der Mikrotubuli an die Kinetochore. Diese Arbeit beschreibt die erste mittels Kryo-Elektronenmikroskopie bestimmte hochaufgelöste Struktur des RZZ Komplexes. Die Organisation der Bestandteile des Komplexes wurde durch Einbeziehung diverser biochemischer Methoden charakterisiert. Diese Arbeit stellt den ersten Schritt für ein strukturelles Verständnis des SAC dar und ermöglicht hierdurch ein besseres Verständnis der komplexen Signalwege, welche die Zellteilung steuern.

2. Molekulare Grundlagen der Batten-Krankheit

Das Ceroid-lipofuscinosis neuronal-3 (*cln3*) Gen kodiert für ein Transmembranprotein (Cln3p). Mutationen dieses Proteins bilden die Grundlage für die Batten-Krankheit, die am weitesten verbreitete aller neuronalen Ceroid-Lipofuscinosen (NCL). Diese vererbbaren neuronalen Krankheiten sind durch den Verlust der Sehkraft, myoklonische Anfälle, Verlust der geistigen Fähigkeiten sowie Störungen der motorischen Fertigkeiten charakterisiert. In dieser Arbeit wird die erste erfolgreiche Expression und Reinigung des humanen Cln3p mittels eines Baculovirus-Insektenzellen-Expressionssystems beschrieben. Initiale biochemische Charakterisierungen der posttranslationalen Modifikationen des Proteins wurden durchgeführt. Weiterhin wurden erste Versuche zum Aufklärung der Proteinfunktion und der Struktur des Proteins begonnen.

Abstract

The work presented in this thesis can be divided into two topics:

1. Molecular mechanism of chromosome segregation

Chromosome segregation is a central process in cell division. Hierarchical protein assemblies called kinetochores navigate this process. They monitor the synchronous separation of sister chromatids in higher eukaryotes by a pathway commonly referred to as spindle assembly checkpoint (SAC). The 800 kDa ROD-Zw10-Zwilch (RZZ) complex is a key player both in SAC and formation of kinetochore – microtubule attachments. This work highlights the first high-resolution structure of the RZZ complex determined by cryo-electron microscopy. The organization of constituent proteins in the complex has been understood by interweaving various biochemical methods. This study is essentially the first step in understanding the SAC from a structural perspective and enables to use hints from the structure in understanding the complex signaling process that drives cell division.

2. Molecular basis of the Batten disease

Ceroid – lipofuscinosis neuronal-3 (*cln3*) gene encodes a trans-membrane protein (Cln3p) mutations in which cause Batten's disease. Batten disease is the most common of all neuronal ceroid-lipofuscinoses (NCL, genetically inherited neuronal disorders) characterized by loss of vision, myoclonic seizures, loss of cognitive function, motor dysfunction. This work reports the first expression and purification of human Cln3p from the baculovirus driven insect cell expression system. Initial biochemical characterization of the protein is also reported. Efforts to understand the protein function have been initialized and variety of techniques used to understand the structure of the protein.

List of Abbreviations

AU	Absorbance Units
BPL	Brain polar lipids
CCD	Charge coupled device
Cdc	Cell division control
Cdk	Cyclin dependent kinase
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cln3p	Ceroid – lipofuscinosis, neuronal-3 protein
hCln3p	Human ceroid – lipofuscinosis, neuronal-3 protein
mCln3p	Mouse ceroid – lipofuscinosis, neuronal-3 protein
rCln3p	Rat ceroid – lipofuscinosis, neuronal-3 protein
DDM	Lauryl-β- D -maltoside
DM	n-Dodecyl β-D-maltoside
DMPC	1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
EGFP	Enhanced green fluorescent protein
EPL	*E. coli* polar lipids
EM	Electron microscopy
ER	Endoplasmic reticulum
Fos-choline®-12	n-Dodecylphosphocholine
Fos-choline®-13	n-Tridecylphosphocholine
GalT	Galactose-1-phosphate uridylyltransferase
GC-MS	Gas chromatography-mass spectrometry
HPLC	High pressure liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mCherry	Monomeric cherry
MNG-3	Maltose neopentyl glycol
$NADPH+H^+$	Nicotinamide adenine dinucleotide phosphate (reduced form)
OG	*n*-octyl-β-D-glucoside
PAGE	Polyacrylamide gel electrophoresis
RZZ	ROD-Zw10-Zwilch
SDS	sodium dodecyl sulfate
4k	Square area with dimensions of 4096 pixels
8k	Square area with dimensions of 8192 pixels

1 Introduction

1.1 Mitosis

1.1.1 General principles

Cell cycle is the process of duplication of a cell into daughter cells. A cascade of indispensable signaling events regulates this process, propelling it irreversibly forward. The point of cell division is to duplicate and distribute the DNA (condensed as chromosomes) into the daughter cells precisely. Any digression or defect otherwise, generally is of irreparable consequence to the cell, usually reflected as cancers [1]. Various proteins regulate the fidelity of the cell cycle; of these, cyclins are of paramount importance. G1/S, S and M are three major classes of cyclins that interact with Cdk (cyclin dependent kinase) to generate cyclin-cdk complexes. Oscillations in the levels of these complexes control the stages of the cell cycle [2, 3]. Mitosis, the shortest and visually most stunning of all the phases in cell cycle, is responsible for the distribution of replicated DNA.

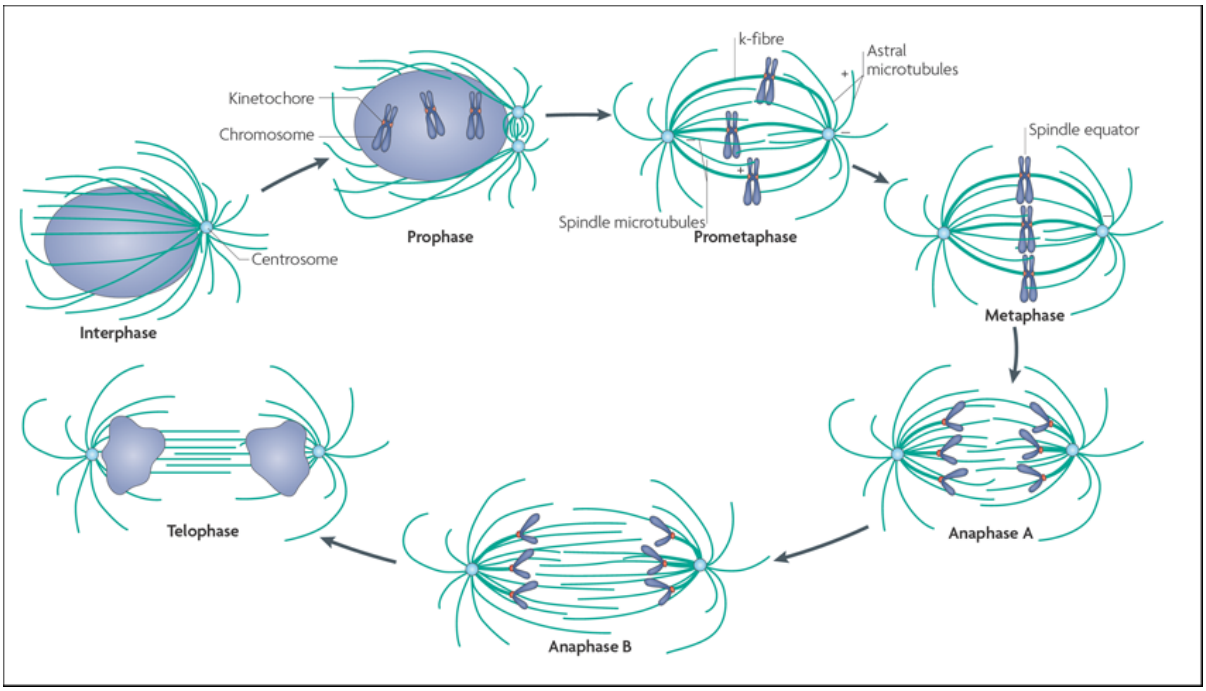


Figure 1.1: Stages in mitosis. Mitosis is comprised of distinct phases. The onset of cell division is marked by duplication and condensation of genetic material into chromosomes. In prophase chromosomes become highly condensed and the centrosomes begin to drift apart. Nuclear envelope breakdown (NEBD) marks the transition from prophase to prometaphase making chromosomes not bound to the nuclear periphery. During prometaphase the spindle microtubules connect with kinetochores on the chromosomes, so that they can align at the spindle equator, a process, which outlines metaphase. Microtubules are indispensible to the whole process. They are bipolar, with their 'minus' end at the centrosomes and 'plus' end engaging either, with the chromosomes (spindle microtubules) or the cell cortex (astral microtubules). Anaphase is defined by separation of the sister chromatids towards opposite ends of the cell. The nuclear envelope starts to re-appear and the DNA starts to unwind, marking the penultimate stage of mitosis. Mitosis culminates when the cell divides, with two daughter nuclei segregated into different cells, which are ready to re-start the process. Source: [4]

Decades of research have led us to a standpoint where the key players have been identified and crucial events outlined. Briefly, mitosis constitutes of five distinct phases, defined primarily on the

spatial organization of chromosomes (**Figure 1.1**)[4]. At the heart of the process lies what is called chromosome segregation, where Cohesin (a multi-protein complex that binds sister chromatids) is cleaved leading to concomitant release of sister chromatids that are subsequently driven towards opposite spindle poles [5].

1.1.2 Kinetochores - molecular assemblies at the helm of chromosome segregation

Kinetochores are macromolecular protein assemblies, which are at the helm of chromosome segregation. These protein assemblies, with loci at the chromosomes, engage primarily in forming stable load bearing attachments to the dynamic ends of microtubules (dynamic polymers of α, β-tubulin). Such attachments underline the process of correct sister chromatid separation [6]. *Saccharomyces cerevisiae* bears the simplest known kinetochore (binds a single microtubule) and contains approximately 40 proteins, which are grouped into 7 different complexes [7, 8]. The kinetochore organization of higher eukaryotes (which contain up to 100 proteins and bind multiple microtubules) has remained widely speculative. On one hand constitutive protein complexes are conserved from yeast to higher eukaryotes, hinting at the fact that eukaryotic kinetochores are possibly multimeric assemblies of yeast constituents [9, 10] and on the other, electron tomography studies on the kinetochores from PtK_1 cells show that the outer plate is protein meshwork which seemed to not exhibit any definite architecture [11]. Single particles of purified yeast kinetochores observed by negative stain electron microscopy revealed a central hub surrounded by several domains [12]. Complete architecture of eukaryotic kinetochore, however, is still obscure.

On organizational level kinetochores are thought to be made up of the inner kinetochore, the outer kinetochore and the corona (**Figure 1.2**). Kinetochores are built up on a homologue of the H3 nucleosome called CENP-A (Centromeric protein-A). CENP-A interacts with various other centromeric proteins such as CENP-C, H, I, K, L, M, O, P, Q, R, S, T, U and W to form what is known as constitutive centromere-associated network complex (CCAN, **Figure 1.2**)[13-15]. The CCAN network constitutes the inner kinetochore region and primarily interacts with the centromeric DNA. The components of CCAN network can be further sub classified, supported by biochemical and functional analysis, to CENP-T/W, CENP-H/I/K and CENP-O/P/Q/R/U complexes. The sub-units of CCAN network are interdependent on each other for their localization at the kinetochore and serve as a platform for other components to rest upon [16].

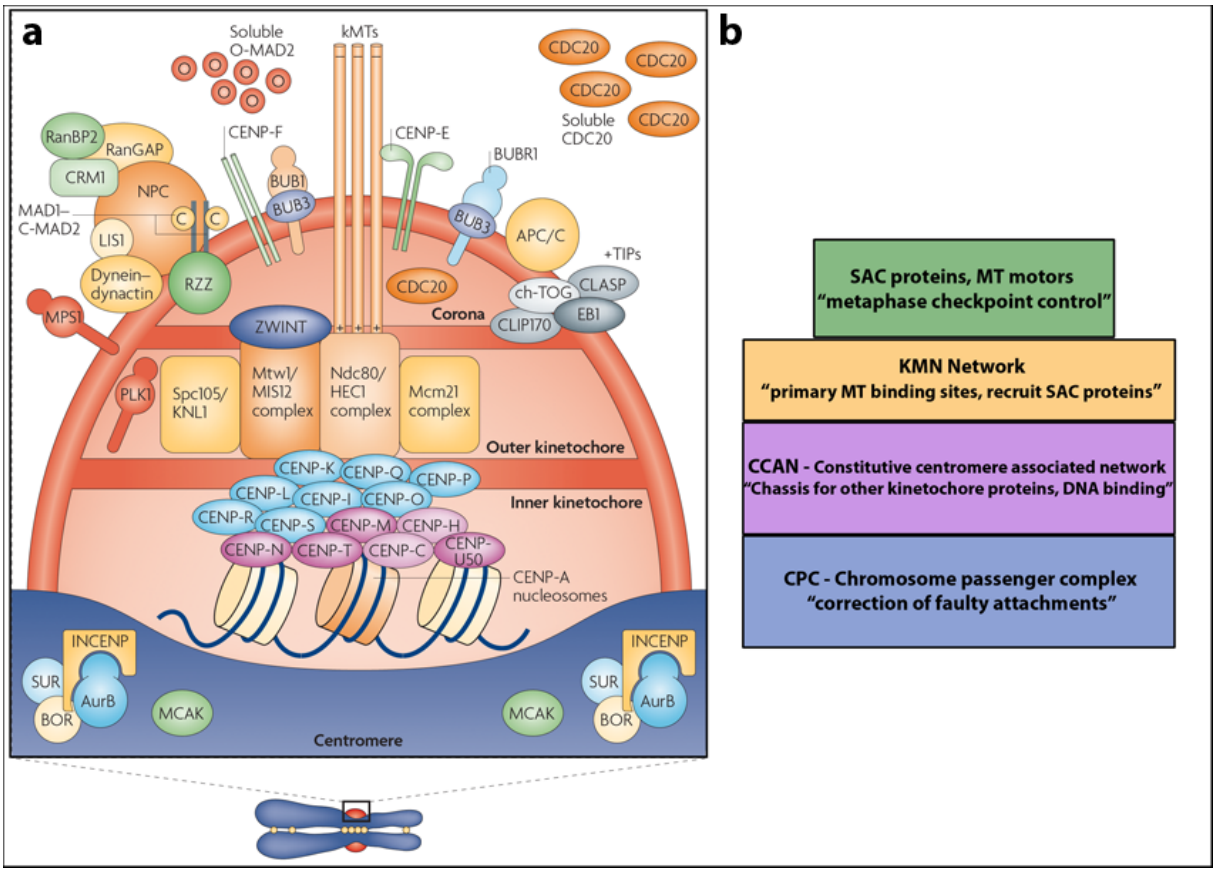


Figure 1.2: Organization of the kinetochore. (a) Eukaryotic kinetochores are hierarchical assemblies consisting of up to 100 proteins and can be segmented into regions called inner kinetochore, outer kinetochore and corona. The Histone H3 homologue, CENP-A is central to the assembly of all kinetochore proteins. It interacts and underlines the assembly of other members of inner kinetochore (highlighted in purple and light blue) , which are collectively known as the CCAN (constitutive centromere associated network). Most of the components of CCAN network, inner kinetochore, remain associated with the kinetochore throughout the cell cycle. The outer kinetochore forms a chassis for assembly of universally conserved complexes and is commonly known as the KMN network. Most of the proteins of the outer kinetochore are recruited during mitosis. Proteins of the outer kinetochore are responsible for primary kinetochore-microtubule interactions and also help recruiting the proteins that form the corona. The proteins of corona encompass diverse roles including spindle assembly checkpoint, (RZZ complex, Cdc20, BubR1, Bub3, Bub1) regulating microtubule binding, motor proteins (dynein-dynactin, CENP-F) etc. The kinetochore centromere interface is populated by the chromosomal passenger complex, with primary kinase functionality, involved in correcting faulty attachments Source: [17] **(b)** organization of the kinetochore at a very rudimentary functional level.

The outer kinetochore region is comprised of the highly conserved KMN network (discussion follows in section 1.1.3), which mediates the interactions with microtubules, and provides a platform for most of the proteins that populate the corona [18]. The corona is the outermost region of kinetochore that hosts proteins and protein complexes which are primarily involved in the checkpoint activity and feedback mechanisms, it also hosts proteins which bind and regulate microtubules [17]. A very important region is also the centromere-kinetochore interface. This space is pervaded by what is known as the chromosome passenger complex (CPC, Aurora-B protein kinase, the inner centromere protein INCENP, Survivin and Borealin), a complex that is critical for correction of faulty kinetochore-microtubule attachments [19]. Recent technological advances in single molecule studies have led to

developments of methods such as 'kinetochore-speckle high resolution co-localization' (K-SHREC) which allows to determine the average separation between the centroids of two different colored fluorescent probes that label different protein domains within the kinetochores of human cells. K-SHREC analysis applied to cells in metaphase where the sister chromatids are aligned in opposite directions led to the mapping of positions of various proteins along the kinetochore axis[20]. Such studies have unveiled spatial organization of kinetochores to a certain extent. One such analysis of cells in metaphase with and without taxol treatment is highlighted in **Figure 1.3** [21]. Cellular complexes seldom come without any critical functional roles. The function of kinetochores is, very rudimentarily, non-faulty binding of microtubules.

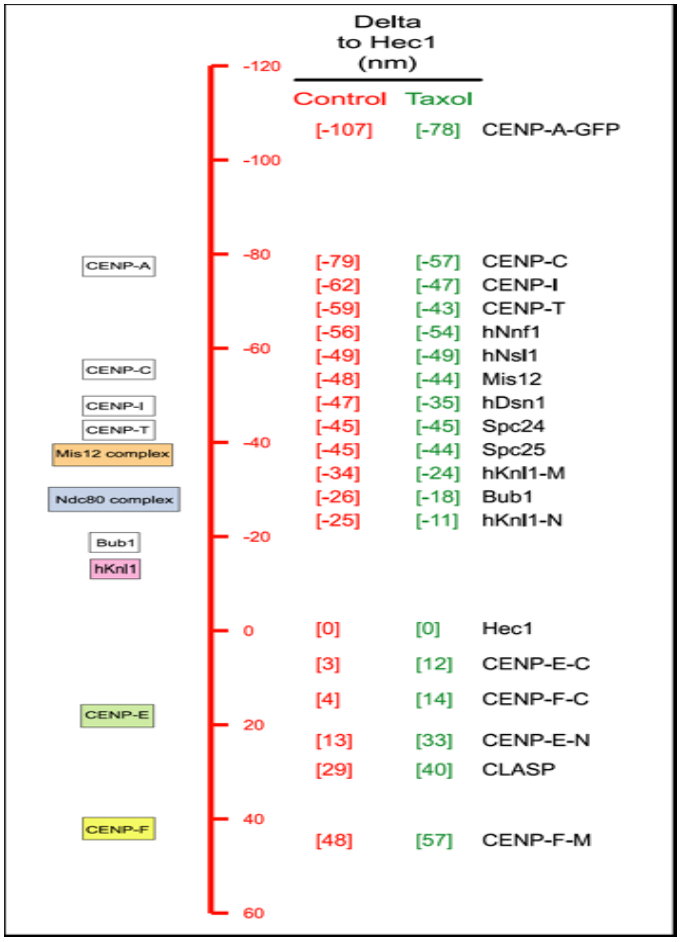


Figure 1.3: Delta measurements in control and taxol-treated HeLa cells.
'kinetochore-speckle high resolution co-localization' (K-SHREC) analysis is a fluorescence based method, where one measures average separation, Delta, between proteins labeled with two fluorophores at a single molecule level. The figure highlights delta measurements for 19 epitopes in 16 kinetochore proteins in control cells (left) and in taxol-treated (right) cells. The scale is calibrated to zero at the N-terminus of Hec1 (Ndc80) centroid, where positive values indicate distances outward (away from the kinetochore and towards the spindle). While negative values are inward (towards the centromeric chromatin). Color-coded boxes highlight specific complex. Such analysis predict the position of epitope within an accuracy limit of <5 nm. Source: [20]

Before continuing any further, two things need to be clarified: One, since ROD and Zwilch (the subjects of this study) are found only in metazoans and higher eukaryotes, most of the concepts introduced and discussions led will be keeping higher eukaryotes in mind. Two, a lot of the current chapter will deal with term 'recruitment'. Recruitment here is referred to in a much broader sense. The proteins that make up and/or associate with the kinetochore do not have any localization sequences/signals. When a protein A is referred to recruit another protein B, it generally refers to the fact that A is at a certain locus (kinetochore) before B and is absolutely necessary for the localization of B. Also, in most

of the cases, the quintessential dependence highlighted in recruitment has not been translated into biochemical interaction.

1.1.3 Kinetochore-microtubule interaction: KMN network

Microtubules are polymers of α/β-tubulin dimers, which shuttle stochastically between polymerization (growth phase) and catastrophe (de-polymerization). They are versatile in their interactions, engaging a diverse spectrum of functionally variant proteins, crucial for a multitude of cellular processes including cell division, migration and differentiation [22, 23]. The following section deals with a subset of such microtubule binding proteins essential for cell division.

Using a synergistic approach of biochemistry and RNAi screening (functional genomics) the KMN protein network was discovered in *C.elegans* [24, 25]. This extremely conserved ten-protein network is what links centromeric DNA to the plus end of bipolar spindle microtubules and regulates functions such as: formation and correction of load bearing kinetochore-microtubule attachments, recruiting and hence controlling the proteins involved in the spindle assembly checkpoint (**Figure 1.4**) [18, 26, 27]. Depletion of the KMN network results in a kinetochore-null phenotype (no load bearing kinetochore-microtubule attachments can be formed) [18, 24, 25, 28, 29]. The network comprises of three sub-complexes: the **KNL1 complex (KNL1-C)**, which is made up of Knl1 (Kinetochore null-1, CASC5 or blinkin in humans, Spc105 in yeast, Spc105R in fly) and Zwint (Zw10 interacting protein); the **MIS12 complex (MIS12-C)**, constituted of four proteins viz. Mis12 (Mis-segregation protein 12, Mtw1 or MIND in yeast), Nsl1 (DC31 or Mis14), Nnf1 (also known as Pnf1) and Dsn1 (also known as Mis13 in humans, Knl3 in *C.elegans*); and also other heterotetrameric **NDC80 complex (NDC80-C)** consisting of Ndc80 (also known as Hec1), Nuf2, Spc24 and Spc25 [18, 24, 29].

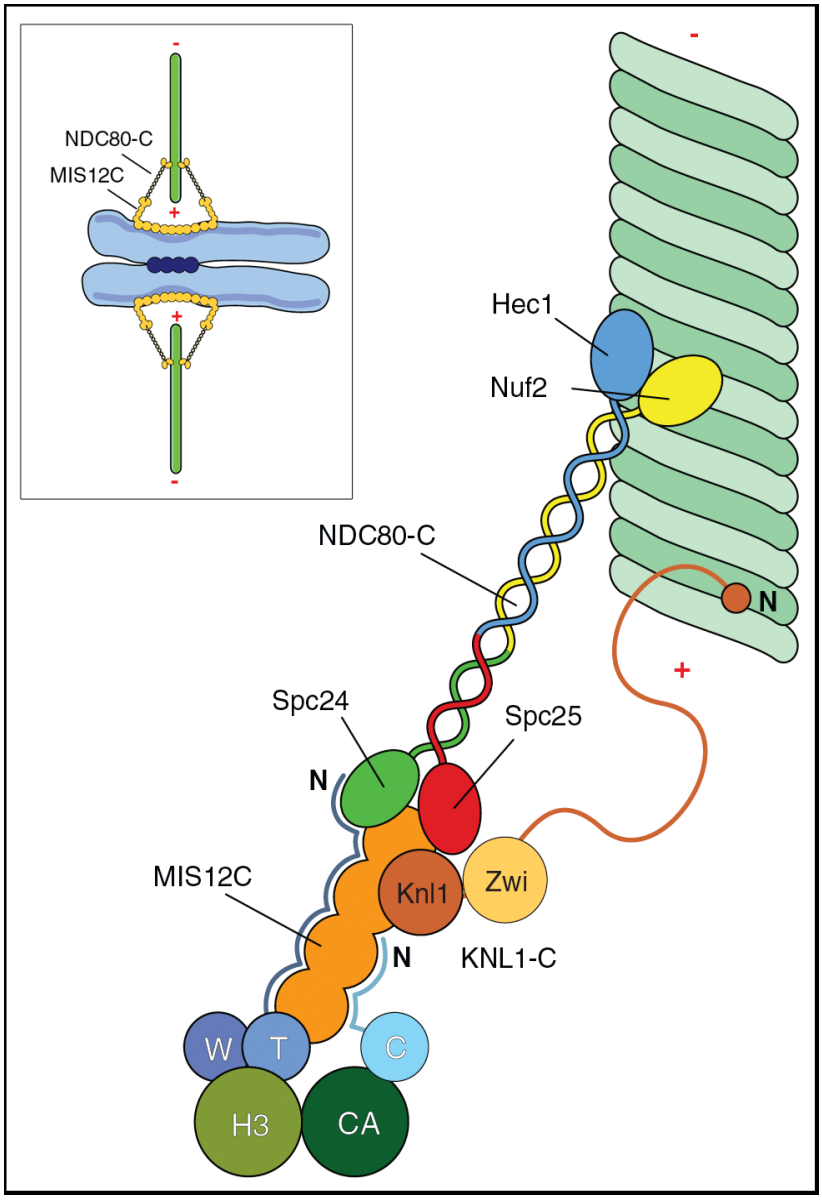


Figure 1.4: Schematic view of the KMN network. The KMN network is a central to microtubule binding and is made up of 10 proteins and is sub-divided into three different complexes: the MIS12 complex (MIS12C), the NDC80 complex (NDC80C) and the KNL1 complex (KNL1C). The KMN network bridges the nucleosomes and the microtubules. CENP-T/W or CENP-C, members of the CCAN network associated with the nucleosomes make contacts with MIS12C. The KNL1 complex comprises of Knl1 and Zwint-1 (Zwi) plays a major role in recruiting most of the other proteins to the kinetochore. NDC80 is a hetero-tetramer (Hec1, Nuf2, Spc24, Spc25). Ndc80 (Hec1) harbors primary microtubule binding sites. Nuf2 and the N-termini of Knl1 have also been shown to bind microtubules, an interaction required for re-enforcing the primary binding. Spc24, Spc25 interact with the MIS12C. 3D structure and architecture of the KMN network is now available [30].
Inset highlights the location of kinetochores on the chromosome and the orientation of microtubules with respect to the chromosome.
Source: [31]

1.1.4 The binding conundrum

Not all protein interactions between the kinetochore and the microtubules result in the formation of load bearing attachments. In a nutshell, binding of the kinetochore to the microtubules is *necessary but not sufficient* for the metaphase-anaphase transition. There are four possible outcomes to the kinetochore-microtubule interaction of which only one is favorable. An ideal scenario is of bi-orientation, where kinetochore pairs from the sister chromatids are bound to microtubule fibers from opposite poles and are in a state of 'tension' (they experience partitioning force, which stretches the chromatin and the kinetochores)[32-34](**Figure 1.5**).

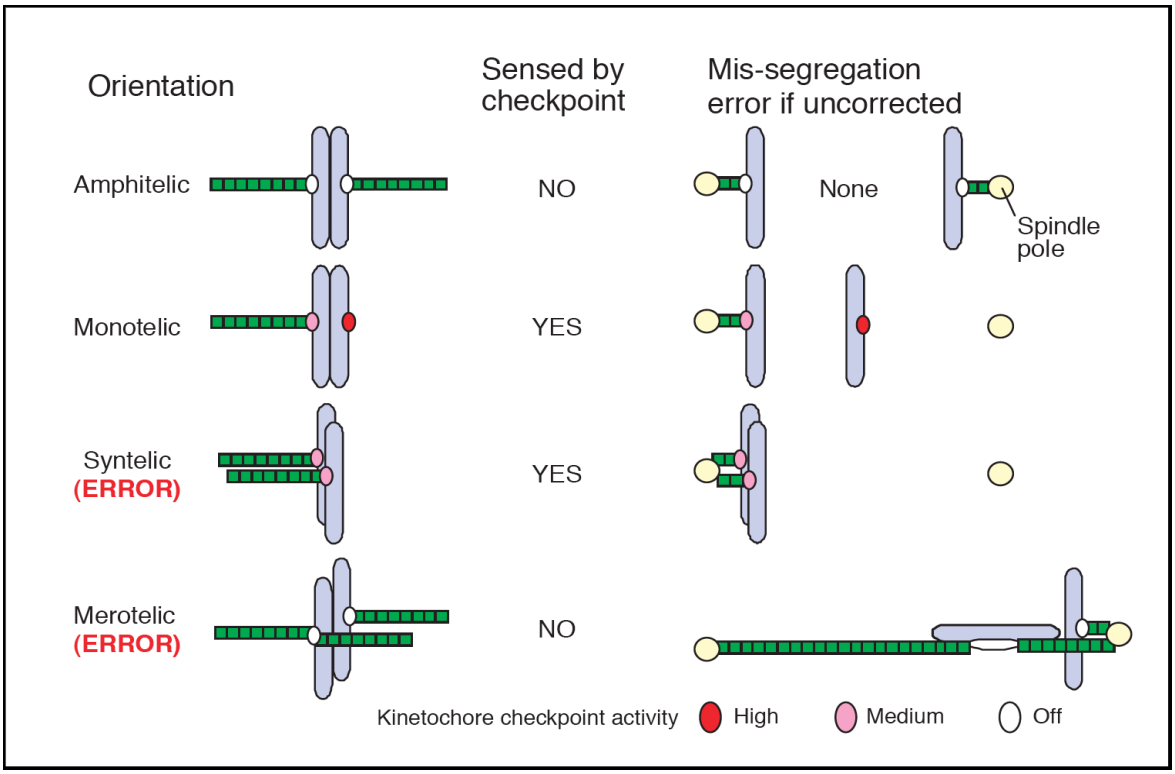


Figure 1.5: Attachment versus tension. Four different possibilities of kinetochore microtubule attachment exist, of which only one is favorable for the metaphase to anaphase transition. The ideal case is when kinetochores of sister chromatid are bound to microtubules from opposite poles. This is also known as amphitelic orientation (bi-orientation). Monotelic attachment precedes bi-orientation generally in prometaphase. The spindle assembly checkpoint keeps the cell from transitioning into anaphase, defects in which lead to abnormal distribution of chromosomes. In syntelic attachment, both sisters in a pair connect to the same pole and in merotely, one chromosome is attached to both poles. Conditions, which lead to loss of bi-orientation generally are unfavorable to the cell. Source: [35]

Dividing cells in the process of undergoing metaphase to anaphase transition have to ensure the following:

1) There cannot be any unattached kinetochore, and
2) The attachments should not be faulty.

The kinetochore hosts two different feedback mechanisms in the cell and oversees the fulfillment of above criteria. Cells, which enter anaphase failing to meet above requirements, often, encounter aneuploidy, mis-segregation or other defects[36].

1.1.4.1 Spindle assembly checkpoint (SAC)

The first feedback mechanism, called the spindle assembly checkpoint (SAC, alternatively metaphase checkpoint) ensures that there are no unattached kinetochores before the cell transcends into anaphase. Certain genes, conserved in all eukaryotes, relay the kinetochore-binding status of the chromosomes in the cell to delay exit from mitosis[37, 38]. Needless to mention, mutations in these genes compromise the fidelity of chromosome segregation and represent a non-utopian case of cell

division, with implications in tumors [39]. One can attempt to define a SAC gene as a gene, mutations in which are of catastrophic consequences to cells due to their tendency to exit mitosis overlooking segregation errors. Gene products of SAC genes include proteins such as Mad1 (mitotic-arrest deficient), Mad2, Bub1 (budding uninhibited by Benzimidazole), Bub3, BubR1, kinases Mps1 (multipolar spindle-1) and Aurora-B [40, 41]. There are other proteins, which are an integral part of the SAC, but are found in metazoans and higher eukaryotes. These include proteins, some of which are focus of the present study: ROD, Zw10 and Zwilch, and others such as p31^{comet}, Plk1 (polo-like kinase-1), MAP (mitogen activated protein) kinase, dynein and dynein associated proteins such as dynactin, Clip170 (cytoplasmic linker protein) and Lis1 (lissencephaly-1) [42-45].

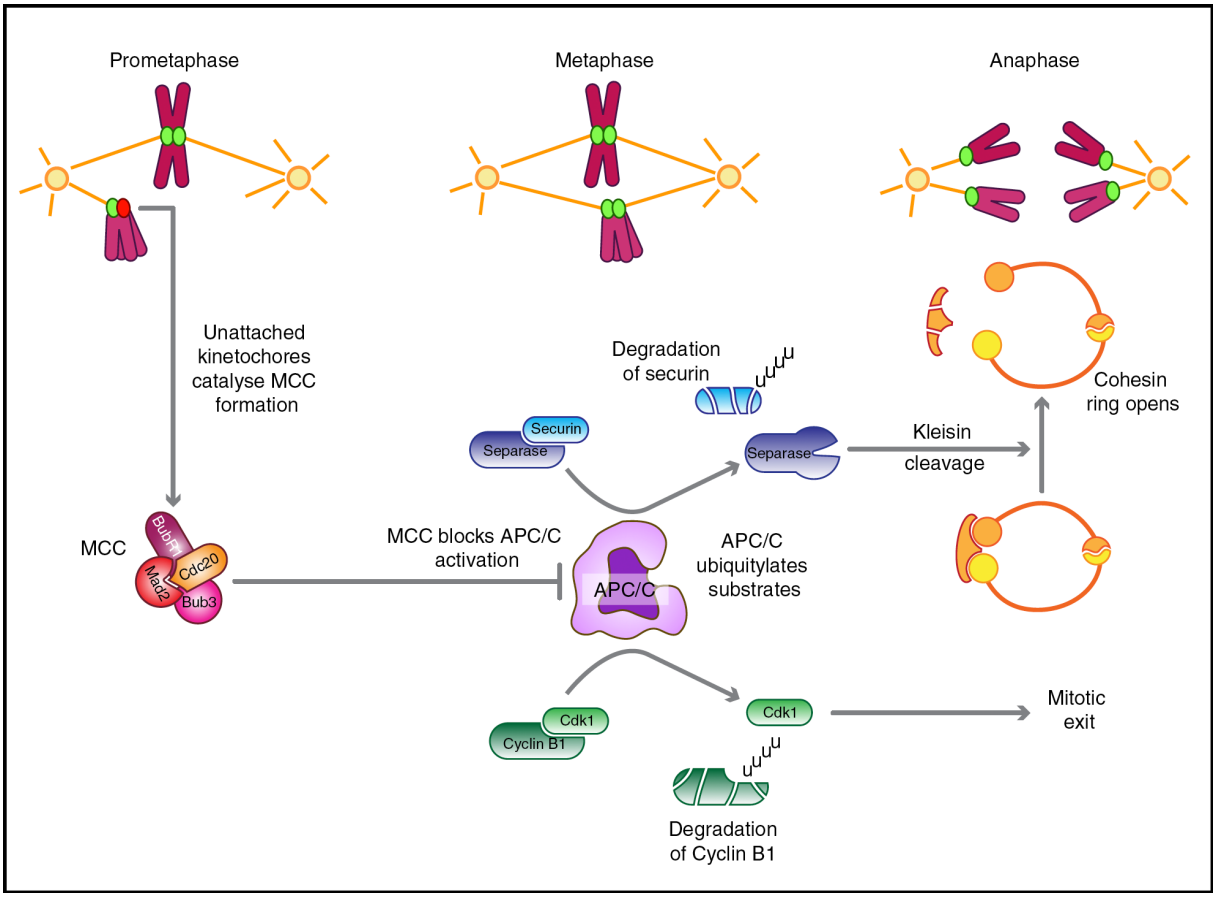


Figure 1.6: Principles of the SAC. Unattached kinetochores catalyze formation of the mitotic checkpoint complex (MCC) composed of BubR1, Mad2, Cdc20 and Bub3, which blocks the APC/C. The MCC is characteristic of early stages of mitosis when the chromosomes are still searching for stable and load bearing microtubule attachments. Once all the chromosomes are aligned with their kinetochores attached to the spindle (metaphase), MCC ceases to exist facilitating the release of Cdc20, which is a cofactor of APC/C. Activated APC/C ubiquitylates its substrates, Securin and Cyclin B1, which are degraded by the proteasome leading to release of Separase and Cdk1. Separase cleaves the Scc1 kleisin subunit of the cohesion ring structure leading to the opening of the ring, allowing sister chromatids to separate, which marks the onset of anaphase. Cdk1 signals the cell to exit mitosis. Source: [46].

The molecular mechanism of SAC can be attributed to the action of proteins Mad2, BubR1 (Mad3 in yeast) and Bub3 which form a complex called as the Mitotic-checkpoint complex (MCC) [47-49]. Note, these proteins are recruited to the kinetochore directly by Knl1. They target and negatively regulate

Cdc20, which is a co-factor of anaphase-promoting complex/cyclosome (APC/C) [50-52]. APC/C is an ubiquitin ligase that, in presence of Cdc20, brings about polyubiquitylation of proteins Cyclin B and Securin, making them susceptible to degradation by the proteasome [52]. This has two-fold effect: one, Securin, which is an inhibitor of Separase (a protease, that targets Cohesin, a protein which binds the two sister chromatids together) which when degraded leads to a pool of active Separase which then goes onto cleaving the sister chromatids and sets them free to be driven pole-wards [53]. Two, degradation of Cyclin B inactivates the master mitotic kinase Cdk1 and hence initiates cytokinesis [52]. In short, MCC targets APC/C mediated ubiquitylation which is indispensable for sister chromatid separation (**Figure 1.6)**. Conversely, when the kinetochores are unattached, Cdc20 remains bound to MCC and hence cells wait to get into anaphase.

How exactly is Cdc20 regulated? The model for regulation of Cdc20 is briefly described below and is generally referred to as the 'template model'. Protein Mad2, an integral part of MCC, populates cells as two different conformers [49, 54] One, open-Mad2 (O-Mad2), which is free or unbound and generally cytosolic and as closed-Mad2 (C-Mad2), which is Mad2 bound to either Mad1 or Cdc20. Mad2 can exist as a dimer, but the interaction is only possible between a molecule of O-Mad2 and a C-Mad2. Both O-Mad2 and C-Mad2 homo-dimers are known not to exist. There is a small amount of C-Mad2 at the kinetochores (in complex with Mad1). This then recruits O-Mad2 to unattached kinetochores during prometaphase. The new O-Mad2, which arrives at the kinetochore, can bind to the Mad1/C-Mad2 already at the kinetochore making a transient but stable complex. During the course of SAC these two populations are non-exchanging. The interaction of the O-Mad2 with Mad1/C-Mad2 can then induce a conformational change onto the O-Mad2 thereby making it susceptible to interact with Cdc20. Hence; Mad1/C-Mad2 acts as a template for conversion of O-Mad2 to Cdc20 bound C-Mad2[55]. This model further postulates that; since the structural conformation of Mad1/C-Mad2 is similar to that of Cdc20/C-Mad2, the latter can auto catalyze (as a positive feedback) conversion of more O-Mad2 to C-Mad2 [17, 56](refer **Figure 1.7**).

Extinction of SAC is propelled by attachment of kinetochore to microtubules. This triggers the C-Mad2/Cdc20 complex breakdown and release of Cdc20, which is now available to bind to the APC/C and thereby facilitates the onset of anaphase as described earlier[52].

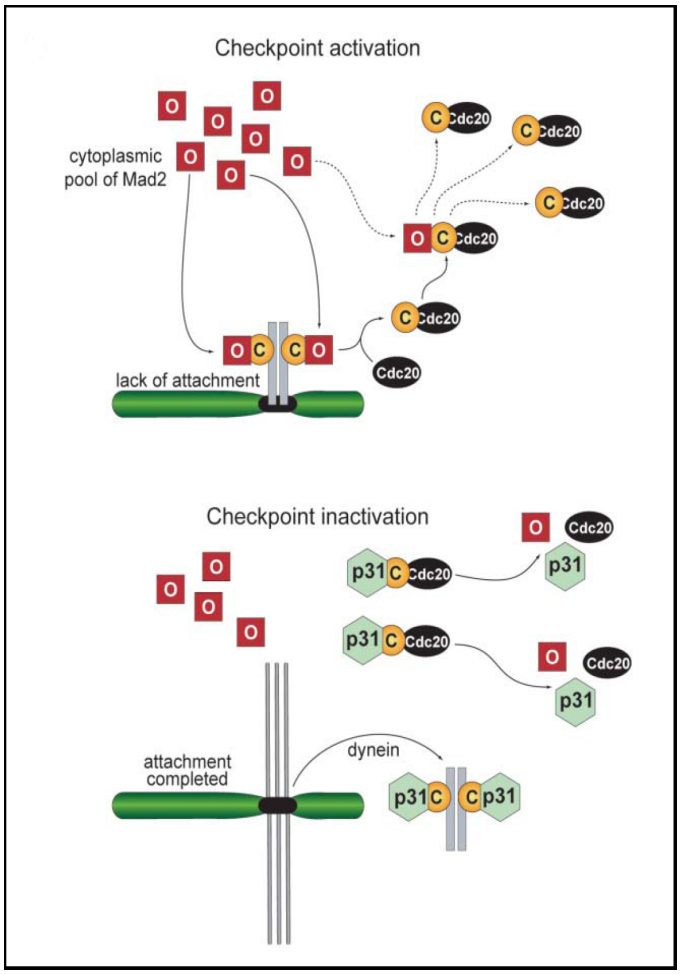


Figure 1.7: The template model. When the kinetochore is unattached (above) during metaphase, the checkpoint is activated. O-Mad2 (unbound, shown in red squares with 'O') interacts with C-Mad2 (highlighted in yellow circles, bound pool of Mad2 present at the kinetochore as Mad1-Mad2) and is converted to C-Mad2 bound to Cdc20. Both the C-Mad2 populations are similar as Mad1 and Cdc20 share the same binding motif and Mad2 adopts the same conformation in both these complexes. Once C-Mad2/Cdc20 is formed it is released into the cytosol where it catalyzes the conversion of cytosolic pool of O-Mad2 to C-Mad2/Cdc20. In this way all the Cdc20 required for the activation of APC/C for the cell to progress into anaphase is preoccupied. Once the kinetochore microtubule attachments are established (represented below) proteins such as $p31^{comet}$ may dissociate the Mad2-Cdc20 complex, leading to release and stabilization of Cdc20, which can then activate the CPC. Also Mad1-Mad2 is siphoned off the kinetochore by dynein-mediated mechanisms. There are also other theories in what leads to the dissociation of the C-Mad2/Cdc20, which will not be discussed here. Source:[54]

The release of Cdc20 from C-Mad2/Cdc20 can be brought about by one of the following ways. First, disappearance of the unattached kinetochores may lead to re-activation of $p31^{comet}$, which interacts with the Mad1/C-Mad2 and hence inhibiting the Mad1/C-Mad2 O-Mad2 interaction [44, 57, 58]. Second, at the onset of anaphase the dynein/dynactin complex siphons off Mad1/C-Mad2 and other kinetochore proteins and hence leaving little or no room for the formation of C-Mad2/Cdc20 [42] and third, Cdc20 may undergo a non-degradative ubiquitylation by the enzyme UbcH10 and this might then favor the release of Cdc20 from the Cdc20/C-Mad2 complex [59]. Lastly, since the recruitment of SAC proteins depends on the phosphor-regulation of Knl1 [60-62], role of phosphatases in extinction of SAC cannot be ruled out[63].

The reader might wonder at this point how RZZ, reported as a SAC gene earlier, contributes to SAC? The explanation is rather straightforward; Mad2, which is absolutely necessary for the SAC activity, is transported to the kinetochores by RZZ. A somewhat detailed explanation follows later in section **1.1.5.2.**

1.1.4.2 Error correction

On a conclusive note, I would like to address the issue of faulty attachments, i.e., what happens if the microtubule attachment is syntelic/merotelic? Or alternatively, how does the cell machinery distinguish between tension (amphitelic/bi-orientation) and attachment (syntelic/merotelic)? Reports of increase in faulty attachments when Aurora B kinase activity was compromised led to its direct implication in being involved with correction of attachment errors [64-66]. Supplementary studies such as the decrease in the number of improper attachments once the kinase activity is restored in Aurora B inhibited cells and similar characteristics of its yeast homologue, Ipl1, cemented its role in error correction [66, 67]. Ipl1 (the only Aurora kinase in yeast) has also been shown to activate the spindle checkpoint by creating unattached kinetochores, thereby actually recognizing wrong attachments and turning them in, to be properly attached [68].

The exact mechanism of Aurora B kinase action remains still obscure. One of the hypotheses, floating for quite some time, is as follows: Aurora B interacts with Borealin, Survivin and INCENP to form what is known as the chromosomal passenger complex (CPC) [69, 70]. CPC pervades the centromere kinetochore interface primarily regulating activities of substrates by phosphorylation (**Figure 1.2**). When the kinetochores are in tension (bi-oriented), they experience a force that tends to stretch apart the sister chromatids and the kinetochore itself and the distance between CPC and its substrates greatly increases [34, 71]. On the other hand when not in tension (syntelic, merotelic attachments) this complex, via Aurora B, modulates the activity of the microtubule binding complexes (NDC80, KMN network) reducing the affinity for interaction [72, 73] (**Figure 1.8**). Few aspects of this model, such as the stretching of centromeric chromatin and the unquestionable importance of Aurora B, have been envisioned. However, this model remains very minimalistic and its validity questionable. One, for example, is if the Aurora B regulation is distance dependent then unattached kinetochores should prolong to exist in the same state as the lack of tension facilitates Aurora B to be in the vicinity of its substrates. Recent studies have also shown that Aurora B might function independent of CPC in recognizing tension across the kinetochore axis.

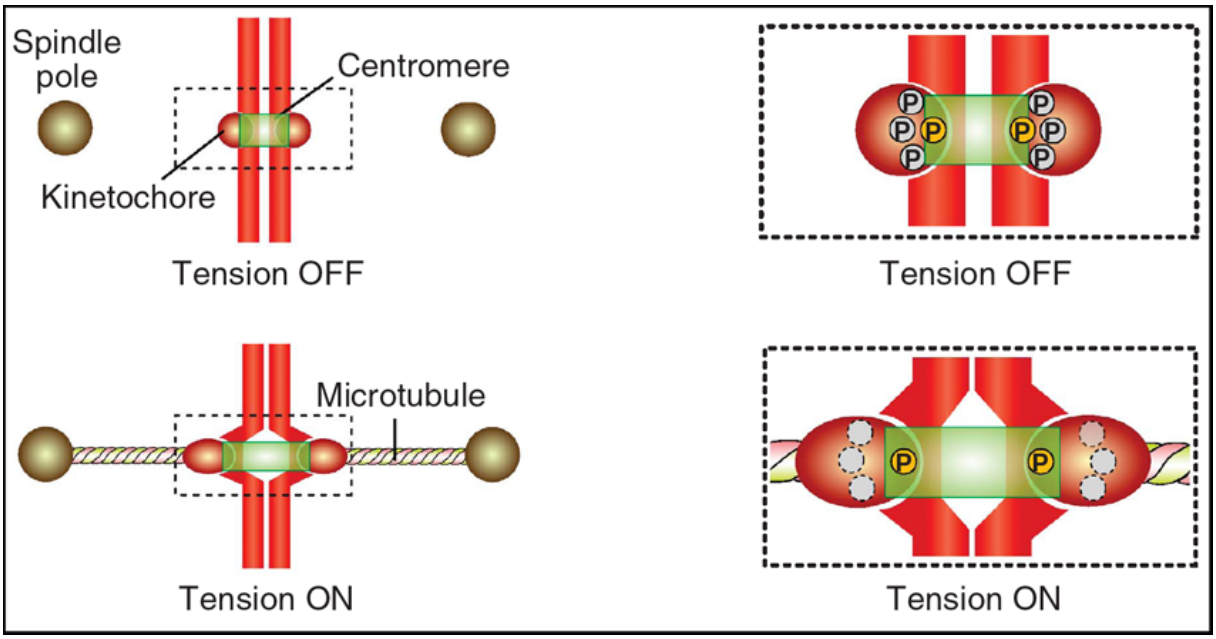


Figure 1.8: Error correction-the hypothesis. Schematic representation of kinetochore-centromere interface in the presence and absence of tension. Tension is by definition a state when the kinetochores of the sister chromatids bind to microtubules from opposite spindle poles. The figure on right is close up representation of boxed area in the left. Some of the substrates at the kinetochore-centromere are phosphorylated at all times (marked by yellow circle with 'P'). Other substrates are phosphorylated only in the absence of tension. When the kinetochores are in tension, the substrates are beyond the distance constraints preventing any phosphorylation. This model postulates that correct microtubule-kinetochore depends on kinase activity of proteins (CPC) present at the kinetochore-centromere interface. Source: [74].

1.1.5 ROD – Zw10 – Zwilch (RZZ) complex

*'…When you are dealt a bad hand of cards (in poker, for example), that's a rough deal. The chromosomes in **rough deal (rod)** mutants were "unfairly" distributed to the daughter cells. The gene also earned its name by making itself extremely difficult to clone (this was back in the days before sequenced genomes, when you had to identify the DNA for your gene the old fashioned way).*
*Do you know why **zw10** is called **zw10**? A rather boring reason: Back in the 1970s someone wanted to estimate how many genes there are in the whole fly genome. So he tried to "saturate" a small interval of the X chromosome (from Zeste to White) with mutations. He found 13 complementation groups. The 10th one was named ZW10. PS, from this he estimated there were about 5000 genes in the genome. Unfortunately, he could only detect genes with phenotypes, of course, so he missed many genes…'*

– Roger Karess, Personal Communication.

Components of the RZZ complex were not all identified at once. The z*w10* gene was first to be identified and its gene product, Zw10, is functionally most diverse of the lot. *rod* was discovered later and the fact that proteins encoded by both these genes, form a stable complex was identified almost 20 years after their discovery. This was the time when Zwilch was discovered and identified to be a component of what has erstwhile been called as the RZZ complex. A timeline of events, from the

discovery of Zw10 as an essential checkpoint gene to the current knowledge of the complex, is summarized in the table that follows.

