The structure-function relationship of the ARISC complexes

Safi Kani Masandi

BSc

MSc

Submitted in accordance with the requirements for the
degree of Doctor of Philosophy

University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

September 2020

ACKNOWLEDGEMENTS

I would like to first of all thank everyone who helped me throughout my PhD. I would like to thank my colleagues in the Zeqiraj lab, Dr Mariam Walden, and Upasana Sykora who started in the lab along with me, and Dr Martina Fogolizzo and Laura Marr. Thank you for teaching me the techniques and skills I have used throughout my project. My thanks also go to the EM team Dr Dan Maskall, Dr Emma Hesketh and Dr Rebecca Thompson for helping me with the many technical problems and questions I had no matter how small. And to Dr Eoin Leen who I sometimes bothered with biochemical questions and/or reagents.

I have made incredible friends during my time as a PhD student. Thank you (soon to be doctors) Eleni Kyriakopoulou, Carla Gomez Santiago, and Dr Nnenna Nwogu. Your friendship throughout the years has been wonderful. I have enjoyed getting to know you both firstly as researchers and then as close friends. During my time at Leeds, I came to realise the importance of having a mentor. To this end, I would like to acknowledge my Msc supervisor Dr Emanuele Paci who supported and encouraged me as his student and continued to do so long after I left his lab. Thank you Dr Yoselin Benitez-Alfonzo and Dr Bernadine Idowu for guiding me through my PhD and imparting wisdom.

On a more personal note, I would like to thank my siblings and my childhood friend Ema Oko for telling me that I am smart enough even when I didn't feel like it.

My final acknowledgement goes to my supervisors Professor Neil Ranson and Dr Elton Zeqiraj.

ABSTRACT

The JAMM/MPN$^+$ protein BRCA1-containing complex subunit 36 (BRCC36) forms two distinct complexes in the cell: the cytoplasmic BRISC-SHMT2 complex, and the related ARISC-RAP80 complex in the nucleus. The BRISC-SHMT2 complex plays a vital role in interferon signalling whereas the ARISC-RAP80 complex regulates DNA damage and repair. Both complexes contain shared subunits (BRCC36, BRCC45 and MERIT40) but differ in their targeting proteins (SHMT2 and RAP80, respectively) and in the MPN$^-$ domain containing proteins (Abraxas2 and Abraxas1, respectively). As an isolated protein, BRCC36 is catalytically inactive and partnering with Abraxas2 results in a BRCC36-Abraxas2 minimally active complex. Abraxas2 supports the activity of BRCC36 through direct interactions by inducing a conformational change within the MPN+ domain of BRCC36 and its active site. Additionally, BRCC36 and Abraxas form a super dimer that is required for deubiquitylase activity and assembly of higher order BRISC and ARISC complexes. The main goal of this thesis is to understand how higher order ARISC complexes assemble and how this affects DUB activity and interactions with the targeting subunit RAP80.

Shortly after the start of this project, a low resolution model of the four membered ARISC complex was published. The model revealed that the complex forms a U-shape architecture, which consists of the BRCC36-Abraxas1 super dimer at the base of the U and BRCC45-MERIT40 at the arms, suggesting a 2:2:2:2 stoichiometry. In this thesis, the structure of the ARISC complexes (ARISC-RAP80 and ARISC) were explored by negative stain and cryo electron microscopy. These structural analyses revealed that the complex displays conformational heterogeneity and exists in higher ordered (ARISC dimer of octamer) and subcomplexes. These results were further supported by mass spectrometry and native PAGE analysis. Negative stain and cryo electron microscopy analysis of the complexes also indicate that at ARISC-RAP80 complex stability was concentration dependent. This phenomenon was also true for the related BRISC-SHMT2 cytoplasmic complex after cryo electron microscopy data was collected at protein concentrations higher than usual and in a molecular crowded environment. Results showed that despite particle overcrowding, data processing was possible and resulted in reduced particle heterogeneity. In addition, preliminary interactions studies to elucidate the function of RAP80 protein revealed that RAP80 is the major interacting subunit for ubiquitin the major contributor for ARISC-RAP80 complex ability to interact with DNA. These results provide a platform for future structural and mechanistic studies into the function of the ARISC-RAP80 complex in the nucleus.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABBREVIATIONS

a-EJ	Alternative end joining
AIR	Abraxas interacting region
APC/C	Anaphase promoting complex
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triose phosphate
ATR	Ataxia telangiectasia and Rad3 related
ATRIP	ATM and Rad3-related interacting protein
BARD1	BRCA1 associated ring domain 1
Bax	Bcl-2-associated x protein
Bdb	Barr body deficient
BER	Base excision repair
Bp	Base pair
BRCA1	Breast cancer type 1 susceptibility protein
BRCA1-A	Breast cancer type 1 susceptibility protein A
BRCC36	BRCA1-containing complex subunit 36
BRCC45	BRCA1/BRCA2 containing complex subunit 45
BRISC	BRCC36-containing isopeptidase complex
CC	Correlation coefficient
CCD	Charge couple device
Ccd25a	coiled-coil domain containing 25
Cdk	Cyclin dependent kinase
CHK	Checkpoint kinase
CPD	Cyclobutene pyrimidine dimers
Cryo-EM	Cryo-electron microscopy
CSN	Cop9 signalosome complex
CtIP	CTBP-interacting protein
CTF	Contrast transfer function
CUL3-KLHL20	Cullin3 Keltch-like protein 20
CYLD	Cyclindromatosis turban tumour syndrome protein

DDR	DNA damage repair
DED	Direct electron detectors
DNA	Deoxyribose nuclear acid
DNA-PKC	DNA dependent protein kinase
DQE	Detective quantum efficiency
DSB	Double stranded break
DSBR	Double stranded break repair
DUB	Deubiquitylating enzyme
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin ligase
EIF	Eukaryotic initiation factor
EM	Electron microscopy
FPLC	Fast protein liquid chromatography
FSC	Fourier shell correlation
G1	Growth phase
HECT	Homologous to E6AP carboxyl terminus
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H4	Histone 4
Histag	Histidine tag
HR	Homologous recombination
HUWE1	HECT UBA WWE domain containing 1
IBR	In between ring
IQF	Internally quenched fluorescence
JAMM/MPN	Jb1/Mov34/Mpr1 Pad1 N-terminal
kDa	Kilodaltons
KenderSOM	Kernal self-organising map

MCD1	Mitotic chromosome determinant 1
MERIT40	Mediator of RAP80 interaction and targeting subunit 40
Mg/ml	milligrams per millilitres
MINDY	Motif interacting with ubiquitin containing DUB family
MJD	Machado-Josephin disease
ML	Maximum likelihood
ml	Millilitres
mM	Millimolar
MMR	Mismatch repair
MRE11	Mitotic recombination 11
MRN	Mre-Rad50-Nbs1
MWCO	Molecular weight cut-off
NaCl	Sodium chloride
NEDD8	Neural precursor cell expressed, developmentally down regulated 8
NF-KB	Nuclear factor kappa- light-chain-enhancer of activated B cells
NHEJ	Non-homologous end-joining
nM	Nano molar
NMR	Nuclear magnetic resonance
OTU	Ovarian tumour proteases
P53	Tumour protein 53
PARP	Poly-ADP ribose polymerase
PCR	Polymerase chain reaction
POI	Protein of interest
PSMD1	Proteasome 26s subunit non-ATPase
PTM	Post-translational modification
PUMA	P53 upregulated modulator of apoptosis
RAP80	Receptor-associated protein 80
RBR	Ring between ring
RING	Really new interesting new gene

RNF8	Ring finger protein 8
ROS	Reactive oxygen species
RPA	Replication protein A
S	Synthesis phase
S1	Ubiquitin binding site 1
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SHMT2	Serine hydroxymethyl transferase 2
SIM	Sumo-interacting motif
SSB	Single-stranded break
SSBR	Single-stranded break repair
SSDNA	Single-stranded DNA
SUMO	Small ubiquitin-like modifier
TEM	Transmission electron microscopy
TEV	Tobacco Etch Virus
TRAF6	Tumour necrosis factor receptor-associated factor 6
Ub	Ubiquitin
UBC13	Ubiquitin conjugating enzyme E2 13
UBD	Ubiquitin binding domain
UCH	Ubiquitin C-terminal hydrolases
UEV	Ubiquitin E2 variant
UIM	Ubiquitin interacting motif
USP	Ubiquitin specific protease
UV	Ultraviolet light
XRCC4	X-ray repair cross-contemplating protein 4
ZUFSP	Zinc finger with UFM1-specific peptidase domain protein
2D	Two-dimensional
3D	Three-dimensional
53BP1	P53 binding protein 1
6-4PPS	6-4 pyrimidone photoproducts

CHAPTER 1
Introduction

1.1 DNA damage repair mechanisms

1.1.1 DNA packaging

Deoxyribonucleic acid (DNA) is the hereditary molecule present in all organisms. The majority of DNA is located in the cell nucleus (termed nuclear DNA), however, DNA is also present in the mitochondria (known as mitochondrial DNA). DNA is composed of nucleotides, each nucleotide is made up of a phosphate group, a sugar group (2'-deoxyribose) and a nitrogenous base. The genetic information in DNA is stored as a code using four nitrogenous bases: Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). The DNA bases pair up with one another, A with T and C with G via hydrogen bonding resulting in the formation of a base pair. Each base is attached to a sugar molecule (Travers and Muskhelishvili, 2015).

Nucleotides are linked to one another by covalent bonds, this results in a long strand, making up the backbone of DNA. DNA mostly exists as a double helix in which the two strands of DNA coil around one another forming the double helix. The formation of this double helix is facilitated by the complementary base pairing of the nitrogenous bases (Travers and Muskhelishvili, 2015).

The haploid human genome contains approximately three billion base pairs packaged into 23 chromosomes. With the exception of the female ova and male sperms, the majority of cells in the human body are diploid consisting of 6 billion base pairs of DNA per cell. When stretched, nuclear DNA is approximately two meters in length. However, the nucleus is only 6 μM in diameter. Therefore, DNA has to be folded in order to be accommodated in cells (Lamond, 2002).

In eukaryotes, DNA in the nucleus is packaged into varying levels of structural order. Histone proteins pack DNA at the most basic level (Worcel et al., 1981). Histones are small positively charged proteins that bind tightly to DNA due to the phosphate group in the phosphate-sugar backbone being negatively charged (Venkatesh and Workman, 2015). There are five histone families: H1/H5, H2A, H2B, H3, and H4. Histones H2A, H2B, H3 and H4 are the core histones, whereas histones H1/H5 on the other hand are the linker histones. Core histones exist as dimers and all histones contain the histone fold, which consists of three alpha helices linked by two loops. The helical structure of histones enables the interaction between distinct dimers, the four distinct dimers come together to form an octameric complex called the nucleosome core (Workman and Kingston, 1998). DNA packaging and ordering by histones results in the formation of a structure termed ‘beads on a string’ (Figure 1).

Chromatin is a mass of genetic material composed of DNA and proteins that condenses to form a chromosome. Nucleosomes, on the other hand, are the basic repeating structural units of chromatin and are composed of DNA segments wrapped around eight histone proteins (Annunziatio, 2015). The nucleosome is approximately 64 Å in diameter and consists of 147 base pairs of DNA wrapped around the histone octamer made up of histones H3-H4 tetramer and H2A-H2B dimers(Simpson, 1978). DNA wraps around the histone octamer 1.65 times in a left-handed super-helical turn (Tessarz and Kouzarides, 2014). The H1/H5 linker histone binds to the nucleosome locking the DNA into place. The histone dimers coordinate three consecutive minor grooves on the inner surface of the DNA spiral through electrostatic interactions and hydrogen bonding (Cutter and Hayes, 2015). The bending of DNA over the surface of histones draws together the phosphate backbone of the two DNA strands on the inside of the spiral, narrowing the major and minor grooves of DNA whilst simultaneously widening the grooves on the outside (Bowman and Poirier, 2015). The bending of DNA in this manner would under normal conditions be energetically unfavourable. It is, however, maintained through charge neutralization from arginine and lysine side chains of the histones (Bowman and Poirier, 2015).

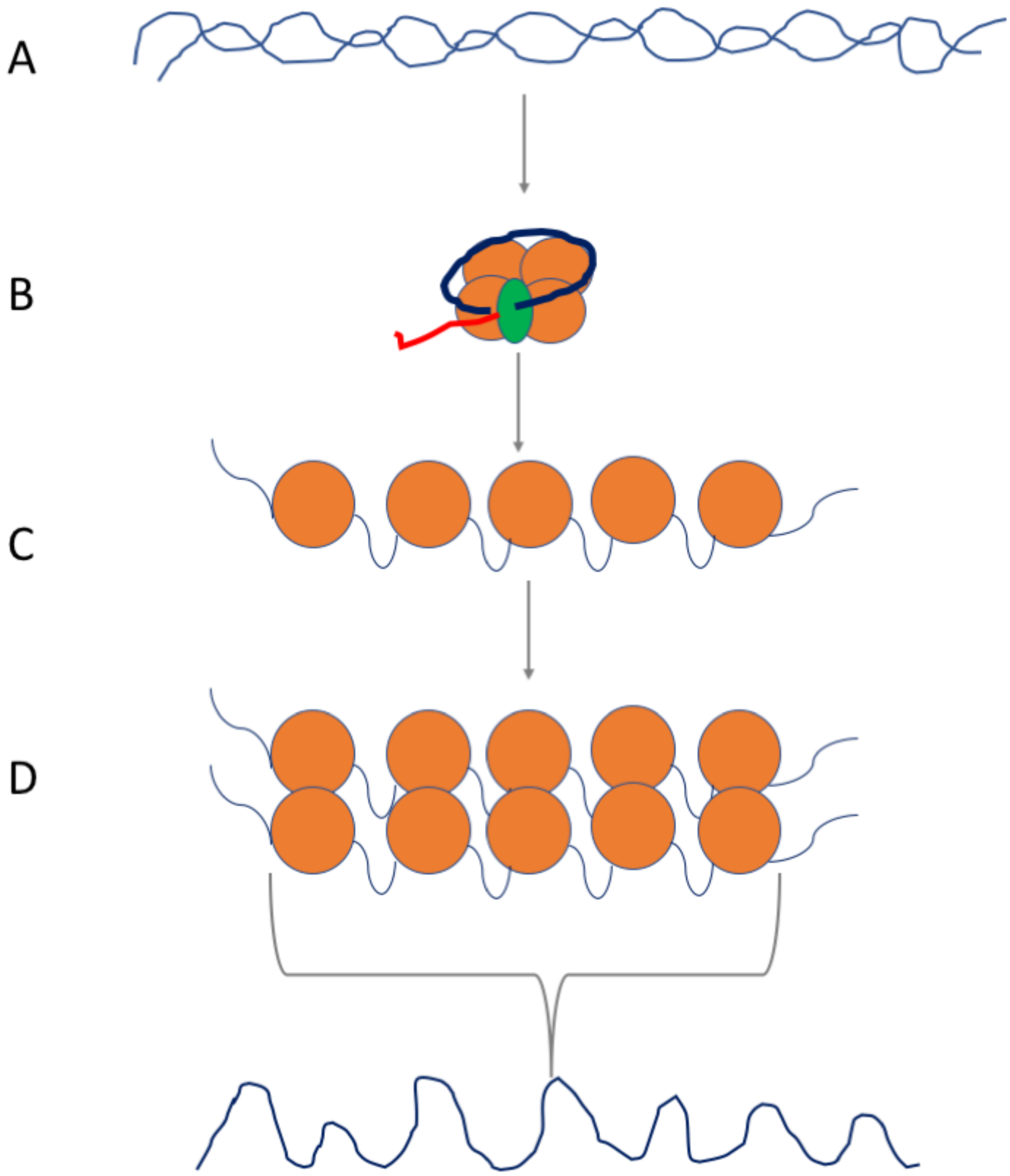


Figure 1: DNA packaging

A) DNA double helix. B) DNA (blue) wrapped around the core histones H2A, H2B, H3 and H4 (orange). Histone H1 is depicted in green and the linker DNA is shown in red. C) the “beads on a string” conformation of DNA wound around nucleosomes. D) Chromatin fibre.

The wrapping of DNA around the histone core sterically excludes other DNA binding proteins. Nucleosomes are however not static but dynamic structures, the DNA sequence, for example, can modulate nucleosome positioning and unwrapping. Post-translational modifications such as acetylation, methylation, phosphorylation, ubiquitylation and ADP-ribosylation chemically alter the histones changing the energy landscape of DNA-histone interactions thereby making the DNA more or less accessible to other binding proteins (Bowman and Poirier, 2015).

The canonical histones used to package DNA can be substituted by histone variants. Histone variants are expressed throughout the cell cycle and replace canonical histones or take their place when nucleosomes are evicted(Weber and Henikoff, 2014). Canonical histones typically differ to histone variants in sequence (Talbert and Henikoff, 2010). Substituting a histone can change multiple histone-DNA and histone-histone interactions, consequently also altering nucleosomes and chromatin fibres. For example, nucleosome structures containing the Histone H1A.Z variants produce more compact chromatin fibres compared to the canonical H2A histone. On the other hand, nucleosomes containing the H2A.Bdb (Barr body deficient) variant wrap DNA less efficiently and do not allow the chromatin to condense as efficiently (Bowman and Poirier, 2015) (Fan et al., 2002). Like canonical histones, histone variants can also be post-translationally modified, further modulating their effect on the stability and dynamics of nucleosomes.

The second mechanism that influences nucleosome stability is the reorganisation of the nucleosome structure by histone chaperones and ATP-dependent chromatin re-modellers. Histone chaperones are acidic proteins responsible for stabilising histones in the absence of DNA. Chromatin re-modellers are central to catalysing changes in the structure of nucleosomes, as under physiological conditions the remodelling would be thermodynamically unfavourable without them (Clapier et al., 2017). Chromatin re-modellers alter the position and composition of nucleosomes by coupling ATP hydrolysis to nucleosome sliding, octamer assembly/disassembly and the exchange of Histones (Burgess and Zhang, 2013).

In addition to altering the composition of nucleosomes, histone variants and post-translational modifications behave as chemical signposts in the cell, guiding factors to specific locations within the genome.

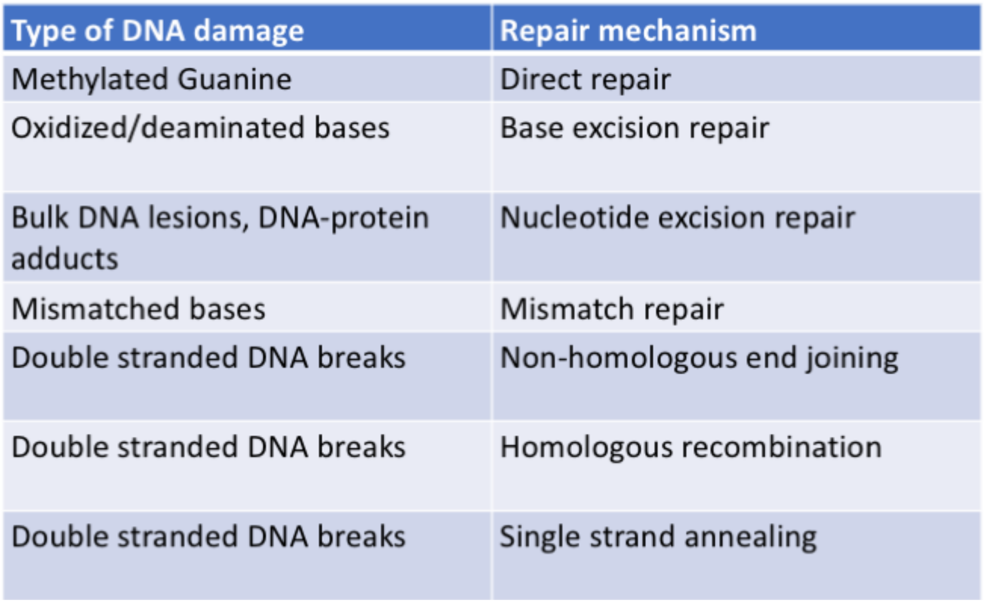
1.1.2 DNA damage

Random genomic mutation can have selective advantage in the evolution of organisms as it results in biological diversity. Some genomic mutations can have adverse consequences; however, it is important to note that some have no effect. If not repaired, or if incorrectly repaired, damaged DNA can result in a cascade of mutations or wider-scale genome aberrations. The preservation of genomic integrity is therefore paramount for proper cellular function and for the viability of organisms (Jackson and Bartek, 2009).

DNA damage can arise as the result of DNA metabolising processes such as replication errors and inaccurate VDJ recombination (Giglia-Mari et al., 2011). During DNA replication DNA mismatches can be introduced and DNA strand breaks can be caused by topoisomerase I and II activity. DNA damage can also result from hydrolytic damage which in turn results in the deamination of bases, depurination and depyrimidination (Giglia-Mari et al., 2011). In hydrolytic deamination, one nitrogenous base is converted into another one. For example, the deamination of cytosine results in the production of uracil. Depurination and depyrimidination results in the removal of purine and pyrimidine bases, this makes the DNA backbone more susceptible to breakage (Giglia-Mari et al., 2011).

Reactive oxygen species (ROS) and nitrogen compounds can be produced at sites of infection and inflammation by immune cells such as macrophages and neutrophils. ROS can attack DNA resulting in the formation of adducts that can block the replication and transcription process and/or impair base pairing. In addition to this, ROS can result in base loss or DNA single stranded breaks (SSBs). SSBs can in turn result in double stranded breaks (DSBs) should they occur in close proximity to one another or if the DNA-replication machinery encounters a SSB (Maynard et al., 2009). DSB can also form as a result of ionizing radiation or by the RAG protein complex when initiating VDJ recombination. Amongst the various types of DNA damage, DSBs are the most deleterious as they affect both DNA strands and can lead to the loss of genetic information (Polo and Jackson, 2011).

Table 1: A summary of DNA damage repair mechanisms



Type of DNA damage	Repair mechanism
Methylated Guanine	Direct repair
Oxidized/deaminated bases	Base excision repair
Bulk DNA lesions, DNA-protein adducts	Nucleotide excision repair
Mismatched bases	Mismatch repair
Double stranded DNA breaks	Non-homologous end joining
Double stranded DNA breaks	Homologous recombination
Double stranded DNA breaks	Single strand annealing

Environmental factors can also elicit DNA damage. Ultraviolet light (UV) is a common exogenous factor. The sun emits three different types of UV radiation: UV-A, UV-B and UV-C. DNA breaks caused by UV-B irradiation has been extensively studied and observed in cells. UV-B radiation can lead to the formation of two common types of DNA lesions: cyclobutene pyrimidine dimers (CPDs) and pyrimidine 6-4 pyrimidone photoproducts (6-4PPs). These lesions may cause secondary breaks and have been associated with transcription and replication blockage. The latter can in turn result in the production of DSBs at sites of collapsed replication forks of CPD-containing DNA (Sinha and Hader, 2002).

1.1.3 Cellular response to DNA damage

Damaged DNA is extremely cytotoxic and failure to repair DNA can result in cell death, genome instability, a wide variety of cancers, neurodegenerative disorders and genetic disorders (Hakem, 2008). To ensure viability, cells have evolved the ability to detect and respond to the initial DNA damage signals and elicit cellular response. The cellular response includes: chromatin remodelling, DNA repair, cell cycle arrest, senescence and apoptosis. These responses are collectively known as the DNA damage response (Norbury and Hickson, 2001).

1.1.3.1 Chromatin remodelling

The repair of damaged DNA requires the coordination of posttranslational modification events and the recruitment of DNA repair proteins at sites of damage. The packaging of DNA into higher-ordered structures can limit cellular responses to repairing DNA by behaving as a barrier preventing access of repair proteins to sites of damage. Repair machineries must therefore be able to first detect DNA damage and then remodel the chromatin structure to allow access to sites of damage, repair damage and finally restore local chromatin organisation. In eukaryotic organisms, histone-modifying enzymes and ATP-dependent chromatin remodelling complexes are central to chromatin remodelling (Liu, B.H. et al., 2012).

Each histone has a flexible N-terminal tail protruding out of the nucleosome. The tails are targeted by histone modifying enzymes to either deposit or remove posttranslational modifications such as methylation, acetylation, ubiquitylation and phosphorylation. The modification of histone via their tails can induce conformational changes in nucleosomes, exposing docking sites for the binding of remodelling complexes (Kouzarides, 2007; Liu, B.H. et al., 2012).

Chromatin remodellers are large multi-subunit complexes conserved throughout eukaryotes. They are characterised by their highly conserved ATPase subunit and can be further categorised into four subfamilies: SWI/SNF (SWItch/Sucrose Non-Fermentable), ISWI (Imitation SWItch), CHD (chromodomain helicase DNA-binding) and INO80(INOsitol requiring 80). See Figure 2 (Clapier and Cairns, 2009). Through the binding and hydrolysis of ATP, chromatin remodellers disrupt histone-DNA interaction resulting in structural changes to the architecture of chromatin. The structural changes of chromatin can take place in the form of: 1) the repositioning or sliding of nucleosomes on DNA to expose specific sites, 2) the removal of histone octamer to expose DNA, 3) exchanging or incorporating histone variants for canonical core histones and 4) removing or exchanging histone dimers to destabilise the nucleosome and unwrap DNA. The remodelling of chromatin and therefore the exposure of damaged DNA initiates the next stage of repair as the exposed DNA can now be targeted by repair proteins (Liu, B.H. et al., 2012).

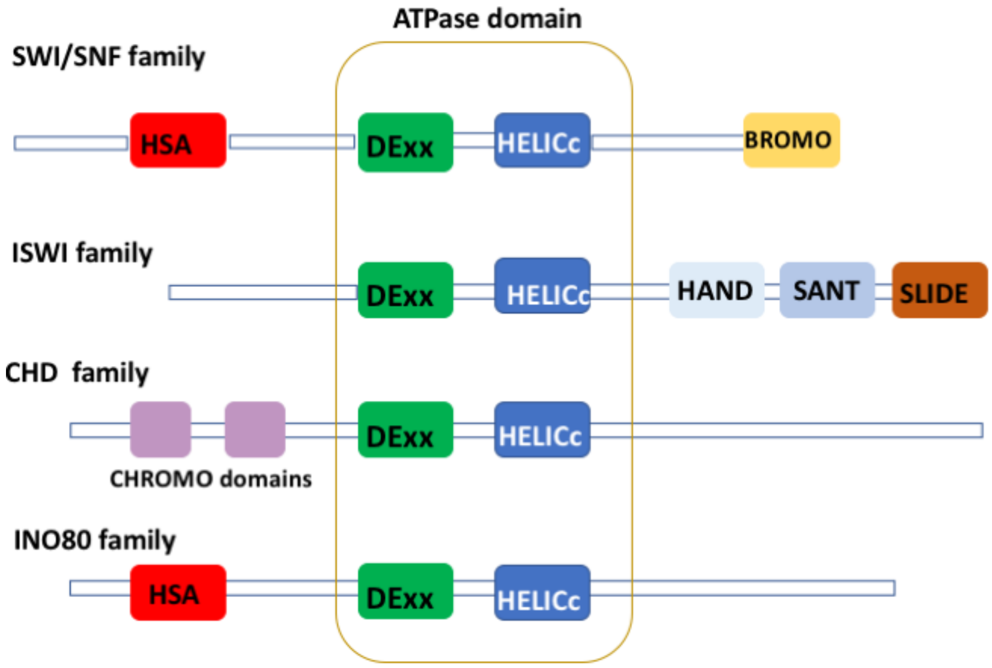


Figure 2: Schematic diagram of chromatin remodeller family

The DExx and HELICc domains are conserved throughout the families. The HAS domain shared between the SWI/SNF and INO80 family is involved in binding actin and actin related family. The BROMO domain recognises acetylated lysine residues in histone tails. The HAND (named because it's 4-helical structure resembles a hand, SANT (Swi3, Ada2, N-cor and TFIIIB) and SLIDE (SANT-like ISWI domain) domain of the ISWI family are 12 helical structures responsible for nucleosome interaction. The CHD family CHROMO domains mediate chromatin interaction by binding to methylated lysine residues in histone tail.

1.1.3.2 DNA repair

The cell has specific repair mechanism for different types of DNA lesions. For example, the mismatch repair (MMR) mechanism is responsible for replacing mis-paired DNA bases with the correct bases and chemically altered bases are repaired by base excision repair (BER) through the excision of the damaged base. SSBs and DSBs are repaired by single-strand break repair (SSBR), nonhomologous end joining (NHEJ) and homologous recombination, respectively.

The DNA damage response (DDR) pathway is mediated by proteins belonging to the phosphatidylinositol 3-kinase-like protein kinase (PIKKs) family, namely ataxia-telangiectasia mutated (ATM), ATM- and Rad3-related (ATR) and DNA-dependent protein kinase (DNA-PK) and by members of the poly(ADP-ribose) polymerase (PARP) family (Harper and Elledge, 2007). The PARP family is composed of 16 members, however, only PARP1 and PARP2 are believed to be involved in DDR (Ali et al., 2016).

Upon DNA damage and following chromatin remodelling, sensor proteins identify DNA damage and activate DDR by recruiting DDR factors to the site of damage. The recruitment and assembly of said factors is reliant on a wide variety of reversible post-translational modifications, namely: phosphorylation, ubiquitylation, SUMOylation, acetylation, methylation and poly-ADP-ribosylation (Oberle and Blattner, 2010). The post-translational modifications behave as specific signals and are recognized by various protein domains. A combination of PTMs are involved in the various DNA damage repair pathways that collectively make up DDR. This introductory chapter will focus solely on the repair pathways elicited by DSBs.

1.1.3.2.1 Double-stranded DNA breaks

There are three independent pathways assigned to repair DSBs: Homologous recombination (HR), Non-homologous end joining (NHEJ) and alternative end joining (a-EJ). Following DSBs, sensor proteins recognise aberrant DNA and activate upstream kinases. However, kinases also play a role in activating sensor proteins in response to DSBs. DSBs are primarily recognised by four sensors: MRN (Mre11-RAD50-Nbs1), Ku70-Ku80, PARP and RPA (Replication protein A) (Ciccia and Elledge, 2010).

1.1.3.2.2 The role of ATM and ATR in DSB repair

ATM and ATR kinases are responsible for the phosphorylation of many DDR proteins at ser/thr-glu motifs (Marechal and Zou, 2013). Following DSBs, ATM and ATR activate CHK11, CHK22 and MK2 kinases by phosphorylating them (Marechal and Zou, 2013). Research has shown that *in vivo* and *in vitro*, the specificities and function of ATR and ATM are distinct. ATM primarily functions in response to DSBs whereas ATR displays broader spectrum of activity, including DSBs. The two kinases are therefore part of two independent sub-pathways (Marechal and Zou, 2013). The MRN complex functions as a sensor in DSB repair (DSBR) and recruits and activates ATM, which in turn phosphorylates its substrates and histone H2AX in flanking nucleosomes. ATM has many substrates, some of which include chk2, p53 and BRCA1 (Shiloh, 2003). The mechanism behind the activation of ATM by MRN complex remains poorly understood. However, the phosphorylation of H2AX is known to be a key factor for the recruitment of many DDR proteins and chromatin remodelling complexes at sites of DSBs (Marechal and Zou, 2013).

In line with its broad spectrum of activity, ATR kinase has been shown to form a complex with ATRIP (Ball et al., 2005). The complex localises to sites of DNA damage, directly interacting with RPA-coated ssDNA. This suggests that ssDNA is a key structure in the activation of ATR at sites of DSBs (Zou and Elledge, 2003). The single-stranded DNA breaks are generated at DSB by nucleases during a process termed re-sectioning. The increasing amount of ssDNA generated at DSBs reduces the efficiency at which ATM repairs DSB as there are less DSB, this demonstrates an on/off regulation of the kinases. The activation of ATR in the presence of both ssDNA and ssDNA-dsDNA junctions at sites of DNA damage is likely to provide a fail-safe mechanism to the DNA damage repair pathway (Marechal and Zou, 2013).

1.1.3.2.3 Role of Ubiquitin and Ubiquitin-like modifiers in DSBs

In DSBR, ATM phosphorylates histone H2A variant H2AX at serine 139. The phosphorylation of H2AX generates a binding site for the scaffolding protein Mitotic Chromosome Determinant 1 (MCD1). MCD1 is subsequently phosphorylated by ATM kinase, which leads to the recruitment of RNF8-UBC13 ubiquitin ligases (Mandemaker et al., 2017). RNF8-UBC13 generate K-63 linked polyubiquitin chains on H1 linker histone which is recognised by RNF168. RNF168 in turn ubiquitylates histone H2A at K-13 and K-15 generating a platform for the recruitment of repair factors with ubiquitin binding domains such as RAP80, 53BP1 and RAD18. Ubiquitin-like modifiers have also been implicated in DSBR (van Attikum and Gasser, 2009). An increasing number of repair proteins were shown to undergo SUMOylation mediated by the SUMO E3 ligases PIAS1 and PIAS5 (Galanty et al., 2009). BRCA1, Rap80, MDC1 and 53BP1 are examples of known repair proteins that undergo SUMOylation (Galanty et al., 2009). The functional role of SUMOylation in the DNA damage repair pathway is not clear, but it is possible that based on the proteins that undergo SUMOylation the PTM regulates protein-protein interaction or aids in the regulation of activity and/or localisation of DNA repair proteins.

1.1.3.2.4 Homologous recombination (HR) repair

The primary cellular action in response to DSB is the localisation of the MRN complex to sites of DNA damage. Followings its recruitment to DSBs, MRN interacts with ATM via its MRE11 subunit and activates it. ATM then phosphorylates downstream DDR factors trigging a down-stream signal transduction pathway preventing the progression of the cell cycle and DNA replication. See Figure 3.

Homologous recombination repair takes place in the S/G2 phase of the cell cycle and requires sister chromatids as homology templates (Li, X. and Heyer, 2008). The pathway repairs double-stranded DNA breaks with high fidelity. The molecular mechanism for HR can be described in three stages: 1) Resection 2) Strand invasion and 3) ligation (Li, X. and Heyer, 2008).

In HR repair, the broken DNA ends are re-sectioned by nucleases such as CtIP into 3'-OH overhang ends (Jasin, M. and Rothstein, R., 2013). The resulting ssDNA tail is recognized and coated by RPA. RPA-bound ssDNA recruits and activates ATR which phosphorylates chk1 to delay the cell cycle (Li, X. and Heyer, 2008). Following this, rad51 is loaded onto ssDNA with the aid of the BRCA1-PALB1-BRCA2 complex. Rad51-bound single-stranded DNA then facilitates the search for a homologous duplex template, and strand invasion and exchange in an ATP dependent manner (Li, X. and Heyer, 2008). During strand invasion and exchange, damaged DNA strands are paired with a homologous region of the undamaged DNA duplex. This forms an intermediate branched DNA structure, termed a Holliday junction. Holliday junction intermediates can slide or branch-migrate along the newly joined DNA. Branch migration involves the extension of the heteroduplex by DNA polymerases beyond the original cross-over point and ligation by DNA ligase-I. The final step of HR involves the cleavage and resolution of the Holiday junction by Mus91 and GEN1 to generate recombinant DNA (Jasin, Maria and Rothstein, Rodney, 2013) (Figure 3).

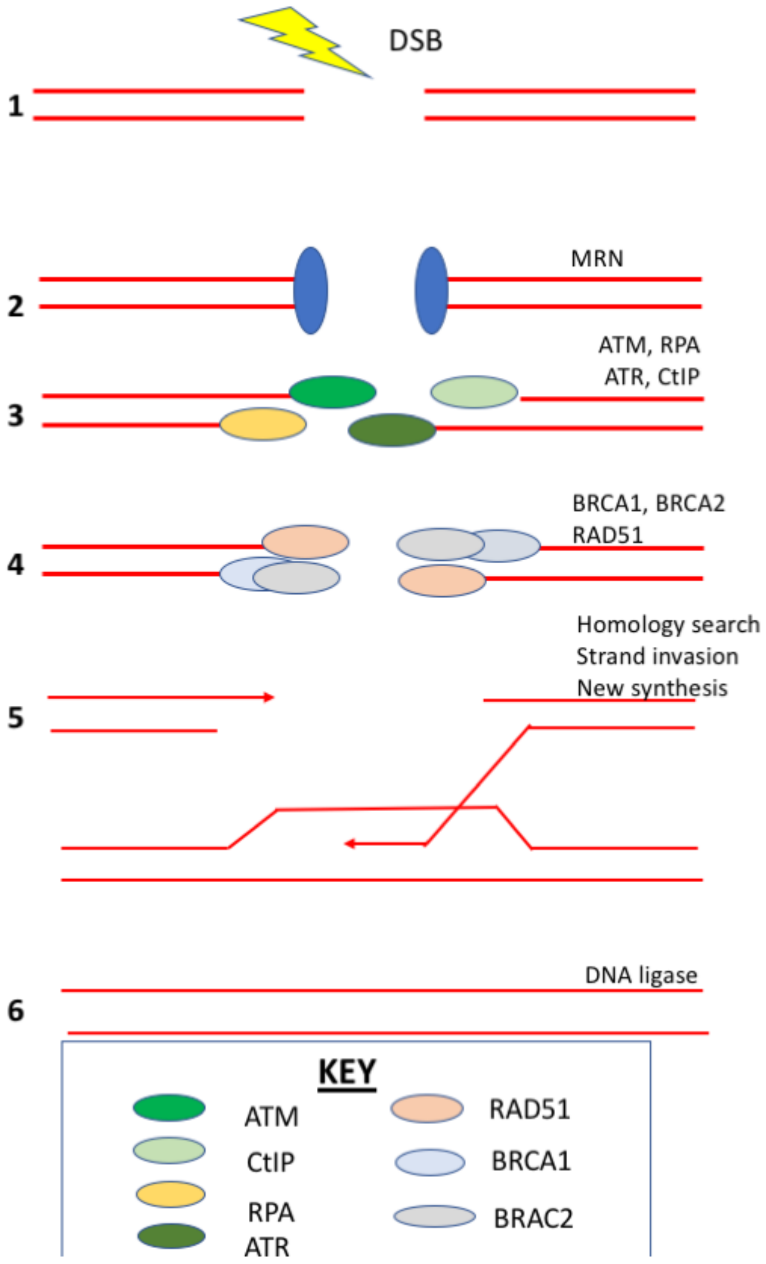


Figure 3: A schematic diagram illustrating the HR repair pathway

1-2) Following DSB, the MRN complex migrates to the site of damage and activates the ATM kinase, which in turn activates the DDR. 3) CtIP-mediated nuclease activity carry 5’ to 3’ end resectioning resulting in the formation of single-strand DNA, which is then coated by RPA and activates the ATR response and facilitates HR repair. 4-6) RAD51 assembles onto ssDNA replacing RPA and performs homology sequence searching and mediates strand invasion. DSBs are repaired by branch migration of the joint DNA molecule, DNA synthesis, ligation and resolution of the Holliday junction.

1.1.3.2.5 Non-homologous end joining (NHEJ)

NHEJ mediates the direct re-sectioning of DNA DSBs. The re-sectioning process of NHEJ differs from that of HR as the HR requires homologous template and occurs in the S and G2 phase of the cell cycle (Weterings and van Gent, 2004). NHEJ repair proteins include DNA ligase IV, XRCC4, DNA-PKc, Ku proteins and Artemis. The pathway can be summarised in three crucial steps: DNA binding and bridging, terminal end joining, and ligation (Weterings and van Gent, 2004). The initiation of NHEJ occurs upon the binding of Ku to DSBs. Ku is a heterodimer composed of Ku70 and Ku80. Ku forms a bridge between damaged DNA ends supporting and aligning the ends as well as protecting them from degradation. Ku effectively aligns DNA in a manner that still allows access to polymerases, nucleases and ligases. The binding of Ku to DSBs results in the recruitment of DNA-PKcs. DNA-PKcs have been shown to phosphorylate DNA ligase IV and XRCC4 in-vitro potentially enabling their interaction with Ku (Weterings and van Gent, 2004).

NHEJ requires blunt DNA ends for ligation to take place. DNA polymerases μ and λ are therefore required to generate blunt ends (Mahajan et al., 2002). Alternatively, the single-stranded ends can be processed by the nuclease activity of Artemis. In order for this to happen, DNA-PKc phosphorylates and binds to Artemis forming the Artemis-DNA-PKc complex (Weterings and van Gent, 2004). The newly formed complex has nuclease activity and trims 5' or 3' DNA overhangs. The final stage of NHEJ is ligation. Once blunt DNA ends have been generated, XRCC4 forms a complex with DNA ligase IV and is recruited to DNA. DNA ligase IV then carries out its ligation activity on the blunt-ended DNA (Figure 4). XRCC4 takes on a regulatory role within the complex and stabilises DNA ligase IV, it also directs the ligase to sites of DSB and helps to stimulate its activity (Weterings and van Gent, 2004). Unlike HR, NHEJ is an error prone repair mechanism.

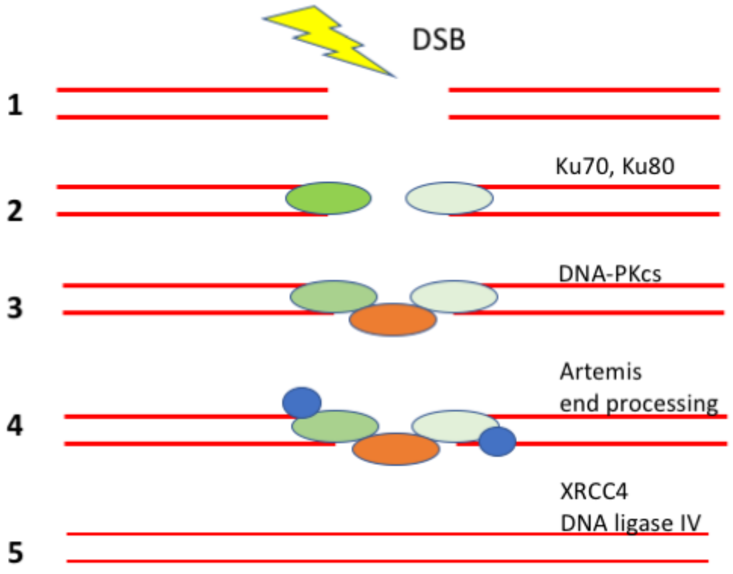


Figure 4: A schematic diagram of the NHEJ repair pathway

1-2) Following DSB, the ku70-ku80 heterodimer interacts with damaged DNA. 3) This facilitates the recruitment of DNA-PKcs. 4) NHEJ requires blunt ends, DNA strands are processed by nuclease activity of Artemis. 5) Once blunt end has been generated, XRCC4 forms a complex with DNA ligase IV and is recruited to blunt DNA. DNA ligase IV directly ligase DSB.

1.1.3.2.7 Cell cycle checkpoints

DNA damage checkpoints are composed of sensors, signal transducers, effectors and mediator proteins that detect DNA damage and initiate signal transduction pathways to slow or arrest cell cycle progression in response to DNA damage (Zhou and Elledge, 2001). Eukaryotic cells have four phases within their cell cycle: G1, S, G2 and M. They also have a single phase outside of the cell cycle termed G0, also known as cell cycle arrest. Progression through the cell cycle is tightly controlled, and proteins involved in regulating the progression through all four phases have been shown to also play vital roles in checkpoint responses (Zhou and Elledge, 2001). The checkpoints of the cell cycle are: G1/S, intra-S and G2/M2 (Figure 5). Although checkpoints have four defined components, there is no defined demarcation between each component.

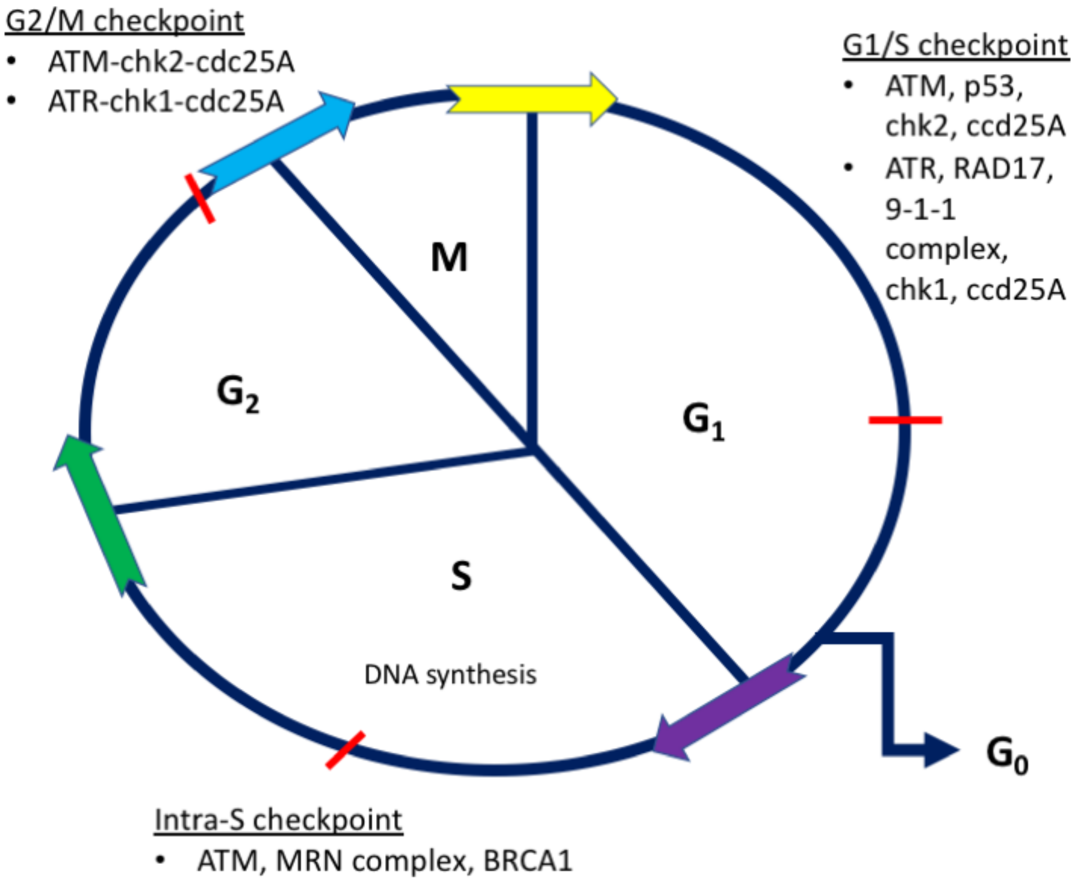


Figure 5: Cell cycle checkpoints

Checkpoints ensure that damaged or poorly replicated DNA is sensed and repaired. Damaged DNA can be sensed via two different pathways in G1 depending on the cause of damage. Ionization radiation damage is sensed by ATM. ATM phosphorylates chk2 which in turn phosphorylates cdc25A. UV light damage is sensed by ATR which phosphorylates chk1. Phosphorylated chk1 in turns phosphorylates cdc25A. G1/S phase arrest is maintained by p53. The intra-S checkpoint ensures that unrepaired or poorly replicated DNA do not progress to G2. ATM, MRN complex and BRCA1 play a central role in the intra-S checkpoint. Depending on the type of DNA damage ATM-chk2-cdc25A or ATR-chk1-cdc25A are activated to arrest the cell cycle in G2.

1.1.3.2.7.1 G/S checkpoint

The G1/S checkpoint prevents cells from entering the S phase in the presence of DNA damage by inhibiting the initiation of replication. Under normal conditions, cells in the G1 phase are committed to entering the S phase at a stage termed the restriction point in mammals (Abraham, 2001). The restriction point precedes the start of DNA synthesis by approximately two hours. During DNA damage, entry to S phase is prevented. DSBs caused by ionization radiation result in the activation of ATM, which in turns phosphorylates target molecules such as p53 and CHK2 (Bertoli et al., 2013). Phosphorylation of ATM substrates then activates two signal transduction events, one to initiate G1/S arrest and the second to maintain it. The phosphorylation of CHK2 by ATM initiates G1/S arrest which in turn phosphorylates CDC25A phosphatases causing its inactivation by nuclear exclusion and ubiquitin-mediated proteolytic degradation. The lack of active CDC25A results in the accumulation of phosphorylated (inactive) CDK2, which is incapable of phosphorylating CDC45 to initiate DNA replication(Sancar et al., 2004).

DNA damage caused by UV-light is sensed by ATR, RAD17-RFCC and the 9-1-1 complex resulting in the phosphorylation of CHK1 by ATR (Parrilla-Castellar et al., 2004). Activated CHK1 then phosphorylates CDC25A leading to G1 arrest. Regardless of whether G1/S arrest is mediated by ATM-CHK2-CDC25A or ATR-CHK1-CDC25A, G1/S checkpoint arrest is maintained by p53. During maintenance phase, p53 is phosphorylated (Sancar et al., 2004). The phosphorylation of p53 inhibits its nuclear export and degradation causing the accumulation of p53 within the nucleus. P53 therefore activates target genes, which bind and inhibit the CDK2-cyclin E complex which is responsible for promoting the S phase, thus maintaining G1/S phase arrest (Sancar et al., 2004).

1.1.3.2.7.2 Intra-S checkpoint

The intra-S checkpoint is activated by damage encountered during the S phase or by unrepaired damage that escaped G1/S checkpoints and leads to blockage in replication. Damage sensors for the intra-S checkpoints encompass large sets of checkpoint and repair proteins (Iyer and Rhind, 2017). For example, during DSBS, ATM, MRN complex and BRCA1 are all required for the activation of the intra-S checkpoint. Biochemical data shows that all three proteins bind DSBs, acting as sensors. ATM, MRN complex and BRCA1 activate intra-S checkpoint by kinase signalling cascades, inhibiting replication and DNA synthesis (Iyer and Rhind, 2017; Sancar et al., 2004).

1.1.3.2.7.3 G2/M checkpoint

G2/M checkpoints prevent cells from undergoing mitosis in the presence of DNA damage. Depending on the type of DNA damage, the ATM-CHK2-C DC25A or ATR-CHK1-CDC25A pathway is activated to arrest cell cycle following DNA damage during G2. In UV-related DNA damage, ATR-CHK1 initiates cell cycle arrest and ATM-CHK2 signals the maintenance of the arrest (Sancar et al., 2004). In other causes of DNA lesion such as ionizing radiation induced DSBs, cell cycle arrest is initiated by ATM-CHK2 and maintained by ATR-CHK1.

1.1.3.2.8 DSB induced apoptosis

Despite all the checkpoints in place to recognise and repair damaged DNA, critical DNA lesions such as DSBs can still trigger apoptosis (Norbury and Zhivotovsky, 2004). ATM and ATR are activated following detection of DSBs and they phosphorylate downstream substrates such as P53. P53 is a known transcriptional activator of pro-apoptotic factors such as BAX, FAS and PUMA. Phosphorylated p53 functions as a cell cycle checkpoint by upregulating p21 which in turn triggers G1/S phase arrest (Roos and Kaina, 2006) (Norbury and Zhivotovsky, 2004). Low levels of DSBs are believed to activate a sufficient amount of P53 required for the upregulation of p21 and therefore G1/S arrest (Roos and Kaina, 2006). High levels of DSBs, on the other hand, result in the accumulation of p53 beyond a predetermined threshold. The high level of p53 is believed to drive the activation of pro-apoptotic genes such as BAX, FAS receptor and PUMA (Roos and Kaina, 2006).

Poorly repaired DSBs are also likely to induce apoptosis. As damaged DNA is cytotoxic, it is reasonable to assume that DSBR pathways are anti-apoptotic. However, NHEJ is an error prone mechanism, it is possible that genomic changes caused during repair by NHEJ may also trigger apoptosis should those changes be deemed lethal by the cell. DNA damage-induced apoptosis is a mechanism used by anticancer drugs such as cisplatin to trigger a cascade of events and induce apoptosis, killing cancer cells (Gonzalez et al., 2001).

1.2 The ubiquitin pathway

Ubiquitylation is an ATP-dependent process executed by three classes of enzymes: 1) ubiquitin-activating enzyme (E1), 2) ubiquitin-conjugating enzyme (E2) and 3) ubiquitin ligase (E3) (Smalle and Vierstra, 2004). The cascade is initiated by ATP-dependent activation of the E1 enzyme, which conjugates the Glycine-76 carboxyl-group of free ubiquitin to its active site cysteine, resulting in an E1-ubiquitin (E1-Ub) thioester intermediate. Ubiquitin (Ub) is then transferred from the E1 enzyme to the active site cysteine residue of the E2, generating the thioester E2-ubiquitin conjugate (E2-Ub). E3 ligases orchestrate the final Ub transfer from E2-Ub to the protein substrate (figure 6). There are four classes of E3 ubiquitin ligases: homologous to the E6AP carboxyl terminus (HECT), really new interesting new gene (RING), ring between ring (RBR) and RING-Cys-relay (RCR)(Pao et al., 2018; Marín et al., 2004; Bailly et al., 1997; Kumar et al., 1997). HECT E3s utilise a two-step mechanism where ubiquitin is first transferred from the E2 to a catalytic cysteine in the E3 ligase, and then to a lysine residue on the substrate. RING E3s, on the other hand, bind to both the E2-Ub and the protein substrate, and facilitate the direct transfer of ubiquitin from the E2 to the substrate (Metzger et al., 2012). RBR ligases are characterised by a RING-HECT hybrid mechanisms involving two ring domains (RING1 and RING2) connected by an In Between Ring (IBR) domain. They carry out their function by facilitating the direct transfer of Ub to the substrate (Dove and Klevit, 2017) (Figure 6). The RCR E3 has two catalytic cysteine residues. The upstream cysteine accepts ubiquitin from the upstream E2 and the residue resides within a dynamic region termed the mediator loop. This loop region was found to be disordered within the obtained structural model. It is proposed that the inherent flexibility of this mediator loop allows the RCR E3 to transfer the attached ubiquitin molecule to a downstream catalytic cysteine residue. The two cysteine residues are found in a novel zinc-binding protein fold dubbed the tandem cysteine domain. The human genome encodes two E1 enzymes, 38 E2s and over 600 E3 ligases. The high number of E3 enzymes ensures greater substrate specificity, allowing for the ubiquitylation of thousands of proteins (Ye and Rape, 2009).

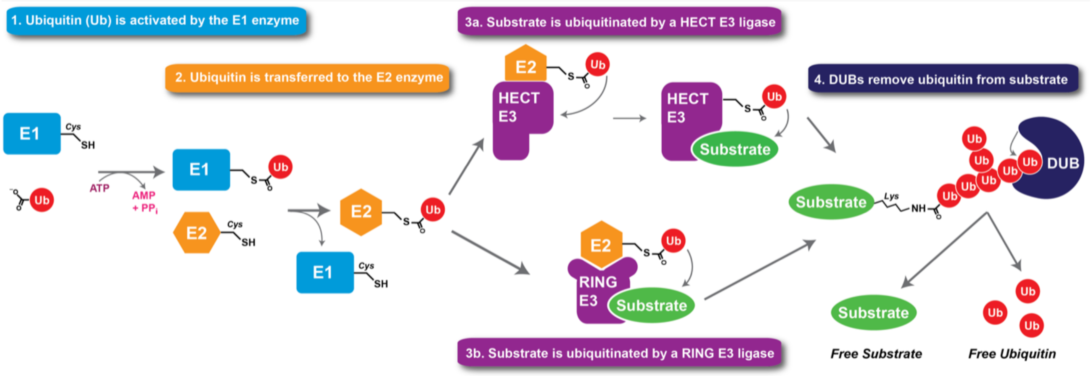


Figure 6: A schematic overview of the ubiquitin pathway

During the ubiquitin pathway, ubiquitin (red) binds to the cysteine residue of the E1 enzyme in an ATP dependent manner. Ubiquitin is then transferred from E1 to the E2 enzyme. The tagging of a substrate by ubiquitin can occur via four pathways. 1) HECT E3 can bind to the E2-ubiquitin complex resulting in the transfer of Ub from E2 to HECT E3, which then tags the substrate with ubiquitin. 2) Alternatively, the target protein and the E2-Ub complex can bind to RING E3 allowing for the direct transfer and ubiquitylation of the substrate. 3) The E2-ubiquitin complex binds to RBR E3, RING2 of RBR forms a covalent bond with ubiquitin and then ubiquitin is transferred to the target substrate. 4) The upstream RCR cysteine binds accepts ubiquitin from E2 ligase, the ubiquitin moiety is then transferred to the downstream cysteine residue. The ubiquitin pathway is antagonised by deubiquitylating enzymes (dark blue), which cleaves the isopeptide bonds recycling ubiquitin and freeing the substrate.

1.2.1 Ubiquitin structure

Ubiquitin is a 76 amino acid protein that is expressed in all eukaryotic cells and adopts a β grasp fold (Figure 7) with a flexible six residue C-terminal tail (Komander and Rape, 2012). The majority of the residues that make up the core of ubiquitin are rigid. However, the β1-β2 loop containing leucine 8 is flexible (Lange et al., 2008). Ubiquitin has multiple interaction surfaces. It is typically recognized through a hydrophobic surface consisting of Isoleucine-44, Leucine-8, Valine-70 and Histidine-68. Isoleucine-44 is recognised and bound by the proteasome and some ubiquitin binding domains (UBDs) (Dikic et al., 2009). Ubiquitin's second hydrophobic surface is centred on Isoleucine-36 and includes two leucine residues (Leucine-71 and -73) located on the ubiquitin tail (Komander and Rape, 2012). Isoleucine-36 mediates the interaction between ubiquitin molecules within chains and is recognized by HECT E3 ligases, deubiquitylases (DUBs) and UBDs (Komander and Rape, 2012). Glutamine-2, Phenylanine-4 and Threonine-12 make up the third ubiquitin patch (Hu, M. et al., 2002). This surface is required for cell division in yeast and also interacts with ubiquitin specific proteases (USPs), a family of deubiquitylating enzymes (DUBs). In higher eukaryotes, Threonine-12, Threonine-14, Glutamine-34, Lysine-6 and Lysine-11 make up a motif termed the TEK box required for mitotic degradation (Jin et al., 2008).