Table 1.1: Timeline in the discovery and current understanding of the RZZ complex

1985	Identification of *l(1)Zw10* as a gene which is essential for proper mitosis in *Drosophila melanogaster* [75].
1989	*Rough Deal (rod*) identified as a gene required for proper mitotic segregation in *Drosophila* [76].
1992	First reports of redistribution of the *zw10* gene product in *Drosophila* during mitosis [77].
1998	Zw10 implicated in the recruitment of dynein and dynactin to the kinetochores[78].
2000	ROD and Zw10 are checkpoint proteins binding to the kinetochore [79]
2001	Zw10 and ROD act together in a complex, which is evolutionarily conserved [80].
2003	Zwilch, identified as a component of trimeric ROD-Zw10-Zwilch complex. First efforts to characterize the RZZ complex [81].
2004	hZwint-1 recruits hZw10 and hence RZZ to kinetochore [82].
2005	Mad2 is recruited to the kinetochores by the RZZ complex [83].
2010	Structure of Zwilch (67 kDa) solved by X-ray crystallography. Also isolation of a sub-complex of RZZ (ROD^{1-350} and Zwilch) [84].
2011	Expression and purification of the RZZ complex from insect cells, using the MultiBac expression system, to near homogeneity (unpublished data)
2013	First 3D structure of the complex determined by single particle electron microscopy at 13.4 Å (unpublished data, current work)
2013	The RZZ complex is implicated in binding and inhibiting the Ndc80 complex. A novel function of RZZ [85]

1.1.5.1 ROD (rough deal)

rod (*rough deal*) gene was identified in 1989 in an independent screen for selection of mutants in *Drosophila melanogaster* that live as larvae and die as pupae in an attempt to discover mutations in genes which encode mitotic functions [76]. The morphology of phenotypes, which survive the pupae stage included defects in eyes, wings and severe defects in the cuticle. The cells which survive mutations in *rod* gene have abnormal anaphases and display chromosome instability in most of the dividing tissues [76].

ROD, like all classical metaphase checkpoint components, is recruited and localizes to the corona of unattached kinetochores during metaphase in all cell types [79]. Once all the checkpoints are quenched, ROD is redistributed along the microtubules; a process called shedding that is mediated by dynein [86]. The departure of ROD from the kinetochores, though in anaphase, varies across species. Studies in *Drosophila* embryos elegantly show that ROD is continuously recruited from the cytoplasmic pool with the half-life of ROD at the kinetochores being 25-45 s [87].

The ROD protein is the largest component of the RZZ complex (240kDa) and its structural characterization is minimal, owing to its obstinacy towards purification. The domain architecture of ROD is highlighted in **Figure 1.9.**

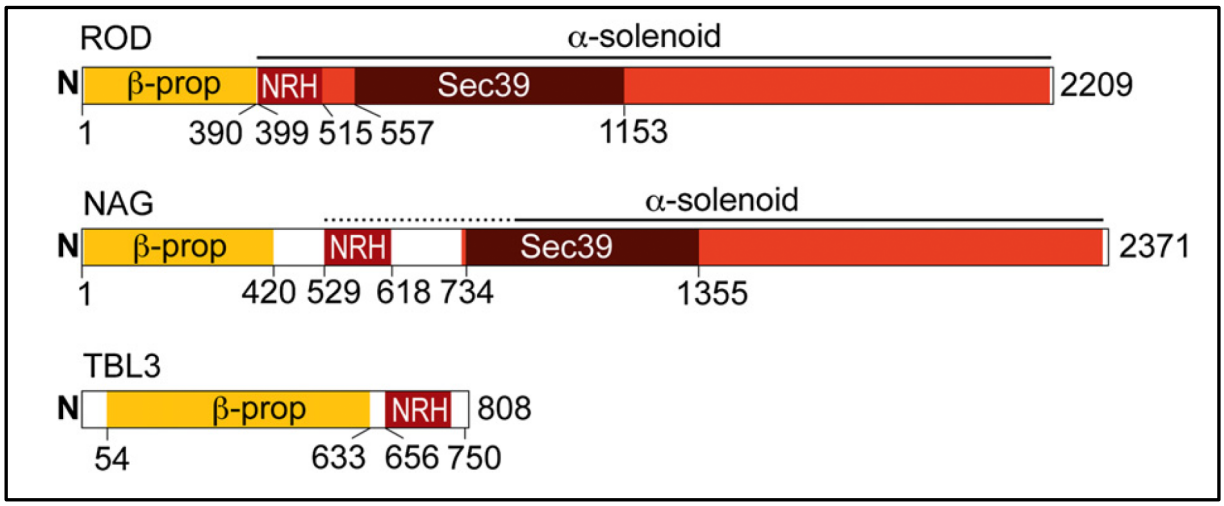


Figure 1.9: Domain organization of ROD. ROD is the largest component of the RZZ complex with a size of 240 kDa. ROD like NAG and TBL3 contains an N-terminal β-propeller domain followed by an NRH region. Both ROD and NAG are predicted to have α-helical stack repeats (α-solenoid), which extend to the C-terminus of the protein. They also contain a Sec39 domain, which is characteristic of COPII vesicles, implicated in transport of substrates. A distant homology search reveals TBL3 as the third family member (probability of homology relationship with ROD is 95%), which however lacks the Sec39 domain.

This is based purely on structure prediction algorithms [84]. It is predicted to have an N-terminal β-propeller domain, with putative WD40 β-propellers that could possibly mediate a variety of interactions. The rest of the protein is predicted to have a secondary structure that is primarily α-helical with a region of sequence homology to Sec39p. ROD is predicted to be structurally similar to NAG (neuroblastoma-amplified gene), a protein that interacts with Zw10 in an independent manner to form a complex together with RINT-1 called the NRZ complex, which plays role in Golgi retrograde

trafficking (homologous to the yeast Dsl1 complex). It is not surprising that the two Zw10 binding proteins ROD and NAG are predicted to be structurally similar. These proteins have a conserved region (novel) called the NAG-ROD homology domain (NRH). This domain has been identified also in other proteins like TBL3 (transducing beta like-3) [84].

1.1.5.2 Zw10

Zw10 is functionally the most versatile protein of all the components that make up the RZZ complex. Most of the interactions mediated by the RZZ complex were discovered and originally identified with Zw10 [78, 88-90]. The gene *zw10* was consequence of a screen, designed to detect errors in the transmission of mitotic chromosomes in *Drosophila* [75]. Integrity of Zw10 (gene product of *zw10*) was soon implicated for accurate chromosome segregation [77] and was found conserved across all eukaryotes (the only protein of RZZ complex found in lower eukaryotes) [91]. The localization of Zw10 parallels the pattern of ROD (as described in section above also in the section dealing with RZZ dynamics) [92]. Zw10 is recruited to the kinetochore by Zwint-1 [82], which is an integral part of Knl1 complex and therefore of the flagship KMN network. In essence, Zw10 links the checkpoint signaling (as a member of RZZ necessary for the SAC) to the structural chassis (the KMN network) of the kinetochore [89].

In mammals and in yeast, Zw10 (and its homologues) forms a second complex together with NAG (Sec39p in yeast) and RINT1 (Tip20p in yeast) called the NRZ complex, which is enriched in Golgi and ER compartments of the cell and interacts with the ER membrane SNARE proteins [93, 94]. In *Drosophila,* loss of Zw10 leads to Golgi fragmentation and loss of Golgi stack integrity [95] hinting at its possible role in ER-Golgi trafficking. Whether *Drosophila* Zw10 mediates anterograde or retrograde trafficking between ER and Golgi remains yet to be answered. Consistent with its role in membrane trafficking, depletion of Zw10 leads to premature halting of furrow ingression and thereby failure of cytokinesis in *Drosophila* spermatocytes[95]. One may be convinced that the failure of cytokinesis in *Drosophila* spermatocytes is just a reflection of lack of dynein/dynactin in Zw10 depleted cells. However studies suggest this is not the case [95].

Structural information from the yeast homologue of Zw10, Dsl1p is available [96, 97]. Crystal structure of N-terminus of Dsl1p (37-355) was determined in complex with Tip20p (RINI1 homologue in yeast) [98] and the C-terminus (333-684) in complex with Sec39p (NAG homologue in yeast. ROD also contains the sec39 repeat, refer to **Figure 1.9**) [96].

1.1.5.3 Zwilch

The interdependencies of ROD and Zw10 proteins on each other for kinetochore localization, the similarities in their mutant phenotypes and their function in dynein transport led to the speculation that they form a complex. This was confirmed shortly after, where these proteins were shown to form a stable complex [80]. By this time at the beginning of last decade, it was also established that the proteins ROD and Zw10 do not effect the kinetochore localization of other spindle assembly

checkpoint (SAC) proteins such as Mad1, hBUBR1 and hBUB1, although dynein, which is known to transport these proteins, gets affected [81].

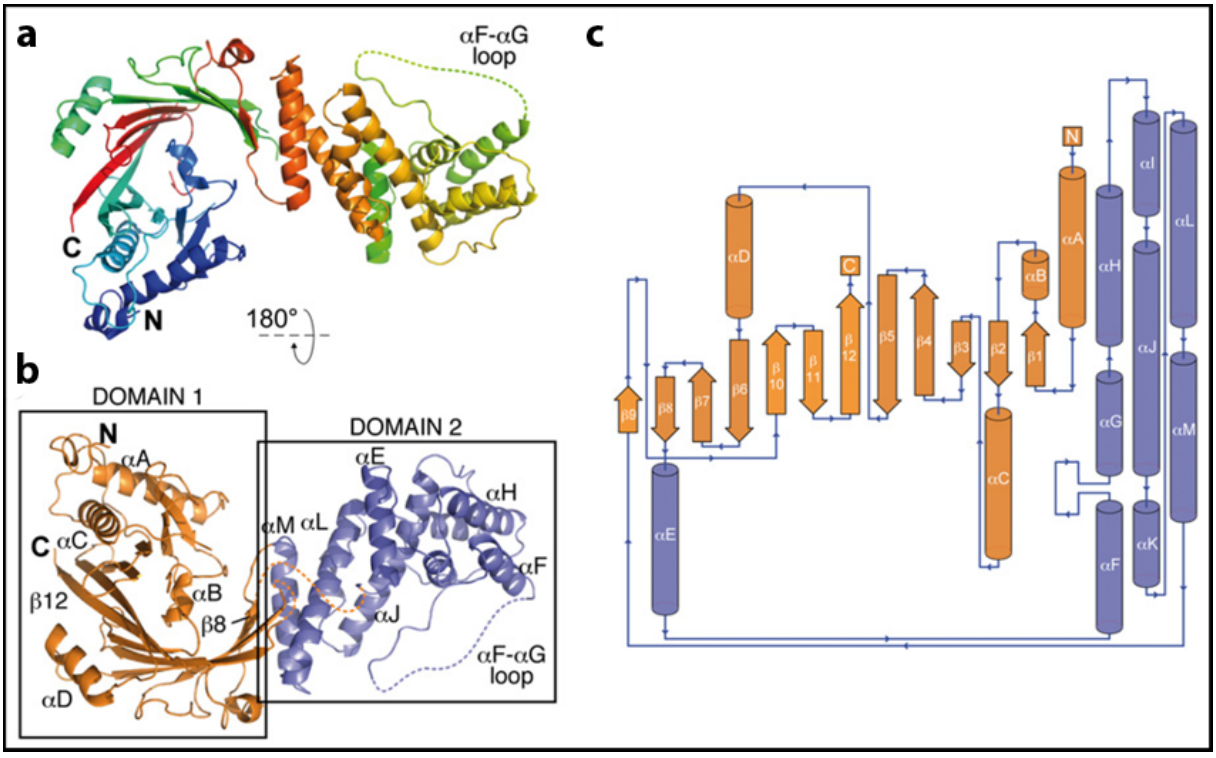


Figure 1.10: Structure of Zwilch. **(a)** Cartoon model of Zwilch. Zwilch is the smallest sub-unit of the RZZ complex, with the size of 67 kDa. It is known to interact with N-terminus of ROD. The structure of Zwilch is not a representation of the full-length protein, but a protein from a construct that contained two interacting segments of it (residues 1-334 and 335-591). However in the representation here the structure is highlighted as if it was continuous with the N-terminus being represented in blue to the C-terminus represented in red. **(b)** Cartoon model after 180° rotation relative to **(a).** The protein can be thought of having two different domains. The one in orange contains mostly β-sheets (domain 1) and the one represented in blue (domain 2) is α-helical. **(c)** Topology map of Zwilch. Source: [84]

The complexity and the numerous functional characteristics of the ROD/Zw10 protein complex led to the wild speculation that these may be part of a larger complex, with yet unidentified proteins. This hypothesis was further fuelled with the observation that Drosophila embryo extracts, when fractionated by size exclusion chromatography and probed with antibodies against DmZw10 (*Drosophila* Zw10) and DmROD (*Drosophila* ROD), led to the observation that both these proteins were found in fractions corresponding to molecular weight of 700-900 kDa [80]. Upon the use of anti-Zw10 antibody immobilized column to purify the complex from *Drosophila* embryo extracts an extra protein band was observed. This newly identified protein was attributed as a product of CG18729, a gene predicted by Berkeley Drosophila project with a molecular weight of 75 kDa (the protein was named Zwilch because of a very trivial reason: the authors undertaking the study when performing initial database queries to identify related proteins, ended up with no results as in 'zilch') [81]. The human Zwilch protein (hZwilch or Zwilch) has a molecular weight of 67 kDa. Subsequently, Zwilch was shown to localize at the outer kinetochore at prometaphase, kinetochore-microtubules during metaphase and again on kinetochores in the anaphase. This trend which is also observed with ROD

and Zw10 proteins and their co-localization at all times during metaphase-anaphase transition cemented the fact that ROD, ZW10 and Zwilch are part of the same complex [81].
Zwilch remains the only protein of known structure that belongs to the RZZ complex. There has been absolutely no study whatsoever in relation to the structure of domains/proteins or the full length RZZ complex. The structure of Zwilch was solved by X-ray crystallography (at 2.55 Å, $P3_221$ space group) [84]. The protein can be thought of consisting of two domains as observed from the X-ray structure and are depicted in **Figure 1.10**. Domain 1 has no sites of strong conservation, where as there are conserved regions in Domain 2, which might act as receptors or binding sites for other proteins.

1.1.5.4 Recruitment of RZZ to the kinetochores

Before moving on to understand the relevance of RZZ in cell division, an imminent question remains to be answered. What protein/protein complexes recruit the RZZ complex to, its locus, the corona of the kinetochore?
The recruitment of various proteins and especially the checkpoint proteins are highly interdependent. Following this line of thought one would expect RZZ localization to the kinetochore follow the same trend. However, it was shown that recruitment of RZZ is independent of Bub3, BubR1 [99], Mps1 [100], Cenp-F [82] or Mad2 [83]. Yeast two-hybrid assays revealed a protein, later named, Zwint-1 (Zw10 interacting protein) while screening for binding partners of Zw10 [88]. Zwint-1 (or simply just Zwint), a 43 kDa protein and a component of KMN network, localizes to the kinetochore prior to Zw10 in prophase and remains at the kinetochore till the onset of anaphase [91]. Biochemical experiments revealed that Zwint-1 harbors a Zw10 binding region in the first 80 amino acids of its N-terminus [82]. Localization experiments also show Zwint-1 and Zw10 co-localize during metaphase making it a possible candidate for recruiting RZZ. Cells depleted of Zwint-1 (siRNA treatment) show a complete loss of Zw10 (and hence RZZ) from the kinetochores hinting at the necessity of Zwint-1 for engaging RZZ at the kinetochores [101]. The converse is not true. Cells depleted of Zwint-1 also exit mitosis in presence of microtubule inhibitors and display characteristics such as entangled chromosomes, un-separated sister chromosomes and aneuploidy phenotypes similar to cells which lack RZZ, indicating that RZZ cannot function if it is not recruited [82]. Zwint-1 is also part of the KNL-1 complex and hence the KMN network, and is recruited to the kinetochore by Knl1 (**Figure 1.4**) [18].
These results identify Zwint-1 as necessary and sufficient for kinetochore targeting and localization of Zw10 and thereby, the RZZ complex. Why and how Zwint-1 gets the signal to recruit RZZ remains yet to be answered.

1.1.5.5 Localization dynamics of the RZZ complex:

ROD, ZW10 and Zwilch form a stable complex at the corona (**Figure 1.2)** of the kinetochore called the RZZ complex. Depletion of any of the components leads to loss of functional complex [81]. The characteristics of their association prior to their localization are unknown. All three proteins are cytoplasmic during interphase [77, 87]. At the onset of prophase, when the nuclear envelope breakdown is triggered, they start accumulating at the kinetochore [90]. Like all other checkpoint

proteins, the amount of RZZ at the kinetochore starts decimating as the cell prepares its transition from metaphase to anaphase. In other words, once the kinetochore microtubules have attached, migration of RZZ along the microtubules towards the poles is observed, predominantly bound to the kinetochores of the migrating sister chromatids [77, 80]. Also, once the kinetochores are engaged in favorable interactions with microtubules, RZZ is ushered away from the kinetochores along the microtubules in a dynein dependent manner, a process referred to as 'shedding' [42, 86]. Once in anaphase, the spatial and temporal localization dynamics of ROD and Zw10 is not coherent and varies for different species; for example, in HeLa cells ROD departs from the kinetochores in early anaphase where as Zw10 in late anaphase [79]. Recent localization studies on RZZ also reveal that there is a continuous flux of material being recruited from the cytoplasmic stock and other being shed along kinetochore-microtubules, mediated by dynein [87]. In essence, the RZZ complex is a short-lived complex localizing transiently to the kinetochore, the fate of its components and their association before prometaphase and beyond anaphase remains poorly understood.

1.1.6 Functions of the RZZ complex

1.1.6.1 Recruitment of dynein

The necessity for motor proteins at kinetochores can be justified with a simple argument that once the microtubules are attached, satiating all the requirements, the sister chromatids need to be driven polewards [102]. Three such well-characterized motor proteins localizing to mammalian chromosomes are dynein [103-105], Cenp-E and Kinesin-1 [17]. Cytoplasmic dynein (minus end directed microtubule motor) is implicated in a multitude of cellular processes and its importance is unquestionable, mainly due to its role in transporting cargo (proteins, vesicles, mRNA) along microtubules. Loss of the cytoplasmic dynein from cells leads to severe defects, ones implicated in cell division are: severe retardation in anaphase migration of the sister chromatids [106], accumulation of proteins which require dynein for removal from kinetochore and hence a delayed metaphase [42, 86] and defects in spindle positioning [107].

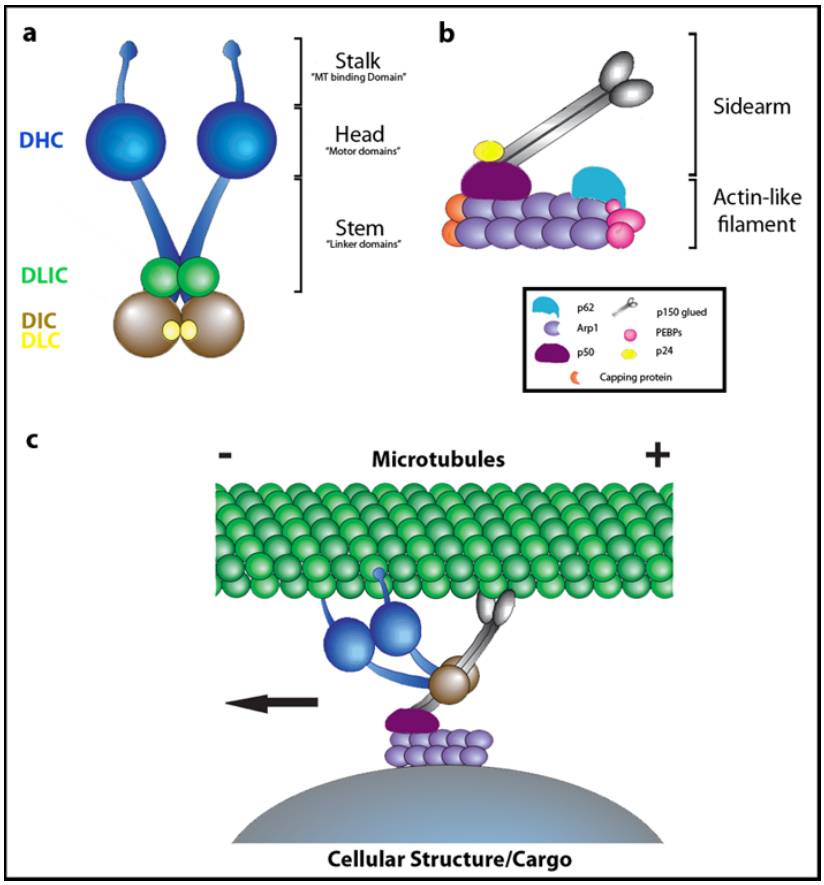


Figure 1.11: Organization of dynein, dynactin and their interaction with an organelle. (a) Cytoplasmic dynein is found in all animal cells and is indispensible for many cellular functions including organelle transport. The molecular mass of dynein is 1.5MDa and contains approximately 12 polypeptide units. It is made up of two identical heavy chains (DHC) responsible for ATPase activity, two intermediate chains (DIC) which help in anchoring dynein to its cargo, and several other light (DLC) and light intermediate chains (DLIC). Architecturally, it can be divided into stalk, head and stem regions. **(b)** Dynactin organization. p150 is generally referred to as dynamitin. **(c)** Model for interaction of dynein/dynactin with a cellular structure. For simplicity, only few sub-units are shown. Source:[108]

The connection between Zw10 and dynein was established while understanding the effects of Zw10 mutations, where loss of Zw10 resulted in plummeting the levels of dynein at the kinetochore [78]. Observing fluorescently tagged dynein heavy chain (DHC, refer to **Figure 1.11**) in absence of Zw10 mutants revealed DHC co-localization with Zw10 in prometaphase and its loss from kinetochore in cells with mutations in Zw10 [78]. Yeast two-hybrid assays led to identification of dynamitin (p50 component of dynactin, refer to **Figure 1.11** for architecture) as Zw10 interacting protein. Studies also showed that DHC and dynamitin co-localize with Zw10 at the kinetochore [78] leading to a prominent discovery that RZZ is very important for recruiting dynein to the kinetochores. Mutations in *rod* or *zwilch* also abolish the recruitment of dynein supporting the above hypothesis [81]. To sum it up, the Zw10/RZZ complex recruits dynein to the kinetochore, which is essential for removal of variety of proteins once the SAC is off including the components of RZZ [42, 86].

1.1.6.2 Transport of Mad2: to and off the kinetochore

Co-localization experiments targeting different SAC proteins led to the revelation that ROD and Mad2 show near perfect overlap of fluorescent signal in prometaphase not only at the kinetochores but also along the kinetochore-microtubule axis at the onset of anaphase [83]. This behavior is different from BubR1 and other SAC components in many ways viz. BubR1 remains associated throughout at the kinetochores and localizes slightly more internally at the kinetochore than ROD [83]. Mad2 is depleted

considerably at the kinetochore in *rod* or *zw10* depleted cells [83, 89]. Mad2 is removed from attached kinetochores transitioning into anaphase by the dynein-dynactin complex, which as mentioned above (**1.1.6.1**) is dependent on RZZ for its kinetochore localization [42]. Taken together, these facts hint at strong dependence and association of Mad1-Mad2 on RZZ during its lifespan at kinetochores. The nature of this association remains mysterious and needs to be understood. Two-hybrid screens, immunoaffinity columns and co-immunoprecipitation experiments have not shed any light on this interaction [78, 81]. Mad1/Mad2 association with the RZZ complex is an important biochemical process, failing which the cell is left with dire consequences to face. Decoding this association remains essential, and this may involve proteins/complexes hitherto unknown.

1.1.6.3 Role in formation of stable end-on coupled attachments

Following the discovery, which implicated RZZ in recruiting cytoplasmic dynein/dynactin to the kinetochore, a new protein named spindly was identified in *Drosophila* [109]. Spindly acts downstream of the RZZ complex and is required for targeting dynein, but not dynactin to the kinetochores. Using a carefully manipulated early *C. elegans* embryo system it was identified that although both RZZ and spindly are required for accurate chromosome segregation, although deletion of spindly has a more drastic effect [110].

Chromosome movement depends primarily on the binding of microtubules to the NDC80 complex forming a load bearing attachment. Loss of NDC80 complex does not exhibit a 'kinetochore-null' phenotype as residual chromosome movements are observed [25] with dynein/dynactin being their source [29, 111]. Loss of RZZ, leads to loss of checkpoint but the formation of load bearing attachment occurs but at a slower rate. Loss of both RZZ and spindly, however, leads to the formation of the load bearing attachment at an enhanced rate. The most important discovery came along very recently where it was identified that RZZ could interact and inhibit the NDC80 complex [85].

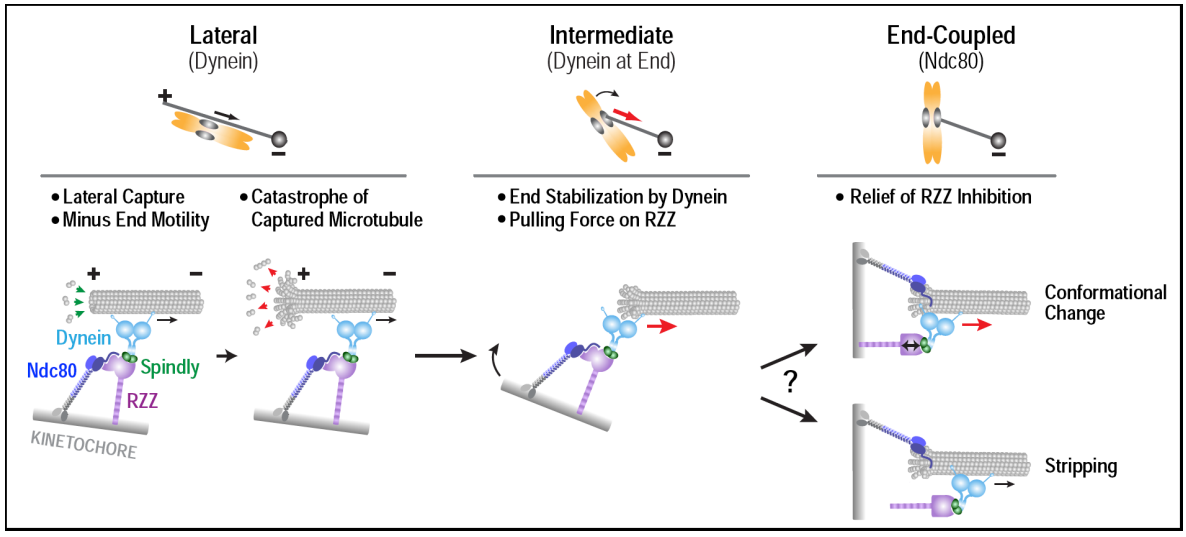


Figure 1.12: RZZ mediated inhibition of NDC80. The model highlights how microtubule is captured by dynein and then transferred to the NDC80 complex to form a stable end on attachment. Microtubules are laterally captured by dynein (left, lateral state) and in this state, RZZ inhibits NDC80 by binding to the 'tail' of NDC80

complex. Given the short half-life of the microtubules, de-polymerization events lead to shortening of microtubules until the point where dynein remains bound which prevents further de-polymerization. This represents the intermediate state, where dynein remains bound at the end. This model postulates that end-coupled force helps in orienting the chromosome and at this point, due to events yet unknown, NDC80 complex is alleviated from inhibition by RZZ. The model further postulates that the alleviation of RZZ mediated inhibition only occurs under conditions when 'merotely' is minimal.

Currently, the state of understanding of this complex problem can be summarized in the figure above (**Figure 1.12**) [85]. Kinetochore dynein (minus end directed motor, depicted in light blue) laterally captures a microtubule (left) and in this scenario Ndc80 remains inhibited by RZZ in a Aurora B independent manner [85]. Since the half-life of microtubules is short, it starts depolymerizing from the 'plus' end and at a certain stage reaches the kinetochore where it is stabilized by dynein (intermediate stage, center in **Figure 1.12**). Following this, loss of the RZZ complex (due to yet unidentified reasons; conformational change, binding of other proteins?) releases NDC80 complex, which can then bind to the microtubule in an end-coupled manner forming stable kinetochore-microtubule interactions. In essence, RZZ remains associated and inhibits NDC80 complex till it is liberated from the kinetochore, so that proper kinetochore-microtubule attachments can occur. The signaling cascade still remains to be deciphered.

1.2 Neuronal ceroid lipofuscinoses

The neuronal ceroid-lipofuscinoses (NCLs) are genetically inherited neurodegenerative diseases. These diseases are characterized at the cellular level by accumulation of auto-fluorescent storage material (ceroid and lipofuscin) in the lysosome of neurons and other cell types. Clinically, these disorders are characterized by seizures, progressive deterioration of cognition (dementia), loss of vision and impairment of motor function (ataxia, spasticity) [112]. NCLs have been classified traditionally depending on the age of onset and ultrastructure of the storage material [113]. Broadly these diseases can be divided into classes such as congenital, infantile (INCL), late infantile (LINCL), juvenile (JNCL) and adult (ANCL). The molecular basis and genes responsible for various NCLs have been identified and are listed in **Table 1.2.** NCLs are not an outcome of a particular subset of proteins. For example, the first two types of NCLs listed in the table are caused due to mutations or defects in the soluble protein PPT1 (palmitoyl-protein thioesterase 1) whereas JNCL occurs due to defects in the membrane protein Cln3p, the function of which is unknown [113].

Table 1.2 NCLs (I, infantile; LI, late infantile; J, juvenile; A, adults; C, congenital)

Name	Age of onset	Gene
INCL, Santavuori-Haltia disease, CLN1	6-24 months	CLN1/PPT1 on 1p32
LINCL, Jansky-Bielchowsky disease, CLN2	2-4 years	*CLN2/TPP2* on 11p15.5
JNCL, Batten disease	4-10 years	*CLN3* on 16p12.1
LINCL, CLN5	4-7 years	*CLN5* on 13q21.1-q32
LINCL, CLN6	1.5–8 years	*CLN6* on 15q21-q23
INCL, CLN7		*MFSD8* on 4q28.1-q28.2
LINCL, CLN8	3-7.5 years	*CLN8* on 8pter-8p22
ANCL, Kufs disease, both autosomal dominant and recessive versions	15-50 years	Additive effect of mutations on all of the above
CNCL, CLN10	On or before birth	*CTSD* on 11p15.5

1.2.1 Juvenile neuronal ceroid-lipofuscinosis

Juvenile neuronal ceroid-lipofuscinosis (JNCL), alternatively known as Batten disease or Spielmeyer-Vogt-Sjogren disease is the most common of all the neuronal ceroid-lipofuscinoses known. Clinical features including loss of vision, myoclonic seizures, loss of cognitive function and motor dysfunction characterize the disease. Distinguishable pathological features, characteristic of this neurodegenerative disorder constellation, are progressive neuronal loss along with accumulation of auto-fluorescent storage material (like lipofuscin) in the cytoplasm of neurons and other cells [114]. The international Batten disease consortium, in 1995, identified a gene mutated in JNCL as *Cln3* (OMIM 607042) [115]. *Cln3* was mapped onto chromosome 16, the precise locus being 16p12.1. It is organized into 15 exons (**Figure 1.14**), with an open reading frame of 1314 bp. *Cln3* encodes a protein with a molecular weight of 48 kDa, which is referred to as Cln3p. Although JNCL is the most common of all NCLs, the molecular basis of the disease is still unknown. This is mainly because Cln3p is a protein with very high hydrophobicity, making it adamant towards characterization.

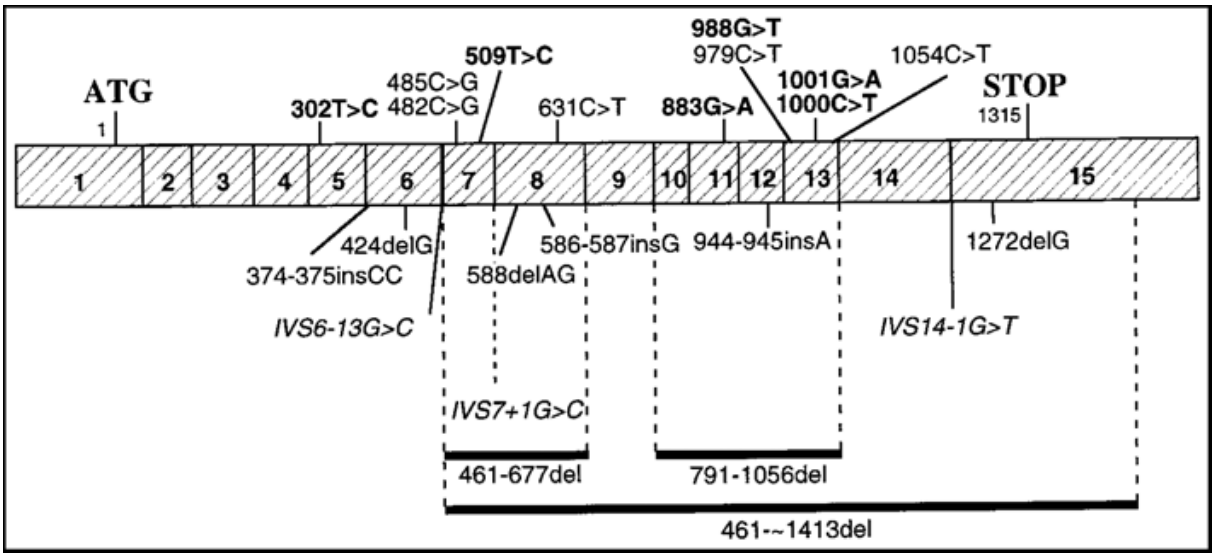


Figure 1.13: Schematic representation of mutations in *cln3.* *Cln3* is made of 15 exons. Many mutations in the gene have been identified causative of JNCL. Approximately 41 mutations in the gene have been identified with JNCL, the most common of which is the 1.02kB genomic deletion leading to the truncation of the protein. The mutations shown in the figure, above the cDNA are point mutations in the ORF and those below are deletions, insertions or point mutations in the introns. Missense mutations are represented in boldface and the point mutations in introns in italics. A complete and updated list of mutations can be found on the NCL mutation database (www.ucl.ac.uk/ncl). Source: [116]

1.2.2 Cln3p

1.2.2.1 Topology and predicted structure

Cln3p is an integral membrane protein with varied predictions regarding its trans-membrane helices. Topology prediction algorithms predict anywhere between 5 and 8 helices spanning the membrane (all α-helical) [117]. Experimental pursuit of topology led to two models, one predicting five [118] trans-membrane helices and other six [119]. The models are similar except for the following major difference: the model which predicts five trans-membrane helices, also predicts that the N- and C-termini of the protein are on opposite sides of the membrane, whereas in the alternative model (with six helices) the N- and C- termini are on the same side of the membrane and are cytosolic. Recently, the six trans-membrane model has gained wider acceptance [120].

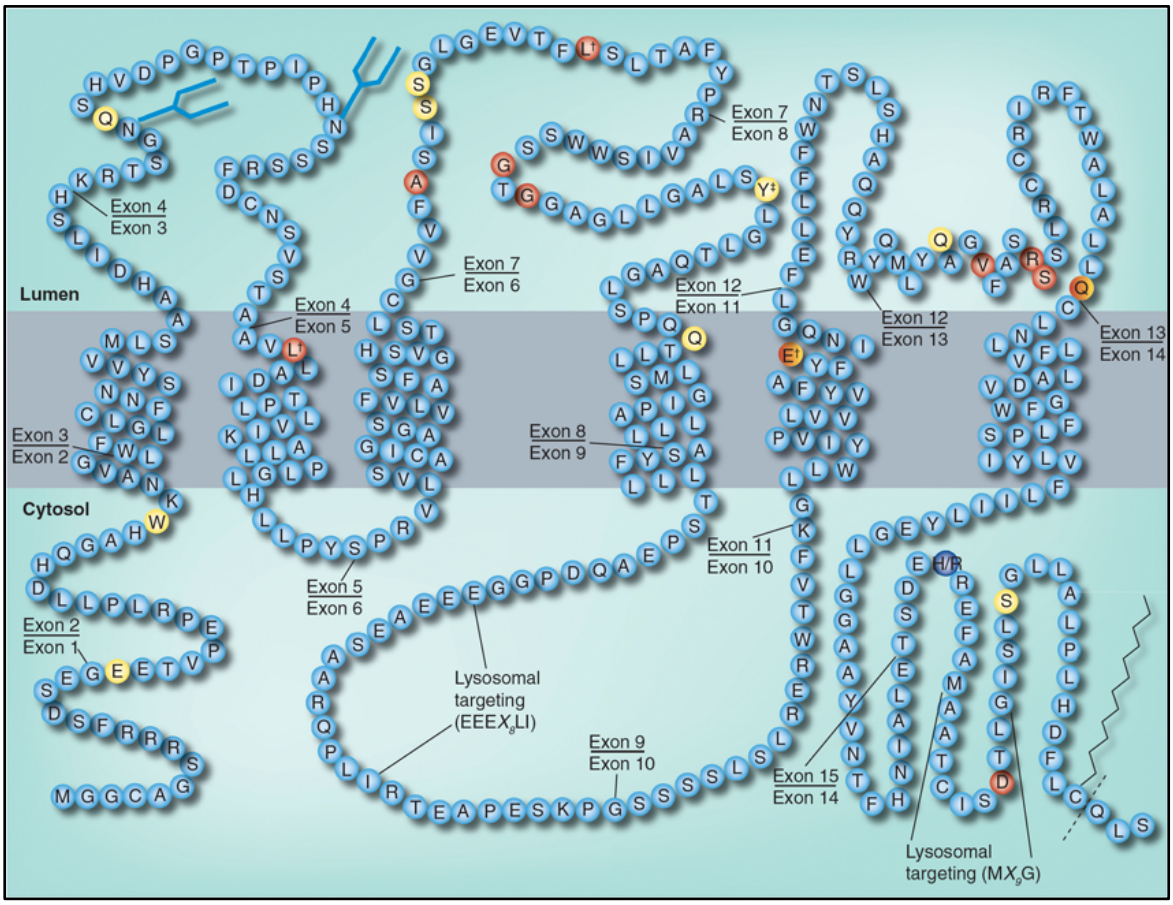


Figure 1.14: Topology of Cln3p. Cln3p is made up of 438 amino acids, most of which are predominantly hydrophobic. Various prediction algorithms have churned out varying results for the topology of Cln3p. The most commonly accepted, however, is the six trans-membrane model. The N- and C-termini are cytosolic. The protein is predicted and shown to undergo a variety of post-translational modifications. N-glycosylation sites are indicated with blue forks, and the zigzag dotted line indicates prenylation (at the very C-terminus). The two lysosomal targeting domains are also cytosolic. Red circles represent missense point mutations and the ones in yellow represent the nonsense mutations. Residue 404 (H/R) represents a polymorphism that has been reported (dark blue). Residues which are colored dually, represent both missense and nonsense mutations. Source: [121]

The topology as predicted by the six trans-membrane model is depicted in **Figure 1.14**. Salient features of this model are as follows:

1. The N- and C- termini of the six-pass trans-membrane protein are cytosolic.
2. An amphipathic helix, which lies perpendicular to the membrane, between helices 5 and 6.
3. Some of the point mutations that cause JNCL lie in the helix mentioned above and most of them occur in loops facing the lumen.
4. All the motifs that direct the protein to the lysosome are cytosolic.
5. And lastly, all the plausible sites for glycosylation lie in the lumen.

The protein Cln3p is known to undergo post-translational modifications (PTMs). PTMs are important for proper folding and functionality of the protein. The protein undergoes N-glycosylation [122, 123], which may be important for its functional role but has been shown of no relevance in trafficking. The protein has ten serines and three threonines that could undergo phosphorylation. It has been shown

to undergo phosphorylation as a posttranslational modification [124], but not mannose-6-phosphate modification. It remains unclear, at this point in time, if the protein undergoes other modifications. Cln3p has a CAAX box (435CQL438S) motif, which has potential to be modified with isoprenoids but it is not sure as of yet if it gets modified by an anchor *in-vivo* [125]. The sites for possible post-translational modifications are highlighted (**Figure 1.14**).

1.2.2.2 Mutations in Cln3p cause JNCL

The NCL mutation database (www.ucl.ac.uk/ncl) lists all the mutations in *Cln3*, causative of JNCL, that have been identified to date. 41 such mutations are known, most prominent of which is a 1.02-kb genomic deletion mutation [115]. Most patients are homozygous for this mutation (74%), but a significant number of patients are heterozygous with 1.02-kb contributing to one and other being a rare mutation (22%). Four large deletions, five small deletions, ten missense, ten nonsense, six splice site mutations and five insertions make up rare mutations and are cause of disease for a significant demographic. A concise representation of mutations in the cDNA is depicted in **Figure 1.13**, and mutations at the level of amino acids are highlighted in **Figure 1.14** (yellow and red colored amino acids representing nonsense and missense mutations respectively). Minimal numbers of patients (4%) are also heterozygous with two rare mutations on both alleles [116].

A closer look at the topology establishes that most of the mutations are on the luminal side of the trans-membrane region, especially in the highly conserved luminal loop 2 [126, 127]. It is widely predicted that the mutations lead to inefficiency in folding and hence affect the stability and trafficking of the protein resulting in lesser stable versions of the protein. Supplementing the prediction above is the observation that mutant protein variants generally have residual activity as compared to wild type protein and can partially rescue C*ln3* null phenotypes [126, 128, 129].

Lastly, it is currently unknown whether mutations involving genomic deletion (like the 1.02kb deletion) encode for a stable protein [130, 131].

1.2.2.3 Localization behavior of Cln3p

One may undermine the importance of this protein, as its functional relevance has not been clearly identified. This is ramification of the fact that functional relevance, mostly, is implicated and inferred upon from: one, the definite sub-cellular localization of the protein and two, presence of conserved motifs/domains or sequences. Various studies have been undertaken to understand where and how the protein localizes in the cell. Not amusingly, in reality, different approaches have led to its identification in different cellular compartments, making things hard to comprehend. The justification for uncharacteristic sub-cellular localization is simply based on the very high hydrophobicity of Cln3p, which makes generation of monoclonal antibodies against the native protein impossible. Instead various peptide antibodies have been used, yielding incongruous results.

Cln3p has been an understudy of numerous localization attempts by techniques such as confocal microscopy, immunofluorescence, cell fractionation and immunocytochemistry. It has been shown to

populate membranes of cellular compartments including: the Golgi [132, 133], ER [122, 133], lysosome [119, 123], plasma membrane [134] and endosomes [119]. It has also been shown to dwell at synaptic vesicles[128], nuclei, cytoplasm [135] and at the presynaptic vesicles [136] in neurons. The pattern of localization has been summarized in **Table 1.3** that follows. What is evident is that distribution of Cln3p depends on the cell/tissue type and the method of study (studying endogenous expression levels vs. expression of tagged proteins. In the table that has been indicated in the right most column where the antibody against a particular region or the tag used for overexpression have been mentioned).

Immunofluorescence experiments coupled with cellular fractionation techniques highlighted lysosome as primary destination of Cln3p [137]. Endogenous protein was also shown to populate lysosomes [134]. Two motifs were identified for the lysosomal targeting of Cln3p: A di-leucine motif at amino acids 425-426 and a, MX9G motif [119]. Deletion of any of the motifs compromises the protein's ability to reach lysosomes, deeming them absolutely indispensable. Finding that Btn1p (yeast homologue of Cln3p) localizes to the vacuole, an organelle that is the lysosomal equivalent, in yeast confirms lysosome as a primary destination for Cln3p [138]. Since NCL is primarily a brain related disease, Cln3p's localization was investigated in neurons and brain tissue. It was found to localize at the synaptic vesicles in murine primary telencephalic neurons [139], at the lipid rafts in bovine brain tissue [140], at the endosomes along neuronal processes and lysosomes in the cell body [119]. Disease causing mutations of the protein when analyzed, in some cases, show a defective lysosomal localization hinting at a possible explanation for the molecular basis of the disease [139].

In essence, Cln3p is a protein that resides mainly at the lysosome, involved in targeting and trafficking pathways. It may either be sorted out of the Golgi or the plasma membrane, following which it gets internalized into the endocytic system and is ultimately integrated into the lysosomal membrane [141].

Table 1.3 Comprehensive overview on the localization pattern of Cln3p

Primary localization (Reference/s)	Cell type/Tissue	(Antibody/Tag)
Mitochondria [142]	Retina sections of mouse	20-aa internal
Golgi [133]	Fibroblasts- HeLa, COS7 cells HeLa	C,N-terminal GFP

Lysosome (late endosomes in certain cases) [119, 122, 128, 134, 137, 139, 143]	HeLa CHO cells NRK, mouse primary neurons A549 cells HeLa, NRK, hippocampal neuronal cells Embryonic testicular cells Mouse liver hepatocytes	aa 242-258 N –terminal GFP aa 4-19(h 345/Cln3) aa (251-265/ Q438) aa(1-33), h385cln3 Q438 and aa (2-18) aa(5-19)(anti-mouse)
Vacuole [138]	Yeast Yeast	N,C terminal GFP C terminal GFP
Neuronal synapses [136]	Mouse brain & primary retinal cultures	h,m 385/Cln3
Cell Membrane, lipid rafts, nuclei [135, 140]	Human autopsy and bovine brain tissue	aa (2-18,64-76, 250-264, 321-340)
Golgi, lipid rafts, plasma membrane, early endosomes [144]	Human neurons hNT2 and fibroblasts	aa (58-77)

1.2.2.4 Functional roles

The precise function of Cln3p has remained an enigma, since its discovery. Various attempts to decode the function of Cln3p have hinted at different possibilities. A lot of interesting findings stem from studying Btn1p (the fission yeast homologue of Cln3p, 39% identical and 59% similar to Cln3p) [126, 138, 145-147]. What follows is a synopsis of the diverse roles that it has been attributed to.

1.2.2.4.1 Role in transport of amino acids, small molecules and in lysosomal pH homeostasis

Structural classification of proteins (SCOP) and Pfam servers predict a similarity of the structure of Cln3p (PF02487) with members of major facilitator superfamily (MFS) (Pfam clan CL0015). MFS is one of the two major superfamilies of membrane proteins known. They transport drugs, sugars, small molecules, metabolites and all kind of ions [148]. Although they contain 12 trans-membrane helices, it is well known that the N-terminal half is similar to the C-terminal half of the protein, a result of internal tandem gene duplication. Not surprisingly Btn1p (fission yeast homologue of Cln3p) has been implicated in the transport of arginine and galactosylceramide [145, 149].

Deletion of *btn1 l*eads to abnormalities in pH and other vacuolar defects, which could be rescued by Cln3p [126]. Elevated lysosomal pH is a hallmark of JNCL and other NCLs [150] Changes in lysosomal

pH affect the intracellular processing of various proteins, for example, amyloid-beta protein precursor, which is involved in Alzheimer's disease [151].

1.2.2.4.2 Role as a fatty-acid desaturase

It was also predicted that Cln3p has a low homology to fatty acid desaturases [152]. This led to the discovery of hitherto unknown activity of Cln3p. It was shown that lysates expressing Cln3p could introduce a double bond at the Δ-9 position of palmitate group [152]. Brain and pancreactic tissues of *Cln3*-knockout mice showed remarkably less desaturase activity towards S-palmitoylated H-Ras protein and S-palmitoylated cysteine [113], making palmitoylated proteins putative substrates of Cln3p.

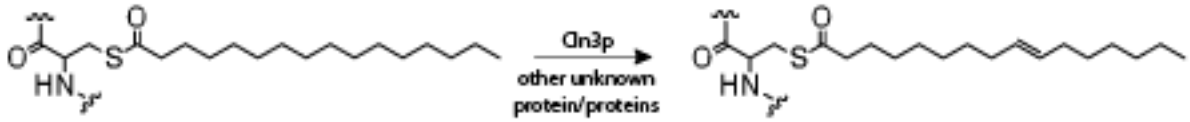


Figure 1.15: Desaturase activity of Cln3p. Lysates and tissue extracts containing Cln3p have been shown to exhibit desaturase activity, where a double bond is introduced in the Δ-9 position of palmitate moiety of the palmitoylated proteins. The reaction however has not been studied with purified components. A schematic is shown.

Many proteins involved in signaling (neurotransmitters, cytokine receptors, H-Ras, and src-kinases) get associated with the membrane as a consequence of their palmitoylation [153]. Change in the nature of the fatty acid (from saturated to desaturated) could possibly lead to alteration in spatial and temporal localization of these proteins and may influence downstream events. Desaturase activity of Cln3p is fascinating find, in lieu of the fact that infantile NCL (INCL) is caused due to mutations in PPT1; a protein that catalyzes removal of palmitoyl groups from post-translationally modified proteins [154].

1.2.2.4.3 Role in apoptosis

Hallmarks of JNCL include massive cortical neuronal death and loss of photoreceptor cells [112] making apoptosis an underlying mechanism for neuro-degeneration amongst JNCL patients [155]. Overexpression of Cln3p rescues cells, subjected to chemical treatment with vincristine, staurosporine, etc., from death [156]. A fascinating discovery that Cln3p is overexpressed, both at mRNA and protein levels, in variety of cancer cell lines including breast, colon, prostate, ovarian, melanoma, neuroblastoma and glioblastoma (also, upto 330% higher than normal in 80% of solid human colon cancer cases) made it a plausible candidate for cancer drug-discovery. When the Cln3 protein levels in cell lines (above mentioned) are blocked, using adenovirus-mediated antisense Cln3 methodology, there is a phenomenal reduction in growth and viability of cells [157]. In essence, antisense-Cln3 methodology could be valid for treatment of some types of cancers and Cln3p could be a target for screening cancer therapeutics.