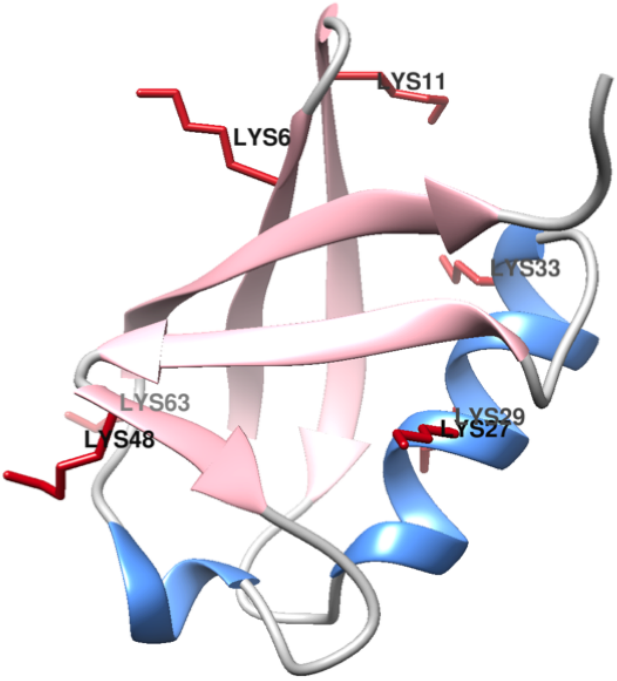


Figure 7: The ubiquitin structure

The ubiquitin structure is characterised a β-grasp ubiquitin fold composed of four β-strands (pink) and an α-helix (blue).

Ubiquitin is conjugated to substrates via one of its seven Lysine residues (K6, K11, K27, K29, K33, K48, K63) or its N-terminal Methionine (M1), all of which form attachment sites for chain assembly (Komander and Rape, 2012).The seven Lysine residues are dispersed throughout the ubiquitin surface and point in distinct directions (Figure 7) (Komander and Rape, 2012). K6 and K11 residues are located near the C-terminal tail and therefore in the most dynamic region of ubiquitin. The two residues may undergo conformational changes during ubiquitin chain formation and/or upon interaction with ubiquitin binding domains. K27, on the other hand, is buried within the core and would require localised structural changes in order to form chains. K29 and K33 are located on the alpha helix, K48 on a loop separating two beta strands and K63 at the loop. Ubiquitin is also subject to ubiquitylation. Its seven lysine residues can serve as substrates for another ubiquitin moiety forming different ubiquitin chains. Ubiquitin's methionine-1 residue located at the N-terminus can form a linear inter-ubiquitin linkage by being conjugated to the C-terminal α-carboxyl group of Glycine-76 (G76) of another ubiquitin protein

(Komander and Rape, 2012). Ubiquitin can therefore form eight different chains with diverse structures, which are involved in different biological pathways.

1.2.2 Ubiquitin chains

Ubiquitin chains can adopt compact or open conformations. K48 and K6 chains, for example, are known to adopt compact conformation (Cook et al., 1992; Varadan, R. et al., 2002).. By contrast, NMR and crystal structures of M1 and K63 ubiquitin chains revealed that they display an open conformation and exhibit high conformational freedom. It is therefore possible that binding partners specific for M1 and K63 ubiquitin chains recognise the chains based on the distance and flexibility between chain moieties. It has been suggested that ubiquitin binding domains of proteins may remodel flexible ubiquitin chains in order to increase interaction interfaces or specificity (Komander and Rape, 2012; Tenno et al., 2004; Varadan, Ranjani et al., 2004).

Ubiquitin can form heterogenous chains, further increasing the scope of their role within cells. Heterogenous chains can be defined as chains containing different lysine linkages (mixed chains) or branched chains. K48 of ubiquitin, for example, can form heterogenous K63 branched chains via the combined activities of the E3 ligases TRAF6 and HUWE1 (Ohtake et al., 2016). These branched chains have been shown to preferentially associate with the proteasome in the cell. However, the role of these chains differs in cells, and they have also been reported to enhance Nuclear Factor кappa-light-chain-enhancer of activated B cells (NF- κB) signalling (Ohtake et al., 2016). K11-K63 chains function non-proteolytically during endocytosis during NF- κB signalling (Wickliffe et al., 2011).

1.2.3 The role of Ubiquitin chains in the cell

The diversity of ubiquitin chains formed in cells is important for both proteolytic and non-proteolytic signalling. K6, K11, K27, K29 and K33 ubiquitin linkages are known as atypical chain types (Ikeda and Dikic, 2008). Atypical polyubiquitin chains are understudied due to a lack of tools that allow linkage-specific detection. By contrast, typical chains are composed of K48 and K63 linkages and are extensively studied (Michel et al., 2017). M1 ubiquitin linkages are created by the linear ubiquitin chain assembly complex (LUBAC), a 600 kDa protein(Kirisako et al., 2006) (Keusekotten et al., 2013).M1 ubiquitylation was previously believed to trigger proteasomal degradation, however it later came to light that MET-1 ubiquitylation is an important post-translational modification in the Nuclear Factor kB (NF-kB) signalling pathway (Hrdinka and Gyrd-Hansen, 2017).

Homogenous K11-linked chains have been shown to help mediate proteasomal degradation of client proteins (Wickliffe et al., 2011). K11-linked chains were first discovered as a product of the human E3 anaphase-promoting complex (APC/C), an essential regulator of cell division and the only known E3 to assemble homogenous K11 ubiquitin chains (Wickliffe et al., 2011). K33-linked polyubiquitylation of coronin 7 by CUL3-KLHL20 ubiquitin E3 ligase has been shown to regulate protein trafficking, in addition to this, K33 chains along with K6 linkages have been shown to increase after DNA damage (Akutsu et al., 2016). This demonstrates that a single type of ubiquitin linkage can have diverse roles within the cell. The contribution of K48 polyubiquitin chains in the proteasomal degradation pathway is well established. However, K48 linkages have also been demonstrated to have non-proteolytic functions within the cell (Mallette and Richard, 2012). K48 ubiquitin chains, like K63 chains, have been linked to the DNA damage repair pathway. For example, K48 polyubiquitin chains have been shown to accumulate at sites of DNA damage, suggesting a role for protein degradation during DNA damage response (Mallette and Richard, 2012). However, the molecular mechanism(s) controlling protein ubiquitylation and degradation following DNA damage remains unclear (Mallette and Richard, 2012).

In the DNA damage repair (DDR) pathway, K63 polyubiquitin chains are known to conjugate on histone H2A. A recent study revealed that K63 linked polyubiquitin chains can bind directly to DNA facilitating DDR (Liu, P. et al., 2018). In their work, the authors have shown that K63 polyubiquitin chains interact with DNA via a motif they termed DNA-Interacting-Patch (DIP), comprising Threonine-9, Lysine-11 and Glutamine-34. They also found that single-stranded DNA (ssDNA) showed slightly more preference to K63 chains than double-stranded DNA (dsDNA). K63 chains, within the context of DDR, serve as a platform for the recruitment of repair factors, such as the RAP80 protein, via the isoleucine44 patch of ubiquitin. Liu et al 2018 examined whether DNA binding to the DIP motif of ubiquitin interfered with the binding of repair factors such as RAP80 to K63 chains. They demonstrated that K63 chains bound to DNA retained their ability to interact with RAP80, suggesting that K63 linked polyubiquitin chains could form ternary complex with DNA, RAP80 or other DNA repair proteins cooperatively promoting DNA repair (Liu, P. et al., 2018).

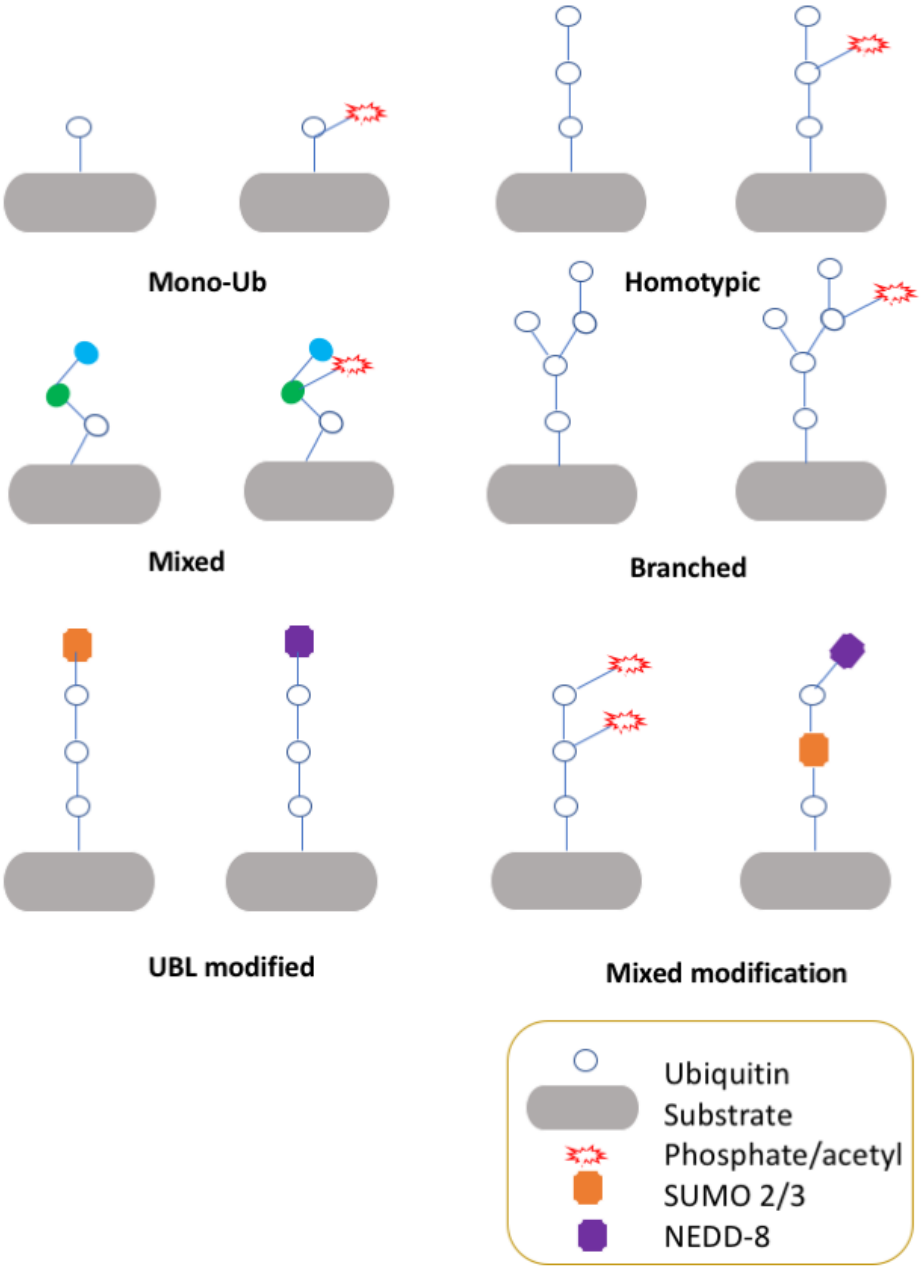


Figure 8: Overview of ubiquitin chain types

Ubiquitin chains present in the cell are extremely diverse. Protein molecules can be tagged with mono or di-, tri- or polyubiquitin chains. The chain can be homotypic, mixed (e.g. K11/K33 linked Ub chain) or branched. In addition to this, the ubiquitin moiety can also be modified by phosphorylation, acetylation, NEDDylation and/or SUMOylation further diversifying the chain. The variety in chain types is essential for the diverse biological pathways ubiquitin plays a role in.

1.2.4 Deubiquitylases

The post-translational modification (PTM) of proteins by ubiquitin is a reversible process orchestrated by deubiquitylating enzymes (DUBs). The human genome encodes over 100 DUBs (Reyes-Turcu et al., 2009), which are classified into seven families (Figure 9) based on sequence and domain conservation: Ubiquitin-specific proteases (USP), Ubiquitin C-terminal hydrolases (UCH), Machado-Joseph disease (MJD), Ovarian tumour proteases (OTU), Jab1/Mov34/Mpr1 Pad1 N-terminal (JAMM/MPN), Motif interacting with ubiquitin-containing DUB family (MINDY) and the newly identified zinc finger with UFM1-specific peptidase domain protein (ZUFSP) (Reyes-Turcu et al., 2009; Kwasna et al., 2018). USP, UCH, MJD, OTU, MINDY and ZUFSP DUBs are cysteine proteases whereas JAMM DUBs are metalloproteases.

DUBs regulate the number and type of ubiquitin chains on substrates in different ways. They can bind to substrates and deubiquitylate them, or they can bind directly to the ubiquitin signal. Some DUBs function as isolated enzymes, whilst others, work as part of bigger macromolecular machines which recruit substrates for deubiquitylation, such as the proteasome (Komander et al., 2009). Other DUBs contain additional protein interaction domains that help to facilitate their interaction with target substrates (Komander et al., 2009). Like E3 ligases, DUBs also exhibit chain type specificity. The recognition of a particular chain type typically requires DUBs to bind two ubiquitin moieties across the active site (Komander et al., 2009).

The activity and specificity of a DUB is dependent on the enzyme's ability to recognise its substrate(s). All DUBs have at least a single ubiquitin binding site , biochemical assays have suggested that the ubiquitin binding sites are responsible for facilitating the formation of the enzyme-substrate complex (Mevissen and Komander, 2017). The binding of ubiquitin to the ubiquitin binding site results in the extension of ubiquitin's C-terminus and its engulfment by DUBs. This activity requires partial unfolding of compact polyubiquitin chains such as K48 chains (Mevissen and Komander, 2017). The extension of polyubiquitin is enabled and dependent on the linkage type and chain length, which in turn affects DUB activity. It is possible that

DUBs with a single ubiquitin binding site indiscriminately remove ubiquitin from target proteins. Some DUBs contain multiple ubiquitin binding sites, enabling them to carry out their deubiquitylating activity with more specificity (Mevissen and Komander, 2017).

DUBs can also contain substrate binding sites within their catalytic domains, allowing them to bind directly to substrates and recognise specific ubiquitylated sites on substrates. Site-specific ubiquitylation studies are difficult and limited by the lack of available reagents (Mevissen and Komander, 2017). However, recent advances in chemical biology are changing this (Mevissen and Komander, 2017; van Tilburg et al., 2016). On the other hand, DUB linkage specificity has been extensively studied and the degree of linkage specificity has been found to vary between DUB families. For example, the majority of JAMM/MPN proteases display specificity for K63 linkages, MINDY DUBs are specific for K48 chains, whereas most USP proteins display little to no linkage preference (Mevissen and Komander, 2017).

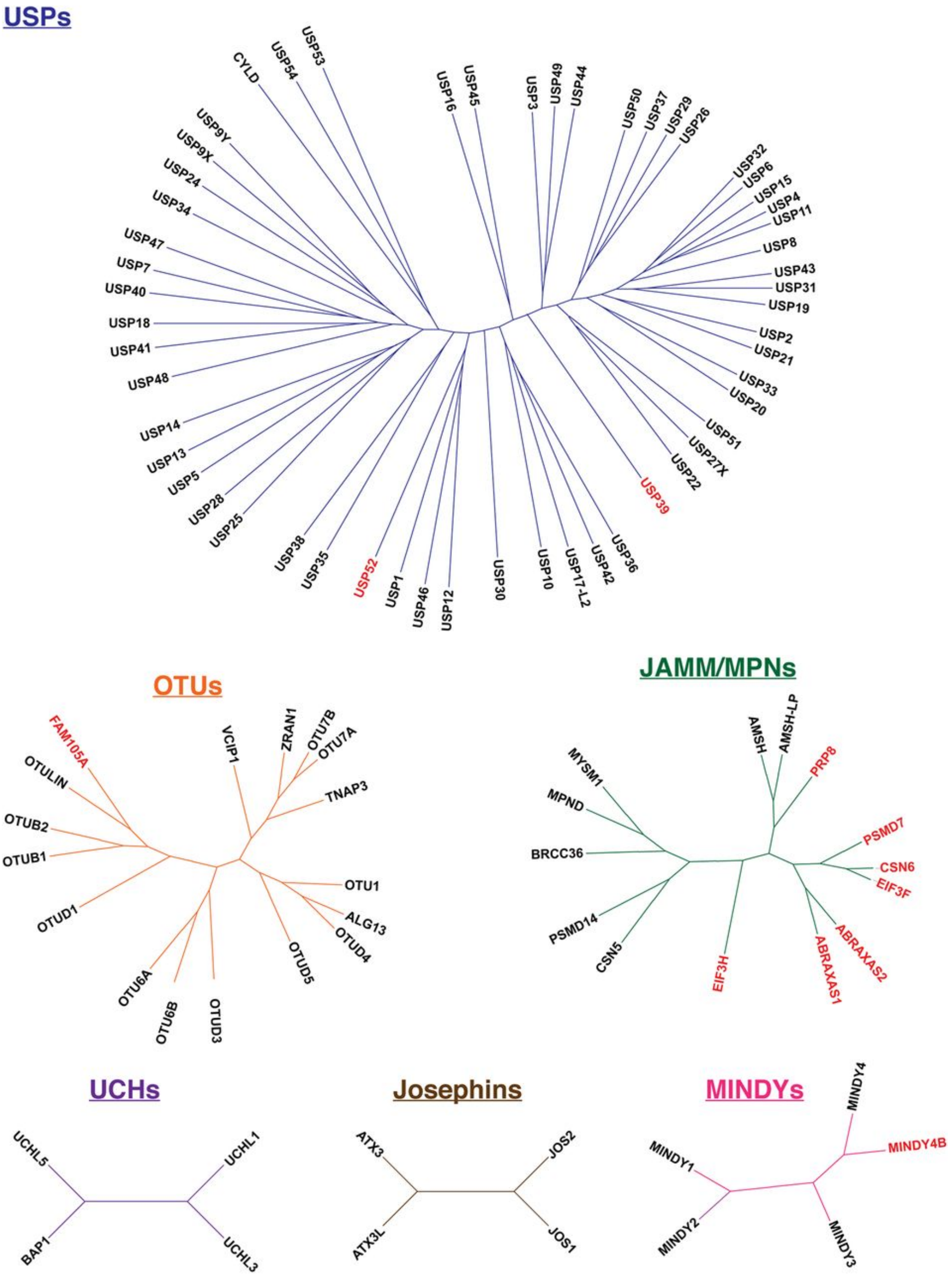


Figure 9: Phylogenetic trees of the six DUB families

Known DUBs are illustrated in the phylogenetic trees. Pseudo/MPN^{-} DUBs are shown in red. The ZUFSP DUB family is not shown as to date there has only been one protein identified in the family. Image obtained from: Walden, M., Masandi, S.K., Pawłowski, K. and Zeqiraj, E., 2018. Pseudo-DUBs as allosteric activators and molecular scaffolds of protein complexes. *Biochemical Society Transactions*, *46*(2), pp.453-466.

1.2.4.1 Mechanism of action

DUBs carry out their activity via endo- or exo-cleavage mechanisms (Mevissen and Komander, 2017). Exo-cleavage is an efficient approach that generates a free substrate and an unanchored polyubiquitin chain, which can be further processed by other DUBs. Endo-cleavage, on the other hand, produces mono-ubiquitin and requires DUBs to engage with a particular chain type multiple times so that each ubiquitin moiety can be progressively removed from the chain. Whether a DUB cleaves endo or exo is dependent on the family to which it belongs to, linkage type, and in the case of endo cleavage whether the ubiquitin chain is distally extended (Mevissen and Komander, 2017). As there are different forms of ubiquitylation, including branched ubiquitin chains, SUMO/ubiquitin hybrid chains, ubiquitin phosphorylation or acetylation, it is also possible that there may be new DUB activities that have yet to be discovered (Mevissen and Komander, 2017).

1.2.4.2 Regulation of Deubiquitylases

In addition to modulating signalling in the cells, DUBS themselves are regulated. (Mevissen and Komander, 2017). DUB regulation can be achieved by controlling their abundance and localisation in cells and/or their catalytic activity (Mevissen and Komander, 2017). In addition, DUBs can be regulated by undergoing post-translational modifications, post-translational processing or by interacting with binding proteins (Mevissen and Komander, 2017). Furthermore, the cellular abundance of DUBs can also be controlled by transcription, translation and degradation. As a result of this, the relative quantities of DUBs in cells and tissue types vary, and like other proteins, DUBs can undergo proteolytic processing, a mechanism that cleaves proteins modifying their activity (Mevissen and Komander, 2017). DUB localisation in cells can be achieved by binding to targeting proteins, via localization signals, or through specific protein:protein interactions that recruit DUBs to defined compartments within the cell (Mevissen and Komander, 2017). Finally, the functional role of DUBs can be positively and negatively regulated by PTMs, such as phosphorylation (Mevissen and Komander, 2017). A key example is the Breast Cancer Type 1 susceptibility protein A (BRCA1-A). For complex assembly and therefore recruitment to sites of double-stranded breaks, Abraxas1, a protein subunit of BRCA1-A is phosphorylated at serine 404 and serine 406 in an ATM and CDK-dependent manner (Wu, Q. et al., 2016)(). The phosphorylation of Abraxas1 facilitates interaction with the BRCT repeats of BRCA1, ensuring the formation of the BRCA1-A complex and enabling its recruitment to sites of DNA double-stranded breaks (Wu, Q. et al., 2016).

The majority of DUBs contain ubiquitin binding sites that can be ubiquitylated (for example when in complex with E3 ligases (Mevissen and Komander, 2017)). The ubiquitylation of DUBs can therefore behave as a mechanism for regulation. Some DUBs contain domains that can recognise and bind ubiquitin-like proteins, allowing them to also be post-translationally modified by ubiquitin variants. SUMOylation, for example, can behave as an inhibitory mechanism (Mevissen and Komander, 2017). The DUB CYLD is susceptible to inhibitory SUMOylation. The SUMOylation of its N-terminus inactivates the C-terminal catalytic domain (Mevissen and Komander, 2017)). Another example of DUB regulation can be learned from cysteine proteases. Cysteine proteases contain an active site cysteine susceptible to oxidation. Oxidation of the catalytic cysteine by reactive oxygen species (ROS) has been shown to inactivate members of the OTU, USP and UCH families (Mevissen and Komander, 2017; Cotto-Rios et al., 2012; Kulathu et al., 2013; Lee et al., 2013). DUBs belonging to the JAMM family are commonly found as part of large complexes. The activity of these DUBs is often regulated allosterically by domains within the enzyme or binding partners.

1.2.5 JAMM/MPN deubiquitylase family

JAMM/MPN deubiquitylases commonly exist as large protein complexes (Walden., et al., 2018) JAMM DUBs vary in their amino acid sequence but share a conserved catalytic JAMM motif and an MPN domain (Komander et al., 2009). JAMM protein domains exist in bacteria, archaea and eukaryotes, however bacterial JAMM domains do not appear to have DUB activity suggesting that during evolution the domain acquired a new function as a Ubiquitin protease (Feng, L. et al., 2010). JAMM

DUBs contain Zn^{2+} in their active site. The zinc ion is coordinated by two histidine residues, an acidic residue, such as aspartic or glutamic acid, and a water molecule. The water molecule is involved in the nucleophilic attack of the scissile isopeptide bond (Shrestha et al., 2014). Out of the 14 JAMM proteins, seven have zinc-binding capabilities and are referred to as MPN^{+} domains. Out of these seven, six exhibit isopeptidase activity against ubiquitin whilst the other requires interaction with MPN^{-} protein for activity. The remaining 7 JAMM proteins lack residues that are capable of coordinating the zinc ion, and are referred to as MPN^{-} or pseudo-DUBs as they are predicted to be inactive (Walden et al., 2018).

1.2.5.1 CSN5 and CSN6

The COP9 signalosome complex (CSN) is made up of 8 subunits (CSN1-8) and is involved in multiple biological processes, including signal transduction, development and the cell cycle. Its most well studied function is the deneddylation of the cullin subunit of cullin-RING ubiquitin ligases (CRLs) (Marchione et al., 2013). CRLs are the largest ubiquitin ligase family and the COP9 signalosome complex helps to regulate CRLs expression (Marchione et al., 2013). The fifth subunit of the complex, CSN5 (also known as Jab1), contains a JAMM domain that is crucial for the activity of CSN5. Mammalian CSN5 was first identified as a protein binding to the transcription factors c-Jun and JunD (Marchione et al., 2013). In addition to its role in deneddlyation, CSN5 has been implicated in various other biochemical functions such as the phosphorylation of c-Jun-NFkB and p53 by CSN-associated kinases, and the control of the intracellular distribution of signalling molecules (Marchione et al., 2013). CSN5, alone, is inactive as a result of its catalytic domains auto-inhibited conformation, however it becomes catalytically active when in complex with CSN6 (Birol et al., 2014). CSN6 is a pseudo-DUB (MPN-) domain protein. Many MPN DUBs contain insertion-1 and insertion-2 sequences within the MPN domain. Insertion-1 has been implicated in recognising the distal ubiquitin in a diubiquitin substrate and in correctly positioning the distal ubiquitin C-terminus over the DUB active site for isopeptide-bond cleavage (Worden et al., 2014). Insertion-2 on the other hand is believed to be involved in determining the specificity of DUBs for ubiquitin chain types (Worden et al., 2014). The CSN5 MPN domain insertions in human are located at residues 97-131 (inertion-1) and 197-219 (insertion-2) (Birol et al., 2014). The binding of CSN6 to CSN5 is believed to induce a conformational change in the insertion 1 region of CSN5, the conformational causes CSN5 to transition from its inactive to active form (Birol et al., 2014).

1.2.5.2 AMSH and AMSH-LP

Associated molecule with a SRC homology 3 domain of signal transducing adaptor molecule, STAM (AMSH) regulates ubiquitin signalling by hydrolysing the peptide bond between ubiquitin and its substrate and/or by cleaving single ubiquitin moieties from K63-linked polyubiquitin chains (Davies et al., 2013). AMSH is an essential regulator of the endosomal sorting complexes required for transport (ESCRT) complexes. The ESCRT machinery is composed of four protein complexes (ESCRT 0-IIII) and the AAA ATPase Vps4. ESCRTs down regulate their signals by modulating endosomal sorting and trafficking of cell-surface receptors to the lysosome for degradation (Davies et al., 2013) . The cell-surface receptors are firstly ubiquitinated and then passed to the ESCRT machinery, which promotes their internalization in endosomes and the subsequent fusion with lysosomes for the proteolysis (Davies et al., 2013).

ESCRT-0 and ESCR-III are the only complexes that can recognise deubiquitinase enzymes. ESCRT-0 achieves this recognition by binding the SH3 binding motif (SBM) of DUBs via its SH3 domain (Davies et al., 2013). ESCRT-III recognition, instead, occurs via its C-terminal MIT interacting motif (MIM) binding to the microtubule interaction and transport (MIT) domain of DUBs (Davies et al., 2013). An X-ray crystal structure of the MIT domain of AMSH bound to the C-terminal MIM fragment of the ESCRT-III complex CHMP3 has been determined (Davies et al., 2013). The structure confirmed that AMSH contains a MIT domain that interacts with the CHMP3 MIM. When in solution CHIP3 MIM is unstructured. The structure revealed that when in complex, CHMP3 MIM adopts a helical conformation. However, CD analysis suggests that parts of C-terminal CHMP3 MIM may be structured within full-length CHMP3, thus contributing to autoinhibition conformation of CHMP3 (Davies et al., 2013).

A clear role of AMSH has yet to be elucidated, but whole genome sequence analysis has however revealed that recessive mutations in the enzyme gives rise to microcephaly-capillary malformation (MIC-CAP) syndrome (Davies et al., 2013). The microcephaly phenotype has been correlated with the accumulation of ubiquitinated proteins indicating a loss of (DUBs) enzyme function (Davies et al., 2013).

The AMSH-like protein (AMSH-LP) has similar characteristics to AMSH. Both proteins are ubiquitously expressed in a wide range of human tissues and contain a nuclear localization signal MPN domain, and a JAMM motif (Kikuchi et al., 2003). Without the nuclear localization signal and the MPN domain, AMSH-LP and AMSH are unable to enter the nucleus (Kikuchi et al., 2003). The AMSH-LP MPN domain shares 75% amino acid sequence similarity with that of AMSH (Kikuchi et al., 2003), and the JAMM motif is completely conserved between the two proteins (Kikuchi et al., 2003). AMSH and AMSH-LP are believed to be functionally similar as they both enhance interleukin 2 mediated C-MYC induction (Kikuchi et al., 2003). However, AMSH-LP does not bind to the SH3 domain of signal transducing adaptor molecule 1 (STAM1) and Grb2 (Kikuchi et al., 2003), suggesting that it may have different functions to AMSH.

1.2.5.3 PSMD14 and PSMD7

The PSMD14 gene in human codes for the protein 26S proteasome non-ATP regulating subunit 14, and is also known as PAD1, POH1 and RPN11. PSMD14 is part of the 19s regulatory cap of the 26s proteasome and plays a vital role in removing ubiquitin from proteins targeted for degradation (Worden et al., 2014). Cryo-EM and X-ray crystallography studies have shown that PSMD14 forms a dimer with PSMD7 (also known as Rpn8) (Worden et al., 2014). The DUB activity of PSMD14 has been shown to be dependent on the hydrolysis of ATP and may be linked to other steps within the degradation process, such as substrate engagement, unfolding and translocation (Worden et al., 2014). A model has postulated that the ATPase-driven translocation of a ubiquitylated substrate pulls the isopeptide bond into the PSMD14 active site for cleavage (Worden et al., 2014).

It is difficult to carry out detailed mechanistic studies of PSMD14 because the process by which it functions is complex and the protein has been found to be extremely difficult to purify (Worden et al., 2014). What is known of PSMD14 is that its catalytic groove lacks features that are usually conserved in MPN DUBs, and that are believed to be imperative for ubiquitin binding or stabilization of the isopeptide bond for cleavage (Worden et al., 2014). Specifically, PSMD14 lacks the aromatic residue that is believed to contact the hydrophobic Ile44 patch of ubiquitin (Worden et al., 2014). Furthermore, it uses a flexible Insertion-1 loop in place of an extended β-sheet to position the C-terminus of Ub for cleavage. In addition to this, insertion-2 does not stabilize the isopeptide bond but instead it interacts with Rpn2, thereby anchoring PSDM14 to the proteasome complex (Worden et al., 2014). PSMD14 therefore differs to other MPN DUBs by the manner in which it interacts with ubiquitin.

1.2.5.4 eIF3F and eIF3H

Eukaryotic translation initiation factor 3 subunit f (eIF3f) and h makes up part of the translation initiation factor 3. Research conducted by Moretti et al identified eIF3f as the deubiquitylase targeting the mammalian Notch receptor (Moretti et al., 2010). The activation of the receptor relies on a variety of cellular events such as metalloprotease cleavage, endocytosis, mono-ubiquitination and gamma-secretase processing. Moretti et al revealed that knocking out the eIF3H resulted in the accumulation of mono-ubiquitinylated Notch and that this accumulation was reversed by murine WT eIF3F. Structural studies have shown that in the translation initiation complex, eIF3f and eIF3H interact with one another through their MPN domains (Des Georges et al., 2015). These studies have also revealed that the two proteins are pseudo DUBs as they lack the catalytic residues responsible for the binding of zinc, and that they contribute to the complex as scaffolding proteins (Des Georges et al., 2015).

1.2.5.5 MYSM1

Myb-like SWIRM and MPN domain containing protein 1 (MYSM1) is a histone H2A DUB, which plays a vital role in early B cell development (Jiang, X.-X. et al., 2011). MYSM1 deficiency interrupts early B cell commitment and defect of B cell progenitors' expression in EBF1(Jiang, X.-X. et al., 2011). Aside from MYSM1, there are another seven proteins with H2A-DUB activity, including six members of the thiol protease family USP and BRCA1-associated protein 1 (BAP1) (Nijnik et al., 2012). Various H2A-DUBs are known to regulate HOX gene expression in Xenopus and Drosophila development, and respond to injury in hepatocytes (Nijnik et al., 2012). However, their roles in mammalian stem cell development and haematopoiesis has yet to be elucidated. By contrast, A. Nijnik *et al* found that Mysm1 plays a critical role in haematopoiesis (Nijnik et al., 2012). They studied Mysm1 gene expression and phenotypic defects in mysm1 mouse lines and found that they were restricted to specific systems and tissues with limited global alteration in cell function and changes in gene expression. Indicating that, unlike other H2A-DUBs with global housekeeping roles, Mysm1 activity is restricted to certain cell types. This theory is supported by the fact that Mysm1 orthologs are only found in vertebrates, while other H2A-DUB, such as the USP22 yeast orthologue, Ubp8 are more widely conserved.

1.2.5.6 Prp protein

Prpf8 is highly conserved in sequence and is known to function as part of the pre-mRNA splicing process where it interacts with pre-mRNA and associates with the spliceosome(Grainger and Beggs, 2005). Prpf8 is found in many conformers of the spliceosome during the splicing cycling and is functional in all three variants of the spliceosome (U2 spliceosome, U12 spliceosome and the trans-spliceosome) (Grainger and Beggs, 2005). Bioinformatics analysis of the Prpf8 protein revealed that it contains a conserved RNA recognition motif (RRM), an N-terminal MPN domain and a putative nuclear localization signal (NLS) (Grainger and Beggs, 2005). Tran et al 2003 and Ambroggio et al 2004 solved structures of the Prp8 MPN domain and found that domain lacked one of the conserved histidine residues required for Zn2+ coordination(Tran et al., 2003; Ambroggio et al., 2003). Based on these studies, the MPN domain of PRP8 may behave as a protein-protein interaction domain and contributes to the structural integrity of the spliceosome complex.

1.2.6 MPN + MPN - pairs

Many of the MPN$^-$ proteins interact with MPN$^+$ domain containing proteins forming a stable complex (Figure 10). For example, the fifth subunit of the COP9 signalosome complex (CSN5) contains a JAMM/MPN domain that is crucial for its activity (Yoshida et al., 2013). However, independently, CSN5 (also known as Jab1) is inactive. This is the result of its catalytic domains auto-inhibited conformation (Birol et al., 2014). When in complex with the 6th CSN subunit (pseudo MPN$^-$ DUB), CSN5 becomes catalytically active (Birol et al., 2014). Interestingly, MPN- proteins can also form a complex with another MPN$^-$ protein. Take for example EIF3F and EIF3H that form the eIF3 complex. The two subunits dimerise with one another and interact with other members of the EIF3 complex (Des Georges et al., 2015). Despite being devoid of catalytic activity their role within the complex as scaffolding proteins is paramount for protein-protein interaction and protein stability(Marchione et al., 2013). PSMD14 (an MPN$^+$ protein) is part of the 19s regulatory cap of the 26s proteasome. The protein plays a vital role in removing ubiquitin from proteins tagged for degradation. Cryo-EM and X-ray crystallography studies have shown that PSMD14 forms a dimer with MPN$^-$ protein PSMD7 (Pathare et al., 2014; Worden et al., 2014; Beck et al., 2012).

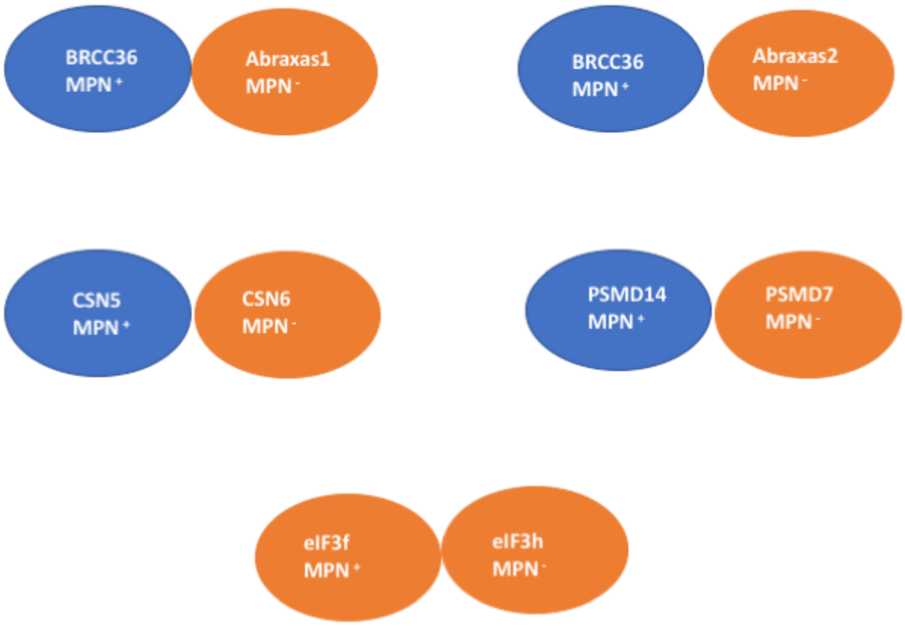


Figure 10: MPN$^+$ and MPN$^-$ pairs

Some MPN+ DUBS require interaction with MPN$^-$ proteins for activity. BRCC36 forms two independent complexes. Alone, the protein is catalytically inactive. It's interaction with Abraxas1 or 2 facilitates the assembly of ARISC and BRISC complex, respectively activating the protein. CSN5 is conformationally inhibited, its dimerization with CSN6 results in a conformational change that renders CSN5 catalytically active. The dimerization of PSMD14 with PSMD7 enables the complex to recognise and deubiquitylate substrates. Unlike the other dimers, the eIF3f-eIF3h dimer is composed of pseudo DUBs. The dimer is catalytically inactive and function as scaffolding proteins.

1.2.7 BRCC36 forms two distinct cellular complexes

The JAMM/MPN$^+$ domain-containing protein BRCA1-containing complex subunit 36 (BRCC36) forms two distinct complexes in the cell: the cytoplasmic BRISC-SHMT2 complex, and the related ARISC-RAP80 complex in the nucleus (Wang et al., 2007; Sobhian et al., 2007; Cooper et al., 2009). The BRISC-SHMT2 complex is composed of BRCC36, Abraxas2, BRCC45, MERIT40 and SHMT2. The ARISC-RAP80 complex, on the other hand, is composed of BRCC36, Abraxas1, BRC445, MERIT40 and RAP80 (Feng, L. et al., 2010).

As an isolated protein, BRCC36 is catalytically inactive (Shao et al., 2009). The regulation of BRC336's DUB activity was first revealed through the crystal structure of the ant BRCC36-Abraxas2 complex (Zeqiraj et al., 2015). The crystal structure (resolved to a resolution of 2.55 Å) showed that ant BRCC36-Abraxas2 forms a tight dimer of heterodimers (referred to as a "super dimer"), which results in a conformational change of the active site within the MPN$^+$ domain of BRCC36. In the structure, both BRCC36 and Abraxas2 bear an MPN domain, and two non-canonical helices ($\alpha4$ and $\alpha5$) within the coiled-coil region. The crystal structure showed that the MPN domains of BRCC36 and Abraxas2 interact with each other. Similarly, the non-canonical helices α4 and α5 of BRCC36 interact with the corresponding helices in Abraxas2, forming a coiled-coil helical bundle (Figure 11). As such, the crystal structure of the super dimer revealed three large networks of hydrophobic interactions involving the MPN core domains, as well as the N- and C-terminal ends of the coiled-coil helical bundle (Zeqiraj et al., 2015).

To better understand how Abraxas2 activates BRCC36, Zeqiraj *et al* solved the crystal structure of zebrafish BRCC36 (Zeqiraj et al., 2015). Comparison of zebrafish BRCC36 with ant BRCC36-Abraxas2 provided further insight into how the interaction of Abraxas2 with BRCC36 enables BRCC36 to carry out its DUB activity. Major differences were found within the active site region of zebrafish BRCC36 (*Dr*BRCC36). For example, the E-loop of zebrafish BRCC36 inactive structure was partially disordered and shifted, with respect to the ant BRCC36 (*cf*BRCC36), when bound to Abraxas2, giving rise to a 4–5 Å displacement of the catalytic glutamate (Glu30 in *cf*BRCC36); Glu27 in *Dr*BRCC36). The distortion could, in part, be responsible for the lack of BRCC36 DUB activity when not bound to Abraxas2. Another notable difference between the inactive ant BRCC36 and active ant BRCC36 was that large portions of the inactive BRCC36 insertion-1 were disordered. The ordered portions of Insertion-1 was also slightly shifted relative to the active ant BRCC36. It appears that the heterodimerization of BRCC36 with Abraxas2 causes structural changes within BRCC36 and is in part responsible for BRCC36 DUB activity (Zeqiraj et al., 2015).

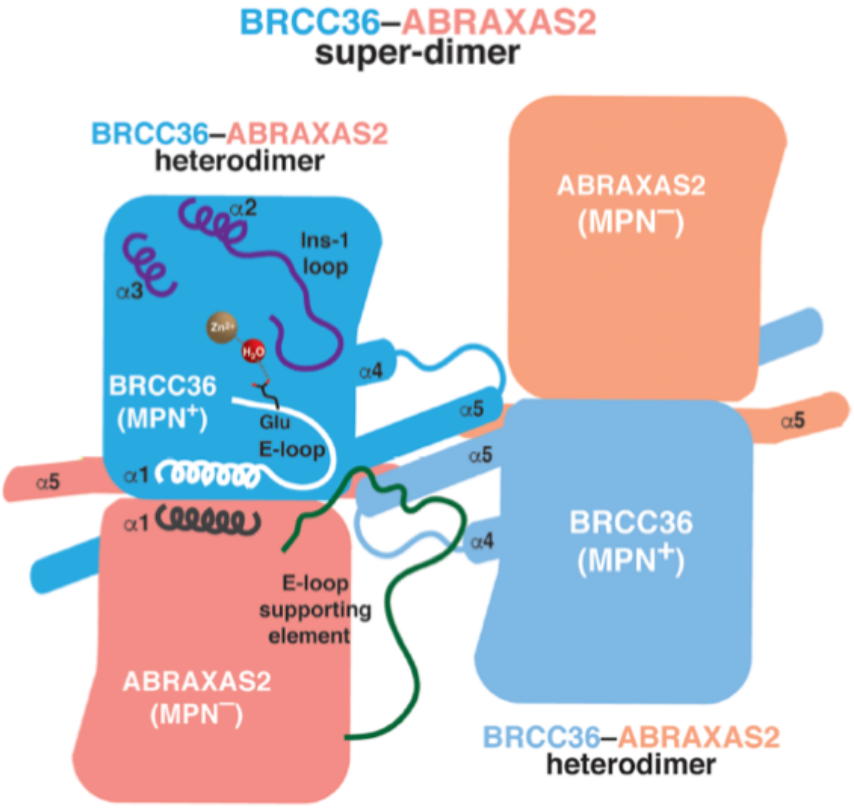


Figure 11: Schematic diagram of the BRCC36-Abraxas super dimer

The C-terminus of Abraxas interacting with the BRCC36 C-terminus leading a helical bundle which In turn drives the formation of the super-dimer.

1.2.8 BRISC-SHMT2 complex

The BRISC-SHMT2 complex is a cytoplasmic DUB and plays a vital role in interferon signalling (Walden et al., 2019; Zheng, H. et al., 2013). Studies have revealed that BRCC36 within this complex is enzymatically inactive and requires interaction with Abraxas2 for minimal DUB activity (Zeqiraj et al., 2015). The activity of the complex was shown to be further enhanced by the binding of BRCC45, while the incorporation of MERIT40 showed little to no enhancement in the activity of BRISC-SHMT2 (Zeqiraj et al., 2015). Interestingly, the binding of SHMT2 was found to completely abolish the deubiquitylating activity of the complex (Walden et al., 2019). Newly published structural and functional information on BRISC-SHMT2 revealed that the interaction of SHMT2 with BRISC blocked the active site of BRCC36 rendering the complex inactive (Walden et al., 2019).

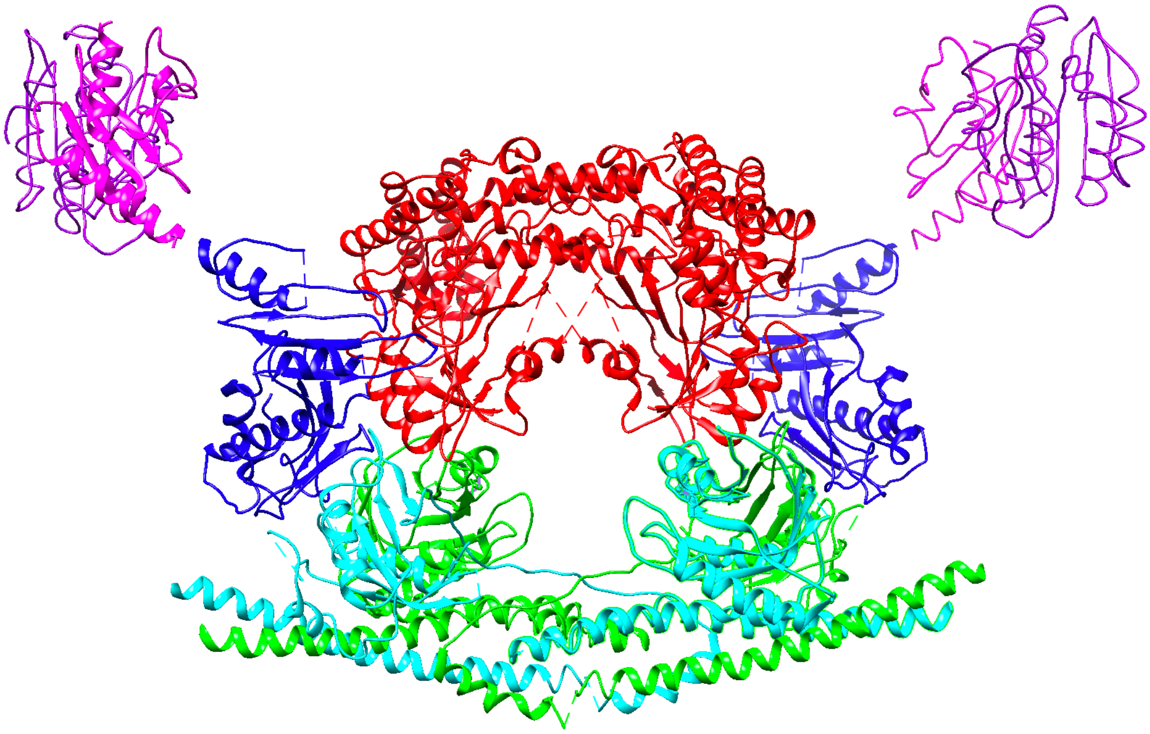


Figure 12: Cryo-EM structure of the BRISC-SHMT2 complex

The cryo-EM structure of the BRISC-SHMT2 complex revealed that SHMT2 forms a bridge across the two BRCC45-MERIT40 ‘arms’ sitting directly above the Abraxas2-BRCC36 superdimer (PDB:6r8f). Analysis of the structure showed that SHMT2 directly interacts with the MPN domain of Abraxas2 and BRCC36 and the UEV domains of BRCC45. SHMT2 is shown in red, MERIT40 in pink, BRCC45 in dark blue, Abraxas2 in cyan blue and BRCC36 is represented in green.

1.2.9 ARISC-RAP80 complex

The ARISC-RAP80 complex is a core component of the DNA damage repair pathway and localises to sites of double stranded DNA breaks (Shao et al., 2009). Despite sharing protein subunits with BRISC-SHMT2, notable differences in ARISC-RAP80 activity have been observed. The work of Patterson-Fortin et al., 2010 showed that BRCC36 requires both Abraxas1 and BRRC45 for minimal activity, suggesting that the RAP80 complex may assemble differently to BRISC-SHMT2 (Patterson-Fortin et al., 2010). At the start of this project there were no structures of the ARISC-RAP80 or BRISC-SHMT2 complexes, which could have helped explain the different BRCC36 activation modes. Kyrieleis et al published a low resolution negative stain EM model of the human ARISC complex revealing the architecture of the complex (Kyrieleis et al., 2016). The model showed that BRCC36-Abraxas1 form a super dimer and that BRCC45 interacts with the super dimer via one of its UEV domains, while it uses another UEV domain to interact with MERIT40. The model showed that the ARISC complex has a stoichiometry of 2:2:2:2 like the BRISC complex (Kyrieleis et al., 2016; Zeqiraj et al., 2015). However, how RAP80 binds and interacts with ARISC, and the stoichiometry of the RAP80 complex remains unclear. BRCC36 in both BRISC-SHMT2 and ARISC-RAP80 display specificity towards K63 polyubiquitin chains, however the ubiquitin interacting motifs (UIM) of RAP80 has been demonstrated to display varying levels of binding specificity towards K6 ubiquitin chains, SUMO and NEDD8 ubiquitin-like chains, or even towards mixed ubiquitin-SUMO chains (Hu et al., 2012).

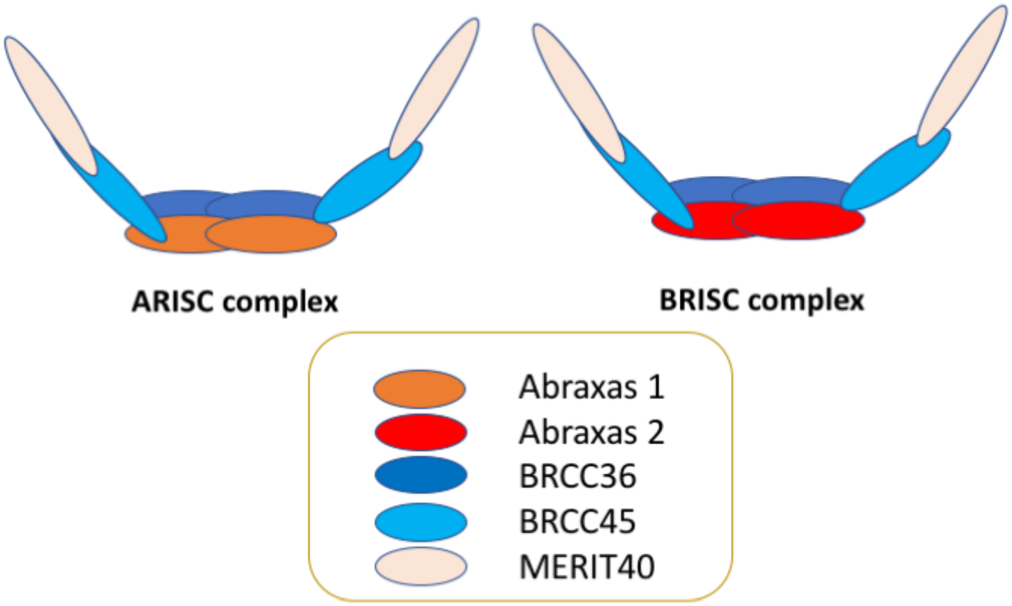


Figure 13: Shared subunits between the BRISC and ARISC complex

The four membered BRISC and ARISC complex share similar subunits, differing only in their MPN$^-$ subunit. When part of a five membered complex, BRISC and ARISC differ also in their targeting proteins SHMT2 and RAP8-, respectively (not shown).

1.2.9.1 The RAP80 protein

RAP80 contains two UIM at position 80-99 and 105-124 and SUMO interaction motif (SIM) domains located at its N-terminus, a central Abraxas1 interacting region (AIR) and two C-terminal zinc finger domains (Wang et al., 2007) (Figure 14). RAP80 is a key component of the complex as it targets the complex to sites of DSBs (Sobhian et al., 2007). The ubiquitin interaction motif is defined as a region of the protein composed of approximately 20 conserved amino acid residues. A domain, on the other hand, is defined as a conserved polypeptide section of a protein that can fold and function independently.

The UIM domains of RAP80 are susceptible to ubiquitylation. Research has shown that in addition to being a substrate for Ubc13, RAP80 is likely to be deubiquitylated by BRCC36 suggesting a potential self-regulation mechanism (Li, Y. et al., 2017; Jiang, Q., 2018). RAP80 UIM domains associate with ubiquitylated H2A and H2B thereby aiding in the recruitment of the complex to sites of DNA damage on chromatin(Lombardi et al., 2017). Structural analysis suggested that one UIM was capable of binding one or two ubiquitin moieties, allowing RAP80 UIM domain to recognise at least two ubiquitins. However, experiments conducted by Wu et al 2009 demonstrated that the deletion of RAP80 UIM completely abolished interaction between UIM domain of RAP80 and ubiquitylated H2A or H2B (Wu, J. et al., 2009). This suggests that the UIM domains function cooperatively and recognise a single ubiquitylated protein. Further structural studies are needed to better understand the molecular details of the interactions between RAP80 and ubiquitylated histones.

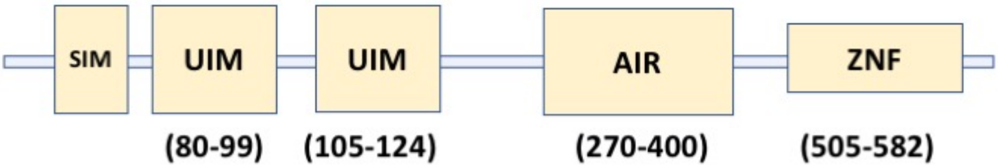


Figure 14: RAP80 domain architecture

RAP80 is a largely disordered and unstructured protein, however, it contains four structured regions; SIM, UIM, AIR and two zinc finger domains.

1.2.9.2 Abraxas1

Abraxas1 contains a JAMM motif, an MPN$^-$ domain, a coiled-coil domain and a pSPxF motif (Wang et al., 2007; Bose et al., 2019). The JAMM motif is composed of 120 amino acids and is found in archaea, bacteria and eukaryotes. The MPN$^-$ domain of Abraxas1 lacks conserved residues that promote zinc binding, a mechanism required for DUB activity (Zeqiraj et al., 2015). Point mutations on Abraxas1 MPN$^-$ domain (W99E) have been shown to result in the loss of interaction with all members of the RAP80 complex (Castillo et al., 2014). MPN$^+$ MPN$^-$ protein interaction is commonly seen in JAMM/MPN proteins and may function as a regulatory mechanism (Walden et al., 2018). The DUB Rpn8, like Abraxas1/2 contains an MPN$^-$ domain. Rpn8 interacts with Poh1 within the proteasome 19s lid using their respective MPN domains. These similarities suggest that MPN$^+$ MPN$^-$ pairing facilitates in the activation of JAMM family DUBs. The RAP80 complex forms an additional complex known as the BRAC1-A complex. BRCA1 associates with BARD1 and is recruited to the complex via the pSPxF motif of Abraxas1 following its phosphorylation by ATM (Rosen, 2013; Kyrieleis et al., 2016).

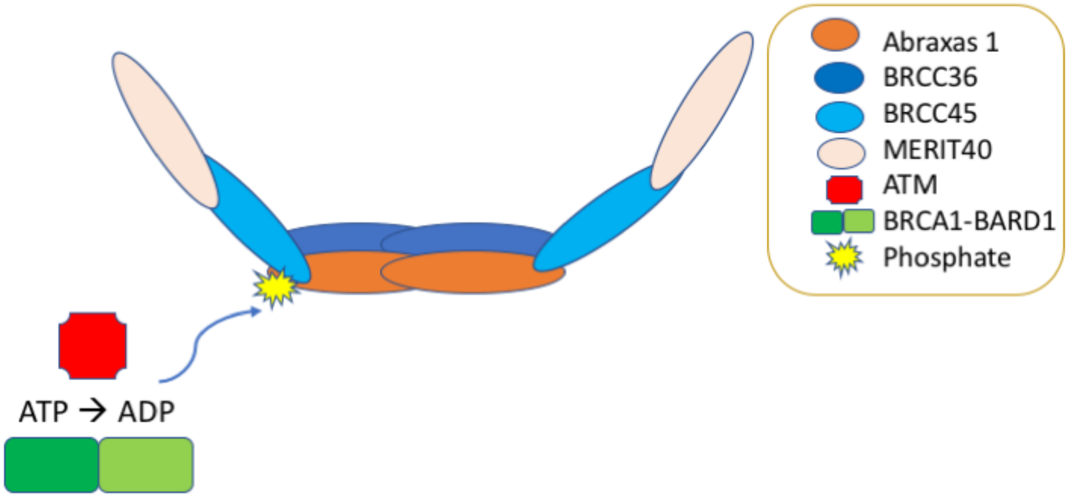


Figure 15: The association of BRCA1-BARD1 with Abraxas1

BRCA1 associates with AIRSC as part of a heterodimer with BARD1. The binding of the BRCA1-BARD1 dimer to the pSPxF motif of ARISC occurs in an ATP dependent manner and requires the phosphorylation of Abraxas1 'tail'.

1.2.9.3 BRCC45

BRCC45 is believed to have multiple functions within the cell. For example, the protein is a death receptor-associated anti-apoptotic protein, and plays a significant role in DNA damage repair (Shi et al., 2016). In the DNA damage repair pathway, BRCC45 is known to facilitate in the deubiquitylation of a key cell cycle regulator, CDC25. BRCC45 was believed to have 2 UEV domains, however the cryo-EM structure of the BRISC-SHMT2 complex revealed the presence of a possible third UEV domain (Walden et al., 2019). UEV domains of BRCC45 are similar to that of E2 enzymes, however they lack key cysteine residues required for the formation of the thio-ester-ubiquitin intermediate (Kyrieleis et al., 2016). BRCC45 interacts with MERIT40 via its C-terminal UEV domain, and with Abraxas1/2 via its N-terminal UEV domain (Kyrieleis et al., 2016; Walden et al., 2019). Mutating three residues of UEV1 domain (WNP) to Alanine was shown to disrupt BRCC45 interaction with RAP80, Abraxas1 and BRCC36, thereby preventing complex formation (Patterson-Fortin et al., 2010).

1.2.9.4 MERIT40

The exact role of MERIT40 within the context of both BRCC36 complexes remain unclear. MERIT40 is believed to be essential in maintaining the integrity of BRISC-SHMT2 and ARISC-RAP80 complexes via its interaction with BRRC45 (Sawant and Varma, 2014). The interaction of MERIT40 and BRCC45 is facilitated by the PxxR motif located at the C-terminus of MERIT40 (Kyrieleis et al., 2016). Research has shown that mutating proline and arginine within the motif disrupts MERIT40-BRCC45 dimerization (Hu, X. et al., 2011). MERIT40 has also been functionally linked to promoting resolution of chemotherapy induced DNA damage and DNA interstrand crosslinks (Fradet-Turcotte et al., 2016).