1.2.2.4.4 Role in post-Golgi anterograde and retrograde trafficking

A considerable amount of endogenous Cln3p localizes to the Golgi [133]. The Golgi is proximally connected to the endocytic pathway, involved in recycling protein and lipid cargo. Loss of Cln3p halts the exit of the mannose-6-phosphate receptor from the Golgi [158], which plays the very crucial role of delivering digestive enzymes to the lysosome. Similarly deletion of *Btn1* (*S. pombe* homologue of *Cln3*) leads to disintegration of Golgi morphology and abnormalities in trafficking of Vps10p (protein involved in vacuole protein sorting) and other proteins [146]. Recent studies have also highlighted the direct involvement of Btn1 on the Golgi SNARE (soluble N-ethyl-maleimide sensitive fusion protein attachment receptor) assembly and there by the endosome-Golgi retrograde transport [159]. Taken together, recent advances indicate strongly that the Cln3p may be involved in post-Golgi trafficking.
Apart from these Cln3p has been shown to interact with plethora of proteins like AP-1, AP-3 (adaptor protein in clathrin-mediated endocytosis), fodrin, Hook1, TPP1 leading to speculations of wider roles and functionalities. Lastly, since JNCL patient cells harbor a defect in maturation of autophagosomes, a step that needs concomitant fusion of autophagosomes with functional endosomes and/or lysosomes, role of Cln3p in autophagy can not be undermined [121]. Cln3p could possibly be involved in interactions with autophagy proteins such as Atg-3 and Atg-7, which are necessary for assembly of inner and outer membranes of nascent phagophore and oversee its maturity.
Cln3p, with its myriad of possible interactions and speculative roles remains fascinating to be understood structurally and biochemically.

1.3 Aim and motivation

The work described in the thesis is assimilated from two different projects. The aim with regard to each of the project follows

Structure of the RZZ complex:

The 3-subunit Rod-Zwilch-Zw10 (RZZ) complex is a crucial component of the spindle assembly checkpoint (SAC) in higher eukaryotes. It is required for kinetochore localization of the Mad1-Mad2 checkpoint complex and the microtubule motor dynein. It therefore contributes not only to kinetochore-microtubule attachment but also to the dynein-dependent stripping of SAC components upon checkpoint satisfaction. How the RZZ complex fulfills these different roles and integrates signals from the kinetochore-microtubule interface remains unclear.
The RZZ complex will be purified in the department of Mechanistic cell biology (AG Musacchio). The quality of protein will be monitored by negative stain electron microscopy after each round of optimization. As soon as the production of RZZ is optimized, I will perform negative stain electron microscopy analysis on the sample. This would then give an idea on how the complex looks, flexibility, symmetry and other important structural aspects of the complex. If the complex takes all the orientations on ice then I would use a common-lines approach to generate initial model, otherwise I would perform a random conical tilt analysis. I will then go on (depending on the nature of the complex) to generate 3D electron density map of the complex. The sample will then be optimized for high-resolution structure determination using cryo-electron microscopy. I will try to understand the organization of proteins in the complex using antibody labeling and cross-linking studies. I will model the various proteins of unknown structure and fit them into the structure. The choice of fitting procedure will depend on the resolution of the structure. Also, if I manage to achieve a structure below sub-nanometer resolution, the density will be modeled using various programs and the implications of the structure will be studied *in vivo*.

Structure of human Cln3p:

Cln3p is the protein implicated in Batten's disease; the most devastating of all the inheritable NCL disorders. The extent and severity of symptoms varies amongst individuals (occurs only in children). It has been nearly two decades since the discovery of the protein and there is little or no information from biochemical perspective. This is predominantly due to the low endogenous expression level and hence, problems with obtaining high yields of properly folded protein. Nevertheless, understanding the function of the protein and its role in cellular pathways remain essential for figuring out ways to alleviate the suffering of patients and finally to find a cure to the disease. Understanding the structure would give valuable insights in the function of the protein and subsequently new structure-based therapeutic strategies can be sought to tackle the disease.

Having this in mind, I need to express the full-length human protein in sufficiently large amounts using various expression systems for protein production. As the bacterial expression system is inept in proper folding of membrane proteins, I will try and optimize expression of protein in insect cell expression system. Fusion constructs with different reporter and purification tags for facilitated tracking and convenient purification will be used. I will use a selection of purification methods to achieve the highest possible purity for three-dimensional x-ray crystallography studies. Besides this, I will also use electron crystallography and lipidic mesophases for crystallization trials. I would also like to address the biochemical function of the protein. Therefore, assays will be set-up to try and understand the protein function. Finally, I would also like to address the oligomerization state and nature of protein in an attempt to understand what role it plays inside the human cell.

.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Table 2.1 List of all the chemicals used for this research work

Name	Manufacturer
Acetonitrile	Sigma-Aldrich
Acrylamide/Bisacrylamide	AppliChem
Ammonium peroxodisulfate (APS)	Serva
Ammonium sulfate	AppliChem
Ampicillin	Gerbu
β-mercaptoethanol	Serva
Bacto-agar	Roth
Bacto-tryptone	Sigma-Aldrich
BCIP/NBT solution	Roche
Bradfored reagent	Roth
Brain polar lipids extract	Avanti polar lipids Inc.
CH_3COONa	Sigma-Aldrich
$CoCl_2$	
C-flat holey carbon grid	Protochip Inc.
CHAPS	Affimetrix (Anatrace)
Chloramphenicol	Roth
Chloroform	J.T. Baker
Complete protease inhibitor cocktail	Roche
Coomassie Brilliant Blue	BioMol Feinchem. GmbH
Cymal-6	Glycon
DDM	Affimetrix (Anatrace)
DLPC	Avanti polar lipids Inc.
DM	Affimetrix (Anatrace)
DMSO	Sigma-Aldrich
DNAse I	Roche
DOPC	Avanti polar lipids Inc.
DOPG	Avanti polar lipids Inc.
Dried milk powder (fat free)	Sigma-Aldrich
DTE	Sigma-Aldrich
DSS (Disuccinimidyl suberate)	Pierce
EDTA	Sigma-Aldrich
E. coli polar lipids extract	Avanti polar lipids Inc.
Ethanol (absolute p.a.)	J.T. Baker
Fetal Calf Serum (FCS)	GIBCO (Invitrogen)
Fluoromount	Serva
Fos-Choline-12	Affimetrix (Anatrace)
Fos-Choline-13	Affimetrix (Anatrace)
Formaldehyde	Sigma-Aldrich
Glucose	Sigma-Aldrich
Glutathione	Roth
Glycerol	Sigma-Aldrich
HCl	J.T. Baker

HEPES	Gerbu Biotechnik GmbH
Holey carbon girds	Agar scientific
Imidazole	Roth
Incidur	Ecolabs
IPTG	Roth
Kanamycin	Gerbu
KCl	Sigma-Aldrich
KH_2PO_4	Sigma-Aldrich
K_2HPO_4	Sigma-Aldrich
Liver polar lipids extract	Avanti polar lipids Inc.
Lysozyme	Sigma-Aldrich
Methanol (absolute p.a.)	J.T. Baker
$MgCl_2$	Sigma-Aldrich
$MnCl_2$	Sigma-Aldrich
MNG-3	Affimetrix (Anatrace)
N_2 (liquid)	Air Liquid
NaCl	J.T. Baker
$NaHCO_3$	Sigma-Aldrich
NaH_2PO_4	Sigma-Aldrich
Na_2HPO_4	Sigma-Aldrich
NaN_3	Sigma-Aldrich
NaOH	J.T. Baker
OG	Affimetrix (Anatrace)
Pefabloc	Roth
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich
Sodium dodecyl sulfate (SDS)	Boehringer
Tetrabutyl ammonium bromide	Merck
Tetramethylethylen diamine (TEMED)	Serva
Tris-HCl	Roth
Triton-X100	Sigma-Aldrich
Tween-20	Calbiochem
Uranyl format	SP1-chem
X-Gal	Sigma Aldrich
Yeast extract	Roth
Yeast nitrogen base	Roth

2.1.2 Devices and instruments

Table 2.2 List of instruments used for the work

Instrument	Manufacturer
ÄKTA purifier - Box, pHIC, UV, P - 900	GE Healthcare
Analytical balance PT1200	Satorius
Automatic pipette Pipetus	Hirschmann Laborgeräte
Automatic pump SP quick	Heidolph
Autoclave LABOKLAV 100	SHP Steriltechnik
Centrifuge Allegra X-22 R	Beckman Coulter
Centrifuge Avanti J-20 XP	Beckman Coulter
Clean Bench, Insect cell S-1000	BDK
Column, SEC, Superose 6 10/300 GL	GE Healthcare
Column, SEC, Superose 12 10/300 GL	GE Healthcare
Column, IEC, Mono Q	GE Healthcare
Column, IEC, Mono S	GE Healthcare
Column, IMAC, XK 16	GE Healthcare
Confocal fluorescence Microscope	Leica
Cryo plunge3	Gatan Inc.
Dryblock, Dri-BLOCK BD-2A, 42°C	
Dryblock, Dri-BLOCK BD-2P, 96°C	
Electron Microscope JEM-1400	JEOL
Electron Microscope 3200 FSC	JEOL
Fluorescence Imaging System Versa Doc	BIO-RAD GmBH.
Fluorescence Microscope	Zeiss
Fluorescence Multititer Plate Reader POLARstar Omega	BMG Labtech
Fractionizer Frac-950	Beckman Coulter
High vacuum coating system MED020	Bal-tec
Ice machine AF 100	
Imaging System RockImager 1000	Formulatrix
Incubator Excella E 24	
Incubator, 30°C	BINDER
Incubator, 37°C	BINDER
Incubator, 37°C	NewBrunswick Scientific
Incubator ED 115	BINDER
Light Microscope	Zeiss
Microwave Continent MW 800 G	IKA Labortechnick
Magnetic stirrer RCT basic	Heidolph
Magnetic stirrer MR 3000	Microfluidics
Microfluidizer M-100 S	Eppendorf research plus
Micropipettes	Biorad
Objective	Leica
PAGE chamber	Biorad

Pipetting robot mosquito HTS	TPP Labtec
pH meter 761 Calimatic	Knick
Plasmid Isolating Robot QIAcube	QIAGEN
Power Source PAC 3000	Biorad
Precision balance TP-124 A	Denver Instruments
Refrigerator, 4°C	SIMENS
Refrigerator, Premium/Comfort NoFrost, -20°C	LIEBHERR
Refrigerator, C-MDF-U 50 V, -80°C	SANYO
Rotor, UC, Ti 45	Beckman Coulter
Rotor, UC, Ti 70	Beckman Coulter
Rotor, UC, TLA 100.4	Beckman Coulter
Rotor, C, JA-10	Beckman Coulter
Rotor, C, JA-25.50	Beckman Coulter
Rotor, C, JLA-8.1000	Beckman Coulter
Safety hood	Waldner service
Scanner CanoScan LiDE 60	Canon
Shaker ST5 CAT	
Shaker KS 130 Basic	IKA
Spectrofluorometric detector RF-551	Beckman Coulter
Tabletop Centrifuge 5415 D	Eppendorf
Tabletop Centrifuge 5415 R	Eppendorf
TITAN Krios	FEI
Thermomixer compfort	Eppendorf
Thermal Cycler PTC-200	MJ Research
Thermocycler primus 96 advanced gradient	peqlab
UV/Vis- spectrophotometer ND-1000 NanoDrop	peqlab Biotechnologie GmbH
Ultracentrifuge Optima L-70 K	Beckman Coulter
Ultracebtrifuge Optima TL 100	Beckman Coulter
Vortex shaker Genie 2	Scientific Industries

2.1.3 Dispensables and kits

Table 2.3 Consumables and kits

Consumable	Manufacturer	Details
96-well crystallization plates Corning CrystalEX Microplate 3550	Corning Inc.	
96-well crystallization plates Corning CrystalEX Microplate 3553	Corning Inc.	
96-well crystallization plates V-Shaped	Corning Inc.	
Amicon ultra-4	Millipore	Cut-off: 10 kDa
Amicon ultra-15	Millipore	Cut-off: 10 kDa
Cell culture flask	Becton Dickinson Falcon	75 cm^2
Cell culture plate	BD Biosciences	10 cm

Crystallography kits	Qiagen	Optimized kits containing various precipitant solutions.
Dialysis bag	Spectrum labs	Cut-off: 12-14 kDa
FuGENE HD	Roche	
Glass slide	Thermo Fisher Scientific	6 x 2 cm
Graduated pipettes 1, 5, 10, 25 mL serological	Sarstedt	
Insect cell culture bottle	Bellco Glass	
Micropipette tips	Sarstedt	10, 100 µL Biosphere
Micropipette tips (ReFill-System)	Nerbe plus	10 µL, 200 µL, 1250 µL
Ni-NTA resin	QIAGEN	50-mL volume
PVDF membrane	Whatman	Polyvinylidene fluoride
Parafilm	Pechiney Plastic Packaging	Parafilm "M"
PD-10 column	GE Healthcare	Desalting column
Qiaprep Spin MiniPrep	Plasmid purification	Qiagen
Reaction tubes	Eppendorf	0.5, 1.5, 2.0 mL
Reaction tubes	BectonDickinson Falcon	15 and 50 mL
VivaSpin 500 columns	Sartorius stedim biotech	Cut-off: 10 kDa

2.1.4 Enzymes and antibodies

Table 2.4 List of enzymes and antibodies used

Enzyme	Manufacturer
γ-phosphatase	AG Musacchio
BSA	Thermo Scientific Pierce
Dnase I	Roche Bioscience
DpnI, FastDigest	Fermentas
MangoTaq DNA Polymerase	Bioline
Phusion High-Fidelity DNA Polymerase	Finnzymes
PNGase F	NEB
Precision Protease	Dortmund Protein Facility
Trypsin	Sigma-Aldrich
Cln3p- Goat polyclonal antibody	Santa Cruz biotech
Mouse anti-His	Sigma Aldrich
Goat anti-mouse (fusion with alkaline phosphatase)	Calbiochem
Polyclonal-ROD antibody	Generated at the antibody facility, Milan (AG Musacchio)

2.1.5 Media

Table 2.5 Media and composition

Name	Ingredients
LB, for 1 L	1% (w/v) bacto-tryptone (10 g) 0.5% (w/v) yeast extract (5 g) 200 mM NaCl (10 g) ddH_20 to 1 L, pH 7.0
Insect cell growth medium	TC100 + 10% FCS
Mammalian cell growth medium	DMEM/F12 1:1 (PAN Biotech) 10 % FCS (GIBCO) 1:1000 Penicillin (5,000 U) Streptomycin (5,000 µg) (PenStrep, Invitrogen)
SOC, for 1 L	2% (w/v) bacto-tryptone (20 g) 0.5% (w/v) yeast extract (5 g) 10 mM NaCl (0.584 g) 2.5 mM KCl (0.186 g) 10 mM $MgCl_2$ (0.952 g) 20 mM glucose (3.603 g) ddH_20 to 1 L, pH 7.0
TB, for 1 L	1.2% (w/v) bacto-tryptone (12 g) 2.4% (w/v) yeast extract (24 g) 4 mL glycerol 15 mM KH_2PO_4 (2,13 g) 72 mM K_2HPO_4 (12,54 g) ddH_20 to 1 L, pH 7.0
2TY, for 1L	1.6% (w/v) bacto-tryptone (16 g) 1% (w/v) yeast extract (5 g) 200 mM NaCl (10 g) ddH_20 to 1 L, pH 7.0
TC-100 medium (for 5L)	5 L autoclaved H_2O 110 g/L TC-100 powder 1.75 g $NaHCO_3$ pH 6.2

2.1.6 Insect and Mammalian cell lines, bacterial strains

Table 2.6 List of strains and cell lines used for the study

Name	Source	General information
BL21 Rosetta (DE3) pLysS (*E.coli*)	Novagen	F^- ompT $hsdS_B(R_B^- m_B^-)$ gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R)
DH10EmbacY	Geneva Biotech	Used for Bac-to-Bac system
N2a	MPI Dortmund	Mouse neural crest derived cell line
OmniMax (E.coli)	Invitrogen	F' {proAB+ laclq lacZΔM15 Tn10(TetR) Δ(ccdAB)} mcAr Δ(mrr-hsdRMS-mcrBC) ϕ80(lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 reAl1 tonA panD
Sf9 (*Spodoptera frugiperda*)	Invitrogen	Immortalized ovarian cancer insect cell line used for baculovirus production
Sf21 (*Spodoptera frugiperda*)	Invitrogen	Immortalized ovarian cancer insect cell line used for baculovirus production
XL-1 Blue (*E.coli*)	XL-1 Blue	F' {proAB+ laclq lacZΔM15 Tn10(TetR)} recA1 endA1 gyrA96 thi-1 hsdR17 supE44 reAl1 lac

2.1.7 Software and scripts

Table 2.7 List of software and scripts used for analysis of data

Software	Source
Chimera	UCSF Chimera
CTFGUI-update-15-10-12.py (script for CTF and defocus estimation)	Dr. Rouslan Efremov and Dr. Christos Gatsogiannis
EPU- For automated data collection	FEI
EMTools- For automated data collection	TVIPS GmBH
ImageJ	http://imagej.nih.gov/ij/
I-Tasser	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/htmL/page.cgi?id=index
PyMOL	Schrödinger, LLC
SPARX/EMAN2	http://blake.bcm.edu/emanwiki/EMAN2
Versa Doc Quantity One 4.6.9	BIO-RAD

2.1.8 Solutions and buffers

2.1.8.1 General buffers

Table 2.8 List of common buffer, solutions and their composition

Name	Composition
Antibody solution (for Western blot)	1:4000 (or as manufacturer specified) 0.5% (w/v) dried milk powder 0.3 mM NaN_3 Fill up with TBS
Blocking solution (for Western blot)	5% (w/v) Milk powder, fill it up with TBS
Lysis buffer (until unless specified otherwise)	50 mM HEPES 500 mM NaCl 10% Glycerol 5 mM TCEP 1 mM EDTA 1 mM PMSF 1 protease inhibitor cocktail tablet/ 50 mL of buffer 5 µg/ml Dnase I
Paraformaldehyde fixation solution	3.6% (w/v) Formaldehyde in 1x PBS pH 7.5
PBS	2.7 mM KCl 1.7 mM NaH_2PO4 10 mM Na_2HPO4 137 mM NaCl pH 7.5 with HCl
Phosphoimager destaining solution	50 mM CH_3COONa, 20% MeCN
γ-phosphatase fixing solution	50 % Methanol, 10% Acetic acid
γ-phosphatase reaction buffer	50 mM HEPES 100 mM NaCl 2 mM DTT 0.01% Brij 35
Phosphorylation (by Mps1) reaction buffer	25mM HEPES pH 8.5 250mM NaCl, 2mM TCEP 10 mM $MgCl_2$ 1µM ATP
Semi-native running buffer	25 mM Tris 187.5 mM Glycine 0.04% SDS
Solubilization buffer for screening	1% of detergent 20 mM Tris-HCl pH 7.5

	300 mM NaCl 10% glycerol
TBS	150 mM NaCl, 10 mM Tris pH 7.5
TBS-TT	250 mM NaCl 10 mM Tris 0.025% (v/v) Tween-20 0.1% (v/v) Triton X-100 pH 7.4
Transfer buffer (for Western blot)	25 mM Tris 192 mM Glycine, 10% (v/v) Methanol
Uranyl formate staining solution	0.0375 g was added to 5 mL of boiling water and stirred for 5 min. Then 6 µl of 5 M NaOH was added and stirred for another 5 min
XL-MS reaction buffer	25 mM Hepes pH 8.5 250 mM NaCl 2 mM TCEP

2.1.8.2 Buffers for purification of Cln3p in Fos-Choline-13

Table 2.9 List of buffers required for purification of Cln3p and their composition

Name	Composition
Pre-solubilization buffer	50 mM HEPES 500 mM NaCl 1% OG, 2 mM TCEP 1 mM PMSF 10% Glycerol
Solubilization buffer	50 mM HEPES 500 mM NaCl 0.5% Fos-13, 2 mM TCEP 1 mM PMSF 10% Glycerol
IMAC (immobilized metal affinity chromatography) Wash A buffer (wash buffer A-fos)	20 mM HEPES 500 mM NaCl 0.05% Fos-13 10% Glycerol 1 mM TCEP 1 mM PMSF 5 mM Imidazole
IMAC (immobilized metal affinity chromatography) Wash B buffer (wash buffer B-fos)	20 mM HEPES 500 mM NaCl 0.05% Fos-13 10% Glycerol

	1 mM TCEP 1 mM PMSF 20 mM Imidazole
Elution buffer (elution buffer-fos)	20 mM HEPES 300 mM NaCl 0.1% Fos-13 10% Glycerol 1 mM TCEP 1 mM PMSF 150-200 mM Imidazole
Dialysis buffer (dialysis buffer-fos)	20 mM HEPES 300 mM NaCl 0.1% Fos-13 10% Glycerol 1 mM TCEP 1 mM PMSF
Size exclusion chromatography buffer (SEC buffer-fos)	20 mM HEPES 300 mM NaCl 0.1% Fos-13 10% Glycerol 1 mM TCEP 1 mM PMSF

2.1.8.3 Buffers for purification of Cln3p in MNG-3

Table 2.10 List of buffers for purification of Cln3p (detergent exchange) and their composition

Name	Composition
IMAC (immobilized metal affinity chromatography) Wash A buffer (wash buffer A-mng)	20 mM HEPES 500 mM NaCl 0.01-0.05% MNG-3 10% Glycerol 1 mM TCEP 1 mM PMSF 5 mM Imidazole
IMAC (immobilized metal affinity chromatography) Wash B buffer (wash buffer B-mng)	20 mM HEPES 500 mM NaCl 0.01-0.05% MNG-3 10% Glycerol 1 mM TCEP 1 mM PMSF 20 mM Imidazole
Elution buffer (elution buffer-mng)	20 mM HEPES 300 mM NaCl 0.05-0.1% MNG-3 10% Glycerol 1 mM TCEP 1 mM PMSF 150-200 mM Imidazole
Dialysis buffer (dialysis buffer-mng)	20 mM HEPES 300 mM NaCl 0.05-0.1% MNG-3 10% Glycerol 1 mM TCEP 1 mM PMSF
Size exclusion chromatography buffer (SEC buffer-mng)	20 mM HEPES 300 mM NaCl 0.05-0.1% MNG-3 10% Glycerol 1 mM TCEP 1 mM PMSF

2.1.8.4 Buffers for purification of RZZ and mini-RZZ

Table 2.11 Buffers for purification of RZZ and mini-RZZ

Name	Composition
Lysis buffer	50 mM HEPES pH 8.5 200 mM NaCl 5% glycerol 2 mM TCEP Protease inhibitor cocktail tablet (1 per 50 mL) PMSF
IMAC (immobilized metal affinity chromatography) Wash buffer	50 mM HEPES pH 8.5 200 mM NaCl 5% glycerol 2 mM TCEP 20 mM Imidazole
Size exclusion chromatography buffer	20 mM HEPES pH 8.5 250 mM NaCl 2 mM TCEP

2.1.8.5 Buffers for purification of Zw10

Table 2.12 Buffers for purification of Zw10

Name	Composition
Lysis buffer	1x PBS pH 8.5 150 mM NaCl 5% glycerol 2 mM TCEP Protease inhibitor cocktail tablet (1 per 50 mL) PMSF
IMAC (immobilized metal affinity chromatography) Wash buffer	1x PBS pH 8.5 150 mM NaCl 5% glycerol 2 mM TCEP 20 mM Imidazole
Elution buffer	1x PBS pH 8.5 150 mM NaCl 5% glycerol 2 mM TCEP 200 mM Imidazole
Ion exchange buffer A	50 mM HEPES pH 8.5 100 mM NaCl 2 mM TCEP

Ion exchange buffer B	50 mM HEPES pH 8.5 1 M NaCl 2 mM TCEP
Size exclusion chromatography buffer	20 mM HEPES pH 8.5 250 mM NaCl 2 mM TCEP

2.1.9 Vectors and plasmids

Table 2.13 List of vectors and plasmids used for the work

No.	Name (as enlisted in the AG Raunser database)	Features
p007	pOPINF_Nterm10HisTagPP	N-10x-His, PP
p012	pOPINE_Cterm_10HisTag	C-10xHis
P068	pOPINE_6His_EGFP_PP_LacZ	N-6xHis, EGFP, PP
p231	pOPINE_EGFP_Cln3mouse	N- short linker, GFP, 6xHis, PP
p232	pOPINE_EGFP_Cln3 human	N- short linker, GFP, 6xHis, PP
p261	pOPINE_EGFP_Cln3 rat	N- short linker, GFP, 6xHis, PP
p262	pOPINF_Nterm10His_Cln3 human	N-10x-His, PP
p263	pOPINF_Cterm10His_Cln3 human	C-10x-His
p265	pOPINF_Nterm10His_Cln3 mouse	N-10x-His, PP
p266	pOPINF_Nterm10His_Cln3 rat	N-10x-His, PP
p454	pFH_Cln3 human	N-6x-His
p496	pOPIN_6His_3C_ short linker_Cln3 human	N- short linker, 6xHis, PP
p497	pOPIN_6His_Cherry_3C_medium linker_Cln3 human	N- medium linker, mCherry, 6xHis, PP
p506	pOPIN_His_Sumo_short linker_Cln3 human	N-Short linker, 6x His, PP, Sumo
p507	pOPIN_His_Sumo_medium linker_Cln3 human	N-medium linker, 6x His, PP, Sumo
p508	pOPIN_His_Sumo_large linker_Cln3 human	N-large linker, 6x His, PP, Sumo
p509	pOPIN_His_Cherry_3C_large linker_Cln3 human	N- long linker, mCherry, 6xHis, PP
p510	pOPIN_His_Cherry_3C_short linker_Cln3 mouse	N- short linker, mCherry, 6xHis, PP
p511	pOPIN_His_ 3C_short linker_Cln3 human	N- short linker, 6xHis, PP
p512	pOPIN_His_ 3C_large linker_Cln3 human	N- large linker, 6xHis, PP
	RFP-GalT	Golgi marker
	RFP-PMP2	Peroxisomal marker
	LAMP-RFP	Lysosomal marker

2.2 Methods

2.2.1 Methods: RZZ complex

Anika Altenfeld or Sabine Wohlgemuth (AG Musacchio) optimized and purified all the proteins for this part of work. I therefore will summarize very briefly the protocols used for generation of complexes.

2.2.1.1 Purification of RZZ complex

A cell pellet of 1 L of insect cell culture expressing Zw10 is suspended into 45 mL of lysis buffer. The cells were lysed by sonication (3 x 30s, at 4°C) and the debris removed by centrifugation at 20,000 rpm for 60 min, 4°C. The supernatant was loaded onto a 5 mL pre-equilibrated Ni-NTA column at the speed of 3 mL/min using a peristaltic pump. The resin was washed with 4000 mL of wash buffer, followed by elution (2 mL fractions). All the fractions were analyzed by SDS-PAGE. The best fractions were pooled (10 mL), concentrated and the pool diluted with 40 mL of ion exchange buffer A. The sample was loaded on to a ResQ (strong anion exchanger) column with a 50 mL super loop (flow rate of 0.5 mL/min). The gradient was increased using ion exchange buffer B. Eluted fractions were analyzed again by SDS-PAGE, pooled, concentrated (MWCO 10 kDa) to 500 µL and loaded onto a Superdex 200 10/300 size-exclusion chromatography column. Part of protein was analyzed by negative stain electron microscopy and the rest was concentrated and subjected to crystallization trials.

2.2.1.2 Purification of R-(nGFP)Z-Z and R-Z(cGFP)-Z

The protocol for purification of R-(nGFP)Z-Z and R-Z(cGFP)-Z (N- and C- terminally labeled Zw10 versions of RZZ) parallels that of RZZ described above. The major difference is that Zw10 is expressed as a fusion protein (respective viruses for fusion proteins are generated) and the tag (6XHis, GFP) was not cleaved in the final step.

2.2.1.3 Purification of mini-RZZ

Individual viruses (only ROD$^{1\text{-}1250}$ was required to be generated, previously generated viruses for Zw10 and Zwilch were used) were generated and the proteins co-expressed in Tnao38 insect cells. The cell pellet was suspended in 100 mL of lysis buffer, sonicated (3 x 35 s, 4°C) and the lysate centrifuged at 25,000 rpm at 4°C. The supernatant was then loaded onto the IMAC (5 mL Ni-NTA column, pre-equilibrated with wash buffer) using a peristaltic pump at a flow rate of 2 mL/min. Following this the column was washed with 400 mL of wash buffer and the protein eluted as 1 mL fractions using the elution buffer. The fractions were analyzed by SDS-PAGE, pooled, concentrated (Amicon, MWCO-10 kDa) to 500 µL and loaded onto a Superose 6 10/300 size-exclusion chromatography column.

2.2.1.4 Purification of Zw10

A cell pellet of 1L of insect cell culture expressing Zw10 is suspended into 45mL of lysis buffer. The cells were lysed by sonication (3 x 30s, at 4°C) and the debris removed by centrifugation at 20,000 rpm for 60 min, 4°C. The supernatant was loaded onto a 5 mL pre-equilibrated Ni-NTA column at the speed of 3mL/min using a peristaltic pump. The resin was washed with 4000 mL of wash buffer, followed by elution (2 mL fractions). All the fractions were analyzed by SDS-PAGE. The best fractions were pooled (10 mL), concentrated and the pool diluted with 40 mL of ion exchange buffer A. The sample was loaded on to a ResQ (strong anion exchanger) column with a 50 mL super loop (flow rate of 0.5 mL/min). The gradient was increased using ion exchange buffer B. Eluted fractions were analyzed again by SDS-PAGE, pooled, concentrated (MWCO 10 kDa) to 500 µL and loaded onto a Superdex 200 10/300 size-exclusion chromatography column. Part of protein was analyzed by negative stain electron microscopy and the rest was concentrated and subjected to crystallization trials.

2.2.1.5 Cross-linking and Mass spectrometry (XL-MS)

For the cross-linking, 1mg of Disuccinimidyl Suberate (DSS) was dissolved in 53 µL of DMSO (Stock 1). 20 µL of stock 1 was dissolved in 250 µL of reaction buffer (25 mM Hepes pH 8.5, 250 mM NaCl, 2 mM TCEP; Stock 2). DSS (stock 2) and the protein were mixed in ratios as indicated in the table, and the reaction volume was maintained at 200 µL. Reactions were carried out either for 30 min. at 37 °C or at 4°C overnight. 22 mL NH_4CO_3 was added to stop the reaction (incubated for 15 min 37°C). The samples were then sent to the laboratory of Dr. Franz Herzog at Gene Center, Munich for further analysis.

2.2.1.6 Phosphorylation of RZZ

The reaction was carried out in phosphorylation buffer (25mM Hepes 8.5, 250mM NaCl, 2mM TCEP, 10 mM $MgCl_2$, 1µM ATP). Briefly, 5 µM of RZZ was incubated with same amount of Mps1 for 30 min at RT. Following this the mixture was centrifuged at 13,400 rpm and the supernatant loaded onto a S6 3.2/15 column and 50 µL fractions were collected.

2.2.1.7 Electron Microscopy

Electron microscopy is a technique where one visualizes objects/samples (proteins, complexes, sections) placed in the microscope by impinging electrons. The optics of imaging is essentially conserved from light microscopy and only differs in the technicalities. The 'lenses' for example in the electron microscope are not curved surfaces but applied electric and magnetic field, which influences the behavior of electrons. In single particle analysis, one essentially tries to formulate the three-dimensional (3D) object from its two-dimensional projections (2D). A classic example of object and its orthogonal projections is shown below (**Figure 2.1**).

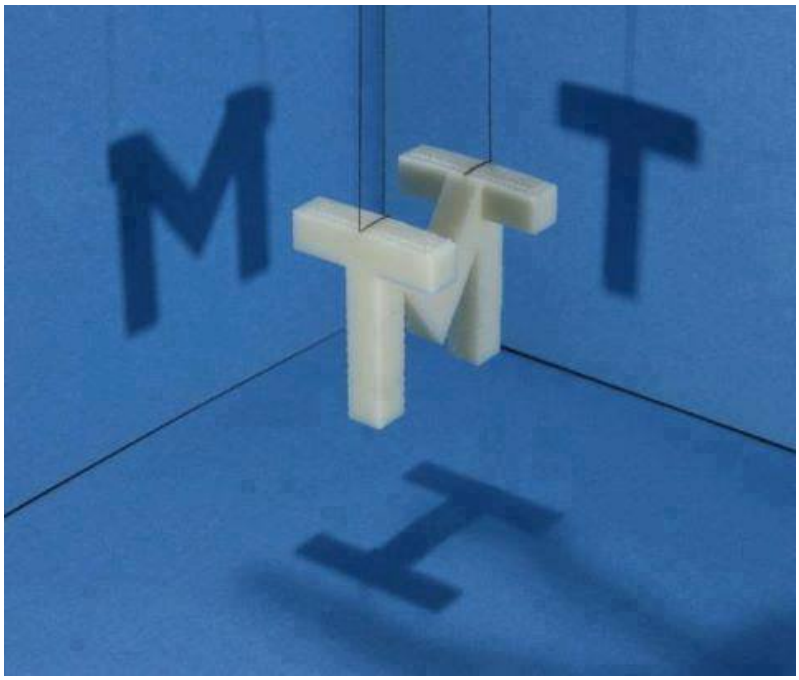


Figure 2.1: Principle of single particle electron microscopy. The object in the figure, depending on how it is projected can have M, T or I as its projections. Once, the angles of projections and their relative orientation with respect to one another are known, one can reconstruct the object. This is essentially the principle behind single particle electron microscopy, where the sample/object can take various orientations on the grid and hence gives rise to variety of projections. By determining the angles (Euler angles) and orientation parameters of these projections the 3D structure of the sample is generated. Source: World wide web.

In reality, things are way more complex as the protein/object can take all possible orientations and there is little information about the relative orientation of all particles. Another major problem is the conformational variability of the sample (where protein or protein complexes exist and adopt distinct conformations in solution but are random in ice). Finally, electron microscopy data is extremely noisy mainly because one cannot use high dose of electrons, due to the susceptibility of biological samples to damage. These problems make structure determination by electron microscopy challenging and tedious. Besides this, one is generally challenged by the size of the sample that can be used for the study. Negative stain electron microscopy (ns-EM) requires the sample be no less than 100 kDa and incase of cryo-EM the sample has to have a size of about 300 kDa.

One generally uses ns-EM for low-resolution information and to generate initial models, if there is no *a priori* information about the sample. In negative stain electron microscopy, a heavy-metal salt solution is used to enhance the contrast of the sample as it forms a thin layer over the sample examined. In Cryo-electron microscopy, the sample applied on the grid is frozen into liquid ethane at relatively high speed leading to formation of vitreous ice, which preserves the 'native' state, and high-resolution features of the sample. Cryo-EM offers the possibility to understand high-resolution information but suffers drastically from lack of contrast in the images. It is of general procedure to visualize and extract as much information as possible from negative stain EM and then follow it up by cryo-EM analysis.

The steps in structure determination by single particle analysis of images from EM (negatively stained or cryo) are more or less same and can be divided into the following steps:

1. Particle selection: single particle images of desired size (depends on the pixel size: for example, RZZ (negative stain) 300 x 300 size boxes were used and for cryo, a box size of 512 x 512 was used) are boxed out from micrograph either manually or automatically.

2. Averaging: images of similar views are aligned and averaged to achieve higher signal to noise ratio. This process involves both rotational and translational shifting of individual particle images, which generally gives information about the 2D views of the sample and its inherent heterogeneity.
3. Determination of initial model/orientation parameters: an *ab initio* estimation of parameters is obtained using the tilt approaches (random conical tilt, orthogonal tilt reconstruction), or common-lines based approaches (angular reconstitution with common-lines). Various methods, operating in real and Fourier space, have been implemented to calculate reconstructions and different software packages make use of these diverse methods. A detailed explanation is beyond the scope of this work, but a comprehensive overview can be found in many reviews [160, 161].
4. Iterative refinement: each image is matched with the re-projected images from the reconstructed volume of the last iteration to determine orientation parameters for new reconstruction until it reaches the convergence.

2.2.1.7.1 Negative staining

4 µL of the sample was adsorbed for 1 min at 25°C onto freshly glow-discharged, carbon-coated grids (Agar scientific; G2400C). The sample was blotted (Whatman no. 4), and the grids subsequently washed with the size-exclusion chromatography buffer (respective buffers in which each of the sample was purified, for complexes and membrane proteins washing with water can lead to unfavorable sample quality) and stained with 0.07% of uranyl formate (SPI supplies), unless stated otherwise. All the images were collected on a JEOL JEM1400 electron microscope equipped with a LaB_6 cathode operated at 120 kV. All the images were collected in low-dose conditions (19 e/Å^2) at a magnification of 67,200 X on a 4k x 4k camera (TVIPS F416).

2.2.1.7.2 2D analysis

For the 2D negative stain analysis, particles were windowed out manually out of the collected images using the e2boxer tool embedded in the EMAN2 suite. Almost all the datasets were subjected to binning for faster and effective use of the cluster space. The pixel size was adjusted correspondingly. Particles were aligned and classified using reference-free alignment and *k*-means classification programs implemented in SPARX. Some datasets were subjected to and analyzed by iterative stable alignment and clustering (ISAC) procedures, also implemented in SPARX. For samples where the averages were not so informative, custom-made masks (using e2display tool in EMAN2) were generated, and local alignments were performed. Following table lists all particles selected for 2D processing for various datasets.

Table 2.14 List of number of particles selected manually for various datasets

Dataset	Particles/boxsize	Essential details
mini-RZZ	3424, 256	Negative stain, defocus (-1.5 µm), JEOL1400, pixel size -2.3 Å/pix.
mini-RZZ (tilt-pairs)	1200, 256	Negative stain, defocus (-1.5 & -1.8 µm), tilt angles 55° and 0°, JEOL1400, pixel size -2.3 Å/pix.
R-(nGFP)Z-Z	6964, 350	Negative stain, defocus (-1.5 µm), JEOL1400, pixel size -2.3 Å/pix.
R-Z(cGFP)-Z	5756, 350	Negative stain, defocus (-1.5 µm), JEOL1400, pixel size -2.3 Å/pix.
RZZ	4000, 300	Negative stain, defocus (-1.5 µm), JEOL1400, pixel size -2.3 Å/pix.
RZZ (tilt-pairs)	1700, 300	Negative stain, defocus (-1.5 & -1.8 µm), tilt angles 55° and 0°, JEOL1400, pixel size -2.3 Å/pix.
RZZ (cryo-positve)	4880, 168	defocus (-2 to 3 µm), JEOL 3200FSC, collected automatically using EMtools software, pixel size-1.25 Å/pix.
RZZ with ROD polyclonal antibody targeted towards the N-terminus	500, 300	Negative stain, defocus (-1.5 µm), JEOL1400, pixel size -2.3 Å/pix.
RZZ (cryo)	35404, 512	Cryo, defocus (-2 to 3 µm), JEOL 3200FSC, collected manually and also using EMtools software, pixel size-1.25 Å/pix.
RZZ (focal-pairs)	8550, 512	Cryo, defocus range (-2.5 to 3 µm and corresponding close to focus images at -0.8-1.5 µm), JEOL3200FSC, collected manually, pixel size -2.3 Å/pix.
RZZ (leiden)	5000, 600	Cryo, defocus (-1.5 to -2.3 µm), TITAN Krios, Falcon II direct detector, 7 frames with total dose 60 e/Å^2, collected using EPU software, pixel size – 1.14 Å/pix.
Zw10	10111, 168	Negative stain, Defocus (-1.5 µm), JEOL1400

2.2.1.7.3 Random conical tilt

For the negative stain 3D reconstructions, tilt pairs were collected at tilt angles of 55° and 0°, and the particle pairs were selected using the e2RCTboxer program embedded in EMAN2 suite. The un-tilted stack of particles was used for reference free alignment and classification. A random conical tilt reconstruction of the tilted particles was calculated from the particles corresponding to the best (un-tilted) class averages by back-projection, followed by back-projection refinement. Symmetry assignment was essential for RZZ complex. 2D averages revealed that the sample had C2 symmetry. Reconstructions were performed in the presence and absence of symmetry and revealed similar density maps. C2 symmetry was used for further, negative stain RZZ refinement steps. The

reconstructions were further refined by iterative projection matching procedure implemented in SPARX, until further convergence was achieved. The resolution of final reconstructions was estimated by the Fourier Shell Correlation criterion ($FSC_{0.5)}$. This was a general approach used for RZZ and mini-RZZ datasets, the details of number of particles and the sizes of windows (in pixels) used for boxing are listed in **Table 2.14**.

2.2.1.7.4 Cryo-EM

Holey carbon grids (Quantifoils, Protochips) were freshly glow discharged (two times on the darker side and once on the lighter side of the grid) and 4 µL of the RZZ was applied. 0.02% (w/v) Triton was added to the sample just before its application onto the grid. After 30 s the sample was manually blotted (Whatman 4) followed by application of another 1.7 µL of the sample that was vitrified by plunging (1 s blotting) into liquid ethane using a Cryoplunge3 (Cp3, Gatan Inc.) The humidity was maintained to be around 90%. Images were collected a JEOL 3200FSC electron microscope equipped with a field-emission gun at an operation voltage of 200 kV, and a defocus range of 1.5-3 µm. An omega in-column energy filter was used with a slit width of 15 eV. The images were recorded with an 8k x 8k CMOS camera F816 (TVIPS) at a pixel size of 1.25 Å/pixel using minimal dose conditions.

2.2.1.7.5 CTF estimation and processing

The Contrast transfer function (CTF) parameters and the defocus were estimated for all the images collected. Images with astigmatism, drift or poor CTF fit were discarded and not used for further processing. Images were then binned four-times to facilitate processing. Single particles were manually selected using the e2boxer program (embedded in EMAN2 suite) and windowed into 128 x 128 pixel boxes (details in **Table 2.14**). 3D reconstructions were created by projection matching procedure implemented in SPARX software package. The negative stain reconstruction was used as the starting template for all the refinement procedures. C2 symmetry was imposed on all the refinement runs, as determined by the random conical tilt approach. Mask necessary for improvement of structure and calculating Fourier Shell Correlation (FSC) was used. It was generated in SPARX by low-pass filtering the density map, binarizing it, and then expanding the binary volume by Gaussian pixels. The resolution was always estimated by $FSC_{0.5}$ criteria.

2.2.2 Methods: Cln3p

2.2.2.1 Transformation into *E.coli*

Freshly thawed 100 µL cell-aliquot (*E. coli* XL-1 Blue cells) was mixed with plasmid DNA (usually 100 – 200 ng) and subsequently incubated on ice for 30 minutes. The cells were then transiently incubated at 42 °C (heat shock) for 1 min to initiate the DNA uptake, following which they were allowed to stand

for another minute. 700 µL SOC-medium was then added to the cells and whole suspension incubated for one hour at 37 °C. The cells were then pelleted (3 min, 6000 rpm) and re-suspended in a very small volume of the supernatant (rest was discarded) and plated on LB agar plates with the appropriate antibiotic resistance marker (125 µg/mL). The plates were incubated overnight at 37°C for colonies to grow.

2.2.2.2 Expression screen in bacteria

A single clone/colony from a LB agar plate was used to inoculate 30 mL of LB medium containing 125 µg/mL ampicillin (w/v, in a 100 mL flask) for an overnight culture. After overnight incubation, two 1-liter flasks one with 250 mL 2TY medium containing 125 µg/mL ampicillin (w/v) and another with 250 mL TB medium containing 125 µg/mL ampicillin (w/v) were inoculated with 2.5 mL of the pre-culture each. The cultures were grown till the desirable OD_{600} was reached following which, expression was induced using IPTG (Isopropyl β-D-1-thiogalactopyranoside) and three 70 mL aliquots of each type of media were made and incubated each at 25°C, 30°C and 37°C. Samples (15 mL) were taken after 1h, 2h, 3h and overnight. Various conditions of expression the O.D and the amount of IPTG used for expression are listed in **Table 2.15** and are labeled from 1-23. The cells were centrifuged and the resulting cell pellet was stored at -20°C, till further use. The cell pellets were, subsequently lysed; membranes extracted and solubilized to identify expression, by western blot analysis.

Table 2.15 List of conditions in screening for expression

	Condition	O.D, IPTG
1	2TY, 25°C, 1 h	0.8, 0.5 mM
2	2TY, 25°C, 2 h	0.8, 0.5 mM
3	2TY, 25°C, 3 h	0.8, 0.5 mM
4	2TY, 30°C, 1 h	0.8, 0.5 mM
5	2TY, 30°C, 2 h	0.8, 0.5 mM
6	2TY, 30°C, 3 h	0.8, 0.5 mM
7	2TY, 37°C, 1 h	0.8, 0.5 mM
8	2TY, 37°C, 2 h	0.8, 0.5 mM
9	2TY, 37°C, 3 h	0.8, 0.5 mM
10	TB, 25°C, 1 h	0.8, 0.5 mM
11	TB, 25°C, 2 h	0.8, 0.5 mM
12	TB, 25°C, 2 h	0.8, 0.5 mM
13	TB, 30°C, 1 h	0.8, 0.5 mM
14	TB, 30°C, 2 h	0.8, 0.5 mM
15	TB, 30°C, 3 h	0.8, 0.5 mM
16	TB, 37°C, 1 h	0.8, 0.5 mM
17	TB, 37°C, 2 h	0.8, 0.5 mM
18	TB, 37°C, 3 h	0.8, 0.5 mM
19	2TY, 20°C, O/N	1.2, 0.5 mM
20	2TY, 25°C, O/N	0.8, 0.5 mM
21	TB, 20°C, O/N	1.2, 0.5 mM
22	TB, 25°C, O/N	0.8, 0.5 mM
23	TB, 37°C for 4 h followed by 25°C O/N	Auto-induction

2.2.2.3 Virus generation and insect cell expression

Plasmid DNA, harboring the gene of interest was mixed with the restricted baculovirus DNA (BACmid), transfection agent and TC 100 media. Strong polyhedrin promoter drives the expression of proteins in the late phase of infection.

The pictorial flowchart highlighting essential steps in generating a working stock of virus is illustrated as **Figure 2.2**. Initial transfection yielded, a P1 viral stock, which was propagated further at low MOIs (0.01 - 0.1) to generate large amount of potent virus. However, for expression high MOI (1-10) was used. MOI can be defined by the equation as follows:

$$\text{MOI} = \frac{\text{Volume [Virus] x Concentration [Virus]}}{\text{Volume[Cell culture] x Concentration [Cell culture]}}$$

Sf21 cells with a density of 0.5 x 10^6 in TC 100 Medium, supplemented with Pen/Strep, Pluronic and 10 % FCS, were used for expression. Insect cell cultures were transfected under the clean bench with the virus:cell ratio of 1:10, 1:100, 1:1000 and 1:10000. The expression was monitored after every 6 h by fluorescence microscopy for proteins expressed as fusions with fluorescent tags. For constructs without a fluorescent reporter tag, the expression was monitored by western blot analysis. Expression is best when cells are infected with virus at MOI of 10 (assuming the virus is very potent, typically a titer range of 10^9 pfu/mL) and after 72 h or infection. Cells were harvested after 72 h and centrifuged at 6,000 rpm for 20 minutes, following which they were washed with TBS. The cell pellet, if not processed directly was flash frozen in liquid nitrogen and stored at -20 °C, till further use.

Figure 2.2: Schematic of virus generation and propagation

2.2.2.4 Lysis and membrane isolation

The Cell pellet was resuspended in lysis buffer (usually 50 mL for a cell pellet from 1L of insect cells), the suspension stirred for 30 min at 4 °C and then lysed using a microfluidizer at a pressure of 700 psi. This step was repeated 3 - 4 times and the temperature was maintained at 4°C. Following this, the cell lysate was centrifuged for 25 min at 15,000 rpm (4 °C). To isolate membranes from the supernatant pool and to get rid of the nuclear and cytosolic fraction, the supernatant from the previous step was subjected to ultracentrifugation (Ti 70 rotor, 50,000 rpm, 90 min, 4 °C). The pellet was homogenized in 50 mM HEPES and used either directly for purification or frozen for subsequent use at a later date. For proteins expressed with fluorescent tags, the membrane would reflect the color of the tag used intensively. Otherwise, the purification was not really good. After incubation for 30 min at 4 °C the suspension was pelleted down again (Ti 70 rotor, 55,000 rpm, 30 min, 4 °C), the supernatant discarded and the pellet was then homogenized in solubilization buffer containing 0.5 % (w/v) Fos-13. The

mixture was gently stirred 3 hours at 4°C. Finally, insolubilized membrane remains were centrifuged down (Ti 70 rotor, 55,000 rpm, 30 min, 4 °C) and discarded.

2.2.2.5 Detergent/solubilization screen

The membranes were isolated as described above and resuspended in TBS. 100 µl of membrane was mixed with 200 µl of solubilization buffer containing respective detergents (1%) and incubated for 2 h at 4°C. For this study, various detergents were tested (DDM (n-Dodecyl-ß-maltoside), DM (n-decyl-ß-maltoside), CHAPS, OG (n-Octyl-β-D-Glucopyranoside), Fos-Choline-12, Fos-Choline-13, Cymal-6 and Triton-X100). A sample was taken from each detergent-membrane mixture (this sample was labeled before centrifugation (BC)). The rest was centrifuged at 55,000 rpm for 30 minutes at 4°C. Another sample was taken from the supernatant of each detergent-membrane mixture and named 'after centrifugation' (AC). Samples were mixed with SDS-loading buffer run on semi-denaturing SDS-PAGE and scanned for in-gel fluorescence. The detergent that gave the same band intensity before and after centrifugation was considered optimal. In order to optimize the solubilization efficiency parameter such as detergent concentration, salt concentration or longevity of solubilization was altered.

2.2.2.6 Purification of hCln3p (EGFP at the N-terminus)

The protocol for EGFP-Cln3p (human, construct p232, AG Raunser database) will be described; other Cln3p constructs were purified along similar lines. EGFP-Cln3p was expressed in Sf21 cells at MOI of 10.