It is possible that the steady state of both BRCC36 DUB complexes are in balance, with the formation of one complex limiting the abundance of the other. Knock-down of Abraxas2 showed increased association of Abraxas1, BRCC36, BRCC45, MERIT40, RAP80 and BRCA1 (Patterson-Fortin et al., 2010). The expression of the MPN-binding partners could be key regulators determining the cellular levels of one complex over the other. Abraxas2 has recently been detected in the nucleus suggesting a possible role for the BRISC complex in the DNA damage repair pathway (Zhang, J. et al., 2014). Although it has not been found to co-localise with DNA damage repair markers upon ionized radiation, it is possible that it could have other functional roles in the nucleus.

1.2.9.5 Implication of the ARISC-RAP80 complex in cancer

Breast cancer is one of the most common cancers in women in the developing world (Bray et al., 2004). A single glutamic acid deletion in the first UIM domain of RAP80 displays reduced ubiquitin binding and DSBs localisation (Nikkilä et al., 2009), leading to impairment in the DNA damage response system and may be involved in the genetic pre-disposition of breast cancer. Germline mutations in a network of genes encoding BRCA1, BRCA2 and their interacting partners have been shown to confer hereditary susceptibility to breast cancer (Solyom et al., 2012).

1.3 Electron microscopy

1.3.1 History of the microscope

The development of the first microscope was attributed to Hans and Zacharias Jansen in 1590. However, major early contributions to the field of microscopy were the result of the work of Robert Hooke and Antoni van Leeuwenhoek between 1660-1670. The 19th century saw the rise of the light microscope. Following the rise of the light microscope, Ernst Abbe and colleagues had advanced the use of the light microscope to its theoretical limits. Light microscopes are limited by a resolution of approximately 0.2 micrometers and a magnification of 1000x. The limited resolving power of the light microscope meant that it was not suitable for the in-depth study of macromolecules. A microscope with an illuminating radiation with shorter wavelength would be more favorable for studying molecules within cells

1.3.2 Development of Electron Microscopy

The electron microscope (EM) was first invented by Max Knoll and Ernst Ruska in the 1930’s in Berlin. The short wavelength of electrons provides electron microscopes with a higher theoretical resolving power compared to the light microscope. The two common types of electron microscope are the transmission electron microscope (TEM) and the scanning electron microscope (SEM). Simply put, the TEM work by transmitting electrons through a specimen and to a detector. SEM, on the other hand, functions by sweeping electron beams across the sample and recording the electrons that bounce back. Scanning electron microscopy techniques therefore allows the user to visualise surfaces of samples whereas the TEM allows users to visualise the whole sample. The attainable resolution by the EM has continually improved over the decades. This has mainly been achieved by increasing the accelerating voltage of the electron beam, improving the electron lens to minimise aberrations, better vacuum systems and brighter electron guns. After the commercialisation of the EM, image-processing techniques were developed allowing for better imaging. Below, I outline some of the techniques used in TEM, namely, negative stain and single particle cryo-electron microscopy.

1.3.3 Negative stain EM

Negative stain EM is a useful technique that allows for the quick observation of macromolecules and macromolecular complexes (Scarff et al., 2018). It also facilitates the assessment of sample purity, homogeneity and concentration. In negative stain EM, the maximum attainable resolution is 12-15 Å (Scarff et al., 2018). The limitation in attainable resolution by negative stain EM is partly due to the envelopment of the particle of interest (POI) in heavy staining. The use of heavy metal stain results in high contrast images in which the POI is less electron dense than the background stain. Support grids used in negative EM typically have a layer of amorphous carbon (Scarff et al., 2018). Polyvinyl (Formvar) or nitrocellulose can also be used as a support substrate. Prior to the application of sample to the EM grid, the grid undergoes glow-discharging, which introduces a negative charge to the grid making them more hydrophilic, allowing for better POI adsorption (Aebi and Pollard, 1987; Scarff et al., 2018). Once a sample has been deposited onto the grid, excess solution is removed with blotting paper (Scarff et al., 2018). The grid can then be washed with water prior to staining. The type of stain utilised largely depends on the sample. It's important that the pI of the protein of interest suits the stain. Uranyl acetate and Uranyl formate are the two commonly used stains due to the fact that they have very fine grains (4-5 Å) and therefore provide higher resolution (Scarff et al., 2018). The use of uranyl acetate as a stain can have drawbacks. The heavy metal stain has low pH and can precipitate at physiological pH, these properties render protein samples more susceptible to precipitating (Scarff et al., 2018) . In addition to this, Uranyl acetate is toxic and mildly radioactive and requires special handling, storage and disposal. For negative stain EM, phosphate buffers and buffers containing high salt concentrations should be avoided as these can result in crystalline precipitates. Reducing agents, detergents, glycerol and high nucleotide levels can also affect staining and should, where possible, be avoided (Scarff et al., 2018). Figure 16 summaries the process involved in negative stain EM sample application.

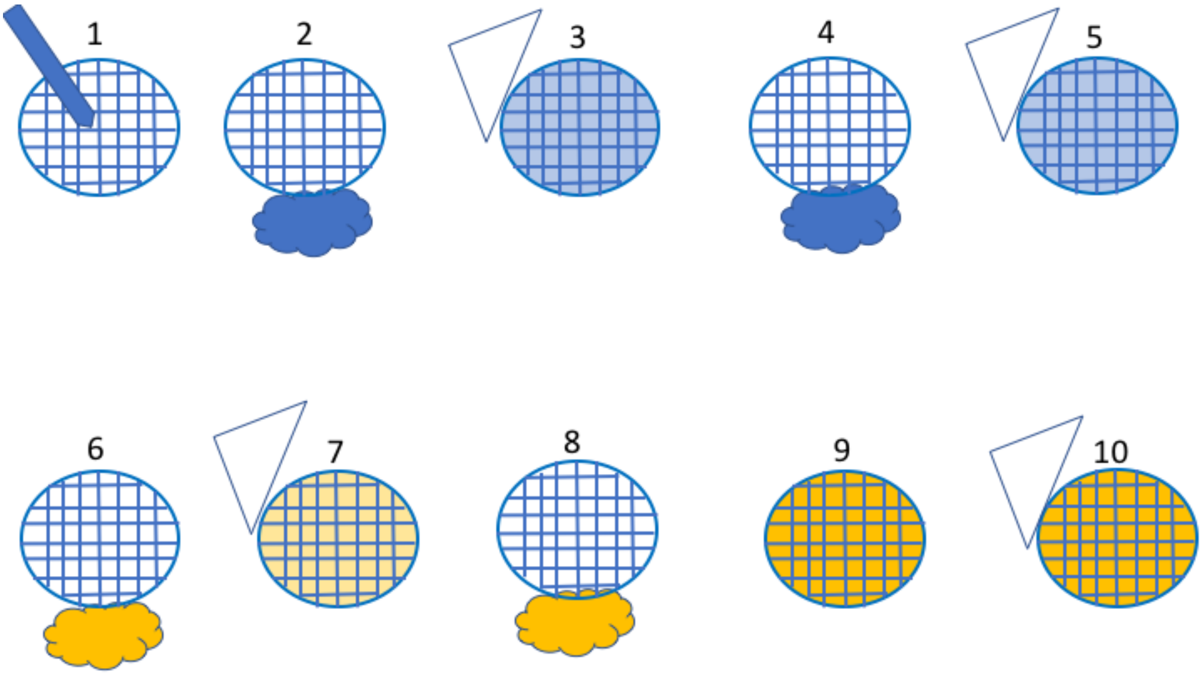


Figure 16: Negative stain EM application

[1] Biological ~3µl sample is deposited onto TEM grid for an agreed period of time and then excess samples are removed by blotting. [2] The grid is then washed in water. [3] Excess water is removed by blotting. [4] Grid is washed in water for a second time. [5] Excess water is removed from grid. [6] Grid is stained with staining of choice. [7] Excess stain is removed by blotting. [8] Grid undergoes second round of staining. [9] The stain is left on the grid for a desired period of time. [9] excess stain is removed and the grid is ready for screening or storage.

1.3.4 Cryo-EM

Cryo-electron microscopy (cryo-EM) is an umbrella term for the sub-disciplines; cryo-tomography, single particle cryo-electron microscopy and electron crystallography (Milne et al., 2013). The technique involves applying samples to an EM grid, freezing the grid at cryogenic temperatures and imaging them using a transmission electron microscope. In biology, cryo-EM has wide applications. The technique can be used to image tissue sections, cells, bacteria, viruses and protein molecules (Milne et al., 2013). Below, I will focus solely on single particle cryo-electron microscopy of protein samples.

Single particle cryo-EM involves computationally selecting individual particles from many micrographs, aligning said particles to generate a two-dimensional (2D) projection image of the particle of interest in different orientations and then combining the different views of the 2D images to generate a three-dimensional reconstruction (Carroni and Saibil, 2016)(Figure 17). In single particle cryo-EM, EM grids are glow discharged and the sample is spread across the holes of the grid. Excess sample is then removed by blotting. The removal of excess sample is commonly an automated process in which the parameters such as blot force and time can be optimised. After blotting, the specimen is plunge-frozen into liquid ethane or liquid ethane and propane mixture generating a thin layer of vitreous ice containing the POI ideally in different orientations (Drulyte et al., 2018) (figure 16). The ice thickness can vary depending on the dimensions of the POI and buffer composition (Drulyte et al., 2018). Following plunge freezing, the grid is stored in liquid nitrogen. The EM grids used for cryo-EM differ to those used for negative stain EM.

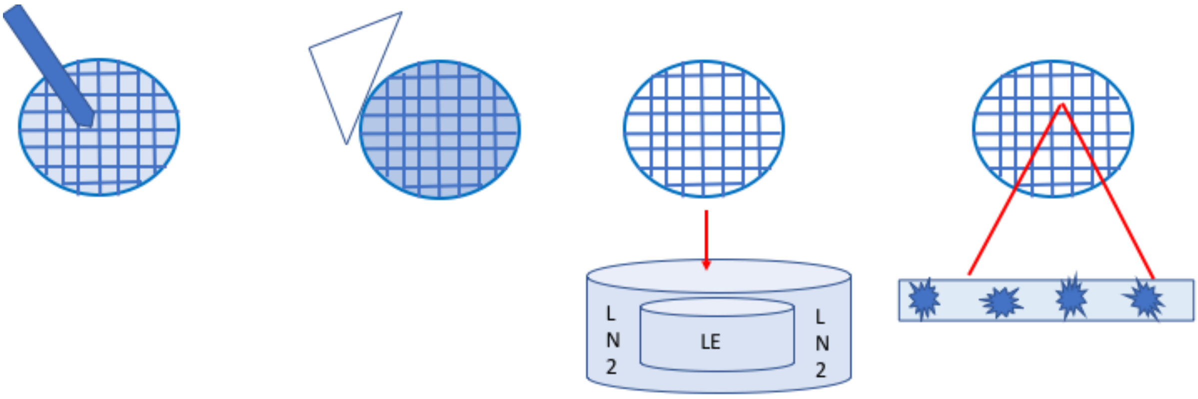


Figure 17: Cryo-EM sample vitrification overview

Biological sample is deposited on the grid, and excess solution is removed by blotting. The grid containing the sample is vitrified in liquid ethane and then stored in liquid nitrogen. Samples should ideally embed in various orientation within the grid holes.

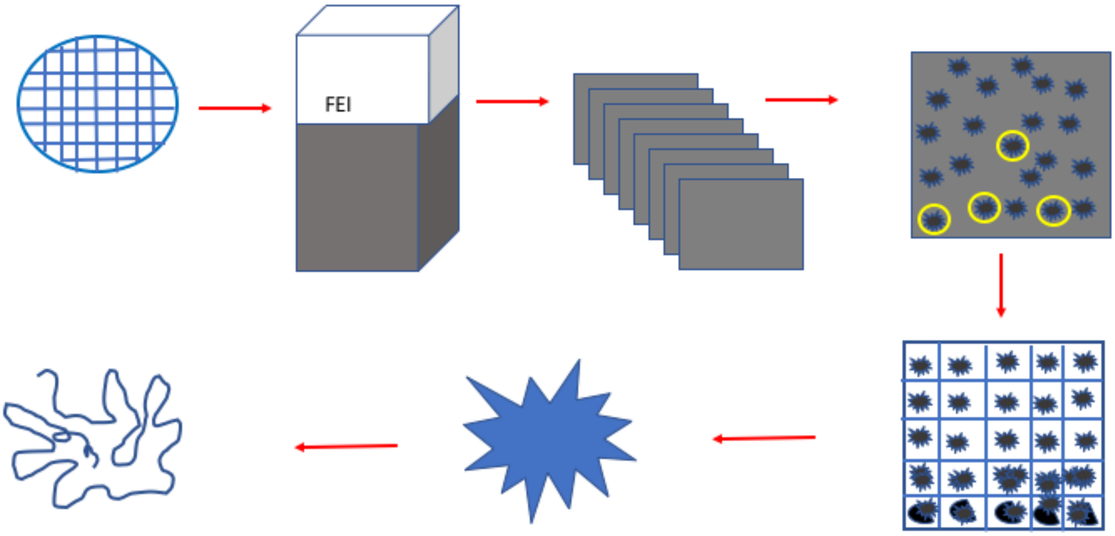


Figure 18: An overview of single particle cryo-EM

The vitrified grid is clipped and inserted into the microscope, and movie frames are collected using data collection parameters of choice. The movies are then processed using an image processing software. Following motion and CTF correction, particles are selected and aligned and classified into 2D classes.2D classification allows for the identification on non-particles and misaligned particles which can be omitted from 3D classification. Suitable 2D classes are used to generate an initial 3D model and a 3D map. After refinement and post-processing, secondary structures can be built into the 3D map.

1.3.4.1 Improvements in cryo-EM

In the past decade, microscope technology has vastly improved. Two of the major developments that helped revolutionise cryo-EM are the development of electron detectors and better image processing (Bai et al., 2015). During data collection, the signal:noise ratio (SNR) of the incoming signal is degraded by errors in the detection process (Bai et al., 2015). The detective quantum efficiency (DQE) of a detector expresses how the SNR is degraded during detection (Bai et al., 2015). An optimal detector would not add any noise and have a DQE value of 1. In practice, detectors add noise and therefore have a DQE value of less than 1. Photographic films were largely popular in cryo-EM because they had a DQE factor of approximately 0.3 and high resolution (Bai et al., 2015). A caveat of the film was that it required developing and scanning. The introduction of charge-coupled device (CCD) camera made data collection more convenient as it introduced automation, eliminating the need to develop and scan film. CCD cameras work well at low energies, however at the higher energies crucial in cryo-EM, their DQE degrades to approximately 0.1 (Bai et al., 2015).

The search for a high-DQE detector resulted in the development of direct electron detectors (DED). In direct electron detectors, incident electrons pass through a thin semiconductor membrane, deposit energy which is in turn detected by electronics embedded on the membrane (Bai et al., 2015). Unlike CCDs, DEDs directly detect incident electrons in a sensitive layer preventing the signal from being scattered into surrounding pixels (Bai et al., 2015). In addition to recording incident electrons, electrons that initially pass through the membrane are backscattered from the underlying support matrix before passing through the membrane and are also detected (Bai et al., 2015). The disadvantage of this is that backscattered electrons can result in noise, degrading the DQE. To improve the DQE, as much of the support matrix is removed as is possible, in a process termed 'back-thinning' (Bai et al., 2015). Newer detectors are less than 50 μM thick, allowing electron to pass through without backscattering (Bai et al., 2015). Such detectors are widely available from Direct Electron (DE), Thermo Fisher (Falcon), and Gatan (K2) (Bai et al., 2015). Backthinned direct electron detectors have a DQE value higher than film detectors. To prevent saturation of the detectors, electron doses used in cryo-EM are spread over multiple frames. DE and Falcon detectors previously generated images by integrating signals from individual movies. A limitation of DE and Falcon detectors and other detectors that integrate signal is that the incident electron can deposit varying amounts of energy and therefore contribute to the output signal (Bai et al., 2015). The aforementioned detectors overcame this limitation by counting individual electron events. Counting mode involves recording movies at a high frame rate of at least 400 frames per second and requires a hardware to produce images in real time. A drawback of using counting mode is that it requires long exposure times which may be an issue when the experimental support is not stable (Bai et al., 2015).

1.3.4.2 Single particle image processing

The pipeline of image processing can be categorised into four different stages: pre-processing, two-dimensional classification, three-dimensional (3D) reconstruction and post-processing (Sigworth, 2016).

1.3.4.2.1 Pre-processing

During data collection, the electron beam can cause particle movement (Zheng, S.Q. et al., 2017). This beam-induced motion blurs the captured image and can significantly reduce the attainable resolution of the data. Beam-induced motion can be categorised into two components: uniform whole frame motion and non-uniform local motion. The latter can vary across the micrograph. Traditionally one of the first steps of image processing is to correct for the motions within the raw frames. Individual particle motion can also be tracked and corrected for during the refinement of the final reconstruction. Stage mechanical shift during the process of collecting data can also result in particle movement. Both beam-induced motion and stage drift can be corrected using programs such as MotionCorr2. The correction of movement has advanced the attainable resolution of cryo-EM as it allows for the recovery of high resolution detail (Zheng, S.Q. et al., 2017).

In addition to MotionCorr, there are various packages available such as unblur and modules within EMAN2 and CRYOSPARC (Punjani, Ali et al., 2017). Although details within each package vary the core principle remains the same in that the movement of a feature is tracked across the movie frames and corrected for (Scheres, 2014). Following motion correction, the contrast transfer function (CTF) of the micrograph is corrected. The contrast transfer function mathematically describes how aberrations in the TEM modifies the image of the sample of interest. The CTF sets the resolution of TEM which is also termed phase contrast of TEM (Carroni and Saibil, 2016). The primary process of CTF correction is to determine the defocus and CTF parameters that each micrograph was collected with (Orlova and Saibil, 2011). Programs such as CTFfind and gCTF calculate the defocus of each individual micrograph and attempt to model the power spectrum (Fourier transformation squared) of the micrograph(Zhang, K., 2016). The oscillating nature of the CTF produces spatial frequencies with zero contrast at any given defocus values and negative values at various points. One method for correcting CTF is phase flipping to remove negative values. A method utilised to remove the zero points is to set a range of defocus vales during data collection so that when the data is averaged there will be no zero points (Orlova and Saibil, 2011).

Micrographs contain a mixture of particles of interest and non-particles such as aggregated proteins, ice contamination, deformed particles and other contaminants. Particle location has traditionally been identified using automated, semi-automated and interactive particle selection programs (Sigworth, 2016). Interactive particle selection requires that the user identifies particles on the micrograph and manually selects them. The program then uses the coordinates generated by selecting the individual particles to extract the small square regions of the micrographs containing the particle. The extracted regions then form a stack of particle images (Sigworth, 2016). Following particle extraction, the particles are aligned and a variety of 2D classes are obtained. The 2D classes can then be used as a template for automated particle selection of the full dataset. Low pass filtered 3D maps can also be used as a reference for automated particle picking. The majority of the available data processing programs include an auto-picking function. The user has the freedom to optimise threshold values eliminating the selection of non-particles (Sigworth, 2016).

A disadvantage to reference based auto-picking is that it can introduce model biased picking and risk eliminating a variety of orientations from being included in 2D classification and thus model building (Sigworth, 2016). Reference based auto-picking can also result in picking in empty micrographs resulting in high noise during the 2D classification stage. Luckily, ‘true’ particles with high signal to noise ratio will predominate in the auto-picking process.

1.3.4.2.2 2D classification

Single particle analysis is an averaging technique. Two-dimensional classification is fundamentally the clustering of similar particle images (Sigworth, 2016). The concept of particle alignment was first introduced by Van Heel and Frank (Sigworth, 2016; Van Heel and Frank, 1981). During 2D classification, there are a set of random images known as the ‘reference’. For each extracted particle image, the probabilities of them aligning with the reference is calculated with respect to its rotation, translation and degree of matching to each reference (Sigworth, 2016). Translated and rotated particle images are formed and their 2D Fourier transforms, weighted by their CTFs are combined to form the overall average image. The newly averaged images are in turn used as references for the subsequent iteration. The process is iteratively repeated approximately 25 times. This process produces averages of multiple images that represent the dataset. The resulting class averages can be a good indication of whether or not the subsequent 3D map will generate a high-resolution model. It is not uncommon for alpha helices to be visible at the 2D classification stage. Alpha helices are often visible at approximately 8 Å. Analysing the class averages allows the user to determine the homogeneity and quality of the dataset. 2D classification is also useful for removing artefacts as they tend to be grouped together and are easy to identify as ‘non-particles’ (Sigworth, 2016).

There are several algorithms available that have been implemented within packages that aligns and classifies cryo-EM and negative EM data. Maximum-likelihood (ML) approach and Kernal self-organizing map (kerdenSOM) are two examples.

The ML approach is utilised to simultaneously align and classify single-particle images (Scheres et al., 2005). In this technique, reference-free class averages are obtained in an unsupervised manner by starting multi reference alignments from random subsets of unaligned data. The ML approach requires that the user adjusts the number of references (K). Theoretically, the number of references should reflect the number of different 2D structures present in the data, however, this number is normally unknown. The user is therefore required to run the algorithm multiple times with different k values. Although there are no maximum K values, the higher the K value the fewer the number of particles will be present in a class. Averaging using a low number of particles results in noisy class averages which in turn results in suboptimal alignments and classification. A minimum of between 200-300 cryo particles per 2D class is recommended. With negative EM data, fewer particles per class is acceptable.

ML primarily functions by comparing all particle images to the reference in all of the available orientations, it then calculates the probability weighted of each aligned possibility. The generated class averages are then used as references for the next iteration. Although the ML approach is widely used, it has two major disadvantages. One disadvantage is that it does not take CTF information into consideration and it assumes white noise in the pixels, the second disadvantage is that it is computationally expensive. In spite of this, the ML approach generates useful results from cryo-EM data. Programs such as RELION utilise the maximum likelihood approach for 2D classification(Scheres, 2010; Scheres, 2012).

KerdenSOM is based on the maximum-likelihood principles(Pascual-Montano et al., 2001; Scheres, 2010). The algorithm can be used to classify pre-aligned images. It does not classify particles in a predefined number of classes, instead KerdenSOM outputs a two-dimensional map of averaged images known as code vectors. This approach is relatively more user-intensive as the user has to decide how many structural classes are present in the map. Unlike ML, the user does not have to give a number of classes. Instead the user determines the size of the output map and the parameters used for its regularization. A large output map can accommodate more structures; however, an excessively large map may contain too few particles contributing to each code vector. Too strong regularization can also impact the map by making it too smooth which in turn makes it difficult to analyse for structural variability in the dataset. On the flip side, too weak regularization will contribute to an organised output map making it difficult to identify structural variation in the data. To obtain optimal results, the algorithm can be repeated multiple times using different regularization parameters and map sizes.

KerdenSOM is has less computational costly compared to ML but is more user-intensive. Both algorithms can be used in conjunction and has been proven to be useful when for example trying to identify stoichiometrical binding of factor to a complex that is only visible in one view/orientation. In such instances, ML can be used to separate top and bottom views of the complex. KerdenSOM can then be implemented to focus on the view displaying the bound factor. KerdenSOM is implemented in packages such as Xmipp and is accessible via the Appion pipeline.

1.3.4.2.3 Three-dimensional reconstruction

Following 2D classification an initial 3D reconstruction of the molecule is obtained using the output from 2D classification. This step can be carried out in a user-free way or may require user input depending on the program selected to process the data. RELION (Scheres, 2012; Zivanov et al., 2018), for example, may require user input to select desirable classes and to reject misaligned classes or classes composed of non-particles in order to obtain the best initial 3D map (Scheres, 2016). CryoSPARC, on the, other hand requires less user input(Punjani, Ali et al., 2017). The program allows users to generate an initial map using the full datasets and can readily identify heterogeneity within the datasets allowing users to process the resulting classes separately (Punjani, A. et al., 2017). This helps to overcome problems sometimes seen during 3D reconstruction where views are often missing or minority structures are ignored.

A good 3D initial map requires the use of 2D classes in various orientations. Selecting different 2D projections will enable the program to generate a map that more accurately reflects the POI. The problem of view distribution can sometimes be overcome by generating an initial map using electron tomography in which series of views from the same area are collected over a whole tilt range, for example -60 to 60 degrees (Drulyte et al., 2018). The initial 3D reconstruction is then used as a reference for the generation of 3D maps in a process called 3D classification. This approach is, however, not optimal when working with smaller structures as data are often missing from the tilt limits and the electron dose accumulates, resulting in sample damage (Drulyte et al., 2018).

During 3D classification, the number of 3D maps generated is determined by the user. 3D classification can be repeated, each time asking for an increasing number of 3D maps. This can enable the user to further identify heterogeneity within the dataset (Scheres, 2016). The process of 3D classification is similar to the approach used during 2D classification in that individual particle images are aligned and tested against the reference image. Following the identification of suitable 3D maps, the maps undergo a refinement process using either itself as reference or the initial 3D reconstructed map. The refinement process generates an estimation of the resolution of the map. The resolution is calculated by splitting the data in half to generate two 3D volumes and calculating the Fourier Shell Correlation (FSC). The resolution of the reconstruction is defined as the FSC value at a given threshold (Liao and Frank, 2010). The thresholds typically used are 0.5 or 0.143. It is often difficult to assess the resolution of the 3D map and false values can be given as the result of over-fitting the data. Over-fitting occurs when the noise is refined instead of the signal giving false resolution reading (Chen et al., 2013). A higher resolution can be determined during post-processing or by carrying out CTF and beam tilt refinement followed by another round of 3D refinement and finally post-processing. CTF and beam tilt refinement works by estimating the defocus and beam tilt values of each particle. Obtaining better CTF values of individual particles as opposed to the entire micrographs may result in better particle alignment and therefore higher resolution.

1.3.4.2.4 Post-processing

The map generated during refinement can sometimes appear to be smoother and less detailed, the resolution obtained following refinement is estimated from the map and may not be as high as expected. This is often due to the loss of high-resolution information during the data collection and the over representation of low resolution information. Radiation damage and specimen movement typically account for the loss of high-resolution information although steps during computational image processing such as the inaccurate determination of orientation parameters are also factors. High resolution information can be restored through a process termed B-factor sharpening or post-processing. Before B-factor sharping a mask is needed that defines the boundaries between the structure and solvent region. The mask is created by low pass filtering the map to 15-20 Å, binarising and extending the map by 3-6 pixels and adding a soft edge. The threshold for binarising the map is established by visualising the 3D map on a program for structural visualisation such as CHIMERA (Pettersen, Eric F., et al 2004) and altering the contour threshold to remove noise. The mask is then used as an input for post-processing. Rosenthal and Henderson 2003 developed a method that allows for objective sharpening of the density map without the requirement of ad-hoc B-factors. Applying ad-hoc B-factors to the reconstructed map is sometimes avoided as doing so can results in the amplification of high-resolution noise as oppose to signal resulting in an over-sharpened map.

1.3.4.2.5 Model building and structural validation

Low resolution maps such as those obtained from negative stain EM give the overall topological information of the molecule. Cryo-EM can also generate mid to low resolution maps of approximately 10-30 Å if the molecule is extremely flexible, not enough data has been collected or if the ice in which the particles are imbedded is too thick. Due to lack of secondary structure and atomic information, structures cannot be built into low resolution map *de novo*. However, if a homologous structure exists or subunits comprising the complex is available, the existing structures can be fitted into the low-resolution map. Programs such as the fit in map option implemented in Chimera are one of the many available tools for generating a model from a low-resolution EM density(Pettersen et al., 2004).

Maps with a resolution of better than 10 Å contain visible secondary structure features such as alpha helices. This allows for a more accurate and flexible fitting of pre-existing structures into the EM map. Although fitting individual structures into EM maps provides further information of complex assembly, crucial information of how the individual structures interact together will be missing. Higher resolution EM maps (sub 4 Å), contain sufficient detail to allow an atomic model to be built into the density *de novo*. Programs such as Coot which was originally designed for building into high resolution crystallography maps now contain features that allow users to build into EM maps *de novo(*Emsley *et al., 2010)*. Following model building into high resolution EM maps, the model must be validated. Model validation is critical as it assess the correctness of the map. Model fitting can be assessed using traditional approaches such as looking at bond angles and length. Unlike crystallography, building atomic models from EM maps is a more recent phenomenon. Cryo-EM is therefore lacking independent validation metrics and tools used in crystallography are often tailored for application in EM. Cryo-EM maps and models can be validated using tools in programs such as PHENIX. Validation traditionally consists of three components: analysis of the experimental data, analysis of the model and analysis of the fit of the model to the data.

Phenix contains multiple tools for validating cryo-EM map and models. The tools are combined into a single tool called phenix.validation_cryoem. The program requires a model, a map and the resolution of the map. Half maps can be provided; however, the user must indicate which map is the full map and which maps are the half maps. Once the analysis is complete, Phenix generates a summary of the results. The summary contains information regarding the model, the data and the model vs the data. Information provided under the model subheading includes compositions (e.g chains, atoms, water, ligands), bonds (RMSD), Ramachandran plot and occupancy. The information summarised under the data subheading includes, box lengths and angles, supplied resolution and estimated resolutions. The model vs Data summary highlights the correlation coefficient (CC) mask, CCbox, CCpeaks, CCvolumes and mean CC for ligands. Additional pages are also created containing more detailed information about each section.

The model-map correlation coefficient (CC) is a metric used to show how well the model fits the map. Calculating the model-map CC requires that the user chooses the CC formula, obtains a model-based map and defines the region of the map to be used to calculate the CC. CCbox uses the entire map as provided to calculate the CC value; this map may correspond to the whole molecule or a portion carved out as a box around selected atom. CCmask, on the other hand, only uses map values inside a mask calculated around the macromolecule, CCvolume and CCpeaks only compare the map regions with the highest density values (Afonine et al., 2018; PHENIX). Although data can be validated using programs, users can take measures early on to determine the accuracy of the data. This can be achieved by analysing the 2D classes and the resultant 3D reconstructed maps to determine whether the 3D map accurately describes the 2D classes. If other structural information is present, they can also be used to validate the reconstructed map and model.

1.3.5 Advantages and disadvantages of EM

EM has many advantages over other structural techniques such as crystallography and NMR, as it allows for the study of larger and more complex molecules (Afonine et al., 2018). The study of larger molecules using NMR often produces a spectrum too complex to analyse, thus making it difficult to obtain structural information. Crystallography on the other hand requires the crystallization of biological macromolecules. However, crystallization of molecules of large molecular weight, or of multi-subunit complexes, is difficult. This limits the samples that can be studied by crystallography. Another advantage that EM has over other structural biological techniques is that it allows for the study of more dynamic and heterogenous samples. For example, rare structures or structures that are dynamic can be readily identified and separated from the rest of the data during 2D classification and the 3D reconstruction stage of image processing. Identifying dynamic regions within a molecule or conformational changes can help to uncover biological and functional information regarding the role of the molecule within its native environment. This role can then be tested biochemically, aiding in the validation of the structure.

The study of heterogenous molecules using EM is not an easy task. The presence of multiple structures within the dataset often requires large or multiple data collection sessions in order to obtain near atomic resolution. This is due to the fact that more particles are required for the alignment of the various structures. Limited particle number will limit the attainable resolution regardless of how large or high quality the dataset is. This is a major disadvantage as data collection is expensive and collecting many datasets will require extensive computational power and user input.

All structural biology methods strongly complement one another. The resolution limits of EM mean that high resolution structures obtained from NMR or crystallography will aid in the model building into the EM map. Conversely, maps of complexes from EM data though missing some high-resolution information will provide information on complex assembly and arrangement. This information is often missing from studying the individual subunits as proteins, for example, upon interaction with another protein often undergoes some conformational shifts. The structure of an isolated subunit may therefore not be presentative of what it will look like in complex.

1.4 Thesis aims

The ARISC-RAP80 complex is a five membered complex, composed of BRCC36, Abraxas1, BRCC45, MERIT40 and RAP80. The complex displays high specificity for K63-linked ubiquitin chains on histone H2A, and plays a significant role within the DNA damage repair pathway by antagonising the actions of RNF8/ubc13. The subunit BRCC36 forms another distinct complex in the cytoplasm composed of BRCC36, Abraxas2 (KIAA0157), BRCC45, MERIT40 and SHMT2. Despite being the catalytic subunit, BRCC36 is inactive as a monomer and requires interaction with other subunits within the complex for activity. When bound to Abraxas2, BRCC36 has been shown to display minimal activity. However, in complex with Abraxas1, the heterodimer is catalytically inactive and requires interaction with BRCC45 for minimal activity. The targeting protein RAP80 and SHMT2 are responsible for targeting ARISC and BRISC to sites of DSB and interferon receptors, respectively. Differences in the activity of both complexes can be partly due to differences between Abraxas1 and Abraxas2. The two proteins display approximately 40% sequence identity. They mainly differ in the C-terminus where Abraxas1 contain nuclear localisation signals that allows the protein to enter the nucleus. The disparities between the two proteins, could affect the way in which BRCC36 is activated. The aim of this thesis was to obtain high resolution structures of the ARISC-RAP80 complex, and to better understand the function of the complex.

Chapter 3 arose from the negative stain EM analysis of the ARISC, ARISC-AIR and ARISC-RAP80 complexes. Negative stain EM data indicated that the complexes exhibited various levels of stability. The presence of AIR and RAP80 contributed to complex stability. As a result of this, the complexes displayed varying levels of structural and stoichiometry heterogeneity.

Chapter 4 further explores the structure of the ARISC-RAP80 complex using cryoEM. Results from these analyses was in line with negative stain EM analysis of the complex. 3D reconstruction of the ARISC-RAP80 revealed structural heterogeneity and allowed for the identification of a novel 10-membered complex and an unknown structure.

In chapter 5, the ARISC complexes are characterised using a mixture of biochemical and biophysical experiments. Results presented in this chapter further suggests that the ARISC-RAP80 complex is structurally heterogeneous and is capable of binding DNA with strong affinity. Interestingly, the ARISC complex and not the AIRSC-AIR complex also displayed DNA binding capabilities, suggesting that Abraxas1 contributes to the complexes ability to interact with DNA.

Altogether, these chapters provide the first structural characterisation of the ARISC-RAP80 complexes and the first cryo-EM maps for ARISC alone or in complex with RAP80. Interestingly, a new conformation of ARISC (an asymmetric dimer of octamers) was identified and its 3D structure revealed for the first time. Finally, this work provides the foundations for future determination of higher resolution ARISC-RAP80 structures and for functional characterisation of these complexes in DNA damage.

CHAPTER 2
Materials and Methods

2.1 Equipment

Transmission electron microscopes were used to obtain low and high resolution structural information of the ARISC complexes. The microscopes and their corresponding camera and detectors as summarised below.

Table 2: Transmission electron microscopes

Microscope	**Camera/Detector**	**Magnification**
FEI TF20	Ceta charged couple device (CCD)	29,000 x
FEI Titan KRIOS 1	Falcon III direct electron detectors (DED)	75,000 x

2.2 Materials

2.2.1 Buffers

Protein samples were purified using the following buffers. All the components and concentration are listed below in table four.

Table 3: Buffer composition

Buffers	**Composition**	**pH**
Gel filtration	25 mM HEPES, 150 mM NaCl, 1 mM TCEP	7.5
High salt	50 mM Tris, 500 mM NaCl, 20 mM Imidazole, 5 % glycerol, 0.075 % β-mercaptoethanol, 1 mM Benzamidine	7.6
Low salt	50 mM Tris, 300 mM NaCl, 20 mM Imidazole, 5 % glycerol, 0.075 % β-mercaptoethanol, 1 mM Benzamidine.	7.6
Elution	50 mM Tris, 300 mM imidazole, 300 mM NaCl, 5 % glycerol, 0.075 % β-mercaptoethanol, 1 mM Benzamidine	7.6
Lysis	50 mM Tris, 300 mM NaCl,	7.6
Dialysis	50 mM Tris-HCL, 1500 mM Nacl, 20 mM Imidazole, 5 % glycerol, 0.075 % β-mercaptoethanol, 1 mM Benzamidine	7.6
DUB assay buffer	50 mM HEPES, 100 mM NaCl, 0.1 mg/ml BSA, 1 mM DTT, 0.03 % Brij-35	7.5

2.3 Methods

2.3.1 Cloning and PCR

2.3.1.1 Cloning of the ARISC complexes

Cloning of the ARISC, ARISC-AIR and ARISC-RAP80 complex was carried out by Dr Elton Zeqiraj at the Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada. The MultiBac system for heterologous expression and production of protein complexes in insect cells (Sf9) was used (Zeqiraj et al., 2015; Berger et al., 2013).

2.3.1.2 300 bp PCR EMSA product

Polymerase chain reaction (PCR) was carried out to obtain a 300 bp DNA product for the electrophoresis gel mobility assay (EMSA). Five 50 μl PCR reactions were set up as follows. Following PCR, the samples were run on a 1% agarose gel. DNA bands were then extracted using the QIAquick gel extraction kit. See table six for primer sequence.

Table 4: 300 bp EMSA DNA product PCR overview

PCR reaction
5 μl 10 x KOD buffer 5 μl 2 mM dNTP 2.5 μl forward primer (20 ng/μl) 2.5 μl reverse primer (20 ng/μl) 3 μl 25 mM mgSO4 1 μl template (Pfacebec HTB vector) 30 μl filtered milliQ water 1 μl KOD polymerase
PCR protocol
Denaturation: 98°C Annealing: 50°C Elongation: 70°C Repeat denaturation, annealing and elongation cycle 30 x Final elongation: 30 seconds 70°C

Cool down at 4 °C

Table 5: List of primers and their application

Application	Template	Sequence
PCR	FastBacHTB Fw	ATT CCG GAT TAT TCA TAC CGT CCC
PCR	pFastBacHTB Rev	CTC TAC AAA TGT GGT ATG GC
Ethanol precipitation	Abraxas1 G399C Fw	AAG ATG AAG TGC TAA TAA TTT GGT GAA TAT TCA CGG TCT CCT ACA TTT
Ethanol precipitation	Abraxas1 G399C Rev	TTC ACC AAA TTA TTA GCA CTT CAT CTT TTC AAT TTC TTC ATC TGT TTC

2.3.2 Expression of the ARISC complexes

Protein expression was carried out by Dr Elton Zeqiraj using the MultiBac expression system.

2.3.3 His-tag protein purification

ARISC, ARISC-AIR and ARISC-RAP80 contain a histidine tag on the BRCC45 subunit. The complexes were purified by Ni^{2+}-affinity purification as follows: Protein pellets were thawed in a water bath and resuspended in lysis buffer. The resuspended pellet was then sonicated using a pulse of one second and a three second pause for a total of 16 minutes (four minute sonication). After sonication, the samples were centrifuged using a JA 25 50 rotor for 25 minutes at 4 °C and a speed of 30,000g. Following centrifugation, the supernatant was removed and filtered using a 0.45 μM filter. The filtered supernatant was then passed through a 5 ml equilibrated HisTrap column using a peristaltic pump, and the flowthrough was collected. The column was then washed with three column volumes of low salt, high salt and low salt buffer, and the flowthrough was collected after each wash. Samples from each flowthrough were collected for SDS-PAGE analysis.

The HisTrap containing the bound protein complex was then attached to an ATKA Fast Protein Liquid Chroamtography (FPLC) purification instrument and the complex was eluted using an imidazole gradient (20-300 mM) with low salt and elution buffer over 25 column volumes. Samples from the fraction were collected and subject to SDS page analysis. Fractions containing the eluted proteins were combined and the His-tag was cleaved using TEV. The Protein-TEV samples then underwent overnight dialysis against 4 litres of dialysis buffer. Following overnight dialysis, the sample underwent nickel subtraction to remove the His tag and TEV. Nickel subtraction was carried out using a new and equilibrated HisTrap column. The dialysed sample was introduced into the column using the peristaltic pump and the flowthrough containing the protein of interest was collected. The column was then washed using low salt, high salt and low salt buffer. After each wash the flow through was collected and samples were taken for SDS page analysis to identify the flow through with the proteins of interest. After complex identification, the sample was concentrated to approximately 1 ml and the optical density and concentration was calculated.

The concentrated protein sample was then further purified by gel filtration using a Superdex 200 10/300 m 24 ml column. Samples from each fraction were collected and run on an SDS PAGE gel to identify the fractions containing the desired proteins. Following the identification of the fractions containing the protein of interest, the fractions were concentrated and stored for downstream analysis.

2.3.4 Negative stain EM

Negative stain EM analysis was carried out to obtain low resolution structures of the complexes of interest. 1 % uranyl acetate was centrifuged for 30 seconds at a speed of 10,000 rpm to pellet the crystals. TEM grids were glow discharged using the PELCO easiglow apparatus for 30 seconds at 10 milliamps. Following glow discharging, 3 μl of 0.008 mg/ml protein was dispensed onto the grid for 45 seconds and excess solution was removed with blotting paper. The grid was then washed three times using 20 μl filtered milliQ water and excess water was removed after each wash by blotting with filter paper. The TEM grid was then stained with 20 μl of 1% uranyl acetate, excess solution was removed by blotting with filter paper. The grid was then stained through a fresh 20 μl drop of uranyl acetate for a second time. The stain was deposited on the grid for 30 seconds and excess solution was removed by blotting with filter paper, and the grid was stored for future screening and data collection

2.3.5 Cryo-EM

Cryo-EM was implemented to obtain high resolution structural information of the ARISC complexes. A comprehensive outline of the technique is located in chapter 1.3.4. Various grid preparation techniques were employed to obtain grids suitable for data collection and image processing- the methods are outlined below.

2.3.5.1 Grid preparation

2.3.5.1.1 Quantifoil grids

Quantifoil r2/2 300 mesh grids were glow discharged at 10 milliamps for 30 seconds using the easiglow Pelco apparatus. The grids were subsequently vitrified using the Vitrobot with the inner chamber set to 4 °C and 100% humidity. The vitrification steps were as follows: 4 μl of protein at 0.08 mg/ml was deposited into the grids; excess sample was blotted off using a force of 3 and for 4 seconds. The grid was then plunged into liquid ethane and then was stored in liquid nitrogen. The same grid preparation method was used for the BRISC-SHMT2 dataset, however, quantifoil r1.2/1.3 grids were used to deposit the protein.

2.3.5.1.2 In-house spray machine

Quantifoil 1.2/1.3 200 mesh grids were glow discharged using the Cressington 208 machine. ARISC-RAP80 samples were diluted to 0.05 mg/ml using gel filtration buffer. The grids were pre-blotted for four seconds using 3 μl of gel filtration buffer. Following this the grids were sprayed with the diluted protein samples for approximately 1.5 seconds. As the grids were being sprayed by the apparatus, they were automatically plunged into liquid ethane. After freezing in liquid ethane, the grids were stored in liquid nitrogen for future screening. See figure 19 for the digital photograph of the in-house spray machine.

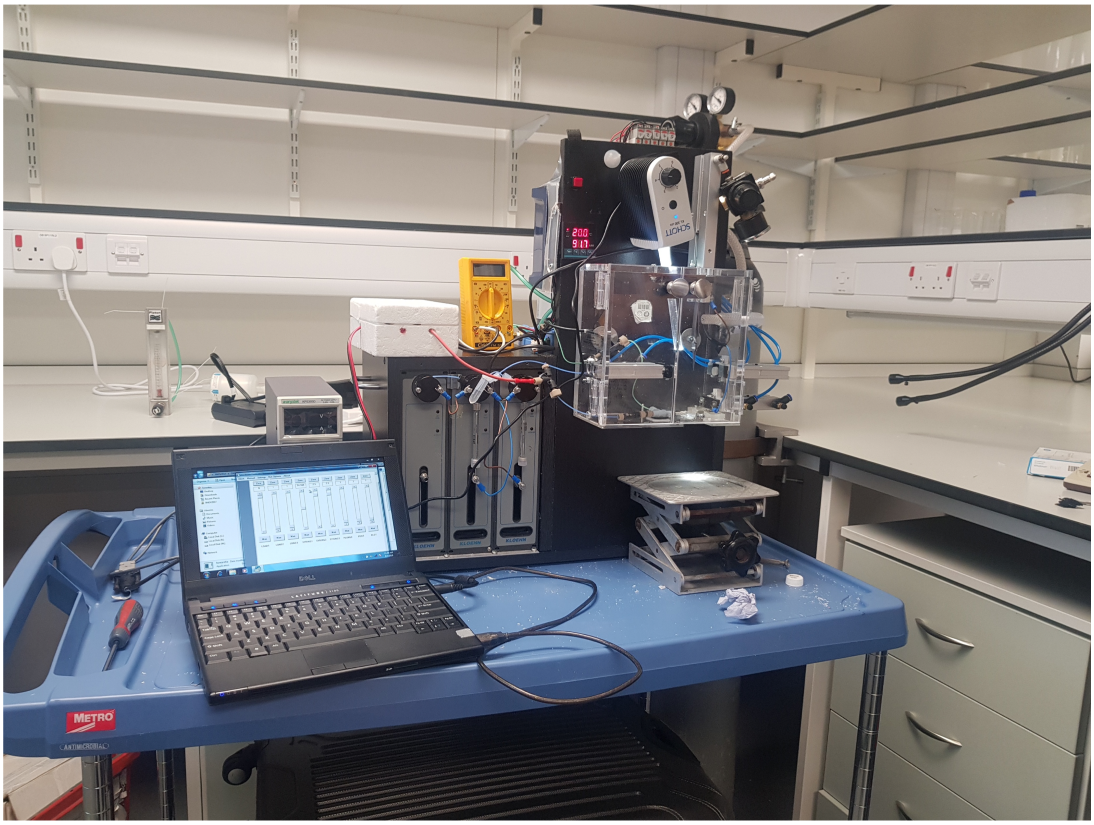


Figure 19: In-house spray and vitrification machine

A digital photograph of the spray vitrification machine. The machine is connected to the laptop which controls the sample spray and vitrification parameters, and an opaque chamber under which the samples are directly sprayed onto the grids.

2.3.5.1.3 Lacey and gold grid optimisation

Lacey carbon film on copper grids and UltrAuFoil r2/2 300 mesh grids were glow discharged for 60 seconds at 20 mA using the Cressington glow discharge apparatus. The grids were vitrified using the Vitrobot. The parameters of the Vitrobot were as follows: 4 °C, 100% humidity, blotting force of 5 and blotting time of 6 seconds. 4 µl of protein samples was deposited into the grid and the excess solution were automatically removed by blotting with filter paper. The grids were plunge frozen into liquid ethane and stored in liquid nitrogen for future screening.

2.3.5.1.4 Gold grids preparation

Gold grids were glow discharged using the Quorum Glow discharge apparatus. The grids were positively glow discharged for 40 mA for 30 seconds. The protein samples were diluted using gel filtration buffer and vitrified using the Vitrobot. See Table 7 for outline of concentrations tested. The parameters used for vitrification were as follows: 4°C, 100 % humidity, blotting time of 6 seconds and a force of 5. The grids were vitrified in liquid ethane and stored in liquid nitrogen for future screening and data collection.

Table 6: Concentration tested for successful data collection

Grid type	**Sample**	**Concentration**
UltrAuFoil	ARISC-RAP80 FR 24	0.95 mg/ml
Lacey	ARISC-RAP80 FR 24	0.95 mg/ml
UltrAuFoil	ARISC-RAP80 FR 24	0.019 mg/ml
UltrAuFoil	ARISC-RAP80 FR 24	0.038 mg/ml
UltrAuFoil	ARISC-RAP80 FR 24	0.095 mg/ml
UltrAuFoil	ARISC-RAP80 FR 24	0.19 mg/ml
UltrAuFoil	ARISC-RAP80 FR 24	0.48 mg/ml

2.3.5.1.4 Data collection

Table 7: KRIOS1 ARISC-RAP80 data collection summary

	DATA	**DATA**	**DATA**	**DATA**	**DATA**
Hardware	Quantifoil	Gold 1	Gold 2	Gold 3	Gold 4
Microscope	KRIOS1	KRIOS1	KRIOS1	KRIOS1	KRIOS1
Detector (mode)	Falcon 3	Falcon 3	Falcon 3	Falcon 3	Falcon3
Accelerating voltage (ke-V)	300	300	300	300	300
Pixel size	1.065	1.065	1.065	1.065	1.065
Data acquisition parameters					
Nominal magnification	75,000	75,000	75,000	75,000	75,000
Spot size	4	5	5	5	
Illuminated area	1.35	1.30	1.30	1.50	1.20
Dose					
Square pixel	1.13	1.13	1.13	1.13	1.13
Dose per physical pixel per second	45.93	45.9	54	57	58.16
Dose per A2/sec	40.64	40.6	47.78	50.4	51.5
Exposure time (Seconds)	2	2	2	2	2
Total dose (e/A2)	81.28	81.2	95.57	100.8	103
Number of fractions	68		66	66	66
Dose per fraction	1.2		1.4	1.32	1.56
EPU parameters					
Defocus range (-µm)	1.6, 1.9, 2.2, 2.5, 2.8, 3.1	1.6, 1.9, 2.2, 2.5, 2.8, -3.0	1.6, 1.9, 2.2, 2.5, 2.8, -3.0	1.6, 1.9, 2.2, 2.5, 2.8, 3.0	2.2,2.5, 2.8, 3.0, 3.2
Autofocus				10	10
Drift measurement	0.05	0.05	0.05	0.05	0.05
Delay after stage shift	5 s	5 s	5 s	5 s	5 s
Delay after image shift	5 s	5 s	5 s	5 s	5 s

Exposure per hole	2	3	4	3	3
Apertures (size in microns)					
C1	2,000	2,000	2,000	2,000	2,000
C2	70	70	70	70	70
C3	2,000	2,000	2,000	2,000	2,000
Objective	100	100	100	100	100

Table 8: KRIOS1 BRISC-SHMT2 data collection summary

	DATA
Hardware	Quantifoil 1.2/1.3
Microscope	KRIOS1
Detector (mode)	Falcon 3
Accelerating voltage (ke-V)	300
Pixel size	1.065
Data acquisition parameters	
Nominal magnification	75,000
Spot size	4
Illuminated area	1.35
Dose	
Square pixel	1.13
Dose per physical pixel per second	48.29
Dose per A2/sec	43.1
Exposure time (Seconds)	2
Total dose (e/A2)	86.2
Number of fractions	71
Dose per fraction	1.2
EPU parameters	
Defocus range (-µm)	1.6, 1.9, 2.2, 2.5, 2.8, 3.1
Autofocus	9

Drift measurement	0.05
Delay after stage shift	5 s
Delay after image shift	5 s
Exposure per hole	2
Apertures (size in microns)	
C1	2,000
C2	70
C3	2,000
Objective	100

2.3.5.1.5 Data processing

ARISC-RAP80 data was processed using RELION 3.0 and cryoSPARC (Scheres, 2012; Zivanov et al., 2018; Punjani, Ali et al., 2017). Cryo-EM data processing can be organised into four phases, see chapter 1.3.4.2 for a detailed outline. The first phase is termed pre-processing. Pre-processing includes importing the data, motion correction, CTF estimation, manual picking, auto-picking and particle extraction. For this project, the software package RELION 3.0 (Scheres, 2012; Zivanov et al., 2018) was used to pre-process all datasets. After importing the raw movies, the movies were corrected for any beam induced or mechanical motions that may have taken place during image acquisition using the MotionCorr implemented program in RELION (Scheres, 2012; Zivanov et al., 2018). Following motion correction, the micrographs were CTF corrected using the gCTF implemented program. Approximately 1,000 particles were then manually selected from each dataset, extracted and then aligned using reference free 2D classification to generate templates suitable for auto-picking. After auto-picking, particles from each dataset were extracted and subjected to a large scale 2D classification job.

ARISC-RAP80 data collected using Quantifoil grids were processed solely in RELION (Scheres, 2012; Zivanov et al., 2018). Data collected from the UltrAuFoil grids were pre-processed in RELION (Scheres, 2012; Zivanov et al., 2018) and then underwent 2D classification and 3D processing in cryoSPARC before being re-imported into RELION for 3D processing (Scheres, 2012; Zivanov et al., 2018; Punjani, Ali et al., 2017). Figure 19 provides an overview of the process and the parameters used at each stage. BRISC-SHMT2 data was processed in RELION 3.0 (Scheres, 2012; Zivanov et al., 2018). The high protein concentration data (0.5 mg/ml) was collected and partially processed by myself. Dr Miriam Walden completed the data processing and combined it with other BRISC-SHMT2 datasets and processed the particles as a single large dataset. Approximately 2,000 particles of each dataset were manually picked and extracted using a box size of 373 Å. The particles were subject to 2D classification to generate initial references for auto-picking. Following auto-picking, the datasets were subject to iterative rounds of 2D classification. High quality 2D class averages were then selected based on visual inspection and used to generate a 3D initial model of the BRISC-SHMT2 complex. The three datasets were combined and subject to 3D classification using the initial 3D model. 3D classification generated eight classes, one of which contained particles comprising the intact complex. The class contained 71,262 particles and was processed as a separate dataset. In parallel, this class along with four other suitable classes were processed a single dataset. This dataset contained a total of 514,930 particles. Both datasets were subjected to a final round of 3D classification to remove 'bad' particles and the best classes from 3D classification were selected and subjected to 3D refinement, per-particle CTF correction and particle polishing. Post-processing was used to appropriately mask the model, estimate and correct for the B-factor of the maps. The final resolution was obtained using the gold-standard Fourier shell correlation criterion (FSC-0.143) (Walden et al., 2019).

Figure 20: Image processing parameters

Pre-processing

RELION 3.0

- **Import**: import movies
- **Motion correction**: 1.065 pixels, 300 kV, 1.56 dose per frame, do dose weighing, B factor 150, number of x and y patches 5,5.
- **gCTF correction**: Spherical aberration 2.7 mm, 300 kV, 0.1 amplitude contrast, 1.065 pixel.
- **Manual pick** : ~ 1,000 particles
- **Extract**: 350 pixels
- **Auto-pick:** 0.09 threshold, 195 Å minimum particle distance

2D classification

RELION 3.0

- **Auto-picking template:** 10 classes, 25 iterations, 300 Å mask diameter

CryoSPARC

- **Large 2D classification:** Extracted particles from auto-pick. 3 rounds of 2D classification.

3D initial model + classification

CryoSPARC

- **Ab-initio reconstruction:** Requested 5 classes
- **Heterogenous refinement:** Refine individual or multiple maps against ab-initio model of choice
- **Local resolution estimation:** Threshold of 0.143

RELION 3.0

- **3D classification:** Imported particles from cryoSPARC 2D classification.

Reference in absolute grey scale: No
Reference been CTF correct: Yes
Iterations: 25
Classes: depends on size of data
Mask: 300 Ås
Regularization Parameter T: 4
C1 symmetry

3D refinement + post-processing

RELION 3.0

- **3D refinement:** C1 symmetry, low pass filtered to 40 Å
- **Mask creation:** Low pass filter to 15 Angstroms, initial binarization threshold depends on map, extending map and soft edge depends on the map.
- **Post-processing:** 1.065 pixel size, estimate B factor automatically, MTF: obtained from RELION wiki online.

2.3.6 Mass spectrometry (MS)

Native MS was used to uncover the stoichiometry of the ARISC-AIR and ARISC-RAP80 complex. The protein samples were diluted to a concentration of 2 mg/ml using gel filtration buffer. Native MS analysis of the ARISC-AIR and ARISC-RAP80 complex was carried out by the MS facility staff.

The analysis of the ARISC-RAP80 complex was further optimised and analysed by Dr Charlotte Scarff using the Thermo Scientific Q Exactive UHMR Orbitrap mass spectrometer. In the final analysis, 2 mg/ml of ARISC-RAP80 was buffer-exchanged into 1 M ammonium acetate using Zebra^{TM} spin Desalting Columns, 7 K MWCO, 0.5 mL. The sample was then analysed neat on the Synapt instrument and buffer exchanged for a second time and diluted 4-fold for analysis on the Orbitrap UHMR system. The mass spectra were obtained on a Thermo Scientific Q Exactive UHMR mass spectrometer. ARISC-RAP80 was introduced into the instrument by direct infusion nanoESI through an in-house prepared gold/palladium-coated borosilicate glass capillary at 1 μM concentration in 1M ammonium acetate. Full scan data were acquired at a resolving power of 12500 in the m/z range 2,000-15,000. The mass spectra presented in the results section are comprised of approximately 150 scans. Data processing was performed with the use of the UniDec Software program(Marty et al., 2015).

2.3.7 Fluorescence activity assay

Fluorescence activity assays were carried out to determine the activity of the purified ARISC-AIR and ARISC-RAP80 complexes, and to assess differences in activity (if any). Prior to comparing the activity of both complexes, a titration profile was established to determine the optimum concentration for observing enzyme activity. The titration profile was obtained using 2.5 nM, 5 nM, 10 nM, 20 nM, and 40 nM of ARISC-AIR and 1 μM internally quenched fluorescence (IQF) K63-linked diubiquitin (LifeSensors). After establishing the optimum enzyme concentration, an activity assay using ARISC-AIR and the IQF substrate was run in parallel with ARISC-RAP80 and IQF di-ubiquitin.

The assays were repeated three times and the data were recorded using a HIDEX Sense microplate reader. The data was analysed using GraphPad Prism.

2.3.7 Gel shift assay (EMSA)

2.3.7.1 Production of 50 bp DNA

Ethanol precipitation was employed to obtain 50 bp PCR product DNA using forward and reverse Abraxas1 primers. The ethanol precipitation protocol was adopted from a published protocol (Green and Sambrook, 2016).