NOTE: *The purification is best if the membranes are intensively green, incase of deviation from norm (may occur due to various reasons passage number of cells, problems with virus) the protocol does not give best results and needs to modified spontaneously and diligently.*

Cells were harvested after 72 h, lysed and the membrane extracted following procedures described above. For small-scale purification, 3 mL of the membrane (in 50mM HEPES) was solubilized in the solubilization buffer in final volume of 15 mL for 3 hours at 4°C with gentle shaking. The solubilized sample was centrifuged at 55,000 rpm for 30 minutes at 4°C, the supernatant filtered (ø 0.2 µm) and added to 1.5 mL of (Ni-NTA or TALON) resin already calibrated with the wash buffer A-fos (with 0.05% Fos-Choline-13). 5 mM imidazole was also added to reduce non-specific binding and the 'resin-solubilized membrane' suspension was incubated with mild agitation at 4°C for 1h. The sample was packed in Econo-Pac chromatography column (BIO-RAD). The resin was washed with 10-20 column volumes (CV, 15 mL in this case) of washing buffer A and with 5-10 CV of washing buffer B (the volume of second wash was reduced if there was considerable amount of protein being leeched off the column). Protein was eluted with 5 CV of the elution buffer in to 0.5 mL fractions manually. Aliquots from the solubilized membrane fraction, flow-through, washing steps and eluted fractions were loaded on a semi-denaturing SDS-PAGE to follow the purification profile of the protein. Gels were scanned for fluorescence with the VersaDoc (mCherry: Filter: 605BP.: Illumination: Blue LED: Gain; 4x;

Binning: 4x4. GFP: Filter: 530BP.: Illumination: Blue LED: Gain; 4x; Binning: 2x2), Coomassie stained or subjected to western blotting (in cases where there were no fluorescent reporter tags).
Large scale purifications, followed similar lines where 5 mL of pre-packed IMAC columns were used, the loading and wash steps carried out using a peristaltic pump and the eluting protein was fractionated using ÄKTA purifier.

2.2.2.7 Detergent exchange

The only case where detergent exchange worked was when the detergent was exchanged from Fos-Choline-13 to MNG-3 on the IMAC column. Briefly, after the membranes were solubilized in Fos-Choline-13, filtered and loaded onto the metal-affinity resin/columns, the wash steps were carried out with a buffer containing 0.05% of MNG-3 (instead of Fos-13, called wash buffer-mng). Similarly, the elution buffer had Fos-Choline-13 substituted with MNG-3 (0.1 %). Everything else was maintained as before.

2.2.2.8 Tag cleavage

The cleavage of tag (EGFP-6x-His or just 6x-His) worked only in the case of constructs with large linker. Dennis Quentin optimized the protocol. The cleavage of tag was carried out in two different ways. One, the eluate (in MNG-3 buffer) was pooled, and a dialysis set up to get rid of the imidazole, overnight. The dialysis buffer was spiked with PreScission protease (generated by DPF, used 0.03-0.06 mg/ 1mg of purified Cln3p). The pooled protein was concentrated (Amicon MWCO: 50kDa) and subjected to size-exclusion chromatography (Superose 6 10/300 GL). Alternatively, elution buffer consisting of protease and no imidazole was used. The cleavage reaction was carried on the column. The flow through was collected and concentrated (Amicon MWCO: 50kDa) and subjected to size-exclusion chromatography (Superose 6 10/300 GL).
Both of the following methods led to precise cleavage of the tag, but best results were obtained with the latter.

2.2.2.9 Chromatography methods: Size exclusion and ion-exchange chromatography

All chromatography based purification methods were performed using the ÄKTA purifier at 4 °C. For size exclusion chromatography either Superose 6 10/30 GL or Superose 12 10/30 GL columns was used. The column was equilibrated with filtered (ø 0.2 µm) and degased SEC buffer. All the SEC buffers used for Cln3p contained detergent. After injecting the desired volume in the loading loop (usually between 100 µL and 1 mL), a preset program with a constant flow-rate of 0.3 mL/min was used. Fractions (0.5-1 mL) were collected automatically. The Adsorption at 280 nm and the fluorescence (detector RF-551, low sensitivity) were monitored (GFP: Ex. 420 nm; Em. 550 nm; mCherry: Ex. 520 nm; Em. 650 nm). The fractions corresponding to the appearing peaks were analyzed by semi-denaturing SDS-PAGE or Western blot.

For ion exchange chromatography Mono S column was used with flow rate between 0.3 and 0.5 mL/min. Before loading the elution fractions from IMAC, they were pooled and desalted to a final salt concentration of 50 mM using PD 10 desalting columns. The protocol for desalting was as per the associated manual (for PD 10 columns). Absorption at 280 nm and fluorescence were monitored similar to SEC chromatography. A salt gradient from 50 to 500 mM salt was used to elute the proteins. Peaks were analyzed either with semi denaturing SDS PAGE or with western blot.

2.2.2.10 Methods for analysis

2.2.2.10.1 Semi-denaturing SDS-PAGE

10-15 % gels were used for all the SDS-PAGE analysis. Before loading, samples were mixed with 4x semi-denaturing sample buffer and gels run with a standard protein weight marker at 150 V. Gels were scanned for in-gel fluorescence (if necessary, for proteins with fluorescent tags) and subsequently stained by Coomassie solution for 10 to 30 min. Excess dye was removed and gels destained by transferring into a 10 % acetic acid solution and was further incubated until solely the protein bands were visible.
In contrast to the normal SDS-PAGE, samples were not heated before loading and loading buffer had minimal amount of SDS to preserve the folding of fluorescent reporter tags (mCherry and GFP) and hence enabling scanning the in-gel fluorescence with the VersaDoc Imaging system. The main drawback of this technique is slight modification of migration behavior, it brings about, of partially folded proteins which makes it sometimes hard to predict the position of the protein, when subjected to biochemical analysis.

2.2.2.10.2 Blue-native PAGE

Samples were mixed with varying concentrations of SDS (0 – 2.5 %) and loaded with a 10x blue-native loading buffer onto a 4 – 16 % Gradient Gel. The gel was run overnight at 4 °C with constant current of 1 - 2 mA (usually 16 h). Following this, the gel was fixed for one hour in methanol containing fixation solution and subsequently destained with 10 % acetic acid (additional Coomassie staining was not necessary due to the Coomassie containing buffer).

2.2.2.10.3 Western blot analysis

Samples were run on a semi-denaturing SDS-PAGE as described above. Following the SDS-PAGE the gel was transferred into the western chamber, above a nitrocellulose membrane and sandwiched in-between the Whatman filter papers. The western blot was run for 1h at a constant current of 50 mA. Prior to this, membrane was activated by incubation in methanol for 1 min and Whatman filter papers were equilibrated in transfer buffer. Pre-stained protein marker served as an indicator for a successful transfer (alternatively, western blot marker was used at times). Subsequently, the membrane was

incubated with blocking solution for 1 h to prevent or minimize unspecific binding of the antibodies. After washing with TBS, the membrane was incubated with primary antibody solution for 1 h at RT (or overnight at 4 °C). Following this, the membrane was rinsed twice with TBS-TT and once with TBS, and incubated with the secondary AB for 1 h. After the incubation of membrane with secondary antibody, it was rinsed four times with TBS-TT and twice with TBS. All the washing steps were carried out for 10 minutes (rocking). Finally to visualize the bands 2 mL BCIP/NBT solution was added. The colorimetric reaction initiated in the previous step was stopped by addition of water. The blot was scanned with a flatbed-scanner.

2.2.2.10.4 Estimation of the concentration of the protein

Protein concentrations were estimated with the Bradford Assay. Briefly, 1 µL of protein solution was mixed with 49 µL of ddH_2O and 150 µL Bradford solution to a final volume of 200 µL in a 96-well microtiter plate. Buffer without protein served as a blank. Samples were thoroughly mixed and the absorbance at 595 nm was measured on a Polarstar Omega Multititer Plate Reader. The final concentration of the protein was estimated by back-calculation from a calibration curve plotted using bovine serum albumin (BSA) as standard.

2.2.2.11 HPLC-MS analysis of Cln3p

Characteristic protein bands were excised using a scalpel and chopped into small pieces. After, brief incubation in 50 µL acetonitrile for 10 min, acetonitrile was removed and gel pieces were re-hydrated in 50 µL of 0.1 M NH_4HCO_3 for 5 min. 50 µL of acetonitrile was added again and pieces incubated for 15 min. The supernatant was discarded and gel pieces dried in the SpeedVac for 30 min. 50 µL of 10 mM DTT/0.1 M NH_4HCO_3 was added to the reaction tube and incubated at 56 °C for 45 min, followed by subsequent cooling of the mixture to RT. Gel pieces were treated again with 50 µL acetonitrile for 5 min and after removal of supernatant, 50 µL of 55 mM iodoacteamide/0.1 M NH_4HCO_3 was added and the tubes incubated in the dark for 15 min. Following steps were repeated again: Removal of supernatant, treatment with 50 µL acetonitrile (5 min), removal of supernatant and drying of gel pieces in the SpeedVac for 30 min. 202 µL digestion buffer was added to the dried pieces. After 45 min incubation on ice, the supernatant was removed and 200 µL of digestion buffer without trypsin was added and further incubated overnight at 37 °C.

On the following day, reaction tubes were centrifuged for 1 min at 13,000 rpm, and the supernatant transferred into a fresh tube followed by addition of 20 µL 0.1 M NH_4HCO_3 and a 15 min-long incubation. Another 20 µL 0.1 M NH_4HCO_3 was added followed by additional 15 min incubation. After drying of the isolated peptides for 60 min in SpeedVac, they were stored in a freezer, till further use. Peptides were re-suspended in 100 µL 0.1 % formic acid and 30 µL was transferred to a HPLC micro vial for HPLC-MS analysis, which was performed by Dr. Petra Janning (Department of Chemical Biology, MPI).

2.2.2.12 PNGase assay

To check for protein glycosylation (which may interfere in various structural studies, at times) and to obtain a homogenous protein population, the asparagine aminase PNGase F is commonly used. Protocol as specified by the manufacturer was followed. The results were analyzed on a semi-denaturing SDS-PAGE.

2.2.2.13 γ-phosphatase assay

Briefly, for 100 µg of protein 60 µL of γ-phosphatase solution was used. Reactions were set up and the products separated via semi-denaturing SDS-PAGE. Subsequently, the gel was incubated with 100 mL of freshly prepared fixation solution for 1 h at RT. This step was repeated followed by rinsing thrice with ddH_2O (10 min). The gel was stained for 90 min in 20 mL Pro-Q Diamond phosphostain solution in the dark followed by incubation with 100 mL destaining solution for 30 min in the dark, for destaining. The destaining step was repeated twice followed by wash with 5 min ddH_2O. The gel was imaged under the Typhoon Scanner. An excitation wavelength of 532 nm, an emission wavelength of 580 nm and a scan width of 50 µm with a sensitivity of 300 or 600 were used. Coomassie staining followed imaging.

Table 2.16 γ-phosphatase reaction

	Cleaved		Uncleaved		BSA	
	conc.: 0.8 mg/mL		conc.: 0.8 mg/mL		conc.: 10 mg/mL	
	+	-	+	-	+	-
protein	50	50	12,5	12,5	1	1
Reaction buffer 10x	10	10	10	10	10	10
$MnCl_2$ (100 mM)	10	10	10	10	10	10
γ-phosphatase	24	0	6	0	6	0
ddH_20	6	30	61.5	67.5	73	79
	100 µL	100 µL	100 µL	100 µL	100 µL	100 µL

2.2.2.14 2D crystallization trials

Two-dimensional crystallization trials were set-up with a protein concentration of 1 mg/mL. Lipid stocks were prepared at 10 mg/mL and dialysis membranes (Spectra/Por® 4 - MWCO 12 000 -14 000) were cut to appropriate size. The lipid (dissolved in chloroform) was transferred to a clean reaction tube and slowly dried under a stream of argon to get rid of all the chloroform. This was followed by addition of protein (always 60 µL, 1 mg/mL as final volume) mixing it gingerly. Protein and lipids were mixed in ratios (w/w), designated as Lipid-Protein-Ratios (LPR). LPRs ranging from 0.1 – 1 were used. After 1h, of incubation in the dark, the protein-lipid mixture was carefully transferred into a dialysis button and sealed with a square piece of membrane (for good dialysis, it is important to avoid air

bubbles). The buttons were placed in a dialysis buffer that lacked detergent and were dialyzed for 1-2 weeks. Sodium azide was added to all the dialysis buffers to prevent bacterial growth.
The samples were prepared and analyzed by negative stain electron microscopy as described in **section 2.2.1.7.1.**

2.2.2.15 3D crystallization trials

Various crystallization suites (also called *screens*) from the manufacturer Qiagen were used: JCSG Core I-IV, JCSG+, MbClass I, MbClass II, pHclear I, pHclear II, Classics I, Classics II, PEGs, PEGs II, PACT. All the screens were set up in 96 well crystallization plates (Corning 3550). The compositions of these suites rely on statistical analysis, aimed at achieving the highest probability of nucleation and crystal growth by interweaving various pH values, salts, precipitants and additives in varying concentrations. The precipitant solution (70 µL) was transferred into the wells semi-automatically, followed by automated pipetting of100 nL of the protein solution (centrifuged to remove aggregates, dispensed using Mosquito robot). 100 nL of the reservoir solution was also pipetted automatically into the wells containing protein solution. Subsequently, the plate was sealed to prevent evaporation and to maintain a closed system establishing equilibrium between fluid and gas phases, crucial for nucleation. The sealed plates were transferred in an automated imaging system (Formulatrix) and were monitored for crystal growth at preset time intervals. Trials were also setup in different Formulatrix systems operating at different temperatures: One at 4 °C, the other at 20 °C. Protein concentration (ranging between 1 and 10 mg/mL) was used.

2.2.2.16 Fluorescence microscopy

2.2.2.16.1 Cell culture

For all experiments Neuronal-N2a (generous gift from Dr. Leif Dehmelt) cells were used. The cells were cultured in the standard mammalian cell growth medium at 37°C in a humidified incubator with 5% (v/v) CO_2 saturation. In order to maintain the cells in culture, they were kept in a 75 cm^2 cell culture flask (BD Falcon) till they reached 90% confluence. At this density, the cells were passaged. The old growth medium was removed from the flask and the cell layer was briefly rinsed in sterile 1x PBS (pH 7.5). In order to detach the cells from the surface 2 mL of trypsin/EDTA solution (PAN Biotech) was added and the cell suspension was incubated for 1-2 minutes at room temperature. Adding 10 mL of complete growth medium containing 10% FCS inactivated the trypsin. The resulting cell suspension was gently pipetted up and down several times such that no cell clusters remained in the suspension. 1 mL of the suspension was transferred to a new flask containing 12 mL fresh standard mammalian cell growth medium. Following this protocol a 90% confluence was reached every 2-3 days. Cells growing at that speed are generally considered healthy for experiments. Cells beyond passage 20 were generally not used any further.

2.2.2.16.2 Fluorescence imaging

For fluorescence imaging, cells were seeded into a 6-well cell culture insert companion plate with lid (BD Falcon) containing cover slips. In each well, 2 mL of standard growth medium was added containing $0.5x10^6$ N2a-cells. The cells were incubated for 24 h to attach to the glass cover slip. After attachment to the surface, transfection was carried out using FuGene (Roche). The transfection was done according the manufacturer's protocol using a 1µL FuGene/0.5 µg DNA ratio. Prior to transfection the media was removed and 1 mL of fresh growth medium without FBS and antibiotics was added.
The cells were fixed using 2 mL of a 3.6% PFA (w/v) solution for each well and incubated for 20 minutes at room temperature. Following this, cells were rinsed with PBS and the glass slip was mounted with Fluoromount (Serva) upside down onto a glass slide (Thermo Fisher Scientific). The slides were stored in the dark at 4°C to avoid any damage done by light (photobleaching) or heat, till further use.

2.2.2.16.3 Microscopy and analysis

Leica SP2 Confocal System (Leica) was used to visualize the fluorescence of transfected N2a cells. The 63× Plan-Apochromat (numerical aperture, 1.4) oil immersion lens objective was used. GFP was imaged with 488 nm wavelength laser excitation and 505~530 nm band pass emission filter. RFP was imaged with 558 nm laser excitation and 580~650 nm filter. A z-stack of the cells was acquired by moving the focal plane from the bottom to the top of the images cells. The pinhole was set to 1 airy to ensure that the acquired image shows a thin plane of the specimen.
For the processing of images the software ImageJ was used.

2.2.2.17 Synthesis of N-acetyl palmitoyl cysteine

500 mg of N-acetyl cysteine was weighed and added to a two-necked round-bottomed flask (RBF) under constant argon pressure. 25 mL of distilled Dichloromethane was added to RBF, followed by addition of Trimethylsilyl Chloride (TmsCl, 428 µL) and the reaction heated to reflux for 2 h. The reaction was cooled to RT and Palmitoyl Chloride was added to the RBF, following this Tri-ethylamine (TEA) was added drop wise for about 2h and the reaction stirred for another hour. The products were analyzed by TLC (Methanol:DCM, 1:19) and separated by flash chromatography and verified by NMR and mass-spectrometry analysis.

2.2.2.18 Desaturase assay

15 nmol of the substrate (*Palmitoyl chloride/Palmitoylated cysteine)* was dissolved in Chloroform:Methanol (2:1, v/v), dried under a stream of nitrogen and re-dissolved in 10 µL of ethanol. 30 nmoles of Bovine serum albumin (BSA, free from fatty acids, SIGMA) is dissolved in 90 µL of

phosphate buffer (pH 7.4, 100 mM) and added to the substrate (in ethanol) under gentle stirring. Solubilized membrane (isolated from insect cells, expressing Cln3p) was added to the substrate and thoroughly vortexed (total volume of the reaction was 300 µL, protein used was 300 µg). After 5 min of pre-incubation at 37°C the reaction was started by addition of 1 µmol of NADH (dissolved in 30 µL of phosphate buffer). The reaction was carried out at 37°C for 1h with gentle agitation. 100 µL of 5M Sodium Hydroxide (NaOH) was added and the reaction terminated by heating for 30 min at 65°C.

The products were extracted into an organic phase, by addition of 300 µL of chloroform. The internal standard (1 nM) was also added at this point. Samples were vortexed for 1min and centrifuged for 5 min, 14000 *g*. The organic phase was extracted and the aqueous phase washed with 250 µL of chloroform. Both the chloroform phases were evaporated under nitrogen and the residue derivatized with Bis-(trimethylsilyl)trifluoroacetamide (BMSTFA) for 1h at RT. 25 µL of Ethyl Acetate (EA) was added and samples were ready for GC-MS analysis.

3 Results and discussion

3.1 Structure of the ROD-Zw10-Zwilch (RZZ) complex

3.1.1 Negative staining and 2D alignment and classification

ROD, Zw10, and Zwilch are proteins that co-localize at the kinetochore during mitosis. They were identified and shown to associate as a complex with molecular weight of approximately 800 kDa (twice the combined size of each monomer), leading to the speculation that they occur and function as a dimer [81]. Re-constitution of the complex, *in vitro,* was deemed impossible for a long time due to inaccessibility of purified ROD. This problem was tackled by simultaneous co-infection of all the baculovirus for insect cell expression (all the purification protocols were established and carried out by Anika Altenfeld or Sabine Wohlgemuth from the department of mechanistic cell biology, MPI Dortmund). I monitored the end product after each round of purification using negative stain electron microscopy (EM), where the sample is stained by a heavy metal salt and viewed under a transmission electron microscope. Ultimately, the complex was purified to near homogeneity, size-exclusion profile of which is shown in **Figure 3.1, a** and the corresponding SDS-PAGE profile in **Figure 3.1, b.** Negative stain EM reveals RZZ complex as an elongated, slender, 'worm-like' particle as shown in the raw micrograph area in **Figure 3.1, c.** I collected micrographs under low-dose conditions and selected single particles (using the 'Boxer' tool in EMAN). I then performed 2D reference free alignment and *k-means* classification on a dataset consisting of 4000 such RZZ particles, representative classes of which are shown in **Figure 3.1, d**. The RZZ complex is approximately 40 nm long and 10 nm wide with an inherent two-fold axis of symmetry. *k-means* classification procedure gave classes with predominant side views. To authenticate the averages from *k*-means I also performed Iterative stable alignment and clustering (ISAC) analysis on the same dataset. ISAC is an alternative to *k*-means clustering and generally helps in isolating a homogeneous subset of images [162]. The results obtained from ISAC analysis corroborated that of *k*-means classification (shown in **Figure 3.1, e)**. In essence, 2D negative stain analysis of RZZ complex reveals it to be a rigid, elongated and symmetrical molecule.

3.1.2 Generation of initial model-RCT

The images obtained from the transmission electron microscope are 2D projections, of the object/specimen, on the image plane. If the Euler angles of each projection are known, then one can generate the 3D reconstruction, invoking the central section theorem (alternatively known as the projection slice theorem[161]). There are two different ways to identify the Euler angles: common-lines based approach, which requires the sample to take all orientations on the grid and the tilt approach, usually applicable for samples with preferred orientations. Random conical tilt method falls into the latter of the categories described above, where two single exposure images are collected at low-dose at two different angles. The 3D reconstruction is generated by filtered back-projection[163].

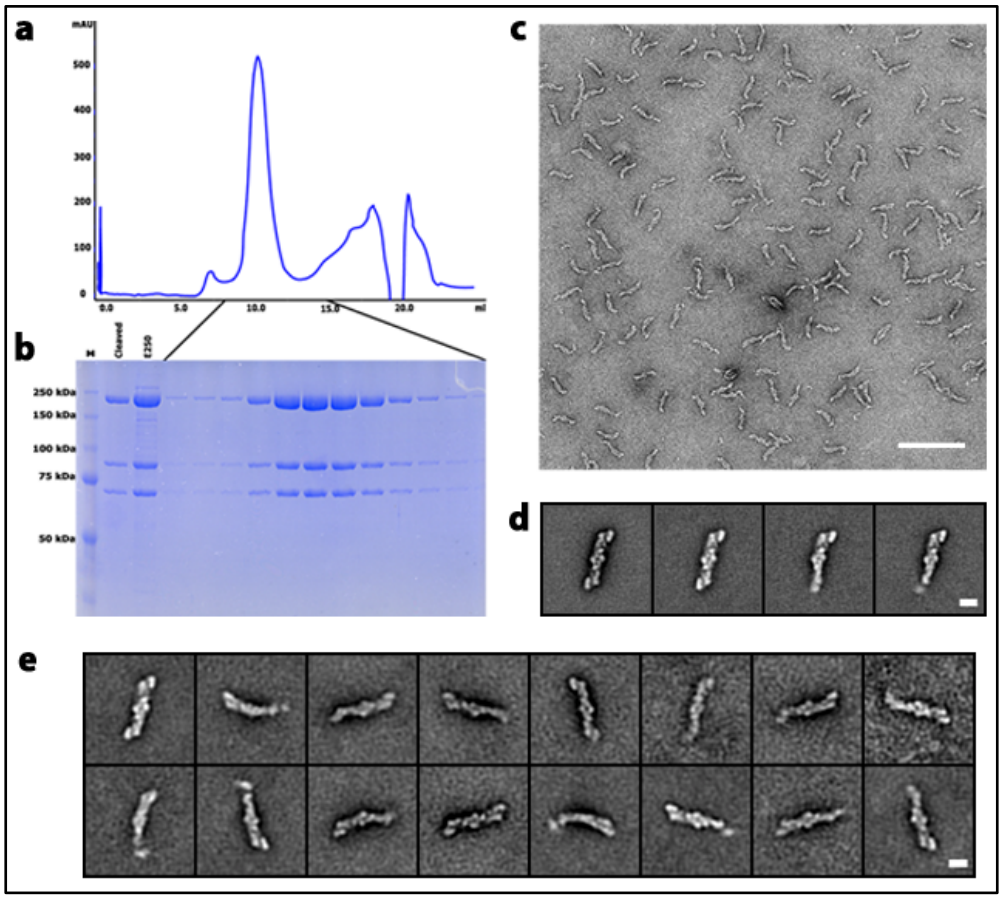


Figure 3.1: Purification and negative stain analysis of RZZ complex (a) Elution profile of RZZ complex from size exclusion chromatography Superose 6 10/300 column. The complex elutes around 10 mL as a mono disperse peak. **(b)** SDS-PAGE profile of the fractions encompassed by the peak in **(a).** Three bands corresponding exactly to the theoretical size of ROD (240 kDa), Zw10 (84 kDa) and Zwilch (70 kDa) are seen. **(c)** Raw micrograph area of negatively stained RZZ complex showing thin and elongated particles. Scale bar, 100 nm. **(d)** Representative class averages of negatively stained RZZ complex obtained from reference free alignment and *k*-means classification. The complex has a clear two-fold symmetry with a center of inversion. Scale bar, 10 nm. **(e)** Representative class averages of negatively stained RZZ complex when processed using iterative stable alignment and clustering (ISAC). Scale bar, 10 nm. The same dataset was used, both for *k*-means and ISAC classification methods. Source **(a)** and **(b):** Anika Altenfeld.

I collected images at angles of 55° and 0° and selected 1700 particle pairs using the e2RCTboxer program embedded in the EMAN2 suite. I used the 0° particle dataset for reference free alignment and *k*-means classification (generated around 8 classes). I chose the class average with best features and generated a 3D reconstruction by filtered back projection (using SPARX) of particles corresponding to the same class from the tilt (55°) dataset. Once I had the back projection I refined it against the un-tilted, 0°, dataset. The back-projection refinement was performed without imposing any symmetry restrictions and the reconstruction showed a clear two-fold symmetry, cementing the fact that RZZ has C2 symmetry (**Figure S1**). I further performed refinement runs imposing C2 symmetry and this yielded us a low-resolution 3D structure. The resolution was estimated by the $FSC_{0.5}$ criterion to be 28 Å**.**

3.1.3 Cryo-positive stain

Negative stain EM is limited in structure determination on the basis of the resolution that one can achieve. The stain solution, which forms a thin layer over the specimen on the grid**,** limits the resolution of the reconstructions to 20 Å. One generally turns to cryo-EM for high-resolution information. Initial attempts to visualize RZZ in cryo conditions using holey carbon grids suffered from major setbacks, as the RZZ complex would prefer to clump on the carbon rather than being in the ice (**data not shown)**. An alternative to such situations is to use holey carbon grids that are coated with an additional layer of carbon[164]. Cryo-negative staining is in essence a thin film vitrification technique where the biological sample comes in contact with the stain for a very short time. This results in exceptionally brilliant contrast of the sample (very high signal to noise ratio). I tweaked the conventional cryo-negative staining procedure in an effort to counter the technically challenging procedure of deposition of an additional carbon layer on the grid. Briefly, I incubated the sample with stain solution and manually blotted it before the grid was plunged into liquid ethane. I was able to record images where individual particles had remarkable contrast. Fascinated by the quality of images I performed preliminary 2D analysis on a dataset consisting of 4880 particles. Representative averages from reference free alignment and classification (*k*-means) are highlighted in the figure below **(Figure 3.2, a)**.

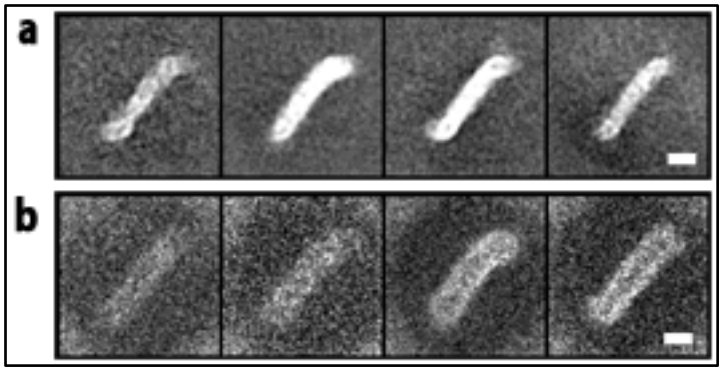


Figure 3.2: Cryo-positive stain (a) Averages from reference free alignment and *k*-means classification. Scale bar, 10 nm. **(b)** Variances of the averages in figure (a) highlighting significant density around the averages, generally arising due to inefficiency in averaging procedures. The reason for such an effect can be explained invoking the concept of positive staining, where it is not uniformly distributed in the sample and therefore leads to blurred averages.

The averages look significantly blurred and featureless. I repeated the analysis couple of times to avoid systematic errors. The results of 2D alignment and classification were disheartening. A closer look at the variances from the *k*-means procedure **(Figure 3.2, b)** led to speculation that the stain was not uniformly distributed on the specimen (negative stain by definition), but was absorbed into the sample and distributed randomly leading to cloudy averages. In essence, my attempt to devise an alternative to the classical cryo-negative stain led to cryo-positive stain instead. Since averages were deemed useless, I did not venture any further with this method.

3.1.4 Cryo-EM

3.1.4.1 2D analysis

I channelized efforts to optimize the ice conditions for imaging RZZ. During the course of this optimization I realized that detergent was absolutely essential to avoid aggregation of RZZ, hinting at the fact that some regions of the structure may be highly hydrophobic. The use of detergent (0.02% Triton X-100 in this case) also ensures uniform thickness of ice on the grid. I also found that much higher concentration of the complex was necessary (approximately 3 mg/mL) as opposed to standard use of 0.1 mg/mL. Another parameter that needed calibration was the ice thickness. Since the JEOL 3200FSC is equipped with an energy filter, I was able to calculate the thickness of ice and choose an optimum value for imaging RZZ.

I recorded images of RZZ at a fixed defocus (3 µm) and in holes with varying thickness of vitreous ice (12-80 nm). I estimated that ice with a thickness of 40-50 nm is best suited for visualizing RZZ complex. Having set all the parameters for data collection right I collected a test dataset of 780 images, from which I manually selected 5997 particles and subjected them to 2D alignment and classification procedures. *k*-means classification gave class averages with distinct features. I also found RZZ taking all orientations in ice.

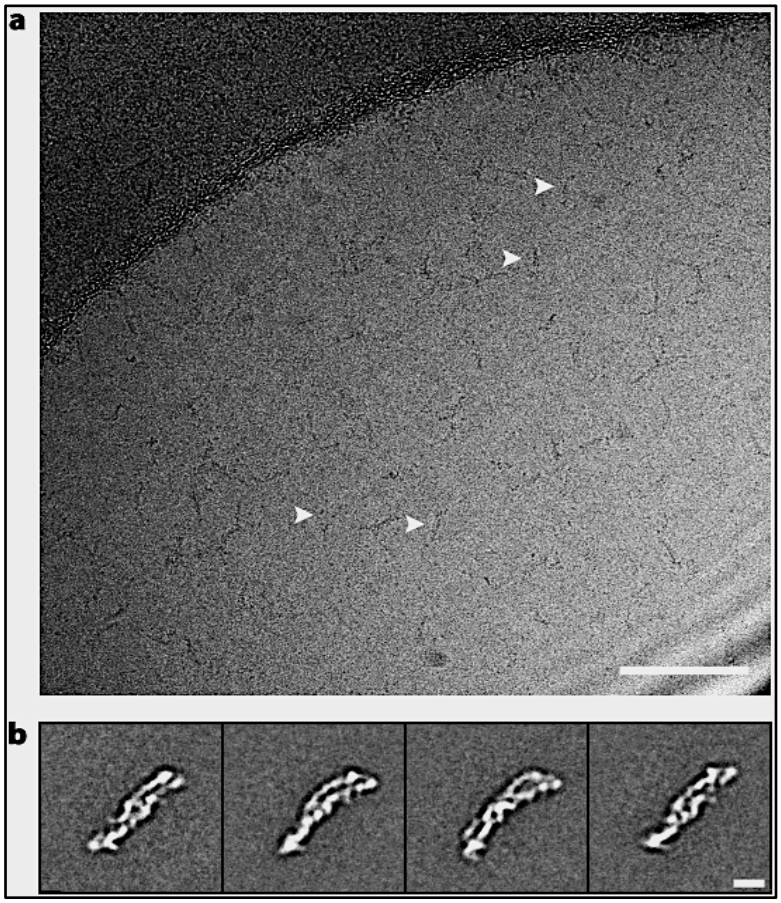


Figure 3.3: Cryo-EM on RZZ complex **(a)** Representative micrograph area of RZZ on holey-carbon grids imaged in cryo conditions. White arrowheads highlight particles. The image was acquired with a JEOL3200FSC microscope equipped with TVIPS 8k x 8k CMOS camera. The sample contained 0.02% Triton X-100 for uniform ice thickness and to prevent RZZ from aggregating. Scale bar, 100 nm. **(b)** Representative class averages as obtained from reference free alignment and *k*-means classification procedures of particles obtained from 5997 cryo-EM images.

Encouraged by the initial results I collected 3300 images with defocus value varying between 2 and 3 µm. A representative micrograph area of RZZ imaged in cryo conditions **(Figure 3.3, a)** and averages that were outcome of 2D analysis are shown **(Figure 3.3, b).**

Correct 3D structure determination relies on the precise estimation and the correction of the contrast transfer function (CTF). The contrast transfer function is defined as Fourier transform of the point spread function (PSF), which the electron microscope (or any optical imaging device) associates with the object during image formation. CTF depends on factors like defocus, spherical aberration coefficient and the wavelength of incident electrons (which depends on the voltage at which the microscope is operated). In CTF correction, one tries to retrieve undistorted object from the image and this process is absolutely crucial for determination of high-resolution structures by cryo-EM. I used customized graphical user interface, built for a program which parallels CTFFIND3 [165], designed in the lab (by Dr. Rouslan Efremov and Dr. Christos Gatsogiannis) to estimate the CTF parameters. The program generates a 1-dimensional fit of the CTF as a function of spatial frequency as shown in **Figure 3.4**. I discarded images where the estimation of CTF was not accurate and also images with drift and astigmatism. Particles were not boxed out from such images. I selected, manually, 35404 particles from 4000 8k x 8k images.

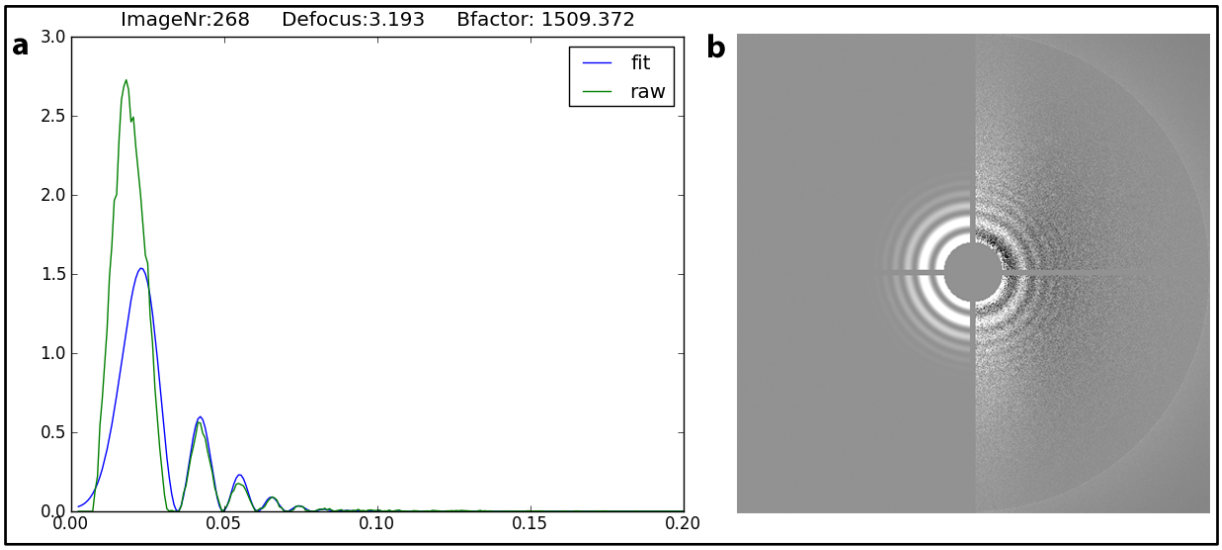


Figure 3.4. CTF fit (a) Plot of the radial average of the average background-subtracted power spectrum (green line) and the fit CTF (blue line). B factor is a coefficient in the Gaussian function describing CTF decay with resolution as exp (-B*q^2), where q is a spatial frequency. Higher B corresponds to faster decay and your high resolution vanishing faster. CTF fitted = CTF*exp (-B*q^2). **(b)** The diagnostic output image for image number 268, showing the average background-subtracted power spectrum on the right, and with the fit CTF on the left.

3.1.4.2 3D analysis

High-resolution structure determination requires very careful and precise analysis of cryo-EM data. I followed the common practice of using the 3D structure obtained from RCT, as an initial reference for the subsequent refinement runs using the cryo-EM data. Since, experimental cryo-EM images are extremely noisy, structures determined by such data may be prone to overfitting, which in mathematical terms is an 'ill-posed' problem. One way to prevent overfitting of cryo-EM

reconstructions is imposing smoothness by using filtering procedures. I applied a low-pass Gaussian filter (70 Å) to the negative stain reconstruction and used it as the starting reference. In brief, I used the projection-matching algorithm implemented in SPARX [166] for 3D refinement runs in an iterative approach to center and orient individual particles against the reference density map. For the first round, I started with a small dataset of approximately 6000 particles and with large pixel values for the translational search parameters along X and Y directions. Also, relatively large angular step value, delta, was used. The resulting 3D density map was used to generate a mask and as a reference for subsequent iterations. I added more particles with each iteration and fine-tuned the alignment parameters until there was no further improvement in resolution (**Figure S3**).

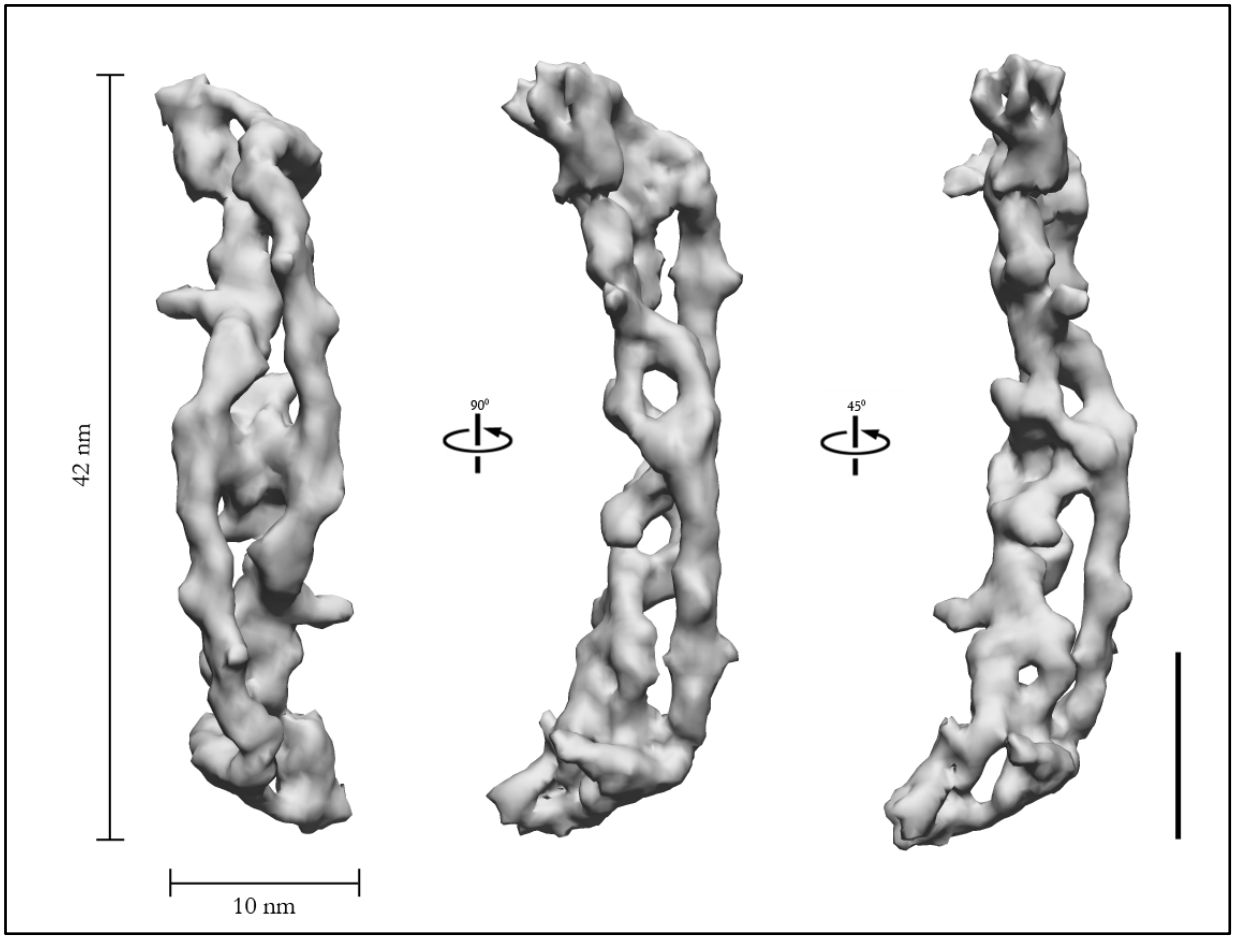


Figure 3.5: Representative views of the 3D structure of the RZZ complex. Scale bar, 10 nm.

RZZ appears as a long, hollow and symmetrical rigid structure with a center of inversion **(Figure 3.5)**. End to end length of the complex is 42 nm and the width is 10 nm. The high-resolution structure, confirms the biochemical finding that RZZ is a dimer. At this point I was perplexed, as I did not know the organization of proteins in the complex and hence, the nature of dimerization. The 3D structure of RZZ depicted above could be an outcome of dimerization, of ROD-Zw10-Zwilch, in one of the following two ways: as a longitudinal monomer, where ROD-Zw10-Zwilch interact along the length of RZZ complex and the complex would therefore be an antiparallel dimer, or as a lateral monomer, where

the interactions between the two monomers is predominantly along the lateral axis. Both the possibilities are highlighted in **Figure 3.7, a** and **b.** A detailed explanation on the nature of dimerization follows later.

Since the resolution of the cryo-EM reconstruction is inversely related to the defocus, images recorded close to focus are expected to contribute towards better resolution. However, images close to focus are deprived of contrast whatsoever. To circumvent this problem I recorded a 'focal-pair' dataset, wherein two images of the same area are recorded, the first close to focus and the latter far away from focus. I used the images far away from focus (higher defocus values) for boxing out the particles and used these co-ordinates to box out the particles from 'close to focus' images. I did not see any significant improvement in the resolution of the structure, as reflected by FSC, during subsequent refinement runs with addition of close to focus particles to the dataset. The resolution of the final 3D reconstruction estimated by $FSC_{0.5}$ criteria is 13.4 Å **(Figure 3.6)**.

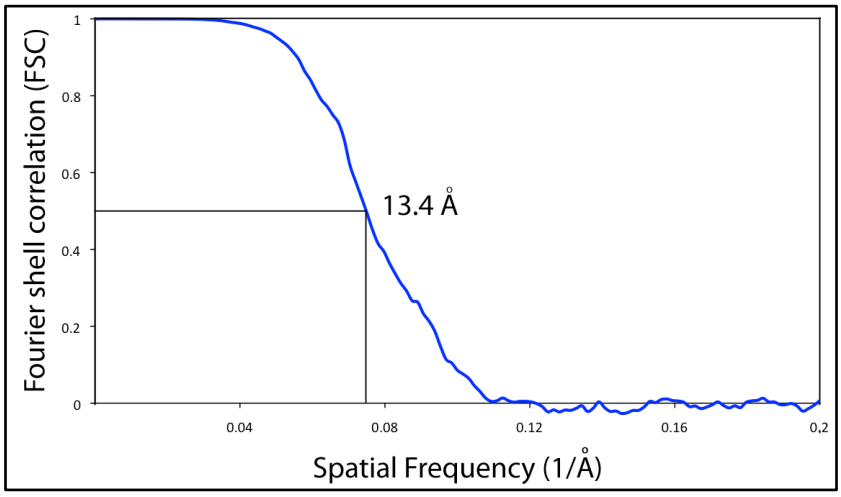


Figure 3.6: Fourier shell correlation. Resolution of reconstructions from electron microscopy is estimated from the Fourier shell correlation (FSC) curves, where FSC is plotted as a function of spatial frequency. The most widely accepted criterion is the $FSC_{0.5}$. The structure by the same criteria stands at 13.4 Å.

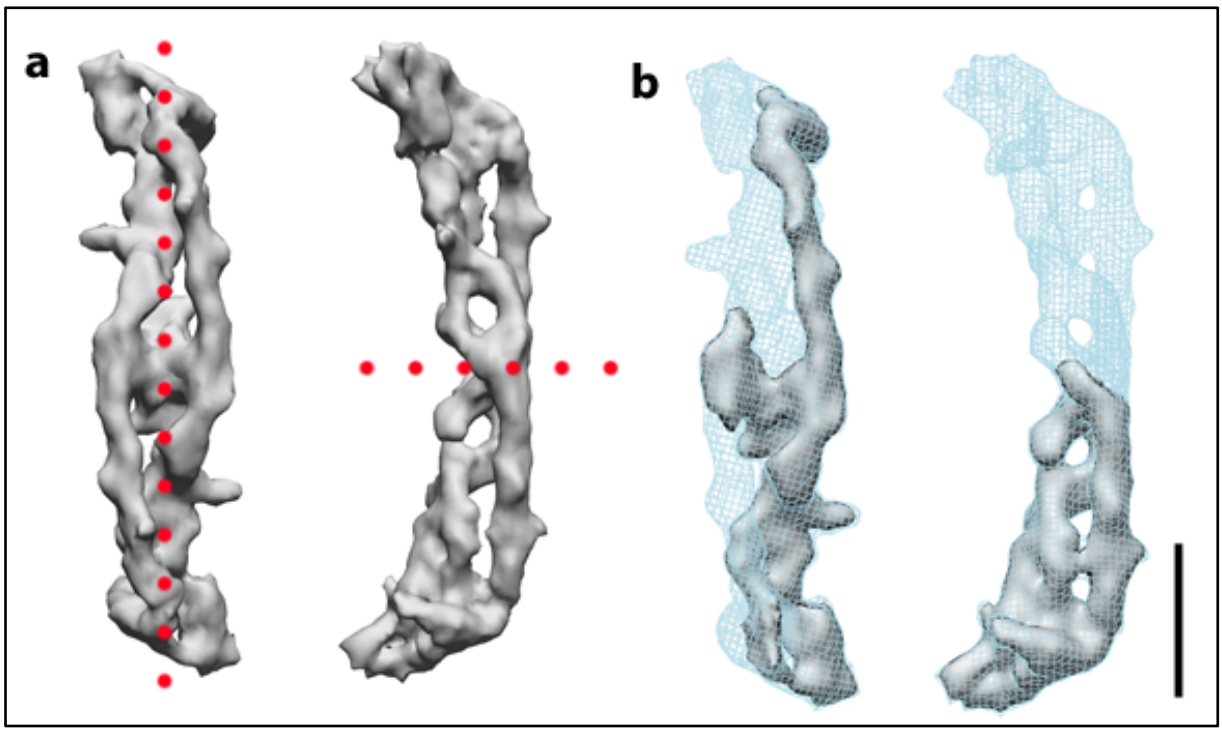


Figure 3.7: Two possible ways of dimerization (a) The C2 symmetry axis is highlighted in dotted red lines. RZZ complex can be a longitudinal dimer or a lateral dimer, depending on how the interface is formed. **(b)** The two possible monomers are highlighted in gray with the complete RZZ density superimposed as blue mesh. Scale bar, 10 nm.

3.1.4.3 Cross-correlation between projections and re-projections

To verify the structure at hand, I calculated cross-correlation between 2D projections (class averages from *k*-means) and the re-projections of the density. I find a high cross-correlation value of about 73% (very good value). Representative projections and re-projections, befitting the views of the depicted structure in **Figure 3.5**, are highlighted in **Figure 3.8**.

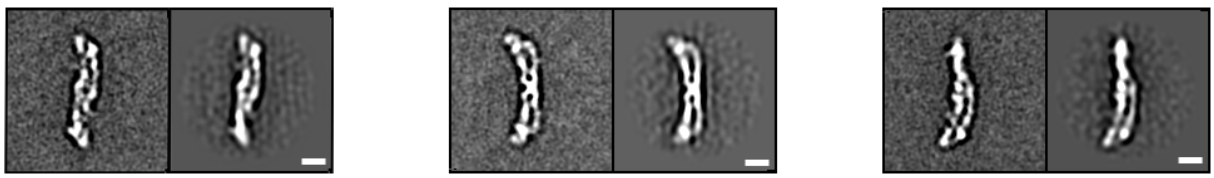


Figure 3.8: Representative projections (left), re-projections (right) pairs. The re-projections were generated using the orientation parameters from the projections. Therefore this is not a perfect method for validating the structure. Nevertheless, the cross-correlation gives a good estimate of how good the structure is. The views represented here are similar to the views of the structure in Figure 3.5. Scale bar, 10 nm.

3.1.5 Topology of RZZ

3.1.5.1 Antibody labeling

I addressed the issue of the organization of proteins in the complex by interweaving biochemical and microscopy methods. First, I targeted localization of ROD by using a polyclonal antibody directed towards its N-terminus where I followed a quick and dirty procedure of incubating the antibody with the complex and immediately subjecting to negative stain EM. I found very few antibody labeled RZZ molecules. One would expect two antibodies per RZZ dimer, but such complexes were seldom found. Also, since the antibody could adopt various orientations, alignment and classification was of not much use. It was to my great advantage that I could spot the antibody labeled RZZ in raw micrograph **(Figure 3.9, c)**.