The Abraxas primers were diluted to a working stock concentration of 2 μM. 25 μl of the forward Abraxas1 primer was mixed with 25 μl of the reverse Abraxas1 primer. The mixture then underwent an annealing reaction in which the mixture was denatured at 95 °C for 5 minutes and annealed at 50°C for 30 seconds. 5 μl of the sample was then run on a 1% agarose gel to verify the size of the PCR product. 1/10th the volume of sodium acetate followed by 2.5 x volume of 95% ethanol was added to the mixture and the reaction was incubated overnight at -20°C. Following overnight incubation, the reaction was centrifuged at 11,000 x g for 30 minutes. After centrifugation, the supernatant was discarded, and the DNA pellet was rinsed with 70 % ethanol and centrifuged for an additional 15 minutes. After centrifugation, the supernatant was discarded, and the pellet was dissolved in filtered milliQ water.

2.3.7.1.1 Gel shift assay with 50 bp DNA

The complexes were diluted to a working stock concentration of 23 μM. Following this, the protein-DNA interaction of the ARISC-RAP80, ARISC-AIR and ARISC, BRCC45-MERIT40 complexes were tested under two different concentrations: 5, and 2.5 μM. The dilutions and concentrations are outlined in Table 10. The protein-DNA mixture was incubated at room temperature for 2 hours. Following incubation, loading dye was added to the sample and the samples were analysed on a 2.5 % agarose gel.

Table 9: 50 bp DNA gel shift assay reaction

ARISC-RAP80
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC-RAP80 complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC-RAP80 complex 2.875 μl milliQ water
ARISC-AIR
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC-AIR complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC-AIR complex 2.875 μl milliQ water
ARISC
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC complex

2.875 µl milliQ water
BRCC45-MERIT40
5 µM reaction 1 µl binding buffer (gel filtration + 5 % glycerol) 5 µl DNA (63 ng/µl) 2.5 µl BRCC45-MERIT40 complex 1.75 µl milliQ water
2.5 µM reaction 1 µl binding buffer (gel filtration + 5 % glycerol) 5 µl DNA (63 ng/µl) 1.125 µl BRCC45-MERIT40 complex 2.875 µl milliQ water
Control 1
1 µl binding buffer (gel filtration + 5 % glycerol) 5 µl DNA (63 ng/µl) 4 µl milliQ water
Control 2
1 µl binding buffer (gel filtration + 5 % glycerol) 4.5 µl ARISC-RAP80 5 µl milliQ water
Control 3
1 µl binding buffer (gel filtration + 5 % glycerol) 4.5 µl ARISC-AIR 5 µl milliQ water
Control 4
1 µl binding buffer (gel filtration + 5 % glycerol) 4.5 µl ARISC 5 µl milliQ water
Control 5
1 µl binding buffer (gel filtration + 5 % glycerol) 4.5 µl BRCC45-MERIT40 5 µl milliQ water
Control 6

1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 4.46 μl glycogen synthase

2.3.7.2 300 bp DNA

The PCR protocol for obtaining the 300 bp DNA is outlined in chapter 2.3.1.2. The complexes were diluted to a working stock concentration of 23 μM. Following this, the protein-DNA interaction of the ARISC-RAP80, ARISC-AIR and ARISC complexes were tested under three different concentrations: 10, 5, and 2.5 μM. The dilutions and concentrations are outlined in table 11. The protein-DNA mixture was incubated at room temperature for 2 hours. Following incubation, loading dye was added to the sample and the samples were analysed on a 2.5 % agarose gel.

Table 10: 300 bp DNA gel shift assay reaction

ARISC-RAP80
10 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 4.5 μl ARISC-RAP80 complex
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC-RAP80 complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC-RAP80 complex 2.875 μl milliQ water

ARISC-AIR
10 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 4.5 μl ARISC-AIR complex
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC-AIR complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC-AIR complex 2.875 μl milliQ water
ARISC
10 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 4.5 μl ARISC complex
5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 2.5 μl ARISC complex 1.75 μl milliQ water
2.5 μM reaction 1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl) 1.125 μl ARISC complex 2.875 μl milliQ water
Control 1
1 μl binding buffer (gel filtration + 5 % glycerol) 5 μl DNA (63 ng/μl)

4.5 μl milliQ water
Control 2
1 μl binding buffer (gel filtration + 5 % glycerol) 4.5 μl ARISC-RAP80 5 μl milliQ water
Control 3
1 μl binding buffer (gel filtration + 5 % glycerol) 4.5 μl ARISC-AIR 5 μl milliQ water
Control 4
1 μl binding buffer (gel filtration + 5 % glycerol) 4.5 μl ARISC 5 μl milliQ water

2.3.8 Photoreactive ubiquitin cross-linking

Photoreactive ubiquitin cross-linking experiments were conducted to determine the subunit(s) capable of binding monoubiquitin. The protocol was adapted from the UbiQ manual. 50 μl of 25 μM UbiQ photoreactive ubiquitin was reacted with the 50 μl of 1.14 μM ARISC-RAP80 complex. The reaction was left to incubate at room temperature in the dark for 30 minutes. After incubation, the mixture was placed 10 cm from a UV light source for 90 minutes. Samples were taken every 30 minutes for analysis on a 4-20 % gradient gel. A total of 1.7 μg of protein was loaded in each well. The reaction was conducted in parallel with a control that comprised of 50 μl of 1.14 μM ARISC-RAP80 complex and 50 μl of gel filtration buffer.

2.3.9 Native PAGE analysis

The experiment was conducted in accordance with the Invitrogen user manual and the protein complexes were diluted to 0.45 and 0.008 mg/ml. The voltage was set to 150 kV and the gel was run for approximately 60 minutes.

CHAPTER 3

Structural and functional characterisation of the ARISC complexes by negative stain EM, Mass spectrometry and fluorescence-based activity assay

3.1 Introduction

The five membered ARISC-RAP80 complex forms an additional complex with BRCA1 and BARD1 termed BRCA1-A complex. BRCA1-A plays a vital role in repairing double-stranded DNA breaks in the homologous recombination (HR) pathway (Hu, Y. et al., 2011; Moynahan et al., 1999). The complex functions by binding to chromatin flanking regions in double-stranded DNA breaks in a phosphorylation and ubiquitylation dependent manner. Phosphorylation and ubiquitylation events are initiated by the MRN complex. Following DSBs, MRN recruits and activates ATM, which in turn, phosphorylates Serine 139 of histone H2AX (Burma et al., 2001). The protein MDC1 senses γ-H2AX via its BRCT repeats and initiates the recruitment of ubiquitin E3 ligases RNF8, RNF168, and the E2 conjugating enzyme UBC13/MMS2. Together, the enzymes ubiquitylate histones H2A and H2Ax in a K63 dependent manner. K63 ubiquitin chains are recognised by the UIM domains of RAP80, which facilitates the recruitment of the BRCA1-A complex to double stranded DNA breaks. K63 chains are catalysed by the JAMM/MPN deubiquitinase BRCC36, antagonising the MDC-RND8-RNF168 cascade and halting DSB repair by HR.

Biochemical analyses of the ARISC complex have shown that BRCC36 is activated when in complex with Abraxas1 and BRCC45 (Patterson-Fortin et al., 2010). The interaction of MERIT40 is thought to be required for the integrity of the whole complex. The Abraxas1 paralog, Abraxas2 forms a distinct complex with BRCC36, known as the BRISC complex. Structural studies of insect Abraxas2 and BRCC36 revealed that the heterodimer forms a tetrameric complex with itself, termed the 'superdimer'. Due to the similarities between Abraxas1 and Abraxas2, the Abraxas1-BRCC36 heterodimer was believed to also form a superdimer (Zeqiraj et al.,2015). Negative stain EM analysis of ARISC complex provided a low resolution description of complex assembly and confirmed that Abraxas1-BRCC36, like Abraxas2-BRCC36, forms a tetrameric complex (Kyreileis et al., 2016). The study found that Abraxas1-BRCC36-BRCC45-MERIT40 form a U-shape architecture in which BRCC45 and MERIT40 form the arms and Abraxas1 and BRCC36 occupy the base of the complex. The analysis of ARISC by negative stain also identified a small population of ARISC particles that self-associate, forming a dimer of octamers, which was further enhanced by covalent cross-linking with glutaraldehyde (Kyrieleis et al., 2016) (Figure 21).

This chapter provides the preliminary observations and low resolution structures of the native (non-crosslinked) ARISC, ARISC-AIR, and ARISC-RAP80 complexes by negative stain EM and investigates the stoichiometry and enzyme activity of the complexes by mass spectrometry and fluorescence-based activity assays.

Unless stated otherwise, protein samples were expressed in insect cells using the MultiBac expression system and cell pellets containing overexpressed complexes were provided by Dr Elton Zeqiraj. The complexes were purified by his-tag purification. The histidine tag was located on the BRCC45 protein subunit. A histidine tag was used because they are commonly used in the production of recombinant proteins since the histidine residues can bind to several types of immobilized ions such as nickel, cobalt and copper and allow for simple detection and purification of his-tagged proteins under specific buffer conditions. The reader is directed to chapter 2.3.3 for a detailed explanation of the purification process.

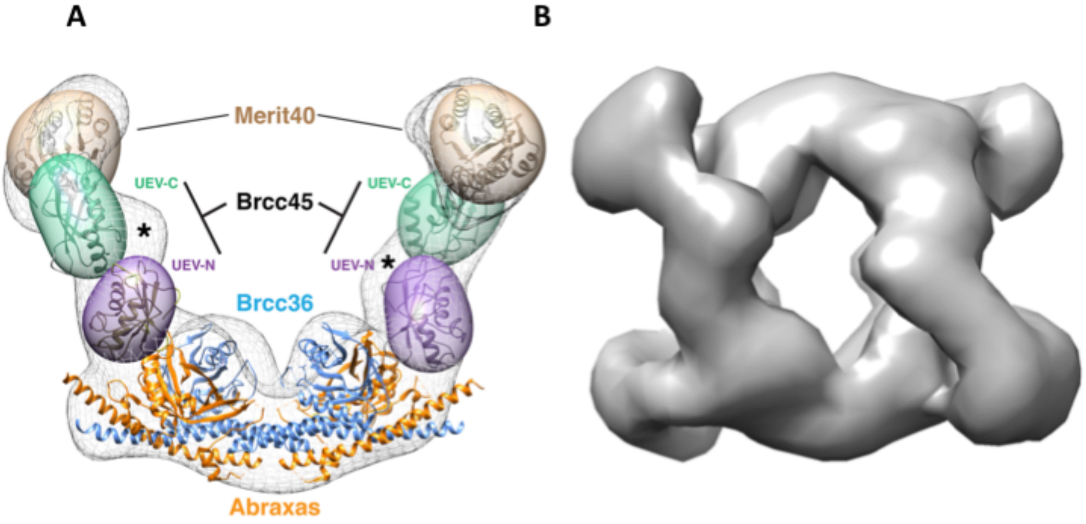


Figure 21: Published model of the ARISC complex at ~20 Å resolution

[A] ARISC forms a U-shaped complex. The crystal structure of Abraxas2-BRCC36 was fitted into the base of the complex and BRCC45 and MERIT40 were modelled into "arm" regions. [B] The ARISC cross-linked dimer identified by Kyrieleis et al. This figure was adapted from Kyrieleis et al 2016.

3.2 ARISC complex purification

Human ARISCΔN complex was purified from pellets as outlined in the methods chapter 2.3.3.1. ARISCΔN complex, which lacks the first 71 residues of MERIT40, was used instead of full length ARISC. This region is heavily phosphorylated and poly-ADP-ribosylated and does not contribute to complex assembly or activity (Zeqiraj et al., 2015). After Ni-affinity purification and washes in low and high salt buffers and elution using an imidazole gradient, the ARISCΔN complex appeared to elute in high concentrations in fractions 13-17 (Figure 22). The SDS-PAGE analysis of the elution fractions suggests that the complex also eluted in fraction 20-28. As the complex appeared to be present at much lower concentration in lanes 20-28, those samples were discarded.

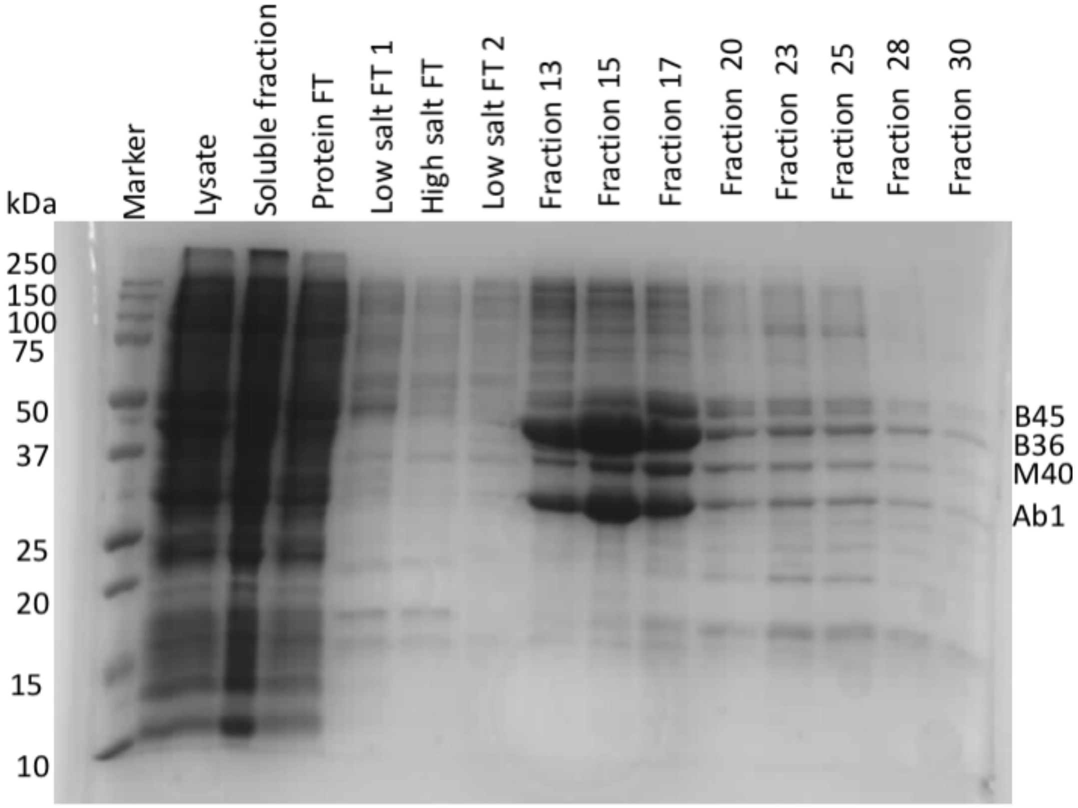


Figure 22: Nickel affinity purification of the ARISCΔN complex

Nickel purification of the ARISC complex shows that the complex had successfully bound to the 5 ml HisTrap column and was eluted in high concentration in fraction 13-19 and at lower concentration in fractions 20-28.

Following nickel purification, fractions 13-19 were combined and concentrated to a volume of 2 ml and protein concentration of 24.7 mg/ml. The concentrated sample was then further purified by gel filtration. Gel filtration results showed that the ARISC complex eluted in fractions 22-28 at high abundance (Figure 23). A second peak containing the BRCC45-MERIT40ΔN complex was observed in fractions 38-40. Fraction 22-28 containing the full ARISCΔN were combined and concentrated to 4.7 mg/ml.

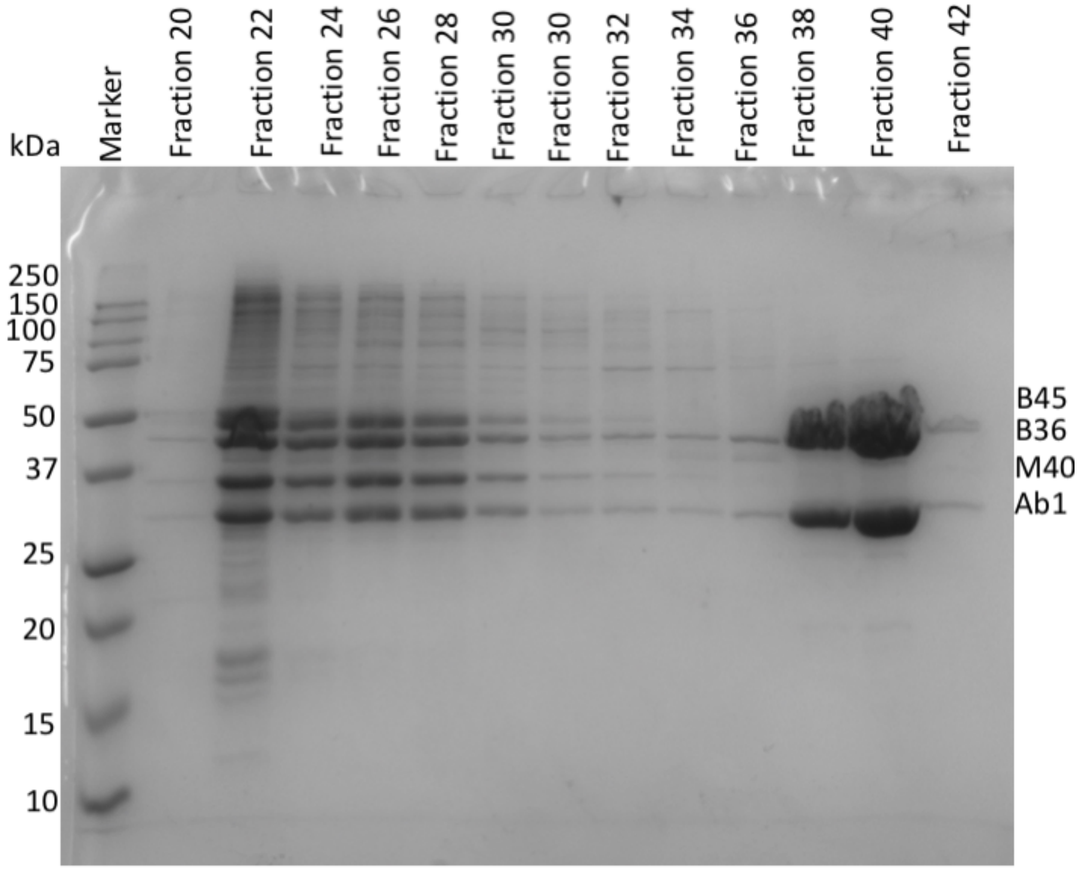


Figure 23: Gel filtration purification of the ARISCΔN complex
Following nickel subtraction, the ARISC complex was further purified by gel filtration and fractions were run on a 15% SDS PAGE gel. Results from the gel shows that the complex eluted in abundance in fraction 22-28, and a lower abundance in fraction 30-32.

3.2.1 Negative stain EM sample preparation

For the remainder of this thesis, the ARISCΔN complex will be referred to as the ARISC complex for simplicity. The ARISC complex was diluted with gel filtration buffer to a concentration of 0.008 mg/ml and directly adsorbed onto glow discharged TEM grids. This sample concentration was found to be the most suitable for studying the complex because at higher concentrations, the particles were too crowded on the micrographs, and further dilution of the sample resulted in complex dissociation. 242 micrographs were obtained using the FEI F20 microscope at a magnification of 29,000 x and a voltage of 120kv. The software package RELION (Scheres, 2012) 2.1 was used to process the data.

3.2.2 Micrograph analysis and particle picking

Analysis of the micrographs revealed the presence of some protein aggregation. The particles on the micrographs were white with slightly poor contrast and varied in size (Figure 24A). After CTF correction, approximately 1,000 particles were manually picked from a subset of randomly selected micrographs, extracted, and aligned using the 2D classification module in RELION (Scheres, 2012) software. The output of 2D classification revealed 2D class averages that varied in shape and size. Some class averages were globular whereas others expressed a relaxed crescent like shape.

Ten 2D classes were used as a template for automated particle picking (Figure 24b), this resulted in the auto-picking of 20,862 particles from the entire dataset. Visual evaluation of the auto-pick job revealed that the templates were suitable for auto-picking. Following auto-picking, the particles were extracted using a mask size of 300 Å. The extracted images were then aligned using reference free 2D classification (Figure 25).

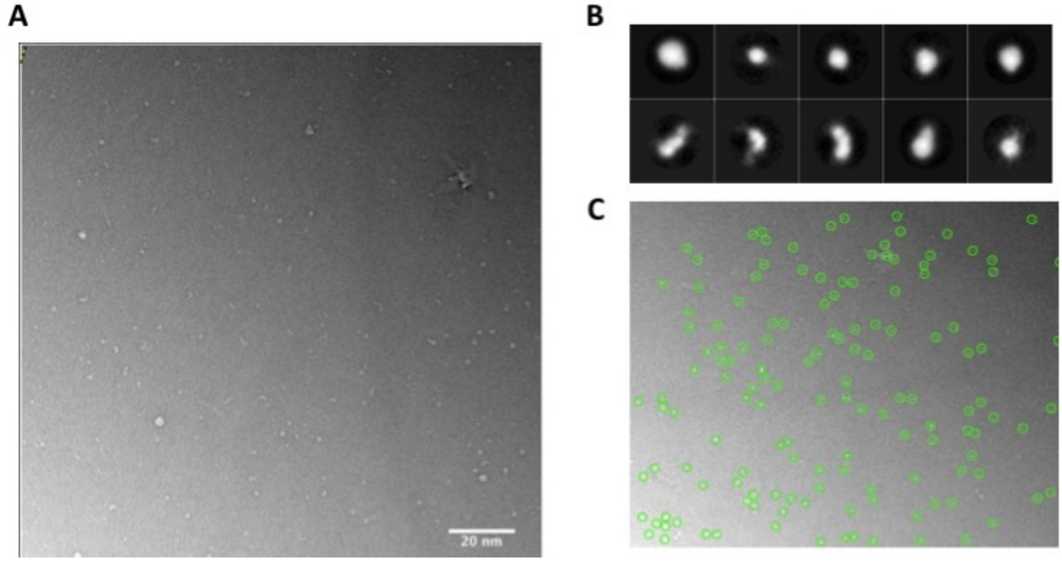


Figure 24: Negative stain EM pre-processing of the ARISC complex

[A] Representative micrograph of the ARISC complex dataset. [B] 2D class averages used as templates for auto-picking. [C] A representative auto-picked micrographs. Analysis of the micrographs confirmed that the 2D class averages were suitable templates for auto-picking.

3.2.3 2D classification

Reference free 2D classification generated white classes averaged with black backgrounds. The majority of the classes were small and spherical. A small subset displayed L-like shape in different orientations or an elongated O-shape. Some of the classes contained distinct two or three interacting subunits. The results of the 2D classification also suggest the presence of protein aggregation, and non-particles.

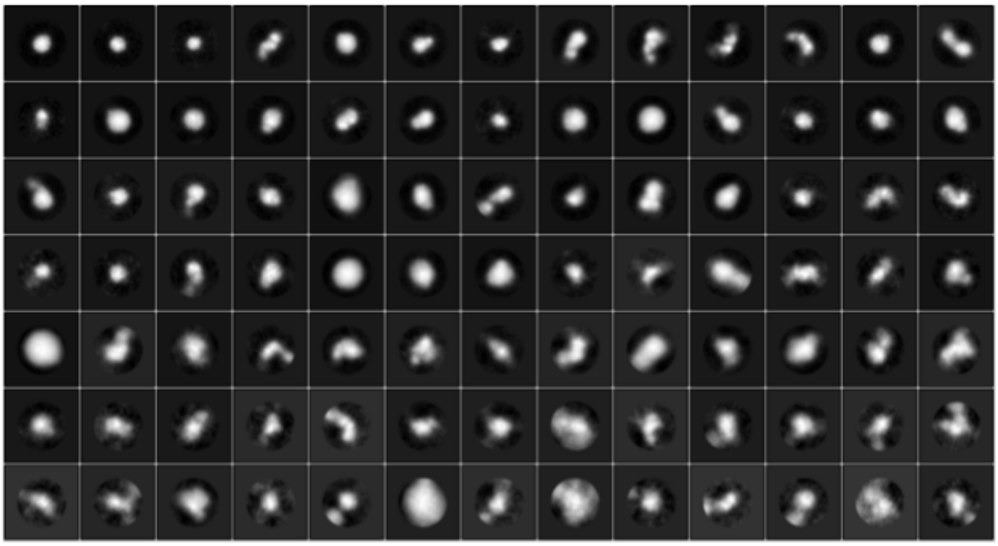


Figure 25: 2D classification of the auto-picked ARISC particles

Reference free 2D classification aligned the extracted particles into class averages. The class averages varied in size and shape. Some classes were small and circular whereas other displayed distinct L or U-like shape.

3.2.4 3D reconstruction and refinement

A subset of 2D classes were selected to generate a 3D initial model in C1 symmetry. To ensure that RELION (Scheres, 2012) generated a model that accurately reflects the data, classes positioned in different orientations were selected. The result from the 3D initial model job revealed a map that closely reflected the 2D classes obtained from reference free 2D classification. This suggests that the 3D map is a suitable representation of the dataset (Figure 26A).

To obtain a more reliable 3D map and maximise the accurate alignment of particles, additional suitable 2D classes were selected from the reference free 2D classification output. Misaligned particles, non-particle, and aggregated class averages were omitted from the selection. 12 class averages containing a total of 4,954 particles were then subject to 3D classification and the C1 3D initial model was used as reference. The published model of the ARISC complex identified the complex as having C2 symmetry. C2 symmetry was therefore applied during 3D classification and four classes was requested.

Results from 3D classification identified four distinct maps (Figure 26B) that shared some similarities at the core of the complex but structurally differed from one another. Class one, for example, contained MERIT40 density at the arms of the complex, that extended in opposite directions whereas class two did not. Class three, on the other hand, was missing density at the arms of the complex and likely contains only the subunits that make up the core of ARISC (Abraxas1-BRCC36). Class four also shared the U-like topology with classes one and two, however, it contained additional densities at the arms that extended and formed a bridge across the complex. It is possible that this class represents the ARISC dimer, also observed by Kyrieleis et al., 2016. As the four maps were structurally distinct from one another, they were refined independently to low resolutions of 34-44 Å.

Overall, results from the negative stain EM analysis of the ARISC complex suggests that under the negative stain condition, the complex dissociates and displays high levels of heterogeneity. However, even at this low resolution, maps from Class 1 and 2 confirmed the U-shaped structure previously described by Kyrieleis et al., 2016.

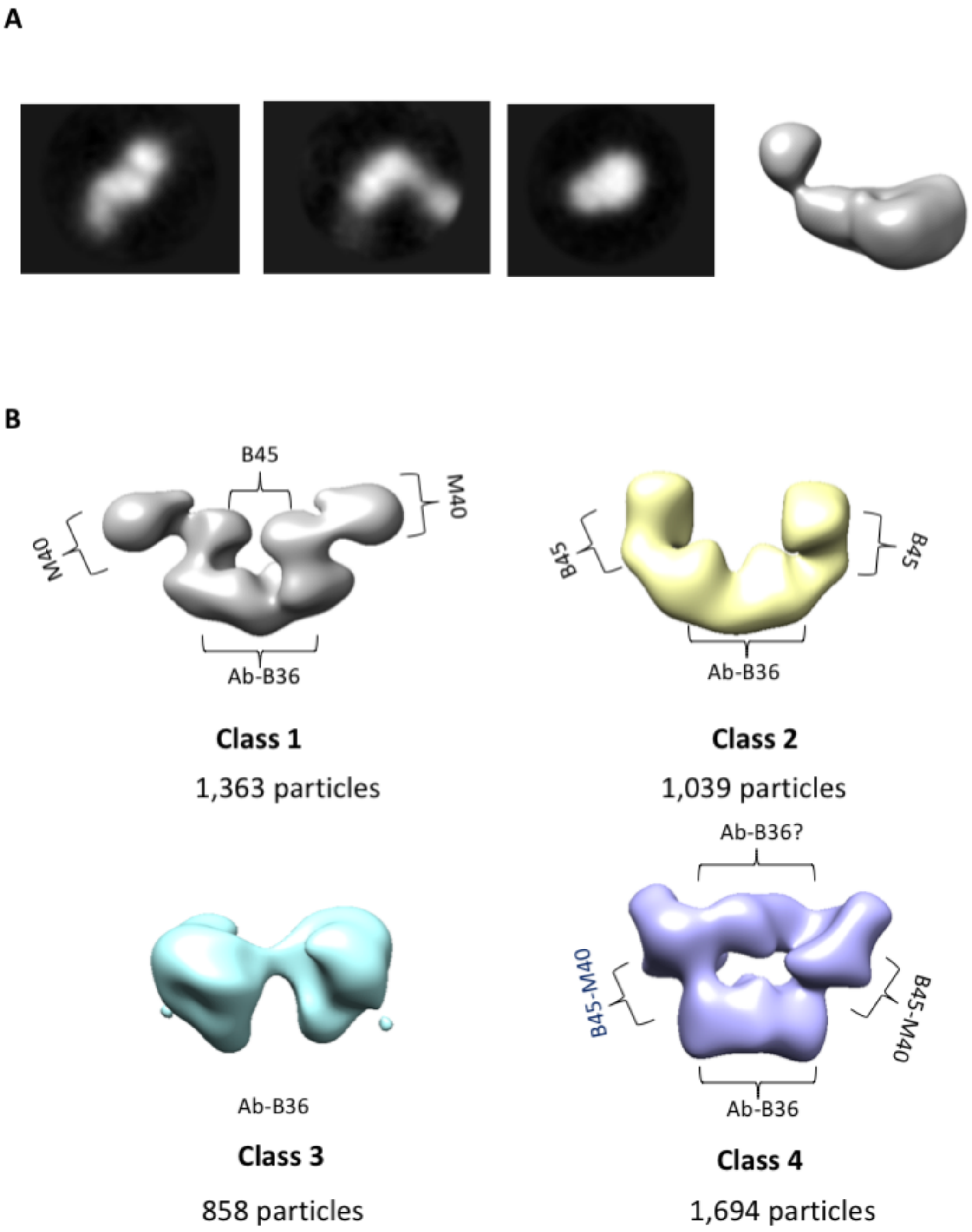


Figure 26: Negative stain EM 3D reconstruction of the ARISC complex

[A] Representative 2D classes selected for the generation of the initial model in C1 symmetry. [B] 3D maps obtained from 3D classification. The maps were refined to resolutions of 35(grey), 35(yellow), 40.8 (cyan blue), 45 (purple) Å.