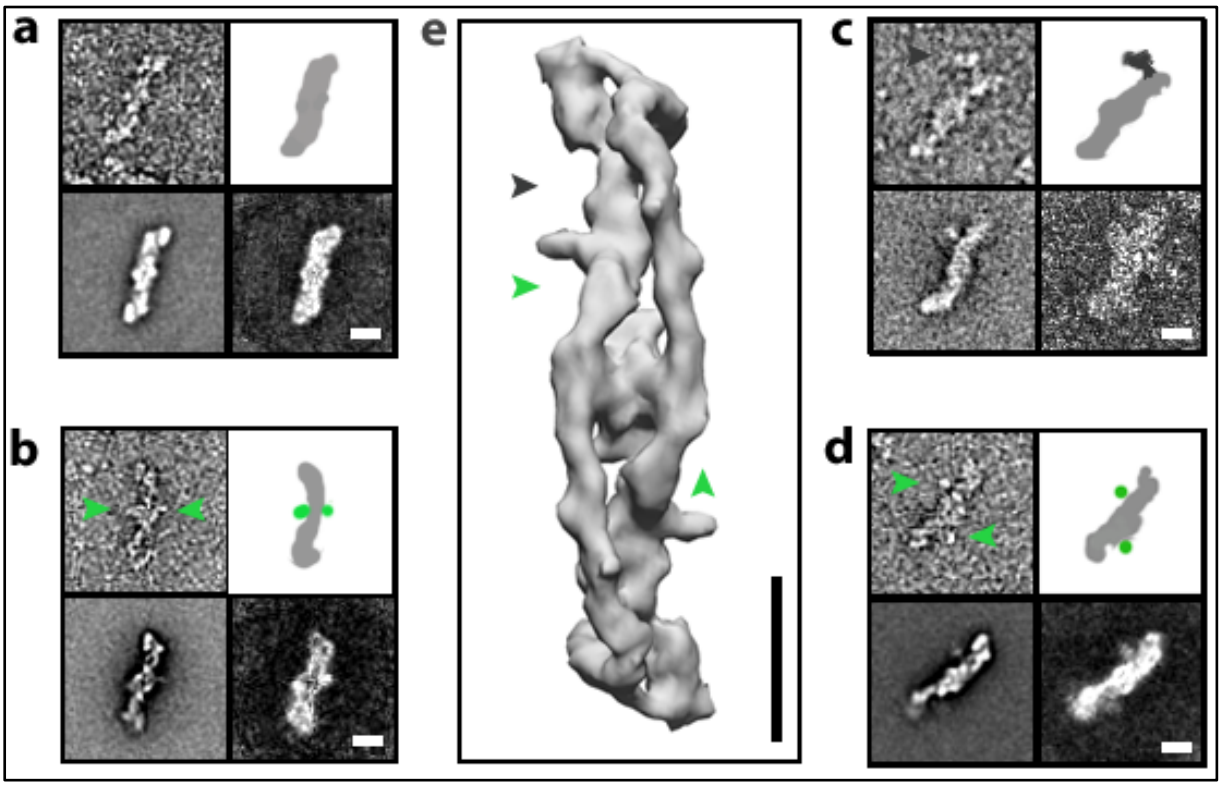


Figure 3.9: Antibody labeling studies on the RZZ complex. Each of the panels in **a, b, c, d** consist of raw particle (top left), a cartoon of the raw particle (top right), representative average from *k*-means classification (bottom left) and the corresponding variance (bottom right). **(a)** RZZ complex **(b)** R-Z(cGFP)-Z, complex where Zw10 is expressed as fusion protein with GFP at its C-terminus **(c)** RZZ in complex with a polyclonal antibody directed towards the N-terminus of the ROD. **(d)** R-(nGFP)Z-Z, complex where Zw10 is expressed as fusion protein with GFP at its N-terminus. **(e)** The loci, as understood from the analysis, mapped onto the RZZ structure. Green arrowheads represent Zw10 and the dark gray arrowhead represents ROD. Scale bars, 10 nm.

Next, I used fusion proteins (generated and purified by Sabine Wohlgemuth, AG Musacchio) of RZZ, where GFP was fused either to the N- or the C- terminus of Zw10 (R-(nGFP)Z-Z or R-Z(cGFP)-Z). I analyzed these complexes by negative stain electron microscopy. I ran into a similar problem as that of antibody labeled RZZ. Though clearly evident from raw images, where the GFP was located, 2D

alignment and classification did not give spectacular results. The main reason might be that since I used polyclonal antibodies and since their binding is not absolutely specific, the averaging procedure did not yield sufficiently good results. I mapped the additional electron density, however, from raw images and variances. I found that Zw10 localizes centrally in the complex, with N-terminus being farther away from the center of the complex **(Figure 3.9, d)**. The density corresponding to GFP in R-Z(cGFP)-Z is more central **(Figure 3.9, b)**. I also found that the N-termini of ROD and N-terminus of Zw10 are proximal in space **(Figure 3.9, e).**

3.1.5.2 hZw10

During the course of investigating the structure of RZZ, I was able to get some human Zw10 protein (purified to near homogeneity by baculovirus expression from insect cells by Anika Altenfeld, AG Musacchio). I performed negative stain EM analysis on the purified Zw10. The averages from *k*-means classification did not have any characteristic shape and were not well resolved **(Figure 3.10, a)**. This was also evident from a look at the variances of the corresponding averages, which revealed the presence of extra density not seen in the 2D averages **(Figure 3.10, b)**. I therefore, performed an Iterative stable alignment and clustering (ISAC) analysis on the same dataset. I discovered that hZw10 comprises of two distinct densities, which hinge upon their interface leading to various conformations in solution as reflected from the averages in **(Figure 3.10, c).** Structural prediction confirms an unstructured region nearly at the center of protein (refer **Figure S9).** Invoking the GFP-fusion labeling studies, I postulate that Zw10 adopts a variety of conformations when in solution, but selectively adopts a particular conformation as a part of RZZ complex. I believe that Zw10 acts as a belt that holds the two anti-parallel ROD molecules together. These results provide first structural insights into the structure of human Zw10.

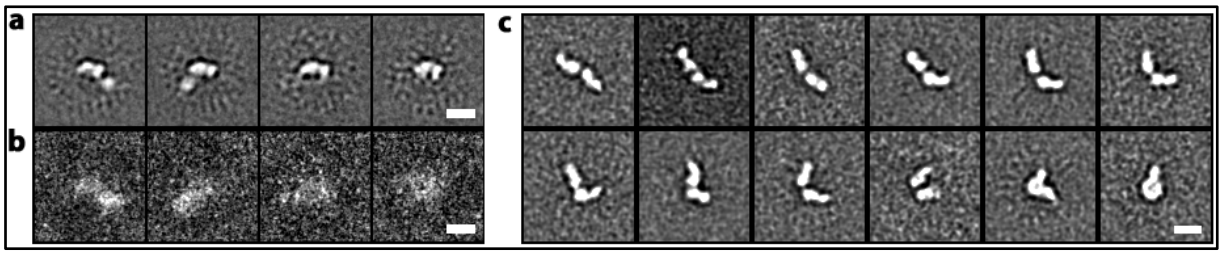


Figure 3.10: Negative stain analysis of Zw10. **(a)** Representative averages of Zw10 as obtained from the *k*-means classification procedure. The complex appeared to be of different sizes in the first look, but the variances highlighted in **(b)** revealed it as a problem surfacing from the *k*-means classification. **(c)** Representative class averages from the ISAC procedure, revealing the inherent flexibility of the Zw10 complex. Scale bar, 10 nm.

3.1.5.3 Crosslinking and Mass spectrometry

Chemical cross-linking coupled with mass spectrometry (XL-MS) presents a low-resolution structural technique for determination of topology of multi-protein complexes [167, 168]. XL-MS taps spatially close lysine residues, which get covalently modified by a bi-functional cross-linking reagent (**Figure**

3.11). With the structure of RZZ in hand this approach was used to identify the subunit interactions and the architecture of the complex. These experiments were performed in the laboratory of Dr. Franz Herzog at the LMU, Munich.

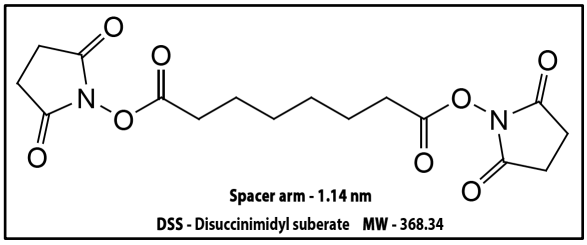


Figure 3.11:
The cross linking reagent used for XL-MS studies was disuccinimidyl suberate (DSS), which has a spacer length (8 atoms) of 11.4 Å.

The analysis identified 106 unique cross-links, which I categorized into two classes: inter (ROD-Zwilch, ROD-Zw10) and intra cross-links (Zw10-Zw10 and ROD-ROD). All the cross-links are listed in **Tables 5.1 and 5.2.** Essential cross-links are depicted in **Figure 3.12.** The very first conclusion from the XL-MS analysis is that Zw10 and Zwilch are not spatially proximal. There are no cross-links whatsoever amongst these proteins.

The ROD-Zwilch cross-links validate previous biochemical experiments, where the N-terminus of ROD was shown to interact with Zwilch [84]. The XL-MS also highlights the fact that both N- and C-termini of ROD interact with Zwilch **(Figure 3.12, a)**.

The second set of inter cross-links are between proteins ROD and Zw10. I find that Zw10 is spatially proximal and interacts with the central region of ROD (850-1670 aa). I also find that certain amino acids on Zw10 (130, for example) cross-links with amino acids 859 and 1665 on ROD **(Figure 3.12, b)**. This could mean either ROD wraps around Zw10 or, alternatively, Zw10 remains sandwiched between two anti-parallel ROD molecules. Since, prediction programs model the stretch of ROD (encompassing amino acids 850-1670) as α-solenoid which are not known to adopt curved structures, I argue that Zw10 is most likely sandwiched between the two ROD monomers. By the same logic I also conclude that the intra ROD cross-links are not from a single polypeptide chain, as it would require significant torsional strain, but from two molecules of ROD which are anti-parallel in the RZZ complex **(Figure 3.12, d)**. I, however, could not extrapolate any significant information from the intra Zw10 cross-links. The analysis of the cross-links was done together with Anika Altenfeld and Dr. Jenny Keller (AG Musacchio).

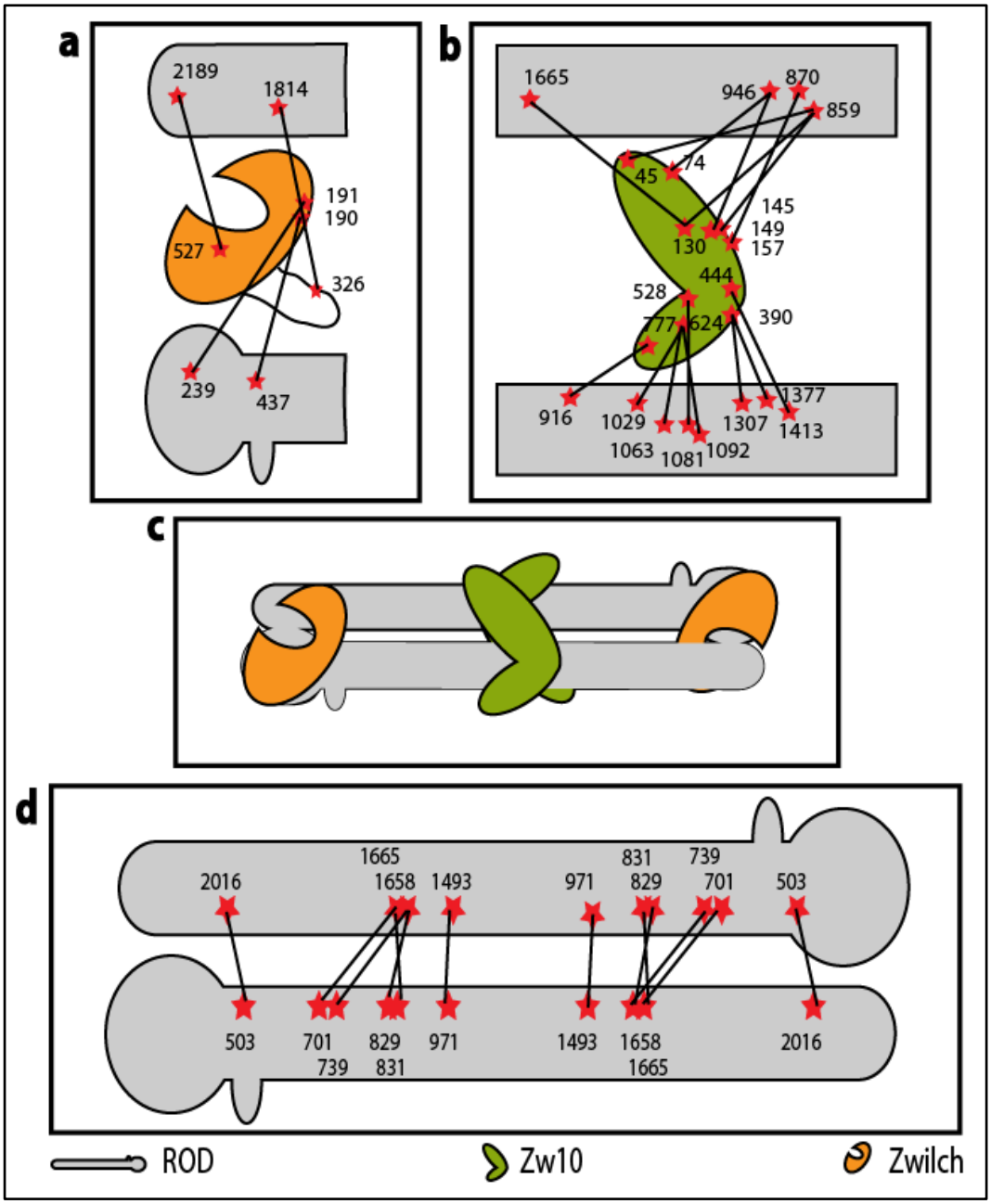


Figure 3.12: Cross-linking mass spectrometry (XL-MS) analysis of the RZZ complex. Representative cross-links between ROD-ROD, ROD-Zw10 and ROD-Zwilch highlighted. XL-MS is a low-resolution technique used for understanding the spatial organization of the proteins in a complex**. (a)** ROD (cartoon in gray)-Zwilch (in orange) crosslinks. **(b)** ROD-Zw10 (in green) crosslinks. **(c)** Cartoon representation on the organization of sub-units in the RZZ complex **(d)** ROD-ROD intra crosslinks. A detailed list of all the cross-links is in the Tables 5.1 and 5.2.

3.1.6 Rigid body fitting

ROD is the largest subunit of the RZZ complex (2209 residues). The first 350 residues are predicted to have a secondary beta-propeller structure organized in putative WD40 beta propellers whereas the rest of the molecule is predicted to have secondary alpha-helical structure and contains 3 regions: an NRH (Nag-ROD homology) domain, a Sec39-like domain followed by an additional 1000 residue segment. Protein homology/analogy recognition engine (Phyre2) was used to generate models of the

WD40 beta-propeller and the Sec39-like domains (Dr. Jenny Keller generated the models for various fragments of ROD). Models satisfying the intra-molecular crosslinks were selected. De novo modeling programs (I-Tasser, CABS-fold, Modweb server) were used to generate models of the NRH and the C-terminal domains [169-171]. Dsl1 has been suggested to be the yeast homologue of Zw10. The crystal structures of Dsl1p are known (PDB IDs: 3K8P, 3ETU). I used these models for fitting into the cryo-EM structure. The crystal structure of Zwilch (PDB ID: 3IF8), as aforementioned, has been known for a while. A gallery of all the models used for fit is highlighted in **Figure S5.** The distance restraints as imposed by cross-linking were applied and I fit the models (by rigid body fitting procedures) predicted for various segments of ROD and the structures of Zw10 and Zwilch. It led to the architecture that is depicted in the **Figure 3.13.**

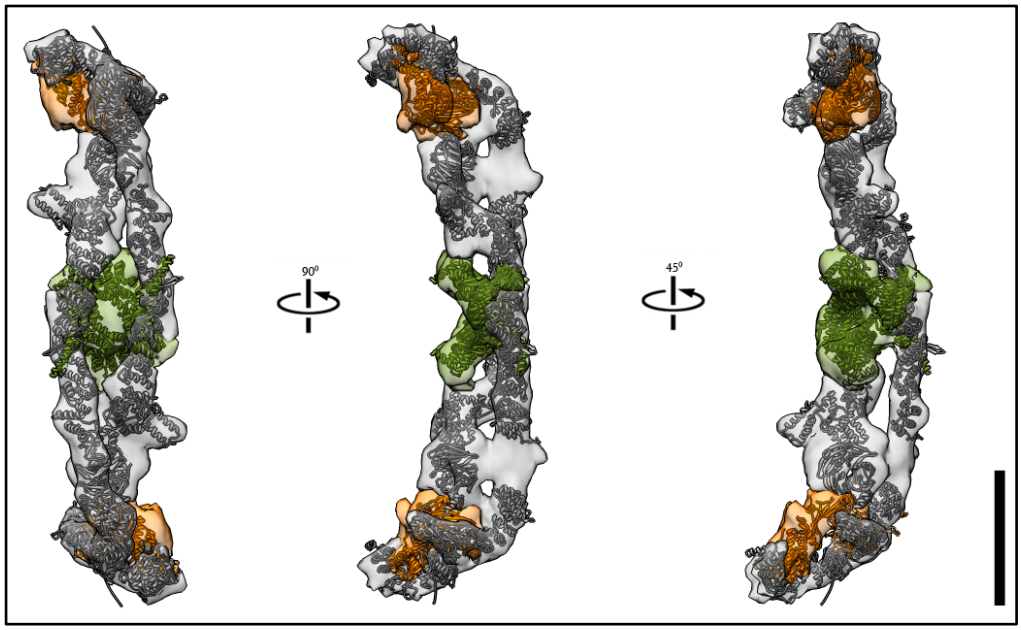


Figure 3.13: Predicted structural models of ROD (see Figure S5), along with the structures of Zw10 (used structure of its yeast homologue Dsl1p, 3ETU and 3K8P chain C) and Zwilch (3IF8) fit into the electron density map of RZZ. Distant restraints were applied, as inferred from XL-MS analysis resulting in the fit, views of which are shown. The volume of RZZ is segmented into colors corresponding different proteins using the color zone tool in Chimera. ROD in gray, Zw10 in green and Zwilch in orange. Scale bar, 10 nm.

3.1.7 Why longitudinal dimer?

I propose that RZZ is a longitudinal dimer due to the following reasons: First, the structural prediction of ROD highlights that it extends as α-solenoid through most of its central region. A Lateral dimer would require a ROD molecule to bend over in a fashion such that the N and C-terminus of the same polypeptide interact, which I believe is energetically unfavorable. Across literature there are seldom examples of α-solenoids adopting a bent conformation. Second, since antibody-labeling studies suggest that Zw10 is predominantly localized to the center, a lateral dimer would mean most of the dimerization occurs via the Zw10 molecules, which I believe is not the case (I do not observe any dimers of Zw10 in negative stain electron microscopy analysis). This is also reflected when Zw10 is

expressed as C-terminal GFP fusion protein. I often find complexes as shown in **Figure 3.14**, where the RZZ dimer seems to breathe out making the two parallel strands of ROD evident.

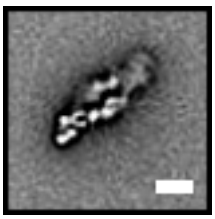


Figure 3.14: Class average as a result of reference free alignment and *k*-means classification procedures. Complex is assembled where the Zw10 is generated as a fusion protein with GFP at its C-terminus (R-Z(cGFP)-Z). Particles that cluster into this class are very few in number. Such classes are obtained along with classes which appear predominantly as highlighted in Figure 3.9,b. Density for the GFP can not be attributed in this. However I see that the complex breathes out making the two anti-parallel ROD strands evident. Scale bar, 10 nm.

3.1.8 Validation of structure

3.1.8.1 Mini-RZZ

To validate the architecture as described above complexes with truncations in ROD were generated. One such complex susceptible to purification and further structural analysis was the $ROD^{1-1250}ZZ$ (hereafter referred to as mini-RZZ, generated by Anika Altenfeld, AG Musacchio). I argued that if proposed architecture is right then deletion of C-terminus of ROD should lead to a collapse in the dimerization of the complex, as it is necessary for the stability of the complex, evident from the crosslinking experiments. It turns out that I was right! Mini-RZZ is difficult to purify, lacks stability, is characteristically shorter and the averages (generated from reference free alignment and *k*-means classification) do not indicate any element of symmetry (**Figure 3.15**).

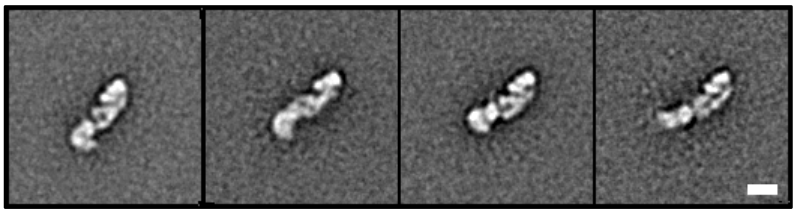


Figure 3.15: Representative negative stain class averages of mini-RZZ as obtained by reference free alignment and *k*-means classification procedures. Scale bar, 10 nm.

To get an overview of the 3D structure of mini-RZZ I resorted to RCT approach. I recorded images at 55° and 0° tilt and boxed particle pairs. The 3D structure was generated by back projection and back-projection refinement, as described earlier (for RZZ, refer **section 3.1.2**). The resolution of structure was 24.6 Å ($FSC_{0.5}$ criterion, **Figure S10, d**). It was clearly evident, that mini-RZZ does not exhibit the characteristic dimerization as RZZ. Characteristic views of the density are represented in **Figure 3.16.**

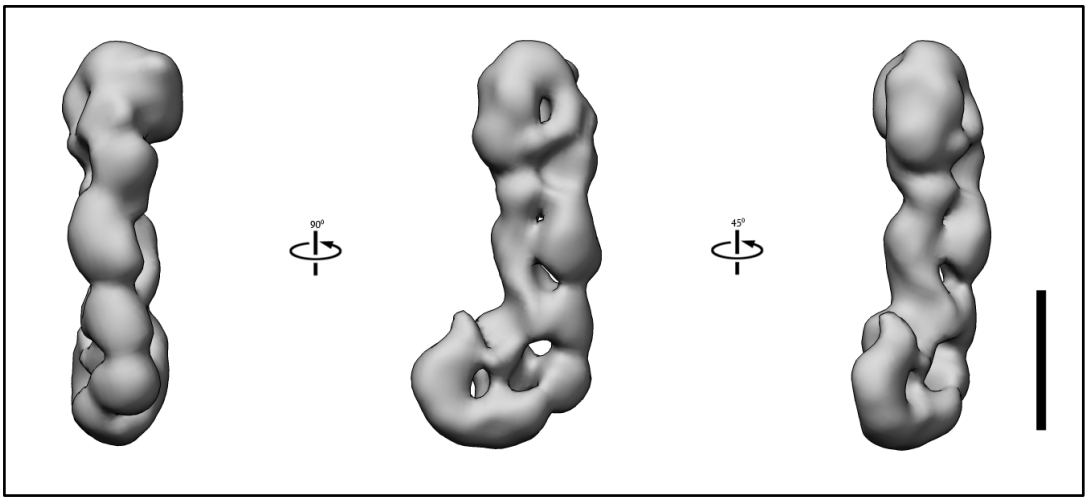


Figure 3.16: Views of mini-RZZ. Structure obtained by back-projection (RCT) and back-projection refinement. Length of the complex is 26 nm. Scale bar, 10 nm.

I was not able to generate difference maps between RZZ and mini-RZZ, mainly because of the fact that the densities were obtained by two different methodologies (cryo vs. stain) and the difference map between the negative stain reconstructions was not particularly useful. However, I fit the density of mini-RZZ into a filtered density of RZZ. This fit does not clearly attribute the missing density in mini-RZZ to the C-terminus of ROD (data not shown).

3.1.9 Salient features of the fitting and proposed architecture

13.4 Å is an intermediate resolution where scouting for secondary structural elements is not only tough but can be misleading. The main limitation with understanding the architecture is that the crystal structures of ROD and Zw10 are not known. Albeit very powerful, the prediction programs may generate models that are far from accurate. However, antibody labeling and cross-linking studies place this analysis on a firm footing, where I can safely highlight the following features.

1. The N-terminus of ROD is predicted to be folding into a β-propeller (**Figure 3.17, a**). The density corresponding to toroid stands out and can be clearly marked. The predicted model (**Figure S5, a**) clearly fits into the density.

2. Zw10 adopts various conformations in solution (also reflected by the inertia to crystallize) but gets locked into one conformation in the RZZ complex and serves as a latch between the ROD monomers. The density corresponding to Zw10 can also be clearly marked out and it is in reasonable agreement to the known structural homologue, Dsl1p (**Figure 3.17, b)**.

3. Zwilch is terminal anchor point in the RZZ complex, and that both N- and the C-terminus of ROD interact with Zwilch (**Figure 3.17, c)**.

4. Most of the ROD protein has a secondary α-helical structure and it adopts an extended conformation in the complex give the distinct shape to the RZZ complex. The two monomers of ROD protein run in an anti-parallel fashion (**Figure 3.17, d)**.

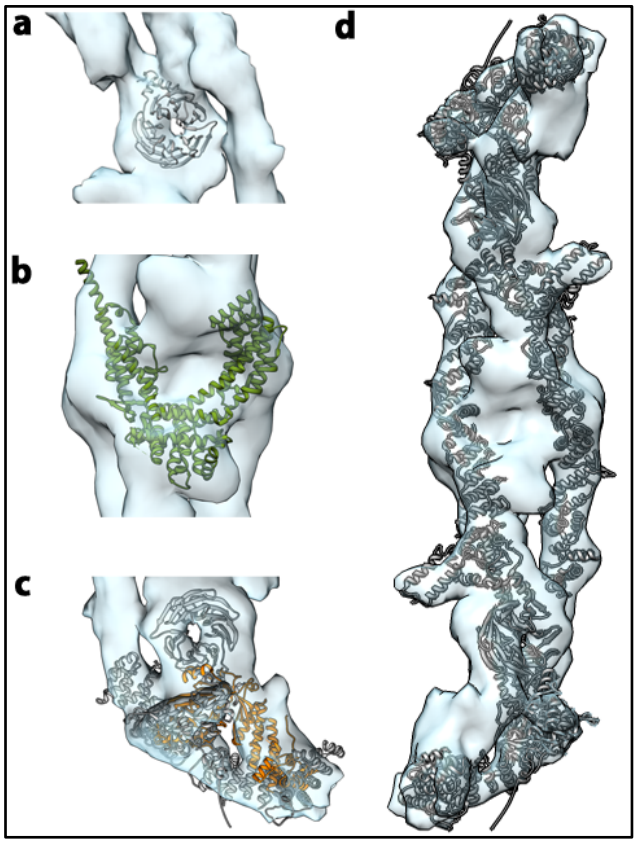


Figure 3.17: Salient features of the structure. (a) The N-terminus of ROD is predicted to be a β-propeller. Density corresponding to this can be clearly visualized in the structure. In fact the architecture of the RZZ complex was built starting from this hint. **(b**) Density corresponding to Zw10. The crystal structures of Dsl1p (in green) fit very well into the density, which clearly seem to correlate with antibody labeling. The two Zw10 molecules lie in an anti-parallel fashion acting as a bridge between the ROD monomers **(c)** Zwilch not only interacts with the N-terminus of ROD, but also the C-terminus of the second ROD molecule. Zwilch seems to be the terminal anchor point of the RZZ complex (crystal structure, 3IF8 in orange). The fitting is not the best, as there seem to be of clashes between ROD and Zwilch molecules. In my defense, the models of ROD fit are from prediction programs that may not be very accurate. A high-resolution cryo-EM structure will facilitate better understanding of the precise interactions. **(d)** ROD molecules make up most of the RZZ complex. Dr. Jenny Keller generated all the structures of ROD. Various elements that were predicted are highlighted in **Figure S7**. The images were rendered using UCSF Chimera.

3.1.10 Why dimerize?

Why is RZZ a dimer? Why should the cell work its machinery in producing a dimer without any functional consequence? Recent reports indicate the role of the RZZ complex in formation of lateral attachments while keeping NDC80C inhibited and then promoting those to end-on attachments [85]. In the study, the authors show that ROD$^{1-400}$: Zwilch (RZ) is sufficient to interact with the NDC80 complex, raising questions about the necessity of the dimer. Dimerization may be due to one of the two following reasons: first, RZZ might be acting as a functional dimer, for example as a receptor for binding the Mad2 dimer or second, dimerization may be the mechanism to inhibit the activity of RZZ, by blocking the WD40 repeats of the ROD and hence preventing it from interacting with other proteins. Although the surface of the β-propeller region is not masked out much due to dimerization it still might be enough for inhibiting activity of the monomeric RZZ complex.

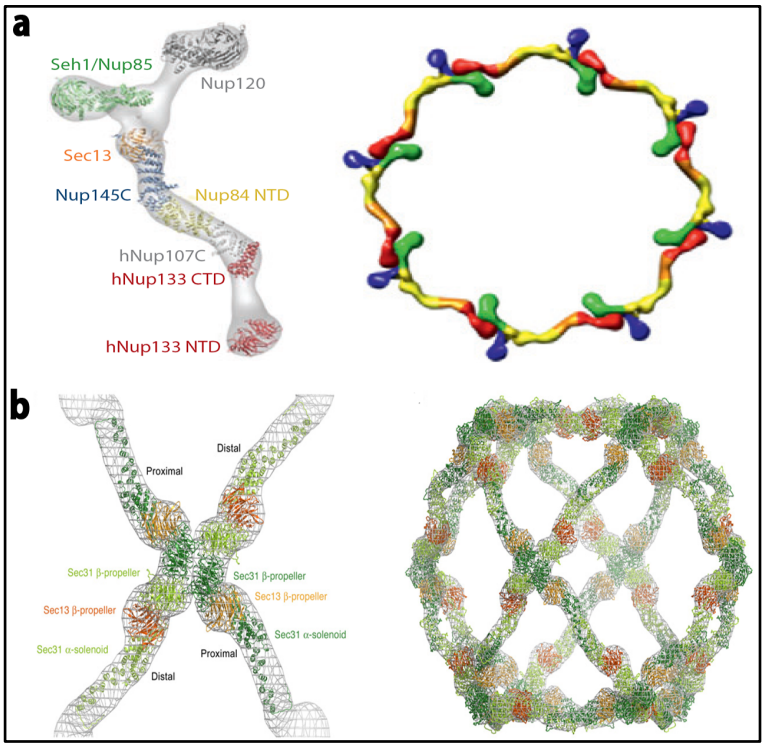


Figure 3.18: β-propeller/α-solenoid domains tend to arrange into higher order structures. (a) The NPC is complex multilayered hierarchical assembly of proteins known as the nucleoporins (Nup). It facilitates and mediates transport between the nucleus and cytoplasm of eukaryotic cells (for sake of simplicity only the architecture of outer ring of NPC is shown). The Y-shaped heptameric complex (composed of Nup85, Nup120, Sec13, Nup145, Nup84, Nup107 and Nup133) interacts in a head to tail fashion to form the outer ring. The key interaction here is the interaction of Nup133 with Nup120. β-propellers occur in almost all the layers of NPC making them indispensible for its architecture[172]. **(b)** COPII is a vesicle coat complex made up of four proteins (Sec23p, Sec24p, Sec13p and Sec31p) and is absolutely essential for the anterograde transport between the endoplasmic reticulum and the Golgi. Again the interactions between the WD40 domains are absolutely essential for the formation of cages.

The latter aspect needs to be reflected upon in detail. ROD is the largest protein in the RZZ complex. It is predicted to have an N-terminal β-propeller followed by an all α-helical secondary structure. I could define and outline the density corresponding to ROD in the structure. I show that it exists as an anti-parallel dimer providing the chassis for the structure of RZZ complex held together by Zw10. The β-propeller/α-solenoid is a key combination occurring recurrently in cellular macromolecular complexes, where proteins engage in forming lattices primarily via interactions mediated by WD40 domains. Two examples of such complexes are the nuclear pore complex (NPC) and the COPII vesicle coat [173] (**Figure 3.18**). The β-propellers (WD40 domains) are far apart in the RZZ complex, and are partly shielded by the C-terminus of the ROD. This may be a possible reason why higher order structures are not seen in these studies.

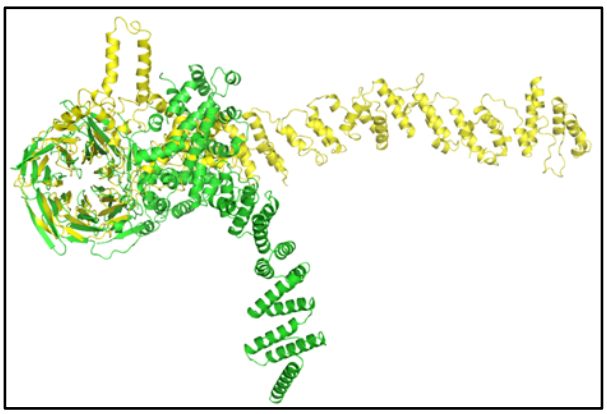


Figure 3.19: Overlay of the predicted structure of the N-terminus of ROD (6-1165, in yellow) with that of Nup120 (yeast homologue of Nup160, in green). The β-propellers are alike (aligned in this figure) followed by a linker domain in Nup120 and NRH domain in ROD, which define the orientation of the α-solenoid. Image rendered by Dr. Jenny Keller using PyMOL.

The oligomeric structures of the NPC and COPII complexes are built up on the β-propeller/α-solenoid framework. In case of the outer ring of NPC, the Y-shaped complex (made up of Nup120, Nup85, Sec13, Nup145, Nup84, Nup107, Nup133) interacts in a head to tail fashion through the β-propeller domain. Similarly in the COPII complex (coat made up of Sec13/Sec31, shown in **Figure 3.18**) the cage is built on the interactions mediated by the β-propeller regions of either Sec13 or Sec31. Structural predictions hint at ROD adopting a similar structure as that of the nucleoporins (an overlay of predicted ROD structure and that of the Nup120 is highlighted in **Figure 3.20**). Multiple sequence alignment between ROD, Sec31 and Nup160, however show little conservation (**Figure S12 and 13**). This can be attributed also to the WD40 domain, which is notoriously known for low conservation. Moreover, it is known for a wide range of proteins that sequence similarity does not parallel structural similarity.

RZZ is cytosolic prior to prophase and little is known about the state of its oligomerization during various phases of the cell cycle. Taking into consideration the fact that components of COPII coat are cytosolic and those of NPC are cytosol bound during certain stages of cell-cycle; it is possible that the components of RZZ, NPC and COPII coat share similar ancestry but have diverged over the course of evolution.

Along this line I would expect the RZZ to form higher order complexes. Sec13/Sec31 protein complex was long known not to form coats without adaptor proteins. However, an entirely serendipitous experiment led to finding that they self-assemble to form minimal coats. RZZ complex, however, does not assemble into higher order structures when treated under similar conditions (**Figure S8**). It can only be speculated at this juncture that an adaptor protein or possibly a modification on RZZ is missing. A promising hint towards this aspect is shift in size-exclusion chromatography profile of RZZ upon phosphorylation by kinase Mps1 (**Figure 3.20)**. Expecting RZZ to form higher order structures is not entirely naïve. This could possibly explain its role in recruitment and transport of kinetochore components like Mad2. One can only speculate the role and necessity of dimerization but precise answering would need a comprehensive effort and detailed study.

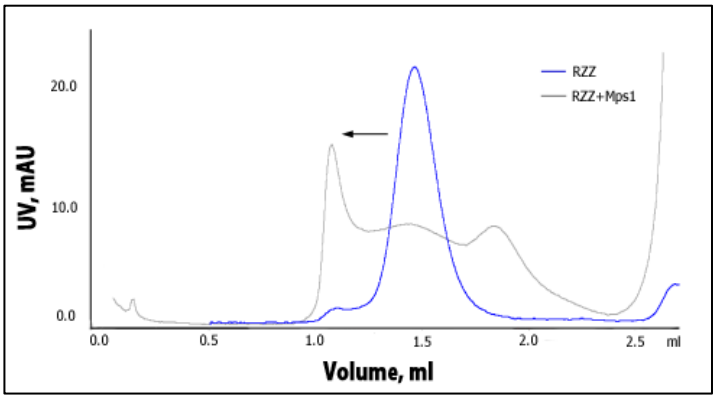


Figure 3.20: Size exclusion chromatography profile of RZZ complex (blue peak) and of RZZ phosphorylated by Mps1. A clear shift is seen as pointed out by the arrow. Possible interaction of RZZ could be explained by electrostatics, where the negative charge introduced due to the phosphate moiety may facilitate it to interact with positive tail of the NDC80 complex. Source: Anika Altenfeld.

3.1.11 Structure in the context of kinetochore

The complexes in kinetochore are heavily interdependent on each other for proper recruitment and localization (**refer to section 1.1.3**). For example, for the SAC, the Mad1-Mad2 pathway is indispensible. This relies primarily on the RZZ complex for recruitment of Mad2 and on dynein-dynactin complex for removal of the components once the checkpoint is quenched. The recruitment of dynein depends on RZZ and that of RZZ on Zwint-1, which in turn requires Knl1 for its recruitment [174]. In essence, these proteins interact with each other, directly or indirectly, validity of which remains to established biochemically. Understanding the spatial and temporal localization of a protein/proteins in a complex *in vivo* is also essential. Efforts directed to understand the positioning of proteins localizing at the kinetochore, along the kinetochore-microtubule axis, led to the first nanometer-scale map as depicted in **Figure 1.3**. In recent developments this analysis (K-SHREC) was extended to proteins involved in the Mad1-Mad2 checkpoint pathway [174]. The essence of this analysis is depicted pictorially in **(Figure 3.21, a)** and can be summarized as follows: RZZ subunits are spatially proximal to the N terminus of Knl1 and Hec1. Mad2-binding domain of Mad1 is close to the C-terminus of Zw10 and Spindly extends from RZZ complex into the outer kinetochore. Of particular interest is the position of the epitopes of RZZ subunits determined from this analysis, according to which the N-terminus of ROD and N-terminus of Zw10 are close to each other (within 3 nm) and that Zwilch is proximal to N-terminus of ROD. This analysis resonates precisely with the architecture as understood from the structural determined, implying that the protein organization is similar in the cellular context as *in vitro.* I find that the structure is in remarkable agreement with the Delta analysis (refer **section 1.1.2**). I compare the inter sub-unit distances as deciphered from my structure with those obtained from Delta analysis in **(Figure 3.21, b)**.

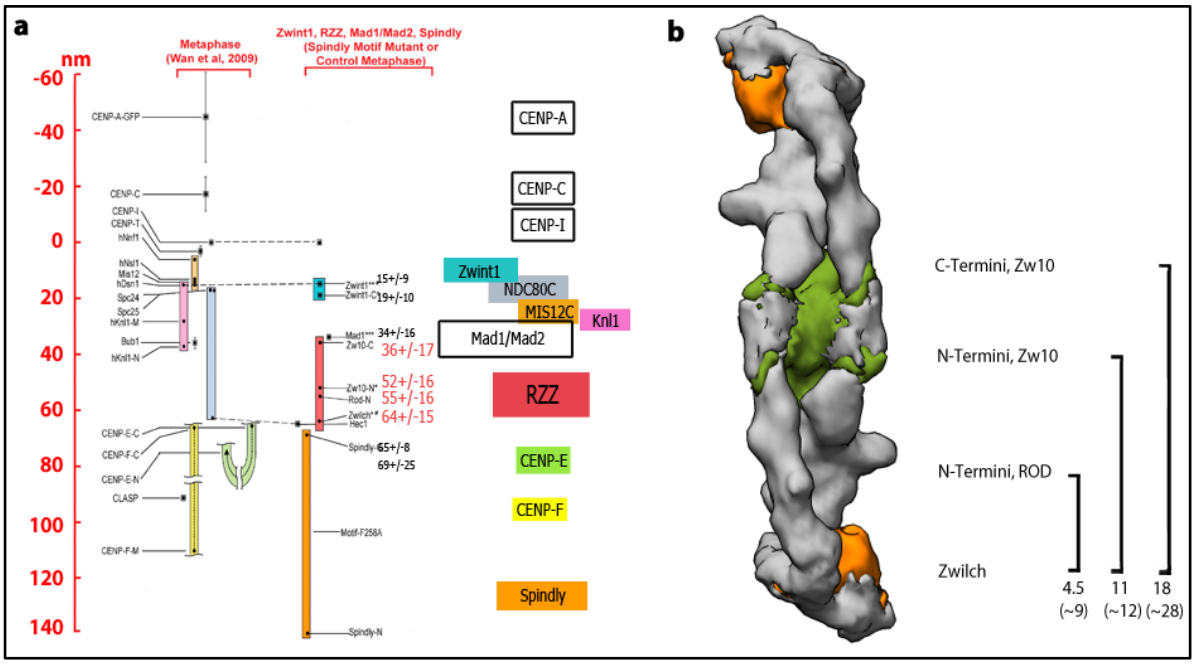


Figure 3.21: Structure in the context of kinetochore. (a) The positions of proteins/epitopes obtained from mean values of Delta measurements. Scale on the left side is set to 0 at the position of CENP-I; positive values are outward (toward the spindle microtubules), whereas negative values are inward (toward the centromeric chromatin). Color-coded boxes indicate complexes and large proteins. Black dots indicate the average Delta values that have been corrected for tilt. The vertical lines indicate the minimum and maximum Delta values measured across the variation in centromere stretch [174]. **(b)** Segmented view of RZZ density. Zwilch (orange), Zw10 (in green) and ROD (gray) are represented. The inter-subunit distance with reference to Zwilch are highlighted on the right. The distances were calculated from the mean position of Zwilch. The approximate values from the K-SHREC are listed in parenthesis below.

The scenario becomes a little bit more complicated in the context of RZZ as a dimer. Firstly, as I argued earlier, the functional RZZ complex at the kinetochore may not be a dimer but this still remains to be established. Secondly, the Delta measurements rely heavily on the accessibility of the polypeptide chain of interest by polyclonal antibodies (during immunofluorescence) leading to high values of standard deviations and a compromise in accuracy. For example, the Delta value for the C-terminus of Zw10 is 36 (+/-17) nm. Thirdly, estimating the position of C-terminus of ROD would be key in understanding the placement of complex in the context of kinetochore. It would give a complete picture as to how RZZ complex is positioned in the kinetochore. In essence, I believe that the inter-protein distances along the kinetochore-microtubule axis, as determined by the k-SHREC, are comfortably placed with the structure, but the scenario gets a tad bit complicated taking dimerization into account.

3.1.12 Road to high-resolution structure

Structure determination by cryo-EM is riding a new high, with near atomic resolution structures feasible for asymmetric and smaller protein complexes [175, 176]. The advent of direct electron detectors has catapulted cryo-EM into spotlight. The power of single particle cryo-EM is limited by

three factors: lack of contrast in the images, introduction of noise and drift. Though there are several other factors, the loss of contrast is primarily a consequence of beam-induced motion of particles (drift) and less efficient detectors (as estimated by the detective quantum efficiency). Thanks to the advancement in technology, the new breed of detectors facilitate the detection of electrons directly rather than converting them in photons for recording the signal (like most of the CCD's in use) and hence have a two-fold advantage. The images recorded have a high signal to noise ratio as there is a considerable reduction of back-scattered electrons and secondly, because of the high DQE the images contain information up to high spatial frequencies. The new detectors also facilitate recording images as frames/movies, which can be used for correction of drift or beam induced motion of particles in ice. The cryo-EM images, which led to the structure presented here, were collected on an 8k x 8k CMOS camera. The structure, although novel, is on the borderline where detection of the secondary structural elements with authority is challenging. I therefore decided to invoke the recent advancement in the detectors to try and achieve sub-nanometer resolution. I collected data on a TITAN Krios, equipped with a back-thinned falcon II (direct detector) at the NeCEN, Leiden. The initial results are very promising (**Figure 3.22)** and I am confident that I can get past the sub-nanometer resolution barrier.

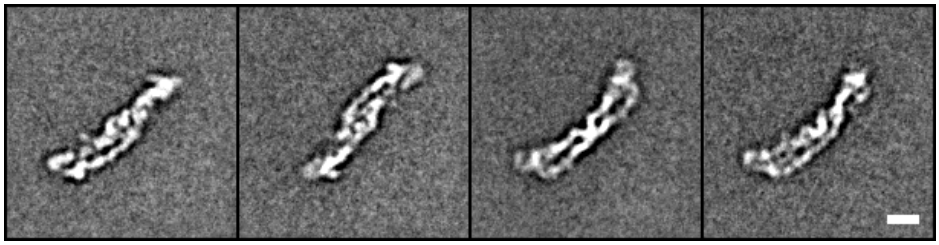


Figure 3.22: Representative cryo-EM class averages of RZZ complex obtained from reference free alignment and *k*-means classification. The complex was imaged on TITAN Krios, equipped with a Falcon II at the Netherlands center for Electron Nanoscopy (NeCEN), Leiden. The averages are from a dataset of 5000 particles. Scale bar, 10 nm.

The fit of various predicted structures in the density is a bit coarse and not ideal. Clashes between residues are evident especially at the ROD-Zw10 and ROD-Zwilch interface. Working towards a structure with better resolution (sub-nanometer) is essential at this stage, as it will facilitate identification of secondary structure elements and proper fitting of the predicted structures of ROD, Zw10 and the structure of Zwilch by flexible fitting methods. A high-resolution structure of the complex is an urgent necessity, as it would open the window to explore and understand various functional aspects of the complex in the cell (*in vivo* studies based on the structure). Ultimately, I intend to understand the molecular basis of RZZ function by targeting various interfaces and domains, as inferred from the structure, and their consequence on the spindle assembly checkpoint.

Finally, one aspect not highlighted in detail is the presence of the RZZ complex and its role in spindle assembly checkpoint (SAC) and lateral attachment of microtubules only in higher eukaryotes. Kinetochores of higher eukaryotic organisms bind more than one microtubule (**Figure 3.23, a**), requiring additional proteins and complexes to ensure the fidelity of chromosome segregation. The signaling in SAC arises because of unattached kinetochores. What signals from the kinetochore ensure

which microtubules to bind and which to avoid? How is the whole process coordinated? The corona of microtubules is mesh-like unstructured region predominantly made up of proteins, which constitute the SAC. Studies show that RZZ can, in a phosphorylation dependent manner due to activity of Mps1 kinase, form higher order structures/aggregates (unpublished data). This may be a hint at how the kinetochore coordinates multiple microtubules dynamics and why the corona of the kinetochore is mesh-like. I believe that phosphorylation, and that it is reversible, brings about a conformational change in the RZZ complex, making it susceptible for oligomerization. This could be also a possible explanation why despite tremendous efforts; reconstitution of interactions of the RZZ complex *in vitro* was not possible **(Figure 3.23, b)**. The role of phosphorylation in the context of kinetochore-microtubule interaction is not new [177]. The nature of change induced upon phosphorylation and the nature of the large aggregates remain to be investigated. What triggers the phosphorylation of RZZ in the cell? What are the binding interfaces on the RZZ complex? What is the structural basis for removal of RZZ from kinetochores? How does RZZ interact with other proteins and protein complexes? What is the structural basis for such interactions? Finally, amidst all the chaos, what triggers the shift of microtubules from lateral to end-on attachment? These are very few questions that need to be understood imminently.

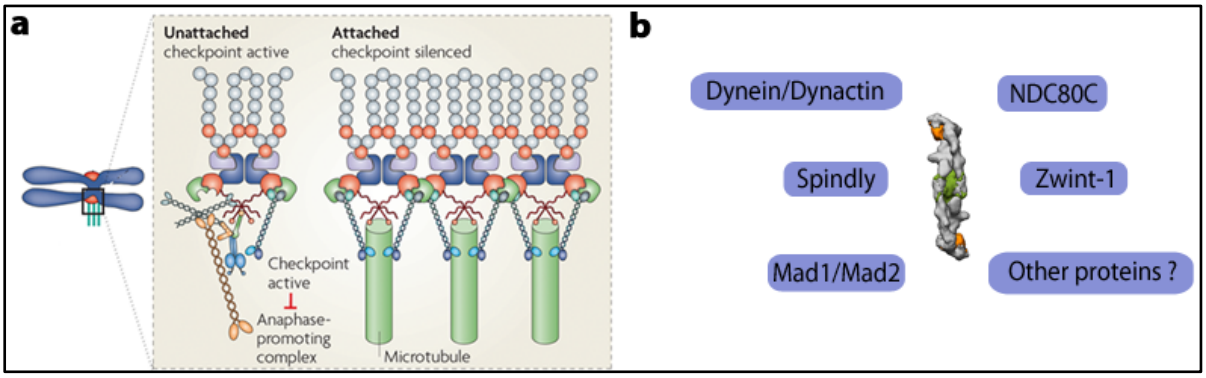


Figure 3.23: Understanding the function of RZZ based on structure. **(a)** Kinetochores that are unattached to microtubules recruit additional proteins to facilitate microtubule interactions and signal cell-cycle arrest. How RZZ coordinates multiple microtubules that bind kinetochores (by lateral capture and then transfer to NDC80 complex) in a dynamic fashion remains to be understood. The higher eukaryotic kinetochores present an extremely complex network of proteins interactions and structures of which remain poorly understood. The structure of RZZ may help in the understanding various aspects of SAC and the kinetochore-microtubule interactions in an indirect fashion. Source: [26] **(b)** Various interacting partners of RZZ. It is still unclear how RZZ mediates these interactions. The molecular mechanisms of these interactions remain to be established.

To summarize, I report the first high-resolution structure of the RZZ complex. I show that RZZ is a dimer. interweaving antibody-labeling studies with XL-MS analysis I highlight the organization of the RZZ complex. The structure is just the tip of the iceberg; efforts to understand how the complex eukaryotic spindle assembly checkpoint is regulated, from a structural perspective, are well underway.

3.2 Cln3p

3.2.1 Cloning and bacterial expression of Cln3p

The first and inevitable step in understanding the structure of any protein is to obtain a relatively large and pure amount of the same. cDNA of *cln*3 (human, mouse and rat) was cloned into pOPINE vectors with a fluorescent reporter (EGFP, fused to the N-terminus) and His (6x) affinity tag (plasmids were named according to the nomenclature of the AG Raunser database as p231, p232 and p261 for mouse, human and rat versions of *cln3* respectively). The plasmids were sequenced to authenticate the proper placement of the *cln3* gene. A great advantage of pOPINE vectors is that they can be used to express proteins in bacterial, insect as well as mammalian expression systems. Sandra Bergbrede devised and cloned the fluorescent fusion constructs.