3.3 Protein purification of the ARISC-AIR complex

The ARISC-AIR complex is composed of BRCC36, Abraxas1, BRCC45, MERIT40 and the 15 kDa protein, Abraxas interaction region (AIR) of the RAP80 protein. ARISCΔNΔC-AIR complex was used in place of full length ARISC -RAP80 and the domain boundaries were guided from the structural studies of the BRISC complex (Walden et al., 2019) and the ARISC complex studies from Kyrieleis et al., 2016. Therefore, ARISCΔNΔC-AIR contains full length BRCC36 (residues 1-316), Abraxas1ΔC (residues 1-265), full length BRCC45 (residues 1-407), MERIT40ΔN (residues 72-329) and AIR (270-400).

Since there was no structural information available on the molecular details of RAP80 interaction with ARISC, I aimed to understand how this minimal RAP80 region, interacts with Abraxas1, and provide further insights on the stoichiometry and assembly of the ARISC-RAP80 complex. Chapter 3.2 showed that the ARISC complex was prone to dissociation. Therefore, negative stain EM analysis of ARISCΔNΔC-AIR also aims to determine whether the binding of AIR to ARISC will improve overall complex stability.

ARISCΔNΔC-AIR protein was provided by Dr Elton Zeqiraj. To obtain cleaner and more homogenous sample, the protein sample underwent gel filtration purification, and samples from the eluted fractions were run on a 15% SDS PAGE gel. Results from the gel shows that the majority of the complex eluted between fractions 25 and 30. However, there were traces of the subcomplexes in all the fractions (Figure 27). For the remainder of this thesis, the ARISCΔNΔC-AIR complex will be referred to simply as the ARISC-AIR complex.

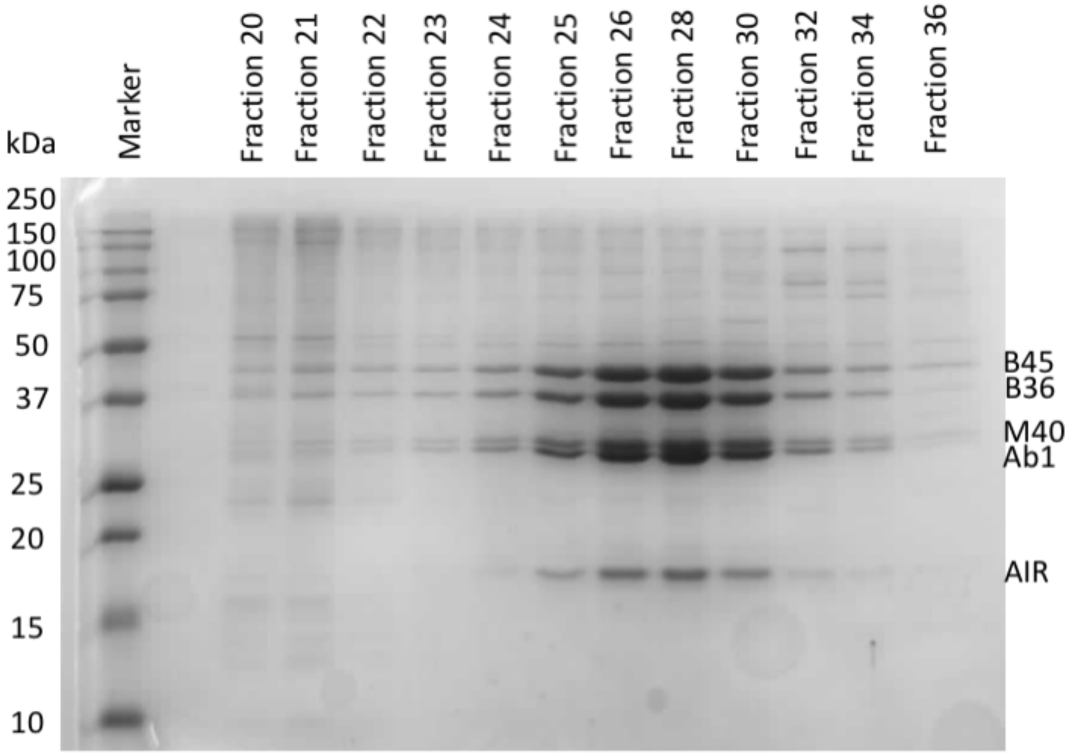


Figure 27: Gel filtration results of the ARISC-AIR complex

Gel filtration results shows that the majority of the ARISC-AIR complex eluted in fractions 25-30. The other fractions contained various subcomplexes of the ARISC-AIR complex.

3.3.1 Negative stain EM sample preparation

0.008m/ml of ARISC-AIR was deposited on glow discharged TEM grids as described in chapter 2.3.4 219 micrographs were collected at a magnification of 29,000x, and a KV of 200 using the FEI F20 microscope. The dataset was processed using RELION 3.0 beta (Scheres, 2012; Zivanov et al., 2018).

3.3.2 Micrograph analysis and particle picking

Visual analysis of the micrographs revealed white particles with good contrast. The particles displayed different shapes and some U-shaped particles were identifiable, indicating complex integrity (Figure 28A). As AIR is a 15 kDa protein, it is impossible to visually determine whether the protein remained bound to ARISC from examination of the raw data. Approximately 1,000 particles were manually selected and extracted to generate 2D templates for auto-picking. The 2D templates contained a single class average that exhibited the expected U-shape of ARISC (Figure 28B). Some class averages appeared to be of the dissociated sub-complexes as they contained only a single arm or were missing MERIT40. Four class averages were globular in shape, two of which were large, indicating BRCC45-MERIT40 dissociation and protein aggregation, respectively.

148,930 particles were auto-picked using the templates. Evaluation of the auto-picked micrograph suggests that the 2D class averages were suitable templates for auto-picking. Following auto-picking, the particles were extracted using a mask size of 300 Å and subjected to reference-free 2D classification.

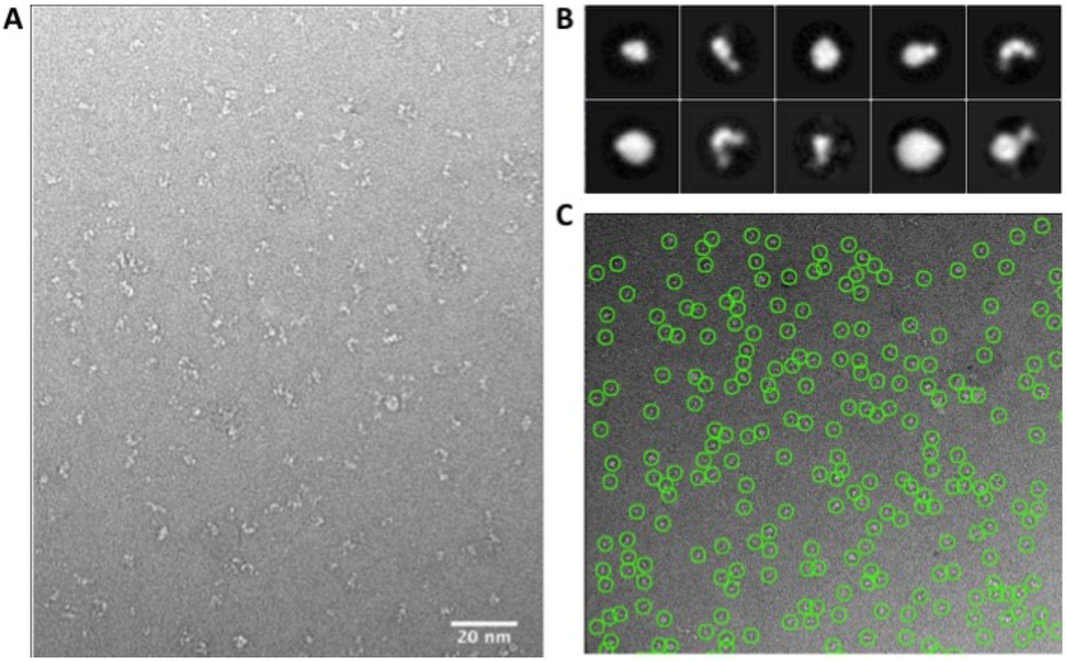


Figure 28: Negative stain EM pre-processing of the ARISC-AIR complex

[A] Representative micrograph of the ARISCΔNΔC-AIR complex dataset. [B] Approximately 1,000 particles were manually selected, extracted and classified to generate ten templates for auto-picking [C]. Representative auto-picked micrographs. Analysis of the micrographs confirmed that the 2D class averages were suitable templates for auto-picking.

3.3.3 2D classification

The 2D classification results revealed generally well aligned particles (Figure 29). The particles displayed different shapes and sizes and were positioned in various orientations. Nine of the 2D class averages shared similar topology with the ARISC 2D classes, however some contained one arm, others two and some appeared to be missing MERIT40 subunits. As AIR is a small protein, it is difficult to determine which class, if any, contained the protein and its location within the complex. Some class averages contained junk and/or misaligned particles. These were easily identifiable as they were fuzzy, exceptionally large and globular.

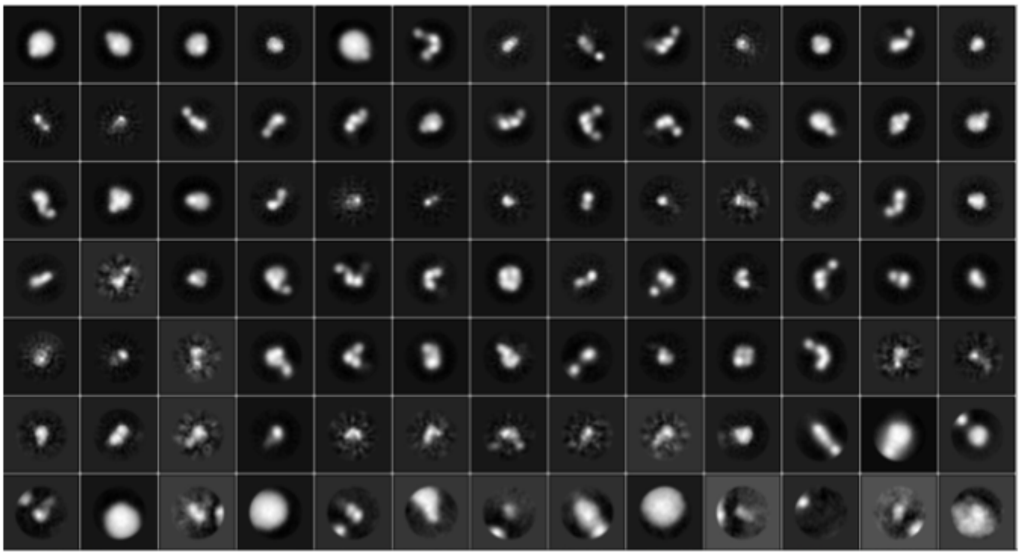


Figure 29: 2D classification of ARISC-AIR complex

Reference free 2D classification aligned the extracted particles into 100 class averages (91 represented here). The class averages varied in shape and size indicating heterogeneity. Some classes were small and circular whereas others resembled the ARISC model. The last row contained non-particles, poorly aligned particles and dissociated subunits.

3.3.4 3D reconstruction and refinement

20 2D classes were selected to obtain an initial model of the ARISC-AIR complex in C1 symmetry. The number of particles in each class were limited to 200. To limit bias and maximise the probability of obtaining an accurate model, 2D classes displaying different views were selected. Results from the 3D initial model job revealed a map that displayed a relaxed U shape (Figure 30A) and shared similarities with the published ARISC model (Figure 21A) and reflected the class averages in the data. This suggests that the initial model is a good representation of the data. To generate a more reliable 3D map, 36 suitable 2D classes containing 87,991 particles were selected from the 2D classification job and were subject to 3D classification. The 3D initial model was used as reference for particle alignment. To aid in the identification of complex heterogeneity, and to accurately separate and align particles, four 3D classes were requested. Results from 3D classification revealed four maps with similar U-like topology of the published ARISC model (Figure 30B) and that like the ARISC 3D maps, displayed structural heterogeneity.

Although the four 3D maps displayed the general topological structure of ARISC, there were some notable differences. For example, MERIT40 in the first class appears to interact with the core of the complex, and class three contains two large densities at the core of the complex between both BRCC45-MERIT40 arms, possibly corresponding to multiple AIR proteins. Out of the four maps, class two and four are the most similar. Differences between the two maps can be located at the core of the complex, where a small protruding density is visible in class two. The four maps were individually refined to a resolution of 25-28 Å and the initial model was used as reference. Due to the small nature of the AIR protein and the low resolution of the maps, it is difficult to accurately conclude that class three contained AIR.

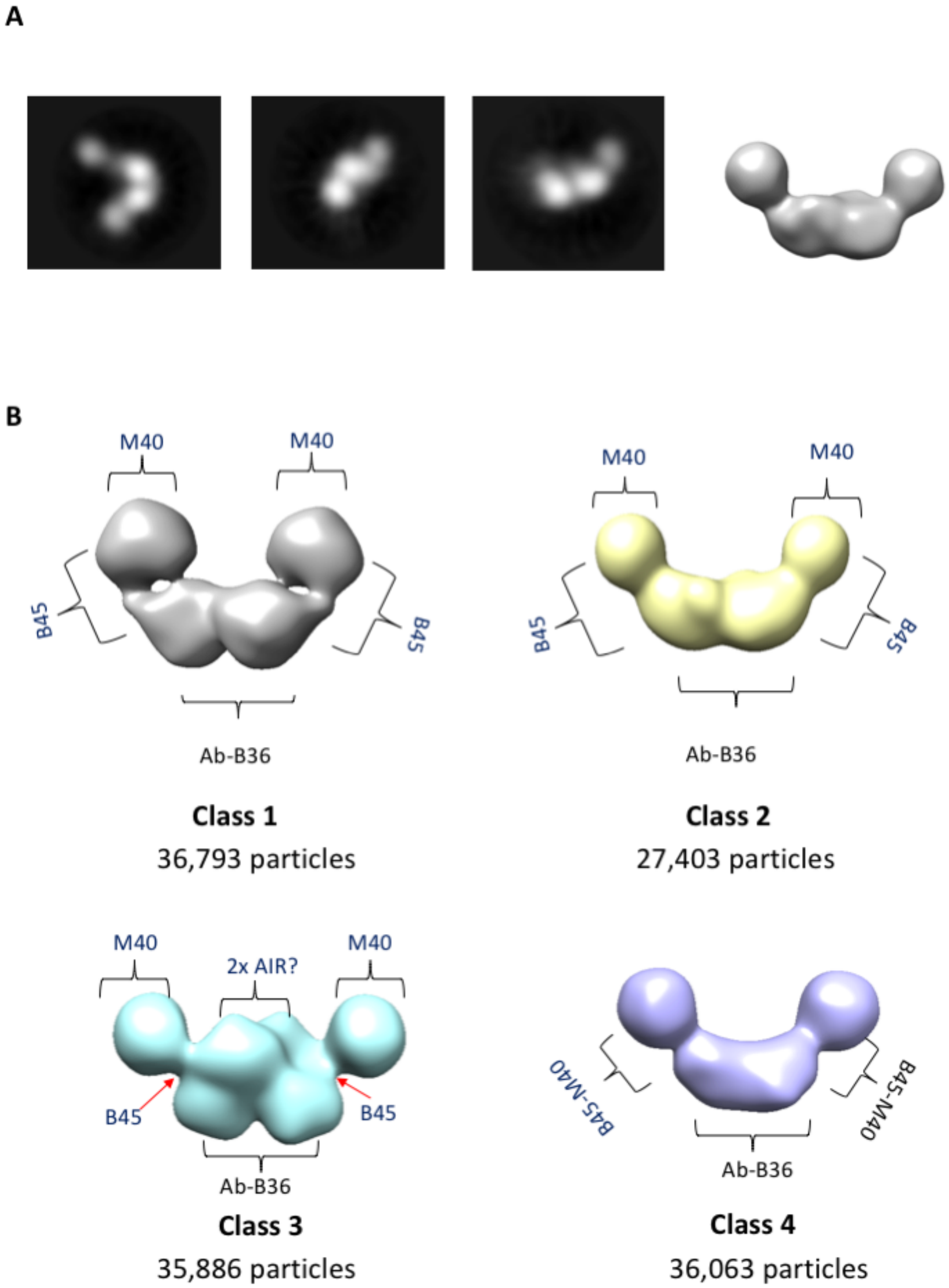


Figure 30: Negative stain EM 3D reconstruction of the ARISC-AIR complex

[A] Representative 2D classes selected for the generation of the Initial model and the initial model in C1 symmetry. [B] 3D maps obtained from 3D classification. The maps were refined to resolutions of 25 (grey), 25(yellow), 27 (cyan blue), 28 (purple).

3.4 Low resolution structural analysis of the ARISC-RAP80 complex

Results from the negative stain EM analysis of the ARISC-AIR complex suggests that the interaction of AIR with ARISC stabilises the complex. Disappointingly, the analysis did not provide clear information on the stoichiometry of the ARISC-AIR/RAP80 complex. AIR is a small protein, despite being mostly unstructured and disordered, RAP80 contains additional structured domains. It has a UIM and SIM motifs located near its N-terminus, and two zinc finger domains at the C-terminus. These regions may interact with ARISC further stabilising the complex and reveal larger bound structures allowing for the identification of RAP80 within the ARISC complex. In this section, negative stain EM was employed to study the structure of the ARISC-RAP80 complex.

3.4.1 Sample preparation

ARISC-RAP80 protein was provided by Dr Elton Zeqiraj. The protein sample was diluted to 0.005mg/ml in gel filtration buffer. ARISC-RAP80 was deposited on glow discharged TEM grids as described in chapter 2.3.4. 333 micrographs were collected at a magnification of 29,000x and a KV of 200 using the FEI F20 microscope. The dataset was processed using RELION 3.0 beta (Scheres, 2012; Zivanov et al., 2018).

3.4.2 Micrograph analysis and particle picking

Analysis of the micrographs revealed particles that varied in shape and largely displayed a globular topology (Figure 31A). Approximately 1,000 particles were manually selected, extracted and aligned using 2D classification. 2D classification generated ten class averages. Out of the ten classes, the third class was the only class that displayed the topology of ARISC, the other nine classes were either globular in shape or were elongated (Figure 31B). The 2D classes were used as templates for auto-picking and 150, 680 particles were auto-picked from the 333 micrographs. Visual evaluation of the auto-picked micrographs suggests that the 2D class averages used for auto-picking were suitable templates. Following auto-picking, the particles were extracted using a mask size of 300 Å and aligned by reference free 2D classification.

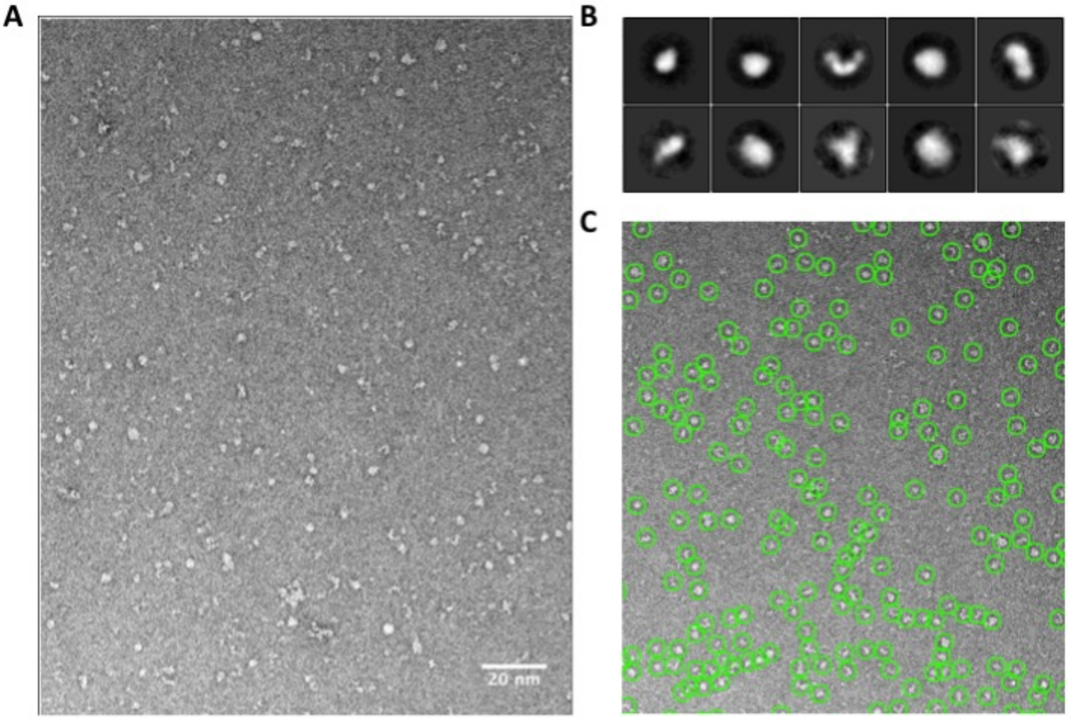


Figure 31: Negative stain EM pre-processing of the ARISC-RAP80 complex

[A] Representative micrograph of the ARISC-RAP80 complex dataset. [B] 2D templates used for auto-picking. [C] A representative auto-picked micrograph.

3.4.3 2D classification

The outputs from 2D classification revealed well-aligned class averages, with the eight classes displaying the U shape seen in the previous negative stain EM datasets. Some classes were small, perhaps containing only two or three subunits, suggesting sample dissociation and heterogeneity. As particles adsorb onto the TEM grid in different orientations, accurately assessing complex dissociation by examining 2D classes is difficult. A small subset of the class averages contained fuzzy edges, this was likely due to particle misalignment. Other classes did not display a distinct shape and were completely fuzzy, these classes contained non-usable particles must likely representing a staining artefact (Figure 32).

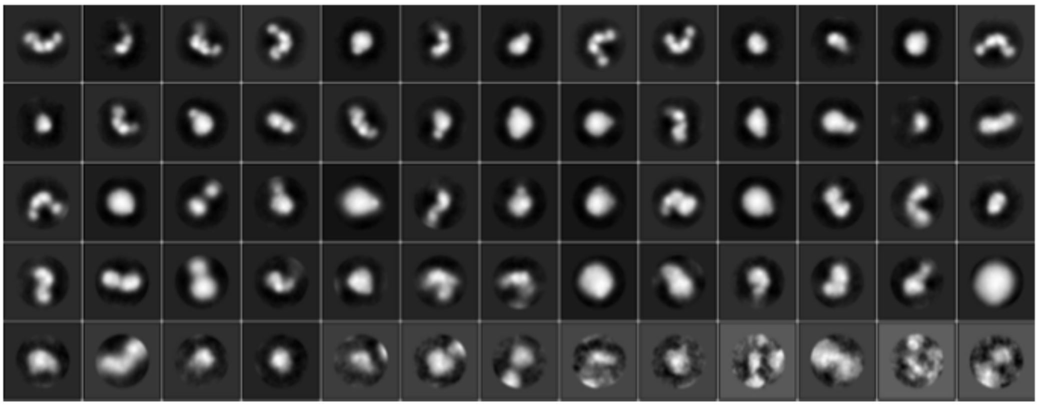


Figure 32: 2D classification of ARISC-RAP80 complex

Reference free 2D classification aligned the extracted particles into class averages. Eight classes displayed the U-like shape characteristic of the ARISC model. Some classes were missing one arm or one MERIT40 subunit, whilst others were significantly dissociated. Class averages on the last row were of mis-aligned particles or non-particles.

3.4.4 3D reconstruction and refinement

2D class averages of the complex in different orientations were selected as input for the generation of an initial model in C1 symmetry. A total of 2,000 particles from 10 classes contributed to the model. The initial model had a curved L like shape as opposed to the U shape seen in the ARISC-AIR data and did not closely resemble the 2D class averages used to generate it (Figure 33A).

After the generation of the initial model, 13 2D classes containing 39,390 particles were selected for 3D classification. 3D classification was then run using the initial model as reference. Four 3D classes were requested to help identify and separate complex heterogeneity. Results from the 3D classification revealed 3D maps that displayed similar topologies (Figure 33B), the 3D maps were asymmetrical, with the fourth map being the most different as it lacked a MERIT40 subunit. Maps two, three and four contained additional densities in the central region of the complex, possibly corresponding to RAP80. Due to the location of the density, it is also possible that they represent a slightly misplaced or flexible BRCC45. The 3D classes closely resembled some of the class averages generated during 2D classification, indicating that they were a true representation of the data. The four 3D maps were individually refined to a resolution of 25-28 Å.

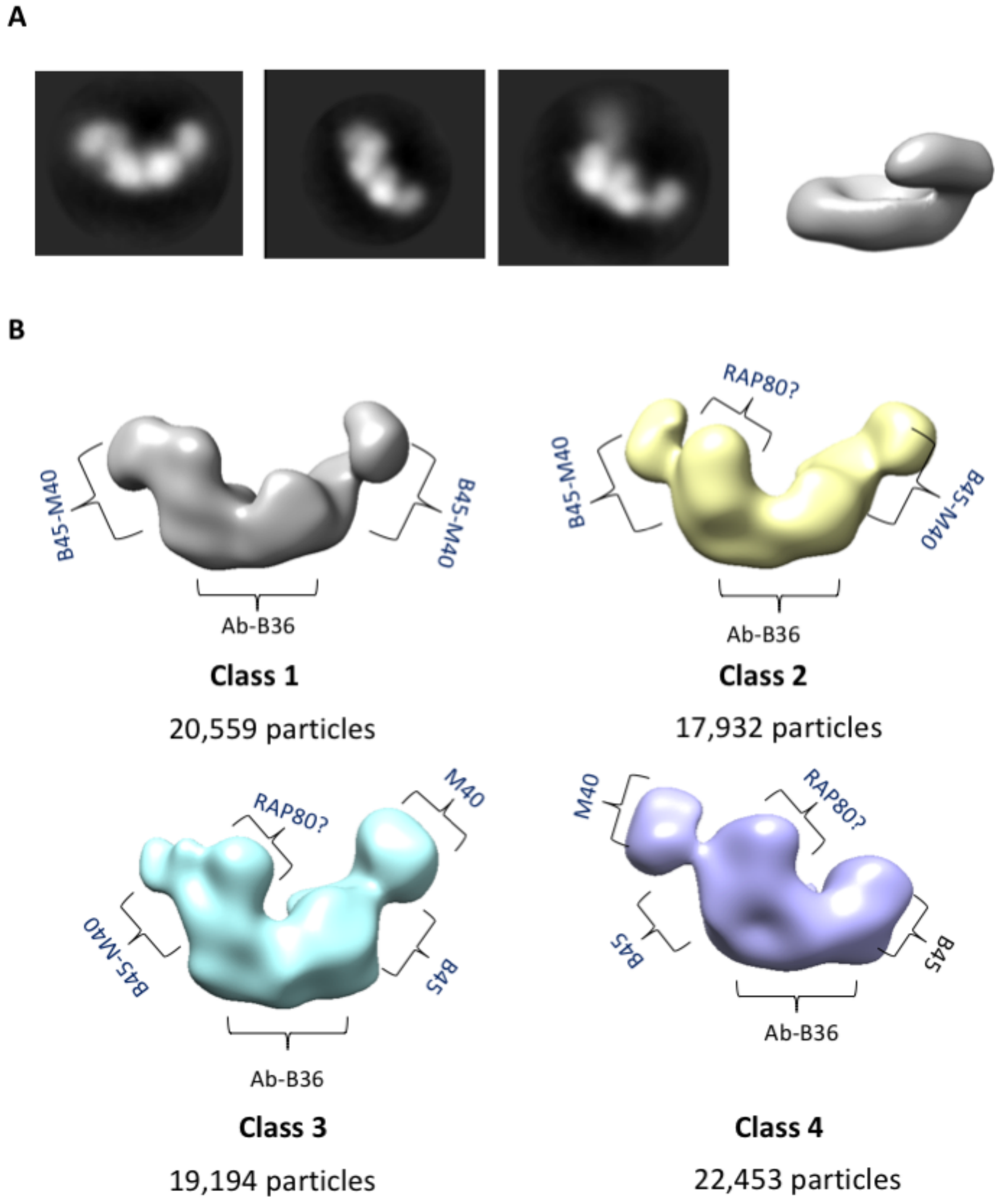