Expression of the protein was screened in various bacterial expression strains (C41, C43 and BL21 (DE3)) under a variety of conditions (23 in total: varying in media, O.D. and time of expression, listed in **Table 2.15, materials and methods**). Since Cln3p is a membrane protein, I adapted the purification protocols to target the membrane pool of the cell lysate. Using EGFP as a reporter tag is of great advantage. The crude lysates can be tested for protein expression by scanning for EGFP fluorescence. In-gel fluorescence image of crude lysate spotted on a petri-plate is shown in **Figure 3.24**, **a.** The semi-denaturing SDS-PAGE profile of the membrane fraction of the lysates with brightest intensity as revealed from (**Figure 3.24**, **a)** is shown in (**Figure 3.24**, **b**).

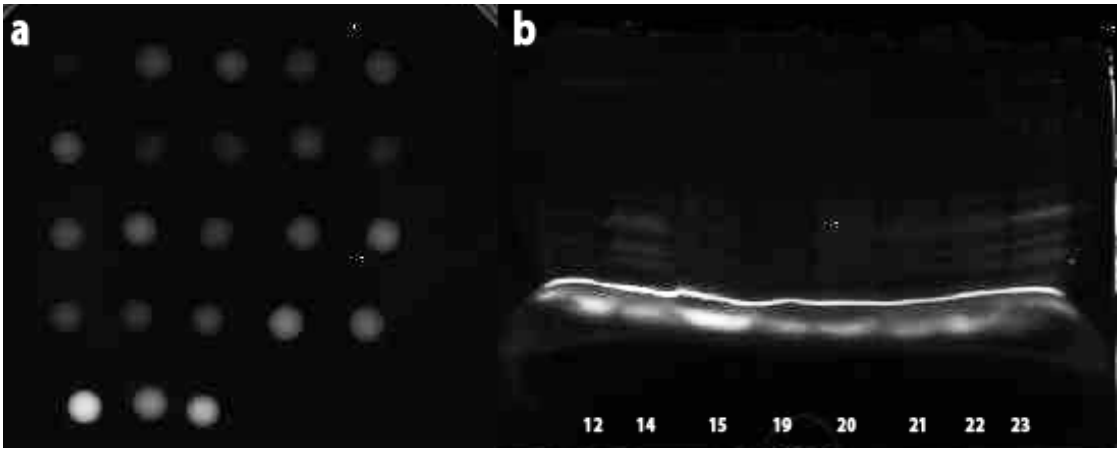


Figure 3.24: Over-expression of Cln3p in *E. coli*. Semi-denaturing SDS-PAGE. The folding of the protein remains generally intact owing to low concentration of SDS in loading and running buffers. Since, the proteins are expressed as fusions with fluorescent reporter tags (EGFP in case of p232), scanning for fluorescence of GFP (ex: 488, em: 509) in the gels leads to the identification of quality and nature of protein bands, a technique known as in-gel fluorescence. **(a)** 4 µL of crude lysate from all different conditions of expression loaded equidistantly on a petri-plate and scanned for GFP fluorescence. Intense white spot indicates the expression of protein (EGFP-hCln3p, p232) **(b)** An in-gel fluorescence image is shown where the most prominent spots seen in (a) were solubilized in solubilization buffer (1% DDM, 500 mM NaCl and 50 mM Tris-HCl, pH 7.5) and loaded onto a 15% semi-denaturing SDS PAGE and scanned for fluorescence. Exposure: 150 s.

The membranes were extracted, solubilized (1% DDM, 500 mM NaCl and 50 mM Tris-HCl, pH 7.5) and analyzed by Western blotting using antibodies directed towards the 6x-His tag. A typical analysis of 23 samples of expression trial of p232 is shown in **figure 3.25.** Unfortunately there is no signal corresponding to the protein, indicating that the protein is minimally expressed in bacteria. I generated various other plasmids harboring no EGFP tag but 10x-His tag (both N- and C-terminus) and screened for expression, but without any favorable results. Since, Cln3p is relatively small (46 kDa) and is predicted to have very short luminal loops, I did not generate any truncated versions of the protein.

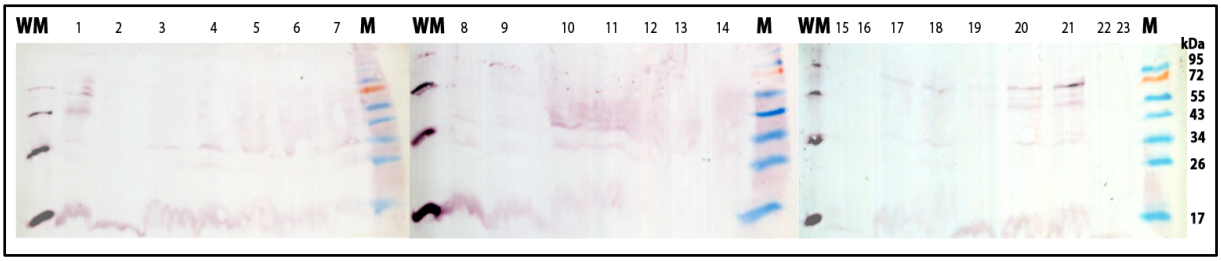


Figure 3.25: Western blot analysis (directed towards the 6x-His) to check for expression of p232 (EGFP-hCln3p) transformed in C43 cells. Varieties of conditions (Table 2.15, materials and methods) were screened but none lead to expression of protein even at minimal amounts. This was repeated for various expression strains such as (BL21-DE3 and C41). WM, Western blot marker. M, PAGE marker.

Overexpression of protein in bacterial system is cheap and effective on one hand but suffers from major setbacks in the context of eukaryotic membrane proteins. It simply lacks the machinery required for the proper folding and stability of eukaryotic proteins making it less suitable for overexpression.

3.2.2 Insect-cell expression of Cln3p

Bacterial hosts for overexpression proved inept for production of Cln3p (all homologues tested). I quickly shifted my focus onto baculovirus driven insect cell expression system. It offers a predominant advantage of overexpressing proteins, which require post-translational modifications in an efficient way. The BACmid (carrying viral genome necessary for the production of virus) and plasmid were co-transfected into Sf21 insect cells. I used a low multiplicity of infection (MOI) for the propagation of virus (0.01-0.1) and high MOI for protein expression (1-10). Once I had the working stock of the virus I checked for expression by infecting the Sf21 cells in a virus/cell ratio of 1:10, 1:100, 1:1000 and 1:10000. I optimized the time and the amount of virus required for efficient protein expression by monitoring the fluorescence of the reporter tag, where the number and brightness of the cells in the fluorescence channel gives an estimate of the quality of expression. Shown below are the DIC (**Figure 3.26, a**) and the fluorescence image (**Figure 3.26, b**) of Sf21 cells expressing EGFP-hCln3p (p232). I generated virus and studied the expression and purification of all the homologues (mouse and rat) of Cln3p. However for the sake of simplicity and to reduce repetition, results of p232 (EGFP-hCln3p) will be shown and discussed hereafter. The homologues behave in a way similar to hCln3p**.**

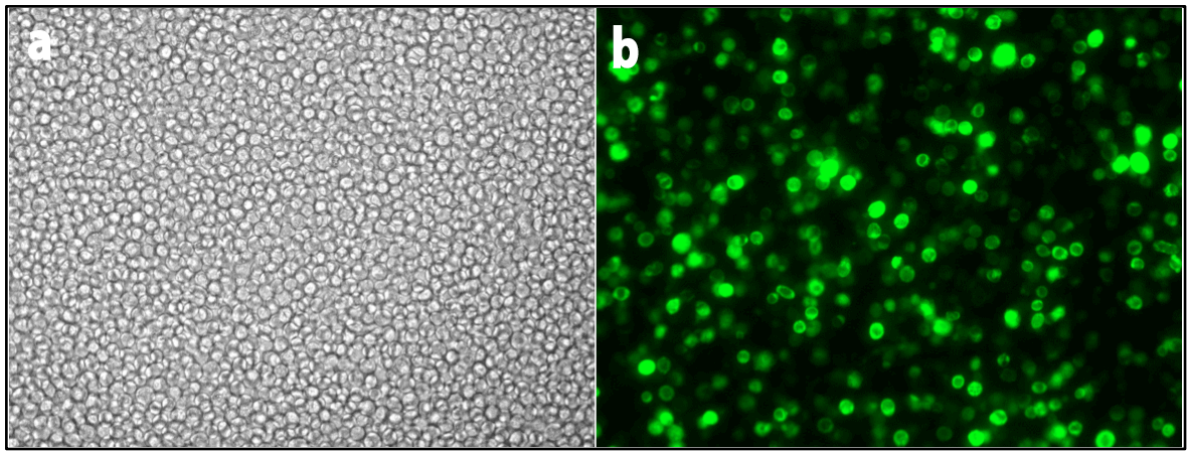


Figure 3.26: Overexpression of Cln3p in Sf21 cells. **(a)** All cells have an equal size and round shape indicating their healthy state. Also, they do not showing any signs of contamination with bacteria or other kinds of microorganisms (e.g. elongated cells or a large number of apoptotic cells). **(b)** Fluorescence image of **(a).** Not all the cells are fluorescent and the expression of the GFP construct varies amongst the cell population, however the expression level in general, is high (transfection is never 100 % complete). Fluorescence intensity is seen nicely located on the cell boundaries, implying that the majority of the overexpressed membrane-resident Cln3p is transported to the plasma membrane. The fact that there is a fluorescence signal is already a first hint that at least the N-terminal part of the protein is folded correctly, what is often related with an entirely functional protein (GFP as a marker for stability).

3.2.3 Purification of Cln3p

3.2.3.1 Detergent screen

Membrane proteins need detergents to get them out of their native membrane and into the solution. There are a variety of detergents on the market and it becomes crucial to select an appropriate one in which the protein is rendered pure and functional. I therefore screened for the solubilization capacity of various detergents. After extraction of the membrane from the insect cells expressing Cln3p I subjected them to treatment with buffers containing detergents of interest. For each of the detergent, I used them well above their critical micelle concentration (CMC) and screened for the protein in the supernatant (soluble fraction) before and after solubilization. I found that Fos-choline-13 is the best-suited detergent for the purification of Cln3p (**Figure 3.27, a**). The level of solubility is low or none in other detergents. Since Cln3p is fused with a GFP reporter tag, I used in-gel fluorescence to detect the protein. This method employs polyacrylamide gel electrophoresis (PAGE) of proteins in semi-denaturing conditions, helping proteins conserve their folded state to a great extent. Gels are scanned for GFP fluorescence prior to Coomassie staining. Since fluorescence is a very sensitive technique one can estimate the quality of protein accurately. I found consistent degradation of protein in all the detergents I screened for. I also found that addition of 10% glycerol was absolutely necessary to stabilize the protein and reduce its degradation during purification (**Figure 3.27, b)**. I therefore, used 10% glycerol in all the subsequent steps after the lysis of cells.

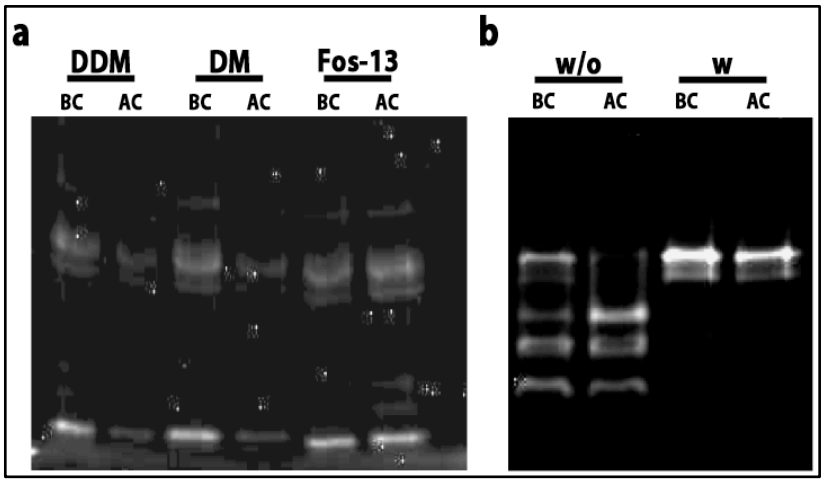


Figure 3.27: Solubilization Screen. **(a)** In-gel fluorescence images of detergent screen comparing Fos-13 with two most commonly used detergents for membrane proteins: DDM, DM. Only in Fos-13 the amount of protein before and after ultracentrifugation (to remove insolubilized membranes BC, AC respectively) remains same. **(b)** In-gel fluorescence image showing the necessity of glycerol in the buffer. A degradation pattern is clearly seen in buffer without glycerol (w/o).

3.2.3.2 Purification of Cln3p in Fos-Choline-13 (Fos-13)

I used affinity chromatography for the purification of the protein as all the constructs I generated had a 6x-His tag. By introducing a pre-solubilization step (I solubilized the membranes with 0.5% OG since Cln3p is not effected by OG), I attempted to get rid of other membrane proteins (membrane anchors and proteins soluble in OG, background) and boost the homogeneity of the Cln3p pool in the membrane fraction. Solubilized membranes (in 0.5% Fos-13) were centrifuged at 100,000 *g* to get rid of the insolubilized proteins. The supernatant was filtered (0.2 µm filters) to prevent clogging of the affinity column by removing very large detergent micelles. Immobilized metal affinity chromatography (IMAC) relies heavily on the interaction of the 6x-his tag of the protein with the metal ion (usually Ni^{2+} or Co^{2+}) and is prone to high amount of non-specific interaction if the expression amount of protein concerned is low (membrane proteins). I screened, therefore, for various resins (NTA, IDA, TALON®) and found that Co^{2+} - TALON® resin is best suited for highest sample purity. It is also noteworthy to mention that the binding affinity of TALON® resin is considerably lower than that of the Ni-NTA and hence the background (non-specific) interactions are also very less. I used in-gel fluorescence again to optimize and assess the purification process. A typical purification profile of EGFP-hCln3p in Fos-13 as monitored by in-gel fluorescence is shown in **Figure 3.28, a**.

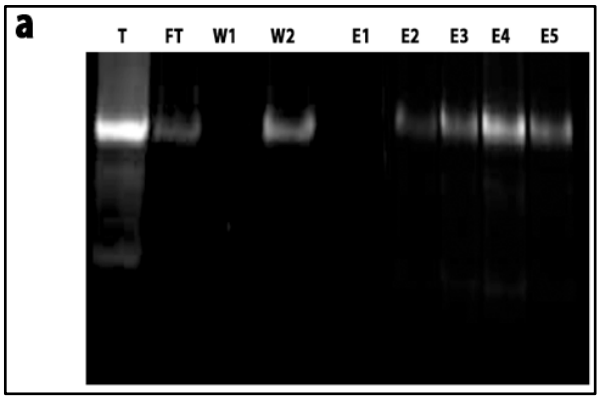
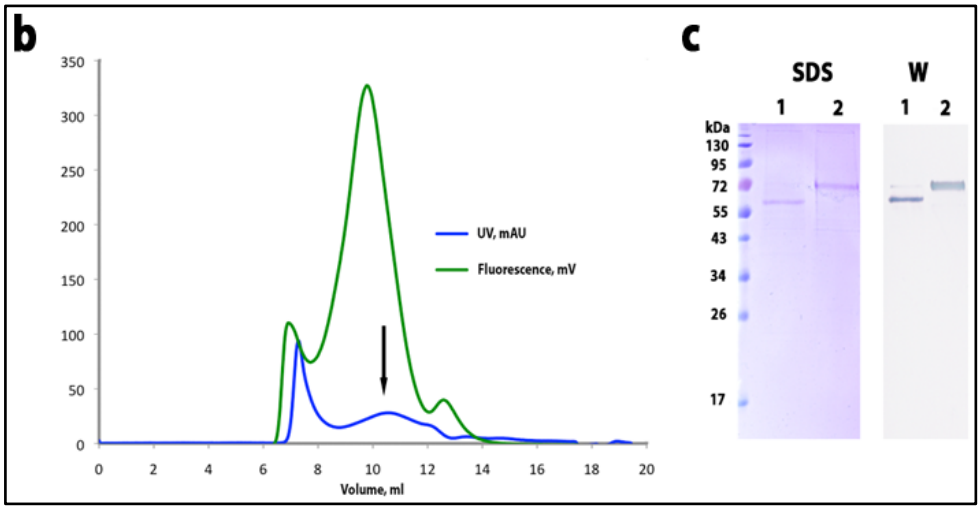


Figure 3.28: Purification of human Cln3p. The solubilization and purification of EGFP-hCln3p was monitored by in-gel fluorescence. **(a)** In-gel fluorescence image of the semi-denaturing SDS-PAGE profile of purification of EGFP-hCln3p. No degradation of the protein is seen. T, total solubilized membrane. FT, flow through. W 1-2, wash. E 1-5, Eluates. The Eluted fractions were pooled and concentrated (using amicon ultra MWCO: 50 kDa) and loaded onto a Superose 12 10/300 column. Exposure 5 s. **(b)** SEC profile of EGFP-hCln3p **(c)** Semi-denaturing SDS PAGE (SDS) and corresponding Western blot (W, towards 6x-His). Lanes 1 and 2 represent protein without and with heating. A clear shift confirms the fact that protein runs slightly above 55 kDa and that it is folded.



Once the solubilized protein was bound to the resin, the resin was washed with buffers containing low concentrations of imidazole to get rid of the non-specifically binding proteins. I eluted the protein in low concentration of detergent (0.1% Fos-13, CMC-0.027%). The protein was then dialyzed overnight (to get rid of the imidazole), concentrated and loaded onto a Superose-12 size-exclusion chromatography column. All the buffers contained detergent (0.1% Fos-13). A typical size-exclusion chromatography (Superose 12 10/300) profile for hCln3p in Fos-13 is shown in **Figure 3.28, b**. A strong UV peak close to the void-volume (devoid of fluorescence) suggests that the protein may have a tendency to oligomerize/aggregate. The unhealthy size-exclusion chromatography profile may also be due to non-homogeneous nature of protein-micelles. The EGFP fusion protein is predicted to have a molecular weight of 76 kDa but since it is folded in the semi-denaturing conditions runs at the size, slightly above 55 kDa. This is confirmed by the overlay of the fluorescence signal, which corresponds to the band around 55 kDa and also by the fact that upon heating the sample (classical SDS-PAGE) the protein runs at the size corresponding to predicted molecular weight of 76 kDa (see **Figure 3.28, c**). Although I report the first expression and purification of hCln3p, it took me quite long to render the

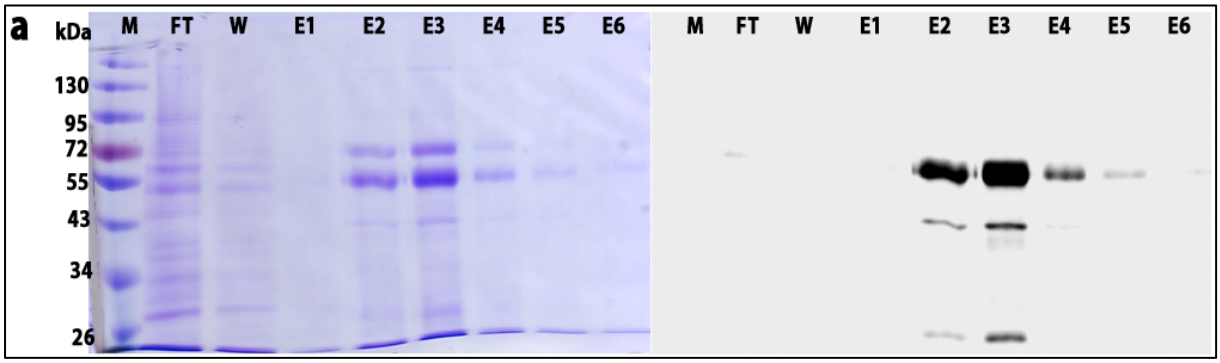
protein in conditions favorable for biochemical and structural analysis. To avoid repetition all the biochemical analysis and data from crystallization trials will be presented at the end.

Presence of His and/or GFP tags hinders crystallization of proteins at times. Therefore, getting rid of the tags is recommended before any such attempts are made. I ran into serious problems while trying to cleave the tag. Also, most of the protein was eluting in the void-volume. Assays to understand oligomerization and the state of post-translational modifications were also uninformative. The 2D and 3D crystallographic trials did not result in any hits. I believe that this was due to the interference of the detergent. Fos-choline-13 is a strong zwitter ionic detergent and seems not to be suitable for structural studies as is reflected by the lack of membrane protein structures solved using this detergent to date. Functional assays also require detergents that do not interfere with the enzymes. Therefore it was necessary to change the detergent. Since, it was not possible to solubilize Cln3p in any other detergent I tried exchanging the detergent once Cln3p was in Fos-13 micelles. I tried detergent exchange (to Cymal 5-8, LDAO, DM, OG, DDM) via gel-filtration, dialysis, on the affinity column, and also using mixed micelles (DDM and Fos-13) but none of these procedures resulted in pure protein. Most of these methods led to precipitation of the protein and therefore those experiments will not be discussed any further.

Just about the time when I was loosing hope, maltose neopentyl-glycol (MNG) amphiphiles were discovered and implicated for facilitating the X-ray structure of G-protein and the G-protein coupled receptor complex [178, 179].

3.2.3.3 Purification of Cln3p in MNG-3

I was not able to solubilize insect cell membranes overexpressing hCln3p in MNG-3 or MNG-3/cholesterol hemi-succinate (CHS) buffer (10:1). Therefore, I resorted to the old approach where I solubilized membranes using Fos-13. I was able to exchange the detergent on the IMAC column. An overview of the purification follows in **Figure 3.29**. The consistent problem with exchanging detergent to MNG-3 was the appearance of a band corresponding to molecular weight of 76 kDa. This band was attributed as the unfolded Cln3p (I ran the gels in semi-denaturing conditions) and confirmed by in-gel fluorescence scan (**Figure 3.29, a** (right) only one band is seen which corresponds to the lower band seen at around 55 kDa in semi-denaturing SDS-PAGE) and heat-denaturing analysis.



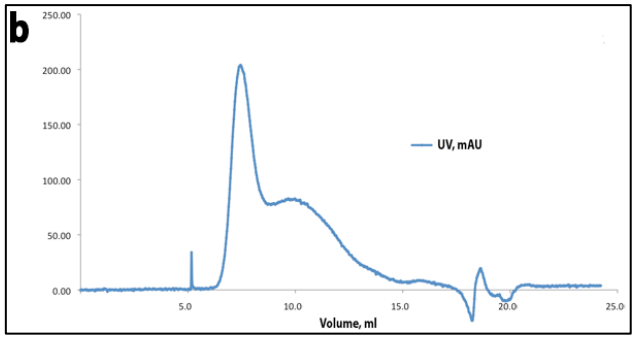


Figure 3.29: Detergent exchange to MNG-3. Solubilization of insect cell membranes was carried out in Fos-13 with subsequent loading onto an IMAC column. The detergent was then exchanged to MNG-3 (0.1% for elution, CMC, 0.001%) while elution. **(a)** A semi-denaturing SDS-PAGE profile of such detergent exchange elution for the EGFP-Cln3p (construct p232, left). Corresponding in-gel flurorescence (right). **(b)** Size exclusion profile of EGFP-hCln3p run after concentrating the elution fractions. M, marker. FT, flow through. W, wash. E1-6, elution.

The gel-filtration profile was quite similar to that of Cln3p in Fos-13. This may be due formation of aggregates. The solution to this problem follows later. Cln3p existing in two different populations in solution (unfolded and folded, as seen from the semi-denaturing PAGE) is not favorable for structural analysis. I therefore opted for ion-exchange chromatography (IEC) to separate the two forms. I screened for conditions for stability of Cln3p in low salt and found that the protein is stable in concentrations up to 25 mM of NaCl (data not shown). For ion-exchange chromatography the elution fractions from IMAC were pooled and desalted using a PD10 to a final salt concentration of 50 mM. I used a cation-exchanger column and the protein was eluted using a salt gradient rising up to 1M of NaCl. The IEC chromatogram revealed two peaks (labeled 1 and 2, **Figure 3.30**) eluting approximately at around 450 mM of salt. Fractions corresponding to both the IEC peaks reveal a band on semi-denaturing SDS-PAGE at the same size of 55 kDa approximately.

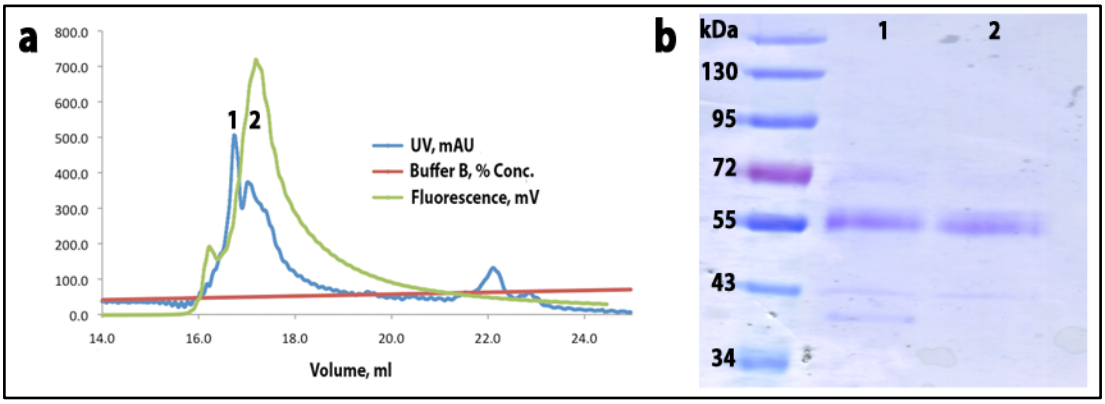


Figure 3.30: Ion-exchange chromatography. Ion exchange chromatography runs were performed on MONO S 5/50 cation-exchange columns. The pI of the protein was estimated to be 5.9 (generally for membrane proteins the estimate is off target) and used a cation exchanger. The resolution of the peaks from IE column was not good. I intended to separate the unfolded protein (runs at 76 kDa on semi-denaturing SDS PAGE) from the folded protein (runs around 55kDa). Although two peaks were observed (1,2) in **(a)** both run at a similar size in **(b).**

A plausible explanation of this behavior may be that hCln3p is oligomeric in solution and different peaks correspond to oligomer and monomer fractions. Surprisingly, I do not find any band at 76 kDa in the semi-denaturing PAGE. IEC experiments to optimize the purity of Cln3p in MNG-3 suffered a setback as I was not able resolve the recurrent peaks in the profile. I also realized that large amount of the protein was lost during the desalting process. I therefore, decided to discontinue IEC and sought

other ways to purify the protein. Size-exclusion chromatography (Superose-12) gave results similar to that of purification in Fos-13, where I find a large void peak followed by a broad peak for the protein (**Figure 3.29, b**).

6x-His and GFP tags can be detrimental in crystallization trials and one is generally advised to get rid of them. Since all the constructs cloned encode for a PreScission protease cleavage site, I tried to cleave the tag off the protein (after elution from IMAC) and use this as a way to enhance the purity of the protein by employing reverse IMAC procedure. The cleavage reaction was carried out overnight at 4°C via dialysis using 12-14 kDa MWCO membranes. The results of one such experiment are highlighted in **Figure S16**. In essence, cleavage by PreScission protease, although essential, was not an efficient procedure. The fault was with the cloning strategy employed; I reasoned that the accessibility of the cleavage site was sterically hindered, and perhaps the protease was not able to access the site which could be possibly embedded in MNG-3 micelles. Besides this, the hCln3p from p232 expressed in insect cells was susceptible to degradation (not helpful for crystallization trials) and the quality of protein varying in different preparations (predominantly depending on the extent of passage of the cells used for expression).

To circumvent these problems I revised the cloning strategy and generated constructs with linkers of varied length spaced between the protein and the tags (refer to **Table 2.13, Materials and methods**). A detailed analysis of 6x-His-mCherry-LINKER-humanCln3p (p509, AG Raunser database) follows next. The fluorescent reporter tag was changed from EGFP to mCherry as the theoretical stability index values for EGFP protein were always predicted low (ProtParam tool, http://web.expasy.org/protparam/protparam-doc.html). The corresponding mCherry fusion protein with linker was predicted to be more stable and therefore used. Sandra Bergbrede devised the cloning strategy and Dortmund Protein Facility (DPF) carried out cloning. Dennis Quentin, as a part of his Masters thesis under my supervision, carried out the experiments.

3.2.3.4 mCherry-fusion construct (p509)

The virus and expression were optimized for the newly generated constructs following protocols established previously for EGFP-Cln3p. The expression of the virus was monitored, as aforementioned (**section 3.2.2**), by fluorescence microscopy (**Figure S15**). I used the established strategy of solubilizing the membranes in Fos-13 and exchanging the detergent over the IMAC column. Semi-denaturing SDS-PAGE and the corresponding in-gel fluorescence scan are in **Figure 3.31, a** & **b** respectively**.** I ran into problems with huge void-peaks in previous purification trials, which led me to believe that probably this was a consequence of reduced ionic-strength of the solution while dialyzing away the imidazole. To counteract the reduction of ionic-strength dialysis buffer was spiked with additional salt (approximately 200 mM). Also, Superose 6 10/30 GL column was used for SEC. The size-exclusion chromatogram revealed a reasonably symmetric peak for mCherry-hCln3p (**Figure 3.31,c**).

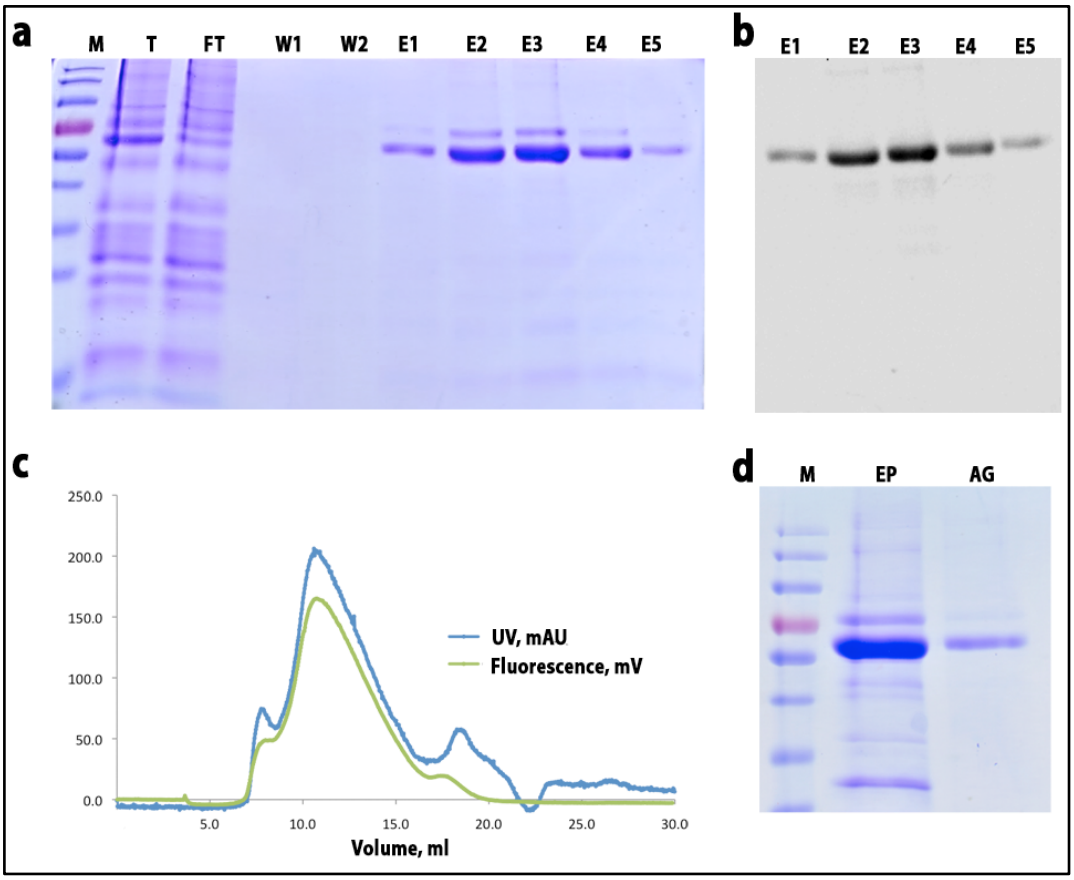


Figure 3.31: Purification of human mCherry-hCln3p (with long linker). Solubilization of insect cell membranes was carried out in Fos-13 with subsequent loading onto an IMAC column. The detergent was then exchanged to MNG-3 (0.1% for elution, CMC, 0.001%) while elution. **(a)** A semi-denaturing SDS-PAGE profile of such detergent exchange elution for the mCherry-Cln3p. **(b)** Corresponding in-gel fluorescence profile of the elution. **(c)** The elution fractions (E1-E5) were pooled and loaded onto a Superose 6 10/300 size exclusion column. Elution profile of one such run is shown. **(d)** Semi-denaturing SDS-PAGE profile of the concentrated elution pool used for SEC and the protein after elution from SEC. M, marker. FT, flow through. T, total solubilized membrane. W1, W2-wash fractions.

The addition of the long linker facilitated facile cleavage of the tag by the protease. During the course of optimization, reverse IMAC after the cleavage was deemed unnecessary. Instead the mixture was loaded onto a Superose 6 10/300 column directly to separate the pure Cln3p from the mCherry tag and the protease. Although this method worked fine best results were obtained following an alternative way: after the binding of the solubilized membranes to the IMAC, followed by subsequent washing steps, cleavage of the immobilized protein is carried out directly on the affinity column by passing the protease containing buffer. Concentrating the flow through and loading it on a Superose-6 column followed this step. The SEC profile in **Figure 3.32** shows minimal void peak followed by a slight shoulder corresponding to mCherry (high fluorescence) that leaches off the IMAC column.

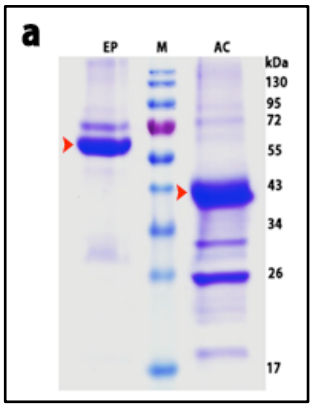
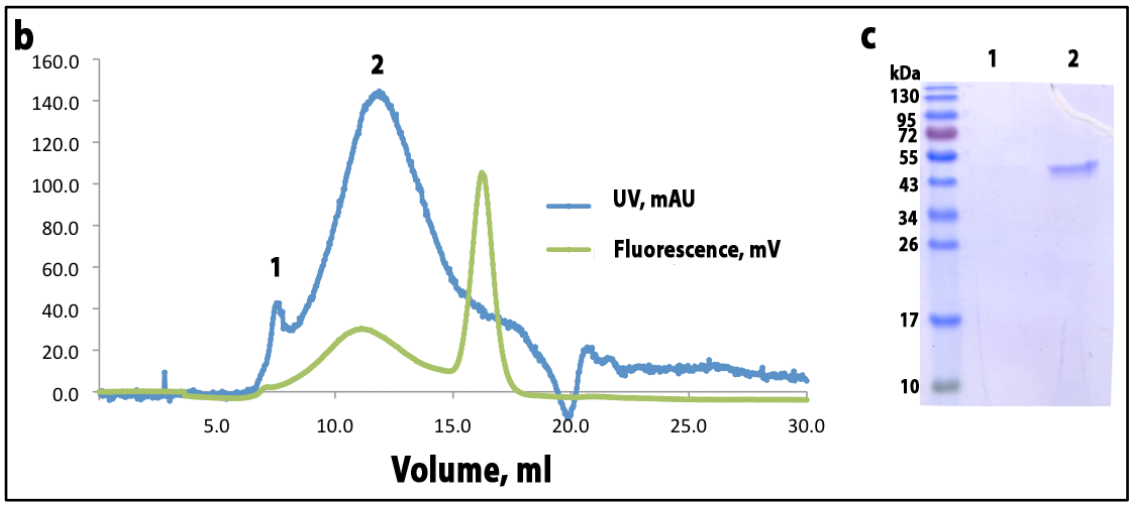


Figure 3.32: Cleavage of the 6x-His-mCherry tag. (a) The cleavage reaction of the Cln3p fusion protein was carried out by PreScission protease. A representative shift (as indicated by red arrows is shown. The reaction goes to near completion. EP, elution pool. M, marker and AC, after cleavage. The protease band is evident at 26 kDa. The cleavage reaction was optimized and found that it was best to cleave the tag on the affinity column and run size-exclusion chromatography of the concentrated flow through. One such typical profile on a superpose 6 10/300 column is highlighted in **(b).** A tiny amount of mCherry leaches off the column and is represented by the sharp fluorescence peak indicated in green. The corresponding semi-denaturing SDS-PAGE highlighting pure Cln3p band is shown in **(c)**



Western blotting and mass-spectrometry analysis confirmed the product of cleavage reaction. The expected band was excised, the protein extracted, tryptically digested and subsequently analyzed via HPLC-MS to unambiguously identify it as Cln3p. The first seven entries were keratin (**Figure 3.33)**, a fibrous structural protein and a known contaminant in mass spectrometry analysis (originates from contact of human skin during sample preparation). However, the only other protein identified was human Cln3p with a number of seven significant sequences. Cln3p is a membrane protein with a number of extremely hydrophobic amino acid stretches that span the membrane; it was not surprising that only a fraction of the theoretically predicted peptides were identified via HPLC-MS. Nevertheless, unbiased search confirms Cln3p as the target protein. Other peptides that were identified during the analysis could be traced to BSA and trypsin. BSA is a common contamination (traces can be found in the HPLC columns where it is used as size standard) and trypsin was added for the digestion of the proteins.

Family	Member	Database	Accession	Score	Mass	Num. of matches	Num. of significant matches	Num. of sequences	Num. of significant sequences	emPAI	Description
1	1	SwissProt_ID	K1C10_HUMAN	626	59020	65	29	31	19	2,17	Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6
2	1	SwissProt_ID	K2C1_HUMAN	592	66170	59	25	34	17	1,17	Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6
2	2	SwissProt_ID	K22E_HUMAN	310	65678	29	14	26	13	0,76	Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2
2	3	SwissProt_ID	K2C6B_HUMAN	143	60315	16	8	13	5	0,33	Keratin, type II cytoskeletal 6B OS=Homo sapiens GN=KRT6B PE=1 SV=5
2	4	SwissProt_ID	K2C5_HUMAN	133	62568	25	9	22	8	0,51	Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3
2	5	SwissProt_ID	K2C6A_HUMAN	116	60293	15	6	14	5	0,33	Keratin, type II cytoskeletal 6A OS=Homo sapiens GN=KRT6A PE=1 SV=3
3	1	SwissProt_ID	K1C9_HUMAN	427	62255	36	14	28	11	0,81	Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3
4	1	SwissProt_ID	CLN3_HUMAN	290	48163	36	21	9	7	0,91	**Battenin OS=Homo sapiens GN=CLN3 PE=1 SV=1**
5	1	SwissProt_ID	ALBU_BOVIN	246	71244	13	7	10	5	0,32	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
6	1	SwissProt_ID	BPIB1_HUMAN	221	52580	20	10	16	9	0,71	BPI fold-containing family B member 1 OS=Homo sapiens GN=BPIFB1 PE=2 SV=1
7	1	SwissProt_ID	TRYP_PIG	155	25078	31	14	8	8	2,38	Trypsin OS=Sus scrofa PE=1 SV=1

Figure 3.33: Overview of the tryptic digestion followed by HPLC-MS analysis on Cln3p. The lane corresponding to Cln3p is highlighted in green.

Other proteins with different reporter tags as well as constructs encoding only His-tag are also being processed. As mentioned earlier I did not generate any constructs with deletions in loop regions or otherwise as Cln3p is predicted to be a very compact protein, most of which lies in the membrane.

3.2.4 Biochemical characterization

All the results in this section are very preliminary and therefore the results may not be conclusive. Repetitions and optimization of experiments are underway.

3.2.4.1 State of oligomerization – Blue native PAGE

The state and nature of oligomerization can be very informative regarding the functional aspects of the protein. Blue-native PAGE was used to analyze if Cln3p forms higher order structures.

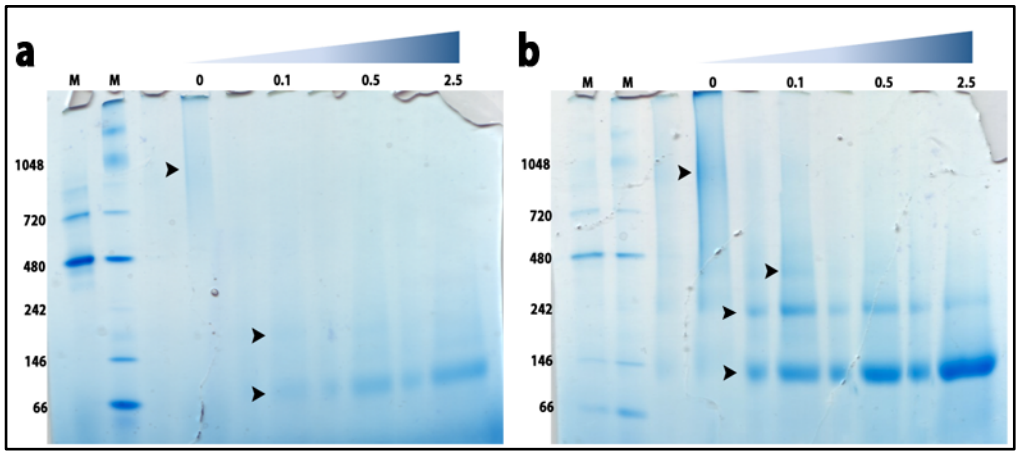


Figure 3.34: Blue-native PAGE analysis of Cln3p. Blue-native PAGE of the cleaved hCln3p in **(a)** and the mCherry fusion protein in **(b).** M: specific blue native marker. Color gradient illustrates increasing SDS concentration in %. In both the cases bands appear at weights higher than the molecular weight of protein (bands are indicated by black arrowheads). Both the proteins are in MNG-3 micelles.

Cln3p (both cleaved and uncleaved versions of protein) has a tendency to aggregate as indicated by the smear and the protein seen in wells of blue-native PAGE. In the presence of SDS (known to disrupt the formation of oligomers/aggregates, used in an increasing amount from 0 to 2.5% (w/v) the intensity of the smear is lost and is compensated by the intensity of the monomeric bands. Representative gels for cleaved hCln3p and mCherry fusion protein can be seen in **Figure 3.34, a** and **b** respectively.

Although not completely precise (due to the smear, caused most likely due to the detergent micelles) it can still be inferred that Cln3p forms some kind of higher order oligomers. The predicted dimer size of the fusion protein should be around 142 kDa. The stability of these oligomers is reflected by the fact that even at high SDS concentration (2.5%), a certain portion was still in higher oligomers are visible. The nature and the precise size of oligomers need to be confirmed with authority. Various bands in the blue-native gel must be analyzed by mass spectrometry to get a clearer estimate of the size. Additional information may be inferred from analytical ultracentrifugation experiments.

3.2.4.2 Post-translational modifications – Glycosylation

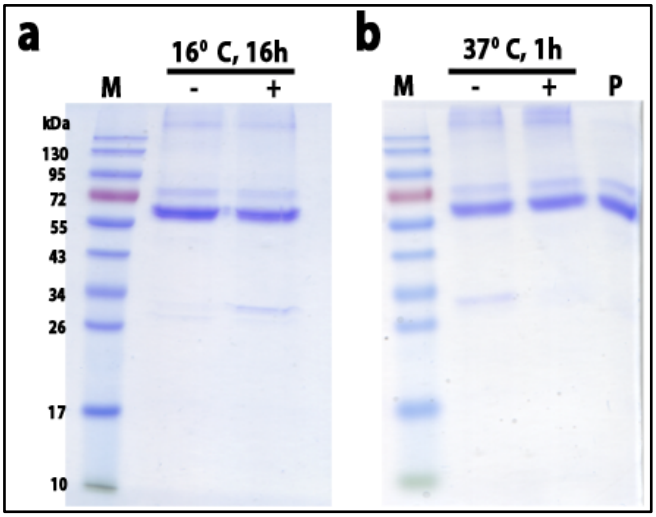


Figure 3.35: Glycosylation analysis of Cln3p. Protocol for analysis of extent of glycosylation differs slightly for membrane proteins. The conditions are generally mild. Two sets of reactions were carried out: at 16°C for 16 h **(a)** and 37°C for 1 h **(b).** Semi-denaturing SDS-PAGE gels for both the reactions are shown. '-', Without PNGase and '+' with PNGase F. M, marker. P, protein not subjected to reaction. The protein used for the assay was mCherry-hCln3p.

Post-translational modifications (PTMs) of a protein can be decisive in protein folding, stability, intracellular localization and catalytic activity. However, for structural studies especially for crystallization glycosylation is not favorable. Also, products of glycosylation can vary depending on the level and degree of glycosylation leading to a mixed protein population. Therefore it becomes essential to know the state of glycosylation of eukaryotic proteins. PNGase F (a de-glycosylating enzyme) to address this question. This enzyme removes N-glycosylation from a protein with a post-translational modification. The protocol for membrane proteins was modified as compared to the conventional proteins as heating leads to aggregation of these proteins in general. The results are shown in **Figure 3.35**.

No change in the size of protein upon PNGase F treatment was seen. This behavior was also independent of the temperature of the reaction. Analysis of membrane proteins for glycosylation can be tricky as the putative sites may be masked by the detergent micelles rendering the assay invalid. Another problem with the set-up of the experiment is that there is no positive control (other membrane protein that undergoes glycosylation). To reaffirm that Cln3p does not undergo glycosylation as a posttranslational modification when overexpressed in insect cells, further analysis is necessary. Experiments such as, use of lectin columns to separate glycosylated proteins (if any) and mutating the glycosylation sites to check for the difference in size of protein expressed, need to be done. A final approach would be to check for sugar modifications by thin layer chromatography (TLC), where one can use spray reagents to identify sugar modifications on proteins.

3.2.4.3 Post-translational modifications – Phosphorylation

Phosphorylation is another very important and interesting post-translational modification, which at times is necessary to trigger the activity of the protein. The cellular machinery for disrupting interactions between proteins or complexes can, at times, employ the introduction of negatively charged phosphate group. Phosphoimaging was used to understand whether Cln3p undergoes phosphorylation as a post-translational modification. This technique relies on the principle of

recognition of phosphorylated amino acids by a specific dye. The protein γ-phosphatase is an enzyme capable of removing phosphorylation, generally used to confirm that the protein has indeed undergone the PTM.

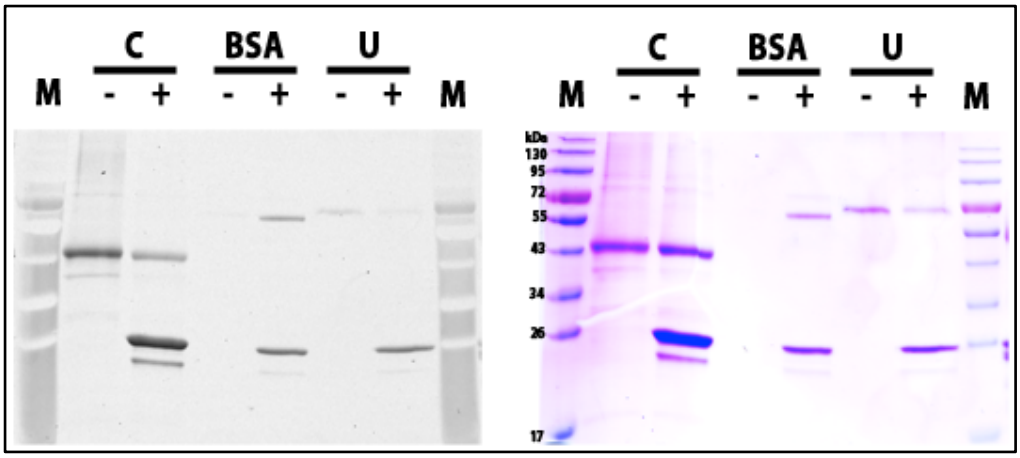


Figure 3.36: Phosphorylation analysis of Cln3p. Phosphoimaging of hCln3p (left) and corresponding semi-denaturing SDS-PAGE on the right. Cln3p undergoes phosphorylation and this decreases in the presence of γ-phosphatase (+). For some unknown reason protein is not visible in the absence of γ-phosphatase in the case of BSA ('-' lane, used as a control). C, cleaved protein. U, mCherry-Cln3p. M, marker. '+', Reaction γ-phosphatase reflected by the band at 26 kda (right). '-', Without γ-phosphatase.

Cln3p is phosphorylated. This is evident from the bands detected in the phosphoimaging, which correspond to the protein in the semi-denaturing SDS-PAGE. Upon the use of γ-phosphatase, there is a significant decrease of the intensity for both the cleaved and the fusion protein. The failure of the control reaction is speculative. These are the newest results towards characterization of the protein and need to be repeated properly. Removal of phosphorylation can used as a strategy for crystallization of the protein as removal of negative charges on the surface may favor protein interactions.

3.2.5 Electron Microscopy

3.2.5.1 Single particle analysis

In an attempt to visualize and understand the structure of the protein, I performed negative stain analysis on EFGP-hCln3p purified in Fos-13 and mCherry-hCln3p fusion protein purified in MNG-3. I was not able to see characteristic particles in both the cases. In the case of Fos-13 very heterogeneous particles and few possible aggregates were seen **(Figure 3.37, a)**. The latter case of MNG-3 revealed 'ribbon-like' filamentous particles and hardly any single particles **(Figure 3.37, b)**. One can only explain this behavior postulating that hCln3p in MNG-3 forms 1D micelles.

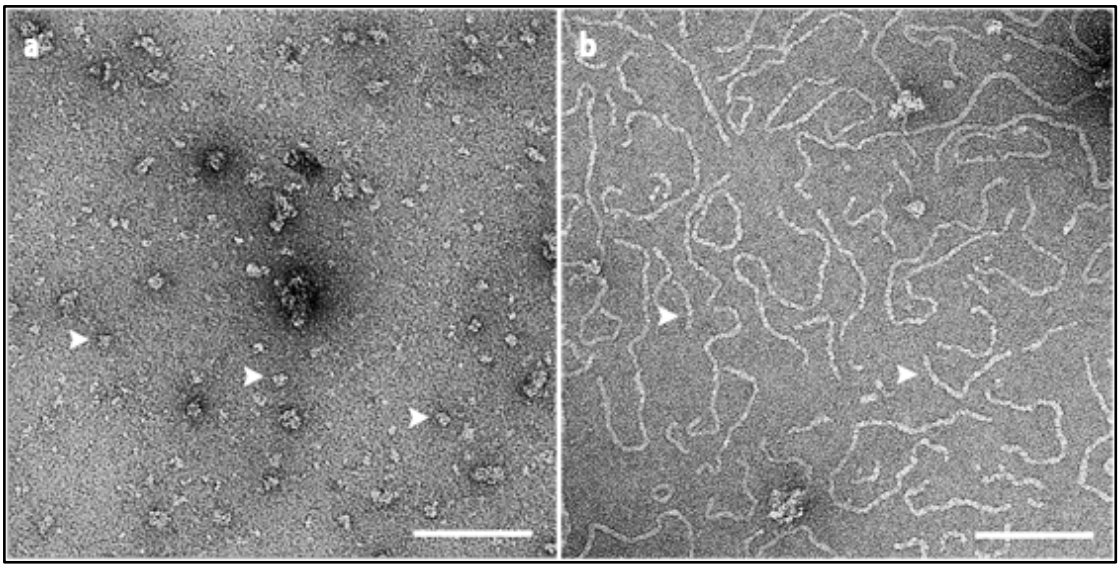


Figure 3.37: Negative stain analysis of hCln3p. (a) Representative micrograph area of EGFP-hCln3p purified in Fos-13. Sample was diluted to 0.01 mg/mL and loaded on to a carbon-coated grid. No characteristic shape and size of particles found. Possible particles are highlighted with white arrowhead. Scale bar, 100 nm. **(b**) Representative micrograph area of mCherry-Cln3p purified in MNG-3. Sample was diluted to 0.01 mg/mL and loaded on to a carbon-coated grid. Long filaments, possibly the protein in linear micelles in MNG-3 are seen. Scale bar, 100 nm

Recent advances in the field of electron microscopy enables one to determine structure of large membrane proteins at atomic resolution [176], where amphipathic polymers (amphipols) are used instead of detergents. Exchanging Fos-13 with amphipols after solubilization of the protein would be ideal starting point for future efforts. Single particle analysis by cryo-electron microscopy (cryo-EM) requires the protein to be around 300 kDa. Cln3p shows a tendency to form higher order structures (the nature these oligomers needs to be studied in detail), which may make it suitable for structure determination by cryo-EM.

3.2.5.2 Electron crystallography

Electron crystallography (2D crystallography) is an alternative to X-ray crystallography particularly used for studying membrane proteins that can organize into 2D arrays in presence of lipids. Membrane proteins are invariably always purified in detergents that are not so conducive for X-ray crystallography studies. In 2D crystallography membrane proteins purified in detergents are incubated with lipids and detergents are gradually removed over a period of time. I used variety of lipids at various LPRs (lipid to protein ratio) and set up trials with all the versions of the proteins purified. Unfortunately, none of the conditions reveal any potential crystals. A snapshot of one such screen set up with the EGFP-hCln3p (p232) is shown in **Figure 3.38.**

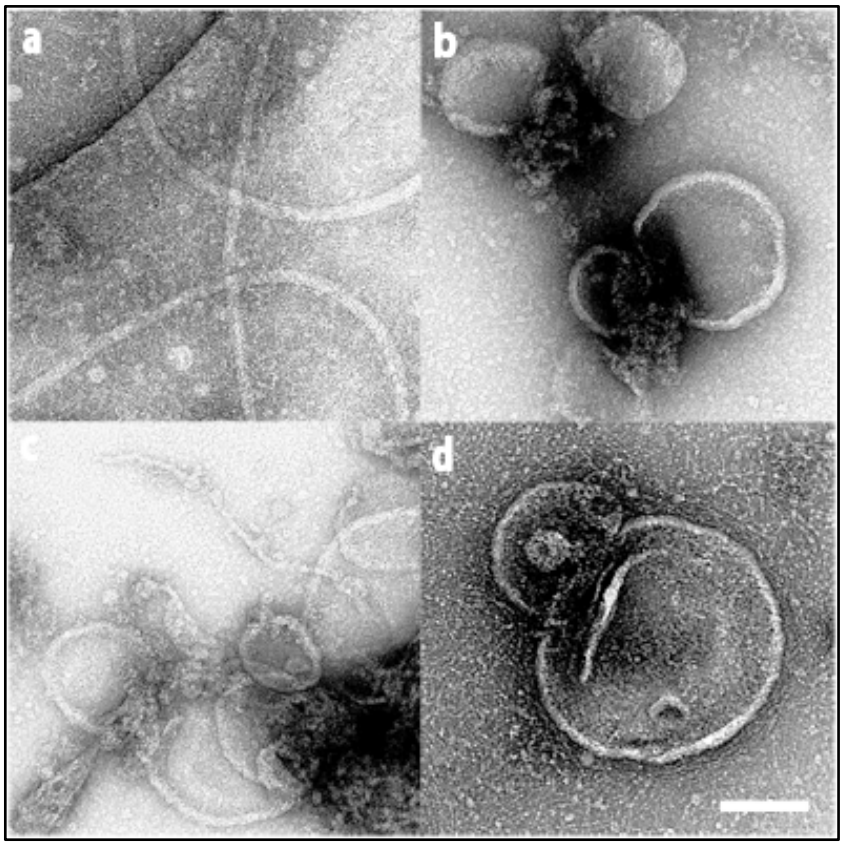


Figure 3.38: 2D crystallography trials on EGFP-hCln3p. Purified protein at a concentration of 1mg/mL was mixed with lipids at various lipid to protein ratios (w/w) and dialyzed to remove the detergent for a span of 2-3 weeks. Results of one such trial of EGFP-hCln3p purified in Fos-13. LPR is 0.5 in all the cases **(a)** DOPC **(b)** BPL **(c)** EPL **(d)** DMPC. Scale bar, 100 nm.

3.2.6 X-ray crystallography

All the purifications that yielded protein in sufficient amounts were used to set up X-ray crystallography trials irrespective of whether the tag was cleaved or not. Various suites available from Qiagen were used. Some of the initial trials revealed promising hits, which need to be optimized further.

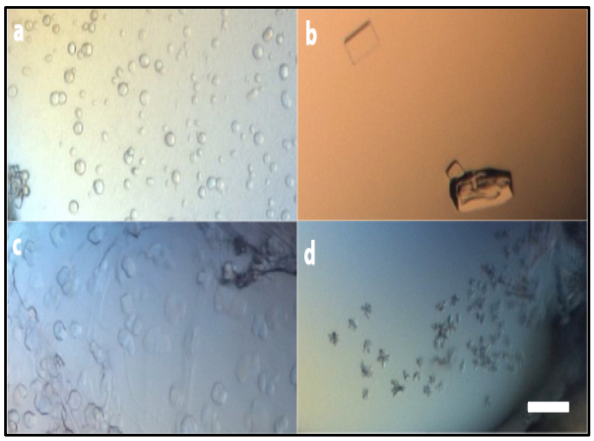


Figure 3.39: X-ray crystallography trials of Cln3p. Representative hits from the various trials set up. The construct, concentration of protein and the conditions are listed. **(a)** mCherry-Cln3p; suite: Classics II; 4 mg/mL; 30% Jeffamine M-600, 0.1 M HEPES, pH = 7 **(b)** EGFP-Cln3p; suite: pHclear II; 1 mg/mL; 1 M LiCl, 0.1 M Sodium Acetate Anhydrous, pH = 5 **(c)** mCherry-Cln3p; suite: pHclear II: 11.5 mg/mL; 30% PEG 6000, 1 M LiCl, 0,1 M TRIS, pH = 8 **(d)** mCherry-Cln3p; suite: JCSG Core IV; 4 mg/mL; 1 M $(NH_4)_2HPO_3$, 0.2 M NaCl 0.1 M Imidazole, pH = 8

3.2.7 Localization studies

Localization of Cln3p has been studied and debated for a long time. The localization of Cln3p relies largely on the method of detection, type of cell lines and the nature of the probe. The human *cln3* gene was cloned into the pOPINE vector harboring an N-terminal EGFP. Curious to identify the localization, since Battens disease affects primarily the neurons, I studied its localization pattern in neuronal N2a cell-line. I also wanted to establish the localization pattern and see whether it gets affected in some of the disease causing mutants I intended to generate. Cln3p has been reported to transcend a variety of intra-cellular organelles, primarily lysosomes and the Golgi apparatus. I checked for localization in these compartments using appropriate sub-cellular markers (lysosomal associated membrane protein (LAMP) as the lysosomal marker, GalT was used as the Golgi marker and peripheral myelin protein (PMP2) as the peroxisomal marker). Cells were co-transfected with 1 µg of the plasmid using lipofectamine. Cln3p localizes primarily to lysosomal compartment in neuronal N2a cells. No localization was seen in the Golgi or the peroxisome. The results are summarized in **Figure 3.40.**

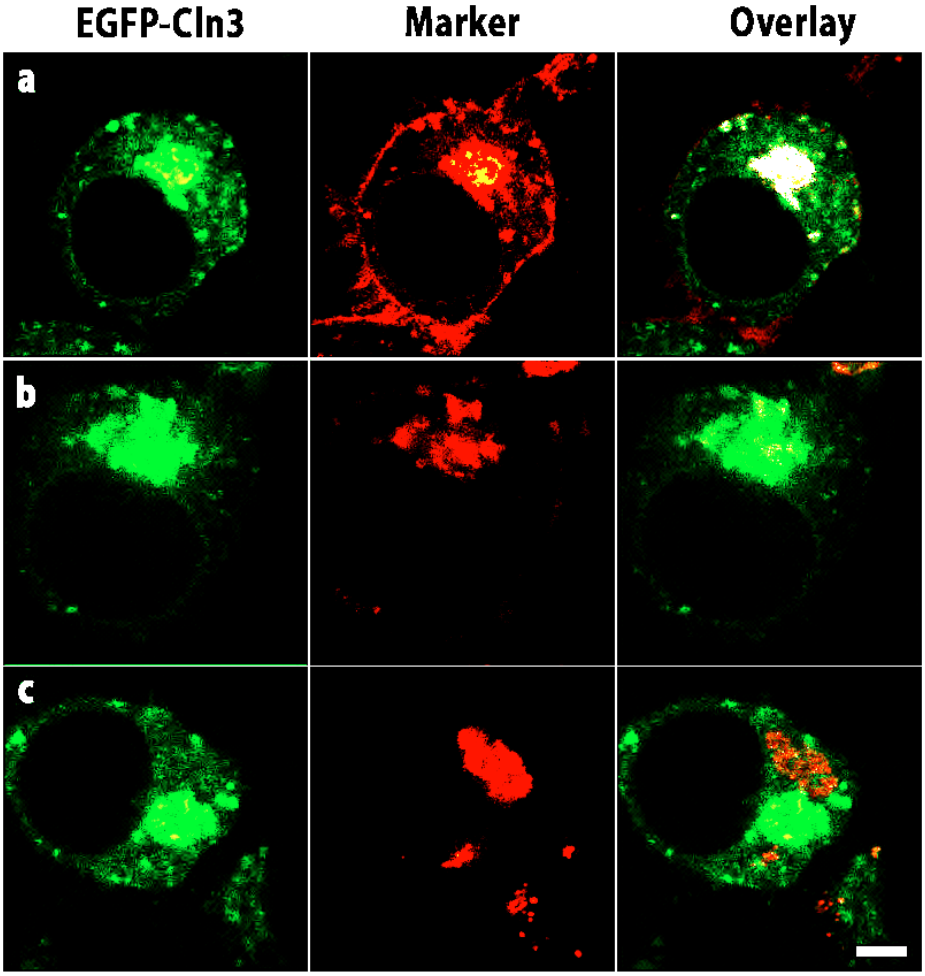


Figure 3.40: Localization of Cln3p. Confocal fluorescence images of N2a neuronal cells transiently transfected with Cln3p (EGFP shown in green and corresponding organelle marker in red. Lysosome associated membrane protein (LAMP) in **(a)** The Golgi marker GalT in **(b)** and peroxisomal membrane protein (PMP2) in **(c)**. An overlay of both the channels is represented in the right. Transfection was carried out with lipofectamine. Imaging was done both on fixed and live-cells to obtain very similar results. Scale bar, 5 µm.

3.2.8 Functional studies: Establishing Desaturase assay

Cln3p has been implicated in desaturase activity of palmitoyl moieties with a preference towards palmitoylated N-Ras [152]. With purified Cln3p, reproducible and in sufficient amounts, I aimed at reconstituting this activity *in vitro*. In collaboration with Dr. Gemma Triola (AG Waldmann, MPI-Dortmund) I started with synthesis of the simpler substrate, N-acetyl palmitoyl cysteine (schematic given below).

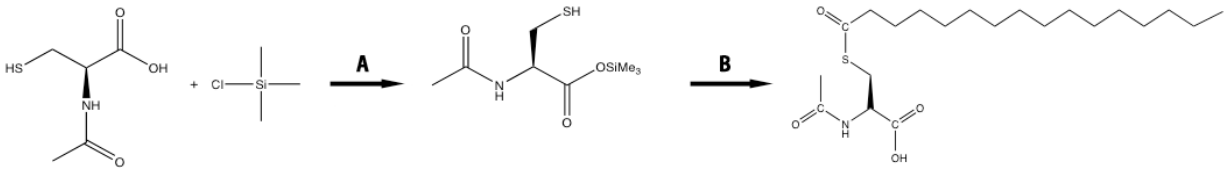


Figure 3.41: Outline of the synthesis of N-acetyl palmitoyl cysteine (NAPC). **(A)** Dichloromethane (DCM), 50^0 C, Reflux **(B)** Palmitoyl Chloride, Triethylamine, 3h, RT, DCM. The heptadecanoic acid derivatives were prepared along similar lines. Work done in collaboration with Dr. Gemma Triola.

Establishing the protocol was necessary. Gas chromatography coupled with mass spectrometry (GC-MS) was chosen as a method of choice for monitoring the reaction. GC is a common analytical method for separating different compounds in a mixture that can be vaporized without decomposition. The reaction is carried out in the aqueous phase followed by the extraction of end products into the organic phase (chloroform). The reaction is terminated by addition of NaOH, which hydrolyses the end product to palmitoleic acid. The retention volumes of palmitic, palmitoleic and pentadecanoic acid (used as an internal standard) can be estimated by running them individually in GC and area under the curve gives us the concentration of the product. Therefore one can estimate, in principle the extent of reaction by running the organic phase in GC. Sensitivity of the GC was calibrated, by running palmitic and palmitoleic acid standards varying between concentrations of 1 nM to 500 µM and it was estimated that the detection works best when the concentration is in the range of 10 µM. The samples and sample mixtures were analyzed in full scan MS mode. The most abundant peaks for each of the sample m/z 299 (pentadecanoic acid), m/z 311 (palmitoleic acid) and m/z 313 (palmitic acid) were determined, following which selected ion monitoring method (SIM) was chosen for more specificity.

To summarize, the protocol for monitoring the desaturase assay *in vitro* has been established (refer **section 2.2.2.18** for detailed description). Initial substrates required for the reaction have been synthesized and attempts to reconstitute and monitor the desaturase assay *in vitro* are well underway. The design of the assay, however, faces two setbacks. One, it is really not known if other proteins, electron-carriers and co-factors are necessary for the function of Cln3p, as there may be missing partners for the red-ox reaction being reconstituted. Two, the use of whole/crude membrane from insect cells overexpressing Cln3p can circumvent the problem of necessity of other proteins but the lipids from the membrane will interfere with the substrate and therefore monitoring and quantification of reaction products becomes difficult. To circumvent these problems non-natural (odd-

carbons) fatty acids like heptadecanoic acid could be used as substrates together with crude membranes. Another alternative would be to use *in vitro* redox systems in which one could couple the oxidation reaction carried out by Cln3p with an electron recycling system (for example glutathione reductase kit, where one could ideally supplement the $NADPH^+$ by Cln3p and its substrate). However, once again *in vitro* systems use either FAD (Flavin adenine dinucleotide) or NAD (Nicotinamide adenine dinucleotide) as substrates and there is no putative binding site for such moieties in Cln3p leading to a large speculation as to whether desaturase activity is the primary function carried out by Cln3p and if loss of this activity underlies the Batten's disease.

3.2.9 Towards the structure of Cln3p

Integral membrane protein structures do not appear every day. The task of purifying these proteins in an environment conducive for structural studies is daunting. An enormous amount of time was spent in rendering Cln3p, protein implicated in Batten's disease, soluble. I report the first purification of Cln3p in amphiphiles and am inching towards determination of its structure. Although initial trials to understand its structure have failed, other methods (such as lipid cubic phase (LCP) crystallography) are yet to be tried. Cln3p is a tricky protein, since its natural substrate is not known. Ligands or inhibitors capable of stabilizing the protein in a specific conformation and facilitating further analysis of the protein must be screened for at Compound Management and Screening Center (**COMAS**, Dortmund). The glycosylation and palmitoylation sites need to be targeted, to mutate them to facilitate crystallography studies. Albeit labor-intensive, antibodies and antibody fragments (nanobodies) that may help facilitate crystallization attempts need to be produced. In antibody-fragment mediated crystallization the available hydrophilic surface area is enlarged to form an improved crystal lattice by promoting protein-protein contacts and providing space for the detergent micelle. Additionally, they restrict flexible protein parts. Alternatively, other scaffolds like affibodies or DARPins (designed ankyrin repeat proteins) that are known to mask residues unfavorable for the formation of crystal contacts remain to be tested.

Single particle electron microscopy (SP-EM) is emerging as a very powerful tool for structure determination. Once thought as a technique suitable only for intermediate resolutions SP-EM is now being routinely used to generate near-atomic resolution structures. Although the number of structures is still low in contrast to crystallography, the potential is immense. Attempts to reconstitute Cln3p in amphipols (amphipathic polymers, which make membrane proteins suitable for cryo-electron microscopy) are well underway (briefly discussed in **section 3.2.5.1**).

On the functional front we are well placed with the design of desaturase assay. The assay needs to tried and tested. Another plausible role of Cln3p may be in arginine transport [145] a possibility that has not been investigated during the course of the work. This needs to be studied in the future by radioactive isotope uptake into proteoliposomes. Alternative way to assay this would be by microscale thermophoresis, where change of the hydration shell of biomolecules due to changes in their structure/conformation results in a relative change of movement along the temperature gradient and is used to determine binding affinities.

Finally, understanding the molecular mechanism of the disease is the key intention. We have come a long way having embarked upon this goal, but obtaining the structure of the protein still remains challenging. Someday, not far away in future we intend to elucidate, the structural basis of Batten's disease.

4 References

1. Gordon, D.J., B. Resio, and D. Pellman, *Causes and consequences of aneuploidy in cancer.* Nat Rev Genet, 2012. **13**(3): p. 189-203.
2. Morgan, D.O., *SnapShot: Cell-cycle regulators II.* Cell, 2008. **135**(5): p. 974-974 e1.
3. Morgan, D.O., *SnapShot: cell-cycle regulators I.* Cell, 2008. **135**(4): p. 764-764 e1.
4. Walczak, C.E., S. Cai, and A. Khodjakov, *Mechanisms of chromosome behaviour during mitosis.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 91-102.
5. Remeseiro, S. and A. Losada, *Cohesin, a chromatin engagement ring.* Curr Opin Cell Biol, 2013. **25**(1): p. 63-71.
6. Wittmann, T., A. Hyman, and A. Desai, *The spindle: a dynamic assembly of microtubules and motors.* Nat Cell Biol, 2001. **3**(1): p. E28-34.
7. Westermann, S., D.G. Drubin, and G. Barnes, *Structures and functions of yeast kinetochore complexes.* Annu Rev Biochem, 2007. **76**: p. 563-91.
8. McAinsh, A.D., J.D. Tytell, and P.K. Sorger, *Structure, function, and regulation of budding yeast kinetochores.* Annu Rev Cell Dev Biol, 2003. **19**: p. 519-39.
9. Zinkowski, R.P., J. Meyne, and B.R. Brinkley, *The centromere-kinetochore complex: a repeat subunit model.* J Cell Biol, 1991. **113**(5): p. 1091-110.
10. Joglekar, A.P., et al., *Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres.* J Cell Biol, 2008. **181**(4): p. 587-94.
11. Dong, Y., et al., *The outer plate in vertebrate kinetochores is a flexible network with multiple microtubule interactions.* Nat Cell Biol, 2007. **9**(5): p. 516-22.
12. Gonen, S., et al., *The structure of purified kinetochores reveals multiple microtubule-attachment sites.* Nat Struct Mol Biol, 2012. **19**(9): p. 925-9.
13. Foltz, D.R., et al., *The human CENP-A centromeric nucleosome-associated complex.* Nat Cell Biol, 2006. **8**(5): p. 458-69.
14. Okada, M., et al., *The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres.* Nat Cell Biol, 2006. **8**(5): p. 446-57.
15. Hori, T., et al., *CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore.* Cell, 2008. **135**(6): p. 1039-52.
16. Perpelescu, M. and T. Fukagawa, *The ABCs of CENPs.* Chromosoma, 2011. **120**(5): p. 425-46.
17. Musacchio, A. and E.D. Salmon, *The spindle-assembly checkpoint in space and time.* Nat Rev Mol Cell Biol, 2007. **8**(5): p. 379-93.
18. Cheeseman, I.M., et al., *The conserved KMN network constitutes the core microtubule-binding site of the kinetochore.* Cell, 2006. **127**(5): p. 983-97.
19. Ruchaud, S., M. Carmena, and W.C. Earnshaw, *Chromosomal passengers: conducting cell division.* Nat Rev Mol Cell Biol, 2007. **8**(10): p. 798-812.
20. Wan, X., et al., *Protein architecture of the human kinetochore microtubule attachment site.* Cell, 2009. **137**(4): p. 672-84.
21. Meraldi, P., et al., *Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins.* Genome Biol, 2006. **7**(3): p. R23.
22. Lyle, K., P. Kumar, and T. Wittmann, *SnapShot: Microtubule Regulators I.* Cell, 2009. **136**(2): p. 380, 380 e1.
23. Lyle, K., P. Kumar, and T. Wittmann, *SnapShot: Microtubule regulators II.* Cell, 2009. **136**(3): p. 566, 566 e1.
24. Cheeseman, I.M., et al., *A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension.* Genes Dev, 2004. **18**(18): p. 2255-68.

25. Desai, A., et al., *KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in C. elegans.* Genes Dev, 2003. **17**(19): p. 2421-35.
26. Cheeseman, I.M. and A. Desai, *Molecular architecture of the kinetochore-microtubule interface.* Nat Rev Mol Cell Biol, 2008. **9**(1): p. 33-46.
27. Varma, D. and E.D. Salmon, *The KMN protein network--chief conductors of the kinetochore orchestra.* J Cell Sci, 2012. **125**(Pt 24): p. 5927-36.
28. Kline, S.L., et al., *The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation.* J Cell Biol, 2006. **173**(1): p. 9-17.
29. DeLuca, J.G., et al., *Hec1 and nuf2 are core components of the kinetochore outer plate essential for organizing microtubule attachment sites.* Mol Biol Cell, 2005. **16**(2): p. 519-31.
30. Petrovic, A., et al., *Modular Assembly of RWD Domains on the Mis12 Complex Underlies Outer Kinetochore Organization.* Mol Cell, 2014. **53**(4): p. 591-605.
31. DeLuca, J.G. and A. Musacchio, *Structural organization of the kinetochore-microtubule interface.* Curr Opin Cell Biol, 2012. **24**(1): p. 48-56.
32. Uchida, K.S., et al., *Kinetochore stretching inactivates the spindle assembly checkpoint.* J Cell Biol, 2009. **184**(3): p. 383-90.
33. Zhou, J., J. Yao, and H.C. Joshi, *Attachment and tension in the spindle assembly checkpoint.* J Cell Sci, 2002. **115**(Pt 18): p. 3547-55.
34. Skibbens, R.V., V.P. Skeen, and E.D. Salmon, *Directional instability of kinetochore motility during chromosome congression and segregation in mitotic newt lung cells: a push-pull mechanism.* J Cell Biol, 1993. **122**(4): p. 859-75.
35. Maiato, H., et al., *The dynamic kinetochore-microtubule interface.* J Cell Sci, 2004. **117**(Pt 23): p. 5461-77.
36. Musacchio, A. and K.G. Hardwick, *The spindle checkpoint: structural insights into dynamic signalling.* Nat Rev Mol Cell Biol, 2002. **3**(10): p. 731-41.
37. Li, R. and A.W. Murray, *Feedback control of mitosis in budding yeast.* Cell, 1991. **66**(3): p. 519-31.
38. Hoyt, M.A., L. Totis, and B.T. Roberts, *S. cerevisiae genes required for cell cycle arrest in response to loss of microtubule function.* Cell, 1991. **66**(3): p. 507-17.
39. Kops, G.J., B.A. Weaver, and D.W. Cleveland, *On the road to cancer: aneuploidy and the mitotic checkpoint.* Nat Rev Cancer, 2005. **5**(10): p. 773-85.
40. Abrieu, A., et al., *Mps1 is a kinetochore-associated kinase essential for the vertebrate mitotic checkpoint.* Cell, 2001. **106**(1): p. 83-93.
41. Kallio, M.J., et al., *Inhibition of aurora B kinase blocks chromosome segregation, overrides the spindle checkpoint, and perturbs microtubule dynamics in mitosis.* Curr Biol, 2002. **12**(11): p. 900-5.
42. Howell, B.J., et al., *Cytoplasmic dynein/dynactin drives kinetochore protein transport to the spindle poles and has a role in mitotic spindle checkpoint inactivation.* J Cell Biol, 2001. **155**(7): p. 1159-72.
43. Tai, C.Y., et al., *Role of dynein, dynactin, and CLIP-170 interactions in LIS1 kinetochore function.* J Cell Biol, 2002. **156**(6): p. 959-68.
44. Xia, G., et al., *Conformation-specific binding of p31(comet) antagonizes the function of Mad2 in the spindle checkpoint.* EMBO J, 2004. **23**(15): p. 3133-43.
45. Minshull, J., et al., *A MAP kinase-dependent spindle assembly checkpoint in Xenopus egg extracts.* Cell, 1994. **79**(3): p. 475-86.
46. Lara-Gonzalez, P., F.G. Westhorpe, and S.S. Taylor, *The spindle assembly checkpoint.* Curr Biol, 2012. **22**(22): p. R966-80.

47. Sudakin, V., G.K. Chan, and T.J. Yen, *Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2.* J Cell Biol, 2001. **154**(5): p. 925-36.
48. Hardwick, K.G., et al., *MAD3 encodes a novel component of the spindle checkpoint which interacts with Bub3p, Cdc20p, and Mad2p.* J Cell Biol, 2000. **148**(5): p. 871-82.
49. Shannon, K.B., J.C. Canman, and E.D. Salmon, *Mad2 and BubR1 function in a single checkpoint pathway that responds to a loss of tension.* Mol Biol Cell, 2002. **13**(10): p. 3706-19.
50. Hwang, L.H., et al., *Budding yeast Cdc20: a target of the spindle checkpoint.* Science, 1998. **279**(5353): p. 1041-4.
51. Fang, G., H. Yu, and M.W. Kirschner, *The checkpoint protein MAD2 and the mitotic regulator CDC20 form a ternary complex with the anaphase-promoting complex to control anaphase initiation.* Genes Dev, 1998. **12**(12): p. 1871-83.
52. Peters, J.M., *The anaphase promoting complex/cyclosome: a machine designed to destroy.* Nat Rev Mol Cell Biol, 2006. **7**(9): p. 644-56.
53. Nakajima, M., et al., *The complete removal of cohesin from chromosome arms depends on separase.* J Cell Sci, 2007. **120**(Pt 23): p. 4188-96.
54. De Antoni, A., et al., *The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint.* Curr Biol, 2005. **15**(3): p. 214-25.
55. Sironi, L., et al., *Crystal structure of the tetrameric Mad1-Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint.* EMBO J, 2002. **21**(10): p. 2496-506.
56. DeAntoni, A., V. Sala, and A. Musacchio, *Explaining the oligomerization properties of the spindle assembly checkpoint protein Mad2.* Philos Trans R Soc Lond B Biol Sci, 2005. **360**(1455): p. 637-47, discussion 447-8.
57. Mapelli, M., et al., *Determinants of conformational dimerization of Mad2 and its inhibition by p31comet.* EMBO J, 2006. **25**(6): p. 1273-84.
58. Westhorpe, F.G., et al., *p31comet-mediated extraction of Mad2 from the MCC promotes efficient mitotic exit.* J Cell Sci, 2011. **124**(Pt 22): p. 3905-16.
59. Uzunova, K., et al., *APC15 mediates CDC20 autoubiquitylation by APC/C(MCC) and disassembly of the mitotic checkpoint complex.* Nat Struct Mol Biol, 2012. **19**(11): p. 1116-23.
60. Liu, D., et al., *Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase.* J Cell Biol, 2010. **188**(6): p. 809-20.
61. Rosenberg, J.S., F.R. Cross, and H. Funabiki, *KNL1/Spc105 recruits PP1 to silence the spindle assembly checkpoint.* Curr Biol, 2011. **21**(11): p. 942-7.
62. Welburn, J.P., et al., *Aurora B phosphorylates spatially distinct targets to differentially regulate the kinetochore-microtubule interface.* Mol Cell, 2010. **38**(3): p. 383-92.
63. Vanoosthuyse, V. and K.G. Hardwick, *A novel protein phosphatase 1-dependent spindle checkpoint silencing mechanism.* Curr Biol, 2009. **19**(14): p. 1176-81.
64. Hauf, S., et al., *The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint.* J Cell Biol, 2003. **161**(2): p. 281-94.
65. Cimini, D., et al., *Aurora kinase promotes turnover of kinetochore microtubules to reduce chromosome segregation errors.* Curr Biol, 2006. **16**(17): p. 1711-8.
66. Lampson, M.A., et al., *Correcting improper chromosome-spindle attachments during cell division.* Nat Cell Biol, 2004. **6**(3): p. 232-7.

67. Tanaka, T.U., et al., *Evidence that the Ipl1-Sli15 (Aurora kinase-INCENP) complex promotes chromosome bi-orientation by altering kinetochore-spindle pole connections.* Cell, 2002. **108**(3): p. 317-29.
68. Pinsky, B.A., et al., *The Ipl1-Aurora protein kinase activates the spindle checkpoint by creating unattached kinetochores.* Nat Cell Biol, 2006. **8**(1): p. 78-83.
69. Vader, G., et al., *The chromosomal passenger complex controls spindle checkpoint function independent from its role in correcting microtubule kinetochore interactions.* Mol Biol Cell, 2007. **18**(11): p. 4553-64.
70. Ruchaud, S., M. Carmena, and W.C. Earnshaw, *The chromosomal passenger complex: one for all and all for one.* Cell, 2007. **131**(2): p. 230-1.
71. Maresca, T.J. and E.D. Salmon, *Intrakinetochore stretch is associated with changes in kinetochore phosphorylation and spindle assembly checkpoint activity.* J Cell Biol, 2009. **184**(3): p. 373-81.
72. Liu, D. and M.A. Lampson, *Regulation of kinetochore-microtubule attachments by Aurora B kinase.* Biochem Soc Trans, 2009. **37**(Pt 5): p. 976-80.
73. Liu, D., et al., *Sensing chromosome bi-orientation by spatial separation of aurora B kinase from kinetochore substrates.* Science, 2009. **323**(5919): p. 1350-3.
74. Santaguida, S. and A. Musacchio, *The life and miracles of kinetochores.* EMBO J, 2009. **28**(17): p. 2511-31.
75. Smith, D.A., B.S. Baker, and M. Gatti, *Mutations in genes encoding essential mitotic functions in Drosophila melanogaster.* Genetics, 1985. **110**(4): p. 647-70.
76. Karess, R.E. and D.M. Glover, *rough deal: a gene required for proper mitotic segregation in Drosophila.* J Cell Biol, 1989. **109**(6 Pt 1): p. 2951-61.
77. Williams, B.C., et al., *The Drosophila l(1)zw10 gene product, required for accurate mitotic chromosome segregation, is redistributed at anaphase onset.* J Cell Biol, 1992. **118**(4): p. 759-73.
78. Starr, D.A., et al., *ZW10 helps recruit dynactin and dynein to the kinetochore.* J Cell Biol, 1998. **142**(3): p. 763-74.
79. Chan, G.K., et al., *Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores.* Nat Cell Biol, 2000. **2**(12): p. 944-7.
80. Scaerou, F., et al., *The ZW10 and Rough Deal checkpoint proteins function together in a large, evolutionarily conserved complex targeted to the kinetochore.* J Cell Sci, 2001. **114**(Pt 17): p. 3103-14.
81. Williams, B.C., et al., *Zwilch, a new component of the ZW10/ROD complex required for kinetochore functions.* Mol Biol Cell, 2003. **14**(4): p. 1379-91.
82. Wang, H., et al., *Human Zwint-1 specifies localization of Zeste White 10 to kinetochores and is essential for mitotic checkpoint signaling.* J Biol Chem, 2004. **279**(52): p. 54590-8.
83. Buffin, E., et al., *Recruitment of Mad2 to the kinetochore requires the Rod/Zw10 complex.* Curr Biol, 2005. **15**(9): p. 856-61.
84. Civril, F., et al., *Structural analysis of the RZZ complex reveals common ancestry with multisubunit vesicle tethering machinery.* Structure, 2010. **18**(5): p. 616-26.
85. Cheerambathur, D.K., et al., *Crosstalk between microtubule attachment complexes ensures accurate chromosome segregation.* Science, 2013. **342**(6163): p. 1239-42.
86. Wojcik, E., et al., *Kinetochore dynein: its dynamics and role in the transport of the Rough deal checkpoint protein.* Nat Cell Biol, 2001. **3**(11): p. 1001-7.
87. Basto, R., et al., *In vivo dynamics of the rough deal checkpoint protein during Drosophila mitosis.* Curr Biol, 2004. **14**(1): p. 56-61.

88. Starr, D.A., et al., *HZwint-1, a novel human kinetochore component that interacts with HZW10.* J Cell Sci, 2000. **113 (Pt 11)**: p. 1939-50.
89. Kops, G.J., et al., *ZW10 links mitotic checkpoint signaling to the structural kinetochore.* J Cell Biol, 2005. **169**(1): p. 49-60.
90. Basto, R., R. Gomes, and R.E. Karess, *Rough deal and Zw10 are required for the metaphase checkpoint in Drosophila.* Nat Cell Biol, 2000. **2**(12): p. 939-43.
91. Starr, D.A., et al., *Conservation of the centromere/kinetochore protein ZW10.* J Cell Biol, 1997. **138**(6): p. 1289-301.
92. Williams, B.C. and M.L. Goldberg, *Determinants of Drosophila zw10 protein localization and function.* J Cell Sci, 1994. **107 (Pt 4)**: p. 785-98.
93. Arasaki, K., et al., *RINT-1 regulates the localization and entry of ZW10 to the syntaxin 18 complex.* Mol Biol Cell, 2006. **17**(6): p. 2780-8.
94. Arasaki, K., et al., *Correlation of Golgi localization of ZW10 and centrosomal accumulation of dynactin.* Biochem Biophys Res Commun, 2007. **359**(3): p. 811-6.
95. Wainman, A., et al., *The Drosophila RZZ complex - roles in membrane trafficking and cytokinesis.* J Cell Sci, 2012. **125**(Pt 17): p. 4014-25.
96. Ren, Y., et al., *A structure-based mechanism for vesicle capture by the multisubunit tethering complex Dsl1.* Cell, 2009. **139**(6): p. 1119-29.
97. Tripathi, A., et al., *Structural characterization of Tip20p and Dsl1p, subunits of the Dsl1p vesicle tethering complex.* Nat Struct Mol Biol, 2009. **16**(2): p. 114-23.
98. Tripathi, A., K.V. Kumar, and S.K. Chaube, *Meiotic cell cycle arrest in mammalian oocytes.* J Cell Physiol, 2010. **223**(3): p. 592-600.
99. Basu, J., et al., *Mutations in the essential spindle checkpoint gene bub1 cause chromosome missegregation and fail to block apoptosis in Drosophila.* J Cell Biol, 1999. **146**(1): p. 13-28.
100. Liu, S.T., et al., *Human MPS1 kinase is required for mitotic arrest induced by the loss of CENP-E from kinetochores.* Mol Biol Cell, 2003. **14**(4): p. 1638-51.
101. Tavormina, P.A., Y. Wang, and D.J. Burke, *Differential requirements for DNA replication in the activation of mitotic checkpoints in Saccharomyces cerevisiae.* Mol Cell Biol, 1997. **17**(6): p. 3315-22.
102. Sharp, D.J., G.C. Rogers, and J.M. Scholey, *Cytoplasmic dynein is required for poleward chromosome movement during mitosis in Drosophila embryos.* Nat Cell Biol, 2000. **2**(12): p. 922-30.
103. Steuer, E.R., et al., *Localization of cytoplasmic dynein to mitotic spindles and kinetochores.* Nature, 1990. **345**(6272): p. 266-8.
104. Wordeman, L., et al., *Chemical subdomains within the kinetochore domain of isolated CHO mitotic chromosomes.* J Cell Biol, 1991. **114**(2): p. 285-94.
105. Pfarr, C.M., et al., *Cytoplasmic dynein is localized to kinetochores during mitosis.* Nature, 1990. **345**(6272): p. 263-5.
106. Hoffman, D.B., et al., *Microtubule-dependent changes in assembly of microtubule motor proteins and mitotic spindle checkpoint proteins at PtK1 kinetochores.* Mol Biol Cell, 2001. **12**(7): p. 1995-2009.
107. Kiyomitsu, T. and I.M. Cheeseman, *Cortical dynein and asymmetric membrane elongation coordinately position the spindle in anaphase.* Cell, 2013. **154**(2): p. 391-402.
108. Yamamoto, A. and Y. Hiraoka, *Cytoplasmic dynein in fungi: insights from nuclear migration.* J Cell Sci, 2003. **116**(Pt 22): p. 4501-12.

109. Griffis, E.R., N. Stuurman, and R.D. Vale, *Spindly, a novel protein essential for silencing the spindle assembly checkpoint, recruits dynein to the kinetochore.* J Cell Biol, 2007. **177**(6): p. 1005-15.
110. Gassmann, R., et al., *A new mechanism controlling kinetochore-microtubule interactions revealed by comparison of two dynein-targeting components: SPDL-1 and the Rod/Zwilch/Zw10 complex.* Genes Dev, 2008. **22**(17): p. 2385-99.
111. McCleland, M.L., et al., *The vertebrate Ndc80 complex contains Spc24 and Spc25 homologs, which are required to establish and maintain kinetochore-microtubule attachment.* Curr Biol, 2004. **14**(2): p. 131-7.
112. Boustany, R.M., J. Alroy, and E.H. Kolodny, *Clinical classification of neuronal ceroid-lipofuscinosis subtypes.* Am J Med Genet Suppl, 1988. **5**: p. 47-58.
113. Rakheja, D., S.B. Narayan, and M.J. Bennett, *The function of CLN3P, the Batten disease protein.* Mol Genet Metab, 2008. **93**(3): p. 269-74.
114. Haltia, M., *The neuronal ceroid-lipofuscinoses.* J Neuropathol Exp Neurol, 2003. **62**(1): p. 1-13.
115. *Isolation of a novel gene underlying Batten disease, CLN3. The International Batten Disease Consortium.* Cell, 1995. **82**(6): p. 949-57.
116. Munroe, P.B., et al., *Spectrum of mutations in the Batten disease gene, CLN3.* Am J Hum Genet, 1997. **61**(2): p. 310-6.
117. Janes, R.W., et al., *A model for Batten disease protein CLN3: functional implications from homology and mutations.* FEBS Lett, 1996. **399**(1-2): p. 75-7.
118. Mao, Q., et al., *Membrane topology of CLN3, the protein underlying Batten disease.* FEBS Lett, 2003. **541**(1-3): p. 40-6.
119. Kyttala, A., et al., *Two motifs target Batten disease protein CLN3 to lysosomes in transfected nonneuronal and neuronal cells.* Mol Biol Cell, 2004. **15**(3): p. 1313-23.
120. Nugent, T., S.E. Mole, and D.T. Jones, *The transmembrane topology of Batten disease protein CLN3 determined by consensus computational prediction constrained by experimental data.* FEBS Lett, 2008. **582**(7): p. 1019-24.
121. Cotman, S.L. and J.F. Staropoli, *The juvenile Batten disease protein, CLN3, and its role in regulating anterograde and retrograde post-Golgi trafficking.* Clin Lipidol, 2012. **7**(1): p. 79-91.
122. Jarvela, I., et al., *Biosynthesis and intracellular targeting of the CLN3 protein defective in Batten disease.* Hum Mol Genet, 1998. **7**(1): p. 85-90.
123. Golabek, A.A., et al., *Expression studies of CLN3 protein (battenin) in fusion with the green fluorescent protein in mammalian cells in vitro.* Mol Genet Metab, 1999. **66**(4): p. 277-82.
124. Michalewski, M.P., et al., *Evidence for phosphorylation of CLN3 protein associated with Batten disease.* Biochem Biophys Res Commun, 1998. **253**(2): p. 458-62.
125. Phillips, S.N., et al., *CLN3, the protein associated with batten disease: structure, function and localization.* J Neurosci Res, 2005. **79**(5): p. 573-83.
126. Gachet, Y., et al., *btn1, the Schizosaccharomyces pombe homologue of the human Batten disease gene CLN3, regulates vacuole homeostasis.* J Cell Sci, 2005. **118**(Pt 23): p. 5525-36.
127. Muzaffar, N.E. and D.A. Pearce, *Analysis of NCL Proteins from an Evolutionary Standpoint.* Curr Genomics, 2008. **9**(2): p. 115-36.
128. Haskell, R.E., et al., *Batten disease: evaluation of CLN3 mutations on protein localization and function.* Hum Mol Genet, 2000. **9**(5): p. 735-44.
129. Golabek, A.A., et al., *CLN3 disease process: missense point mutations and protein depletion in vitro.* Eur J Paediatr Neurol, 2001. **5 Suppl A**: p. 81-8.

130. Cotman, S.L., et al., *Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth.* Hum Mol Genet, 2002. **11**(22): p. 2709-21.
131. Kitzmuller, C., et al., *A function retained by the common mutant CLN3 protein is responsible for the late onset of juvenile neuronal ceroid lipofuscinosis.* Hum Mol Genet, 2008. **17**(2): p. 303-12.
132. Haskell, R.E., T.A. Derksen, and B.L. Davidson, *Intracellular trafficking of the JNCL protein CLN3.* Mol Genet Metab, 1999. **66**(4): p. 253-60.
133. Kremmidiotis, G., et al., *The Batten disease gene product (CLN3p) is a Golgi integral membrane protein.* Hum Mol Genet, 1999. **8**(3): p. 523-31.
134. Mao, Q., H. Xia, and B.L. Davidson, *Intracellular trafficking of CLN3, the protein underlying the childhood neurodegenerative disease, Batten disease.* FEBS Lett, 2003. **555**(2): p. 351-7.
135. Margraf, L.R., et al., *Tissue expression and subcellular localization of CLN3, the Batten disease protein.* Mol Genet Metab, 1999. **66**(4): p. 283-9.
136. Luiro, K., et al., *CLN3 protein is targeted to neuronal synapses but excluded from synaptic vesicles: new clues to Batten disease.* Hum Mol Genet, 2001. **10**(19): p. 2123-31.
137. Ezaki, J., et al., *Characterization of Cln3p, the gene product responsible for juvenile neuronal ceroid lipofuscinosis, as a lysosomal integral membrane glycoprotein.* J Neurochem, 2003. **87**(5): p. 1296-308.
138. Pearce, D.A., S.A. Nosel, and F. Sherman, *Studies of pH regulation by Btn1p, the yeast homolog of human Cln3p.* Mol Genet Metab, 1999. **66**(4): p. 320-3.
139. Jarvela, I., et al., *Defective intracellular transport of CLN3 is the molecular basis of Batten disease (JNCL).* Hum Mol Genet, 1999. **8**(6): p. 1091-8.
140. Rakheja, D., et al., *CLN3P, the Batten disease protein, localizes to membrane lipid rafts (detergent-resistant membranes).* Biochem Biophys Res Commun, 2004. **317**(4): p. 988-91.
141. Storch, S., S. Pohl, and T. Braulke, *A dileucine motif and a cluster of acidic amino acids in the second cytoplasmic domain of the batten disease-related CLN3 protein are required for efficient lysosomal targeting.* J Biol Chem, 2004. **279**(51): p. 53625-34.
142. Katz, M.L., et al., *Immunochemical localization of the Batten disease (CLN3) protein in retina.* Invest Ophthalmol Vis Sci, 1997. **38**(11): p. 2375-86.
143. Kida, E., et al., *Analysis of intracellular distribution and trafficking of the CLN3 protein in fusion with the green fluorescent protein in vitro.* Mol Genet Metab, 1999. **66**(4): p. 265-71.
144. Persaud-Sawin, D.A., et al., *A galactosylceramide binding domain is involved in trafficking of CLN3 from Golgi to rafts via recycling endosomes.* Pediatr Res, 2004. **56**(3): p. 449-63.
145. Kim, Y., D. Ramirez-Montealegre, and D.A. Pearce, *A role in vacuolar arginine transport for yeast Btn1p and for human CLN3, the protein defective in Batten disease.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15458-62.
146. Codlin, S. and S.E. Mole, *S. pombe btn1, the orthologue of the Batten disease gene CLN3, is required for vacuole protein sorting of Cpy1p and Golgi exit of Vps10p.* J Cell Sci, 2009. **122**(Pt 8): p. 1163-73.
147. Haines, R.L., S. Codlin, and S.E. Mole, *The fission yeast model for the lysosomal storage disorder Batten disease predicts disease severity caused by mutations in CLN3.* Dis Model Mech, 2009. **2**(1-2): p. 84-92.

148. Pao, S.S., I.T. Paulsen, and M.H. Saier, Jr., *Major facilitator superfamily.* Microbiol Mol Biol Rev, 1998. **62**(1): p. 1-34.
149. Rusyn, E., et al., *CLN3p impacts galactosylceramide transport, raft morphology, and lipid content.* Pediatr Res, 2008. **63**(6): p. 625-31.
150. Holopainen, J.M., et al., *Elevated lysosomal pH in neuronal ceroid lipofuscinoses (NCLs).* Eur J Biochem, 2001. **268**(22): p. 5851-6.
151. Golabek, A.A., et al., *CLN3 protein regulates lysosomal pH and alters intracellular processing of Alzheimer's amyloid-beta protein precursor and cathepsin D in human cells.* Mol Genet Metab, 2000. **70**(3): p. 203-13.
152. Narayan, S.B., et al., *CLN3P, the Batten's disease protein, is a novel palmitoyl-protein Delta-9 desaturase.* Ann Neurol, 2006. **60**(5): p. 570-7.
153. Resh, M.D., *Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins.* Biochim Biophys Acta, 1999. **1451**(1): p. 1-16.
154. Hofmann, S.L., et al., *Neuronal ceroid lipofuscinoses caused by defects in soluble lysosomal enzymes (CLN1 and CLN2).* Curr Mol Med, 2002. **2**(5): p. 423-37.
155. Lane, S.C., et al., *Apoptosis as the mechanism of neurodegeneration in Batten's disease.* J Neurochem, 1996. **67**(2): p. 677-83.
156. Puranam, K.L., et al., *CLN3 defines a novel antiapoptotic pathway operative in neurodegeneration and mediated by ceramide.* Mol Genet Metab, 1999. **66**(4): p. 294-308.
157. Rylova, S.N., et al., *The CLN3 gene is a novel molecular target for cancer drug discovery.* Cancer Res, 2002. **62**(3): p. 801-8.
158. Metcalf, D.J., et al., *Loss of the Batten disease gene CLN3 prevents exit from the TGN of the mannose 6-phosphate receptor.* Traffic, 2008. **9**(11): p. 1905-14.
159. Kama, R., et al., *The yeast Batten disease orthologue Btn1 controls endosome-Golgi retrograde transport via SNARE assembly.* J Cell Biol, 2011. **195**(2): p. 203-15.
160. Penczek, P.A., *Image restoration in cryo-electron microscopy.* Methods Enzymol, 2010. **482**: p. 35-72.
161. Penczek, P.A., *Fundamentals of three-dimensional reconstruction from projections.* Methods Enzymol, 2010. **482**: p. 1-33.
162. Yang, Z., et al., *Iterative stable alignment and clustering of 2D transmission electron microscope images.* Structure, 2012. **20**(2): p. 237-47.
163. Radermacher, M., et al., *Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli.* J Microsc, 1987. **146**(Pt 2): p. 113-36.
164. Adrian, M., et al., *Cryo-negative staining.* Micron, 1998. **29**(2-3): p. 145-60.
165. Mindell, J.A. and N. Grigorieff, *Accurate determination of local defocus and specimen tilt in electron microscopy.* J Struct Biol, 2003. **142**(3): p. 334-47.
166. Hohn, M., et al., *SPARX, a new environment for Cryo-EM image processing.* J Struct Biol, 2007. **157**(1): p. 47-55.
167. Herzog, F., et al., *Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry.* Science, 2012. **337**(6100): p. 1348-52.
168. Leitner, A., et al., *Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics.* Mol Cell Proteomics, 2010. **9**(8): p. 1634-49.
169. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nat Protoc, 2010. **5**(4): p. 725-38.

170. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. **4**(3): p. 363-71.
171. Blaszczyk, M., et al., *CABS-fold: Server for the de novo and consensus-based prediction of protein structure.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W406-11.
172. Grossman, E., O. Medalia, and M. Zwerger, *Functional architecture of the nuclear pore complex.* Annu Rev Biophys, 2012. **41**: p. 557-84.
173. Stagg, S.M., et al., *Structural basis for cargo regulation of COPII coat assembly.* Cell, 2008. **134**(3): p. 474-84.
174. Varma, D., et al., *Spindle assembly checkpoint proteins are positioned close to core microtubule attachment sites at kinetochores.* J Cell Biol, 2013. **202**(5): p. 735-46.
175. Bai, X.C., et al., *Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles.* Elife, 2013. **2**: p. e00461.
176. Liao, M., et al., *Structure of the TRPV1 ion channel determined by electron cryo-microscopy.* Nature, 2013. **504**(7478): p. 107-12.
177. Foley, E.A. and T.M. Kapoor, *Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore.* Nat Rev Mol Cell Biol, 2013. **14**(1): p. 25-37.
178. Chae, P.S., et al., *Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins.* Nat Methods, 2010. **7**(12): p. 1003-8.
179. Rasmussen, S.G., et al., *Crystal structure of the beta2 adrenergic receptor-Gs protein complex.* Nature, 2011. **477**(7366): p. 549-55.

5 Supplementary information

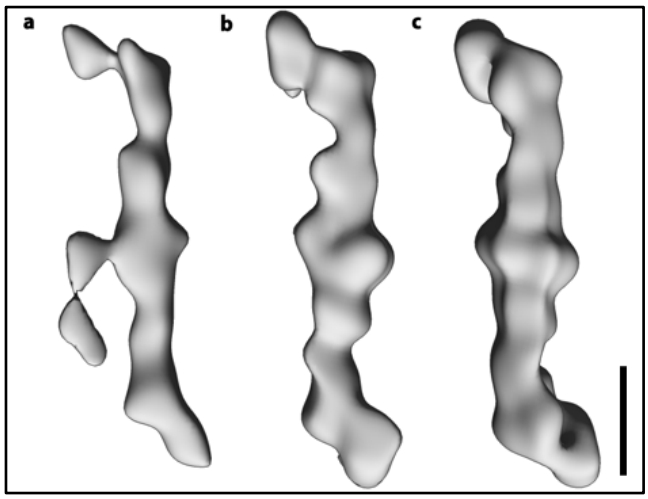


Figure S1: Steps in the generation of initial model of RZZ by Random Conical Tilt. (a) Back-projection of the best average from the tilted stack. The noise and dust is hidden, using the hide dust tool in Chimera. **(b)** After back projection refinement, without imposing any symmetry. **(c)** After back projection refinement imposing C2 symmetry. Scale bar, 10 nm.

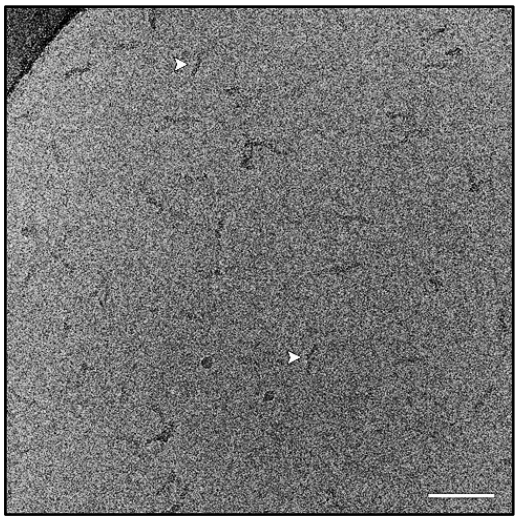


Figure S2: Representative cryo-positive electron micrograph area, recorded on JEOL 3200FSC equipped with TVIPS 8k x 8k camera. White arrowheads highlight particles. Scale bar, 100 nm.

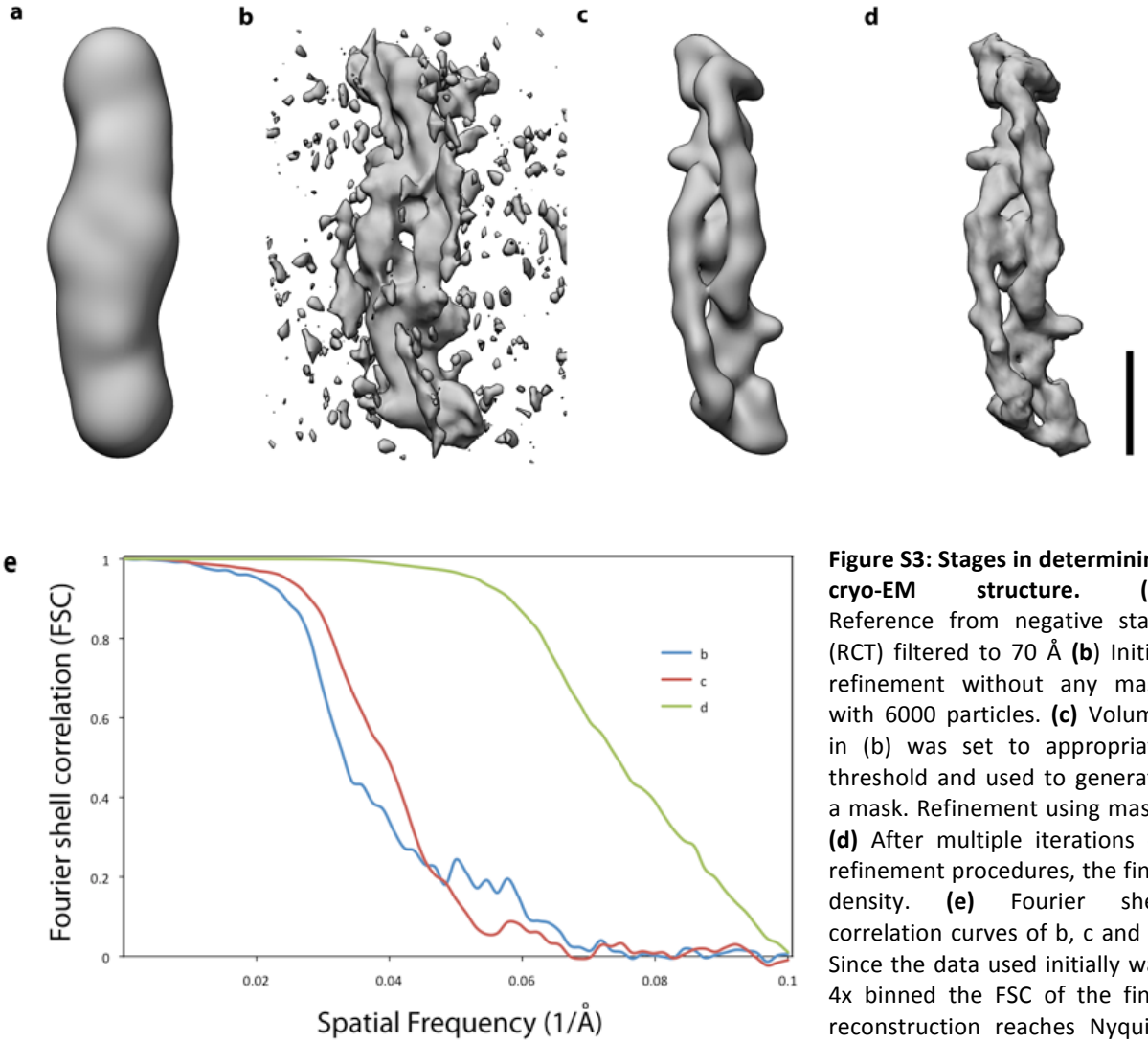


Figure S3: Stages in determining cryo-EM structure. (a) Reference from negative stain (RCT) filtered to 70 Å **(b**) Initial refinement without any mask with 6000 particles. **(c)** Volume in (b) was set to appropriate threshold and used to generate a mask. Refinement using mask. **(d)** After multiple iterations of refinement procedures, the final density. **(e)** Fourier shell correlation curves of b, c and d. Since the data used initially was 4x binned the FSC of the final reconstruction reaches Nyquist limit. Scale bar, 10 nm.

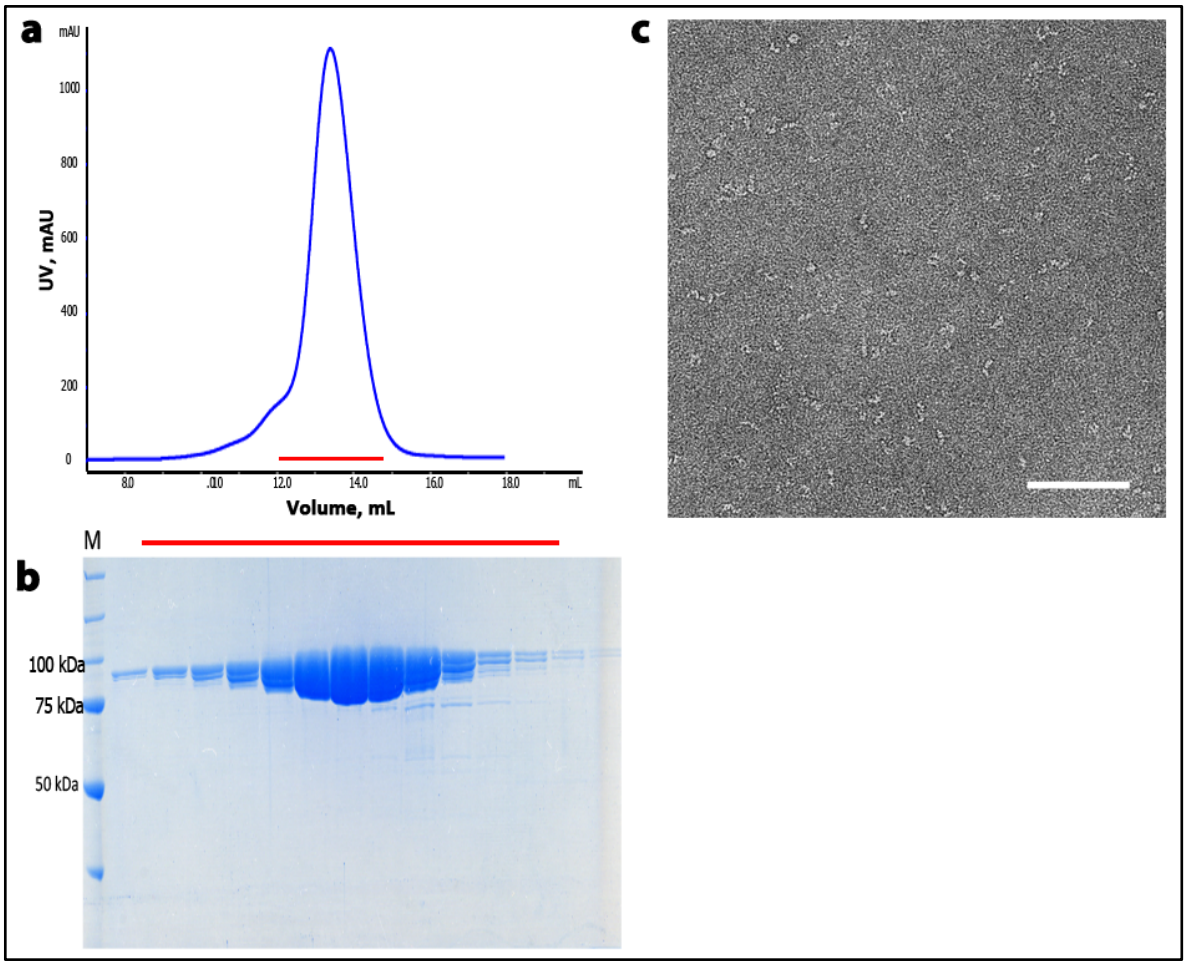


Figure S4: Purification and negative stain EM of Zw10. (a) Size-exclusion chromatography profile of Zw10. **(b)** SDS-PAGE of the fractions encompassed by the peak from SEC indicating highly abundant and pure protein. **(c)** Representative negative stain micrograph area representing Zw10 particles. Scale bar, 100 nm.

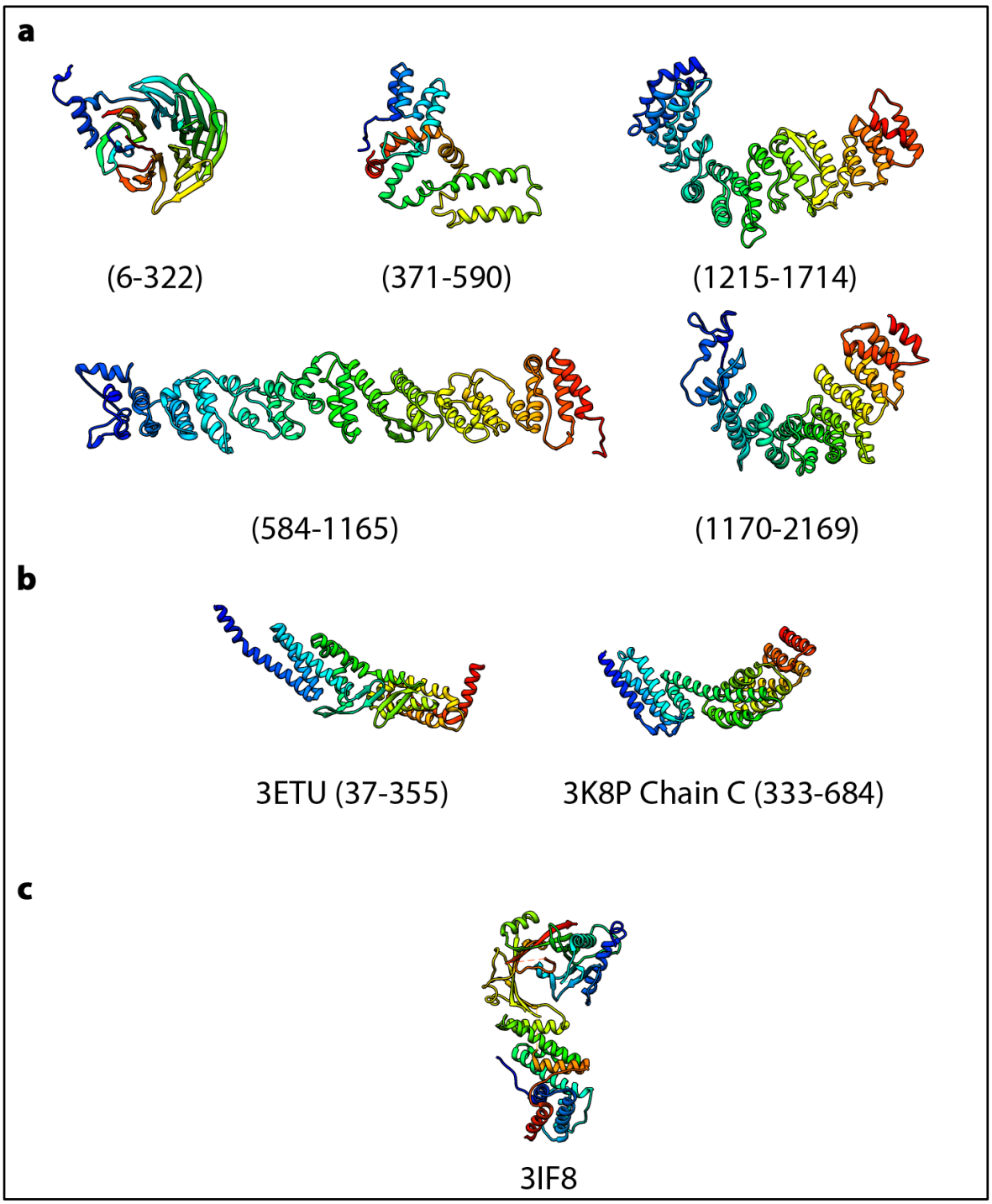


Figure S5: Structures/models of components of RZZ complex (in rainbow representation N-terminus (blue)-C-terminus (red). (a) Models of ROD generated by structure prediction program (I-TASSER) by Dr. Jenny Keller. **(b)** Structures of yeast homologue of Zw10 (Dsl1p). N-terminus (37-355), PDB-ID: 3ETU and the C-terminus (333-684), PDB-ID: 3k8P. **(c)** Structure of Zwilch, PDB-ID: 3IF8

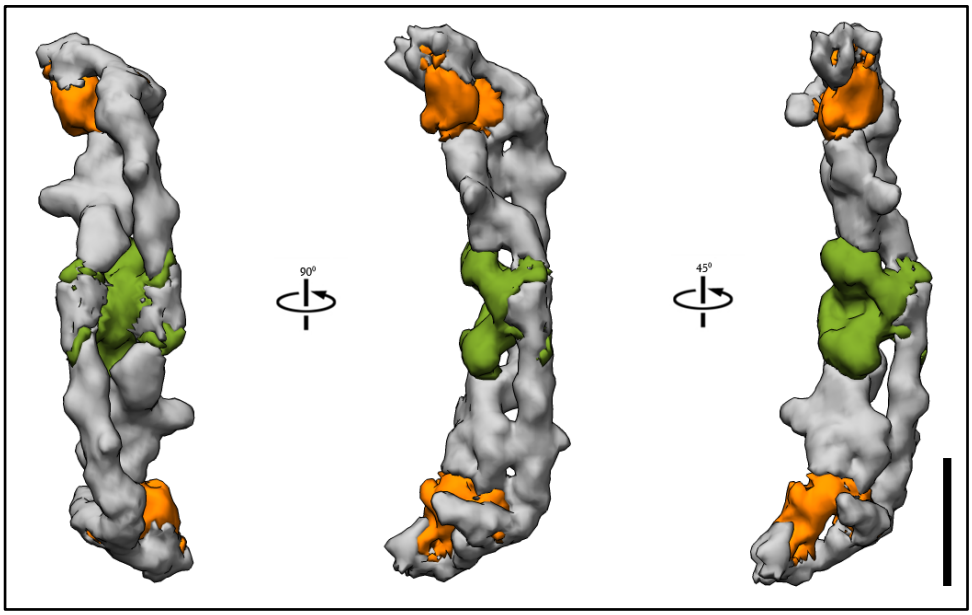


Figure S6: Segmented views of the RZZ complex. ROD (gray), Zw10 (green) and Zwilch (orange). Scale bar, 10nm.

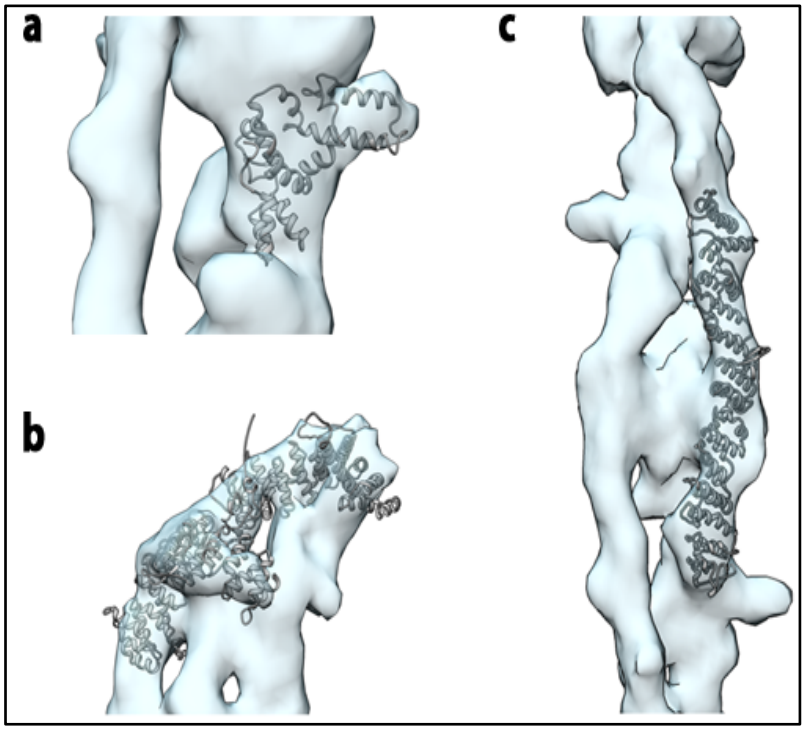


Figure S7: Fitting of various domains of ROD into the cryo EM structure. (a) The NRH domain, amino acids (371-590). **(b)** The C-terminus of ROD extends to tip of the structure, amino acids (1215-1714, 1770-2169). **(c)** The α-solenoid in ROD, amino acids (584-1165)

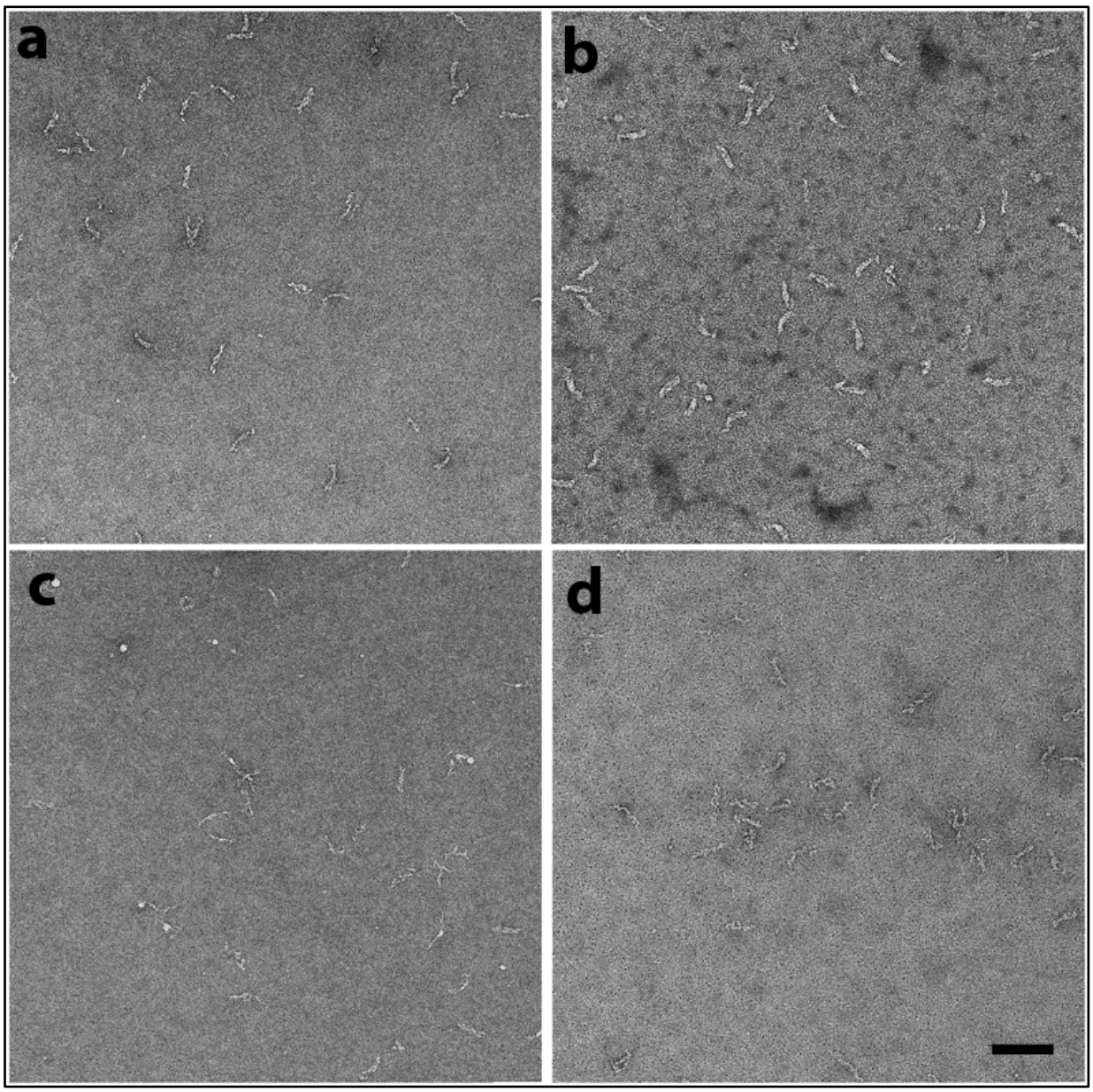


Figure S8: Negative stain micrograph area representing RZZ complex in four of the many conditions tried for spontaneous assembly. In most of the cases the sample loses its characteristic shape. Scale bar, 100 nm. **(a)** 50 mM HEPES pH 8.5, 100mM NaCl, 1mM TCEP **(b)** 50 mM HEPES pH 8.5, 500mM NaCl, 1mM TCEP **(c)** 50 mM MES pH 6.3, 100mM NaCl, 1mM TCEP **(d)** 50 mM MES pH 6.3, 500mM NaCl, 1mM TCEP

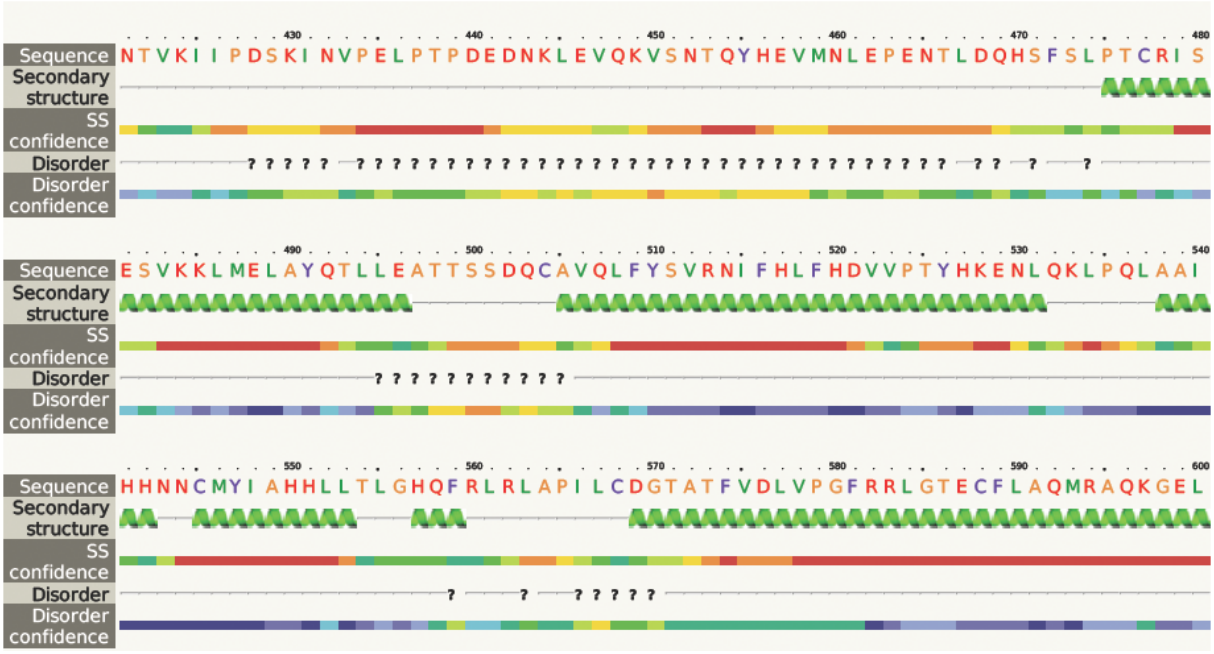


Figure S9: Secondary structure prediction of Zw10. Protein stretch from amino acids 421-600 is highlighted. Most of the protein is predicted to have secondary α-helical structure, with intermittent regions of disorder. The largest such stretch with a very high disorder confidence is central to the protein (amino acids 428-470, approximately). The secondary structure was predicted using Phyre-2. The confidence in prediction is depicted as color spectrum from red (high confidence) to blue (low confidence).

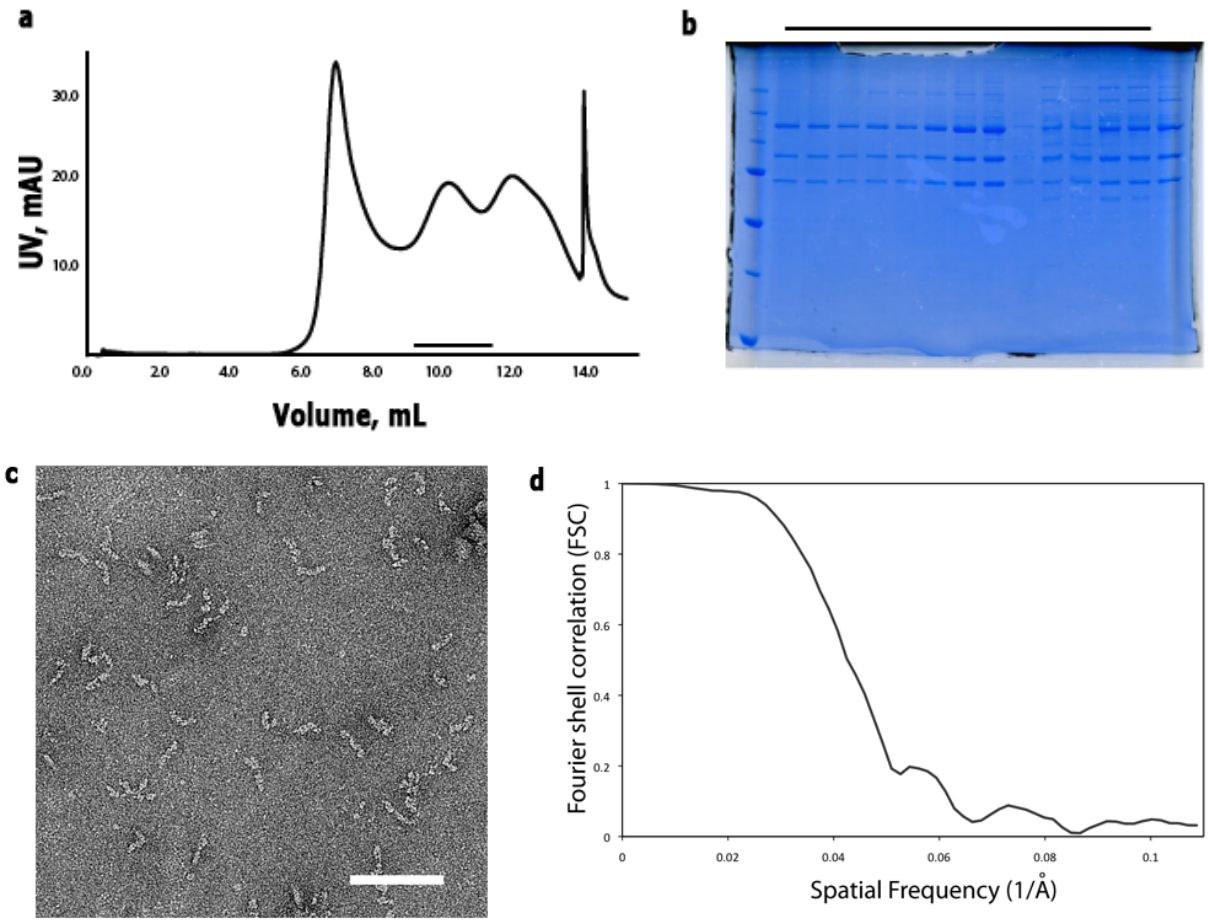


Figure S10: Purification and negative stain EM of mini-RZZ. (a) Size-exclusion chromatography profile of mini-RZZ. **(b)** SDS-PAGE of the fractions encompassed by the peak from SEC indicating instability of complex, reflected from the degradation. **(c)** Representative negative stain micrograph area representing mini-RZZ complex particles. Scale bar, 100 nm. **(d)** Fourier shell correlation curve for the structure of mini-RZZ complex. $FSC_{0.5}$: 24.6 Å.

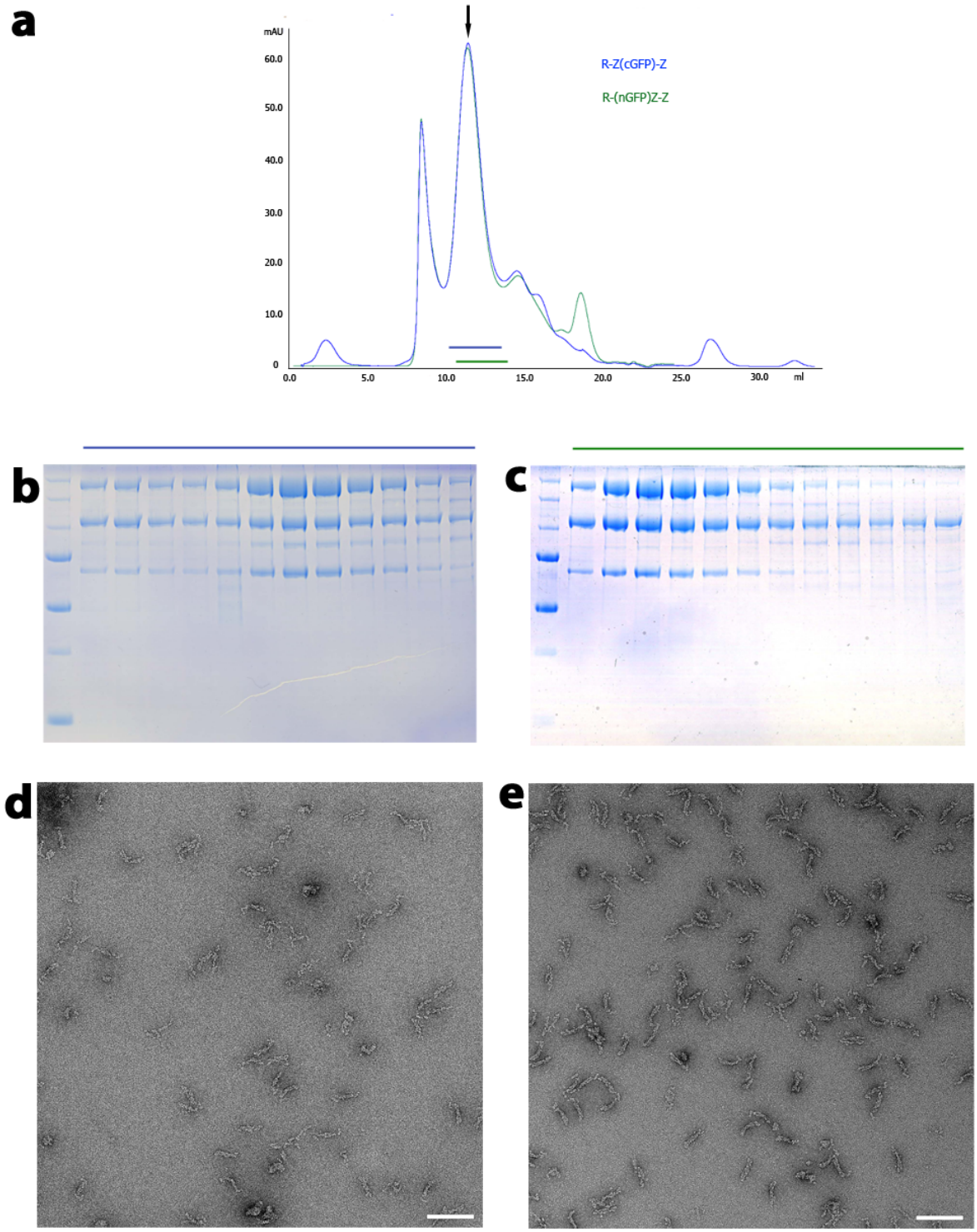


Figure S11: Purification and negative stain EM of R-(nGFP)-Z and R-Z(cGFP)-Z (a) Size-exclusion chromatography profile of R-(nGFP)-Z and R-Z(cGFP)-Z **(b)** and **(c)** SDS-PAGE of the fractions highlighted by the blue and green lines of R-Z(cGFP)-Z and R-(nGFP)-Z **(d)** and **(e)** Negative stain micrograph area representing R-Z(cGFP)-Z and R-(nGFP)-Z particles respectively. Scale bar, 100 nm.

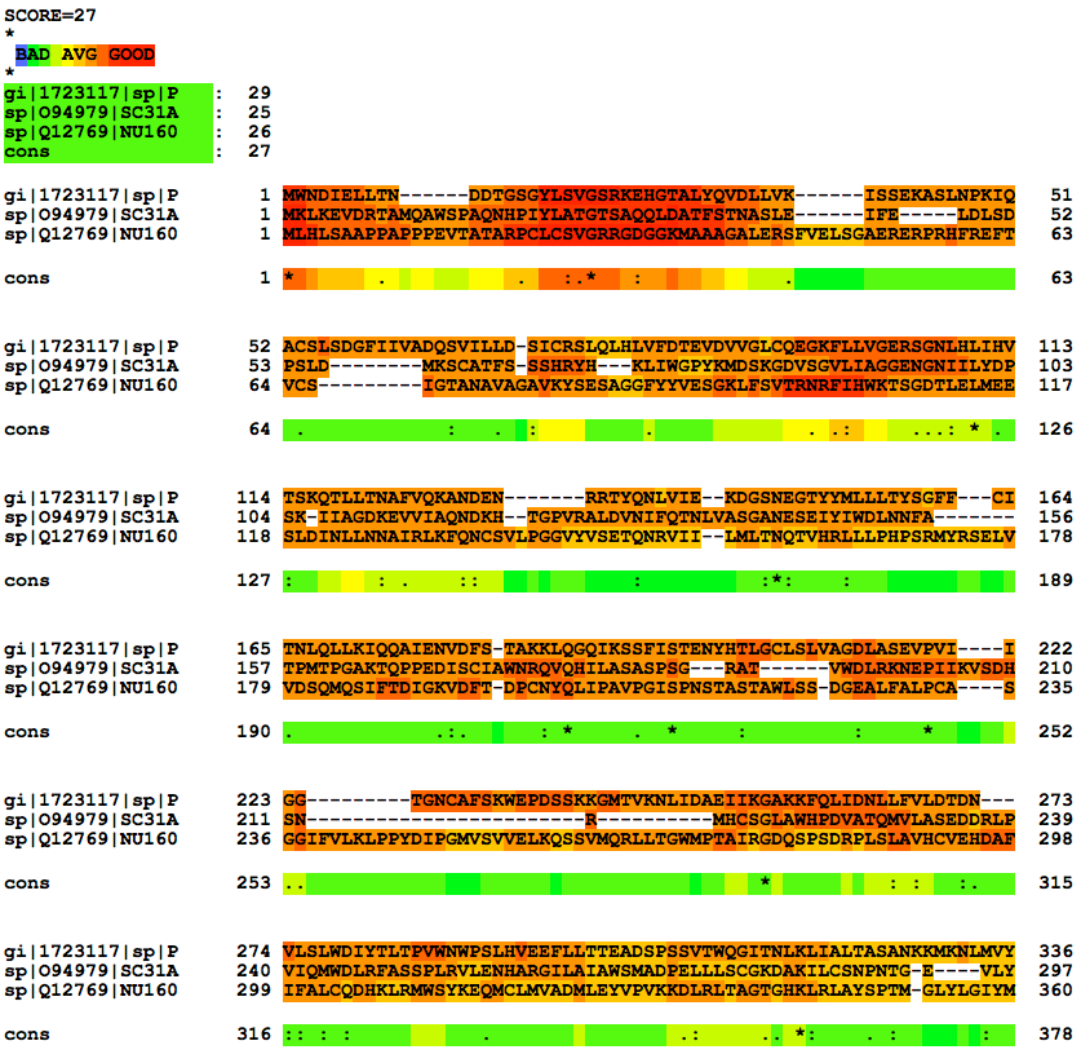


Figure S12: Multiple sequence alignment (MSA) of N-terminus of ROD (1-330) with the N-termini of Sec31 and Nup160 proteins generated using T-Coffee online MSA server. All the three proteins are known or predicted to contain an N-terminal β-propeller. The alignment score (27) is bad and is not surprising as WD40 repeats are known for low sequence similarity even though they are structurally similar.

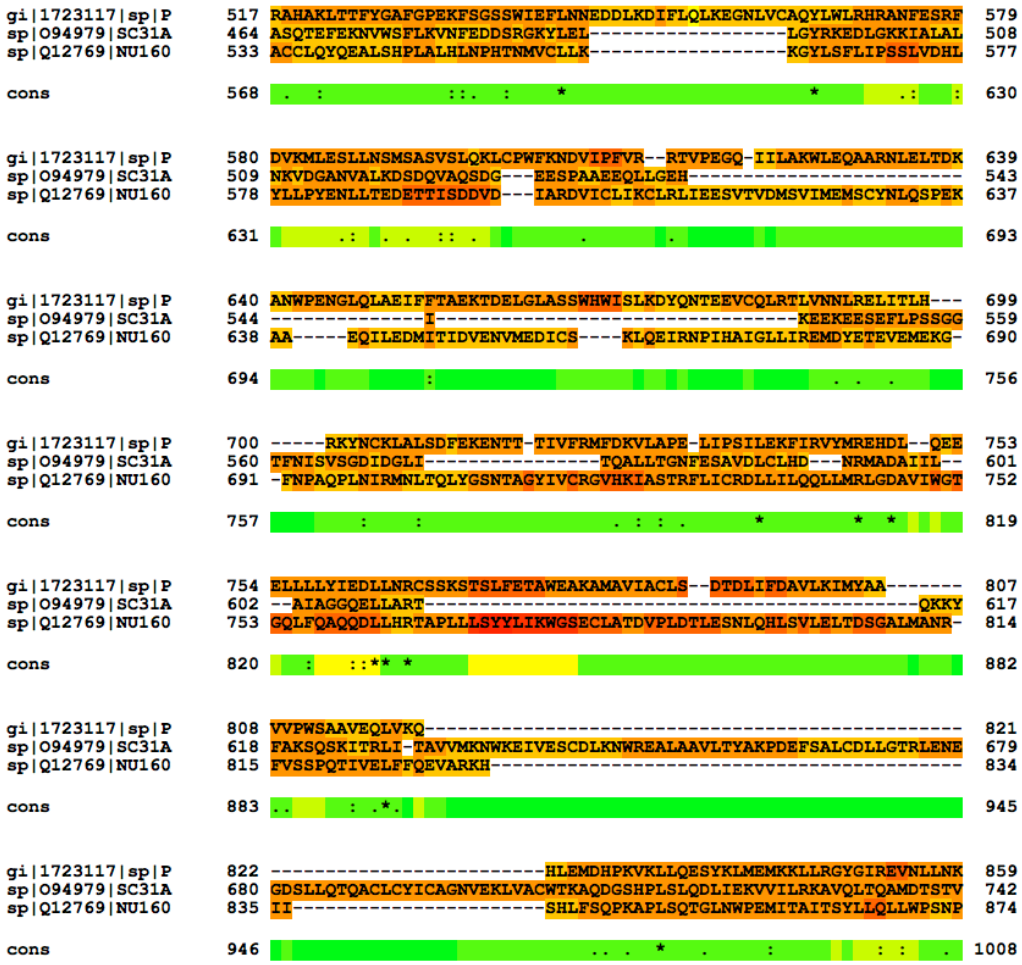


Figure S13: MSA analysis of ROD, Sec31 and Nup160 proteins similar to one in Figure S12. Alignment shown is for predicted α-solenoid region of the proteins. As before (Figure S12) the consensus score is pretty low.

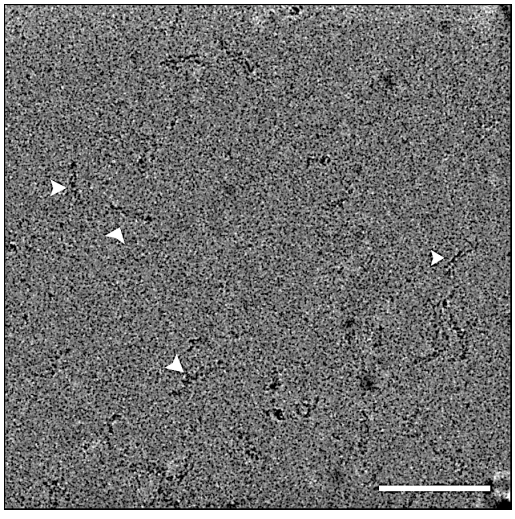


Figure S14: Representative cryo electron micrograph area (an average of seven frames, unaligned), recorded on TITAN krios equipped with FALCON II direct electron detector camera. White arrowheads highlight particles. Scale bar, 100 nm.

Table 5.1 Cross-links between ROD-ROD (intra) detected from XL-MS

ROD	ROD
24	328
24	404
43	116
43	377
52	527
127	240
240	406
328	377
406	430
430	464
503	2016
513	521
513	1031
701	739
701	1658
705	726
739	1658
770	843
829	1646
829	1648
829	1658
829	1665
831	1648
831	1665
838	859
843	859
916	971
946	957
946	1021
946	1029
946	1031
946	1051
957	1051
964	977
971	1493
1021	1051
1021	1086
1021	1076
1029	1051
1029	1063
1029	1076
1029	1081
1029	1085

ROD	ROD
1031	1051
1031	1063
1031	1076
1031	1081
1031	1085
1051	1081
1051	1086
1076	1101
1081	1086
1081	1092
1085	1092
1307	1316
1307	1318
1333	1367
1413	1493
1493	1560
1640	1648
1640	1731
1645	1653
1646	1653
1646	1665
1648	1653
1653	1665
1658	1756
1728	1737
1728	1738
1731	1738
1737	1768
1738	1768
2093	2135
2093	2136
2106	2145
2135	2143

Table 5.2 Cross-links between ROD-Zw10 (left, inter) and ROD –Zwilch detected from XL-MS

ROD	Zw10
859	45
859	129
859	130
859	145
870	157
916	777
946	74
946	143
946	149
946	152
1029	624
1031	624
1063	624
1081	528
1092	624
1307	390
1377	390
1413	444
1430	35
1665	130

ROD	Zwilch
239	191
240	191
437	190
705	326
739	326
946	201
1814	326
2189	527

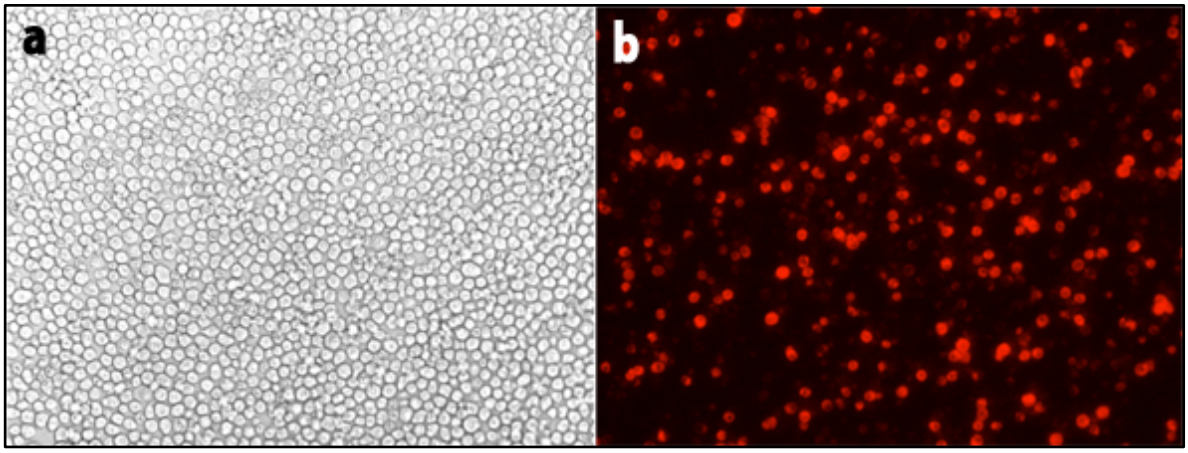


Figure S15: Overexpression of mCherry-Cln3p in Sf21 cells. (a) DIC image (b) fluorescence image.

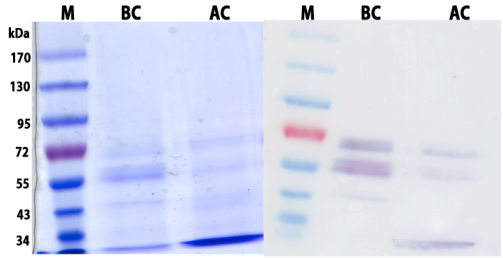


Figure S16: Cleavage of tag from EGFP-Cln3p protein (p232 construct). Semi-denaturing SDS-PAGE (left) and corresponding Western blot (right). BC, before clevage. AC, after cleavage. M, marker.

Acknowledgements

This work would not have been possible without the help of a few very extraordinary people. I really appreciate and express my heartfelt gratitude to everyone who has been a part of this journey, which I embarked upon four years ago.

At the outset, I would like to thank Prof. Dr. Stefan Raunser for his unabated support and guidance throughout this work. I am really grateful for his patience and understanding in giving me the time and space to learn and try out new things.
I would like to thank Prof. Dr. Roger S. Goody for allowing me to be a part of his department and accepting to be the first referee.
I would like to thank Prof. Dr. Roland Winter for kindly accepting to be my second referee.
I am thankful to Dr. Leif Dehmelt, for agreeing to preside over my thesis defense as 'wiss. Mitarbeiter'.
I owe sincere gratitude to Prof. Dr. Andrea Musacchio, for introducing exotic collaboration projects and very fruitful discussions.

My deepest thanks to collaborators Dr. Arsen Petrovic, Dr. Gemma Triola, Dr. Annemarie Wehenkel, Dr. Jenny Keller, Anika Altenfeld for providing materials, extensive discussions and knowledge transfer.
I would like to thank the 'raunser-lab' for making this journey memorable. It was a privilege to be a part of this team. Special thanks to Dr. Christos Gatsogiannis, for teaching me everything (almost) I know about single particle EM; to Dr. Anne Kuhlee for teaching me everything (almost) I know about MS-Word; to Sandra Bergbrede and Dr. Oliver Hofnagel for all the help and assistance with the wet lab and EM work; to Dennis Quentin for carrying on the legacy of Cln3p.

Surprisingly, the time spent away from the lab was a lot of fun. Who would have thought of that? Words cannot express how grateful I am to Anna Bruns, Anchal Chandra, Stefan Schoebel, Gregor Bahrenberg, Siva Jegananthan and Georgios Konstantinidis for all the wonderful times. Thank you very much.

I could probably write another thesis on the good and crazy times spent together with Yashi and Mamü. Thank you guys for being there, for everything, at all times. I feel tremendously blessed to have friends like you.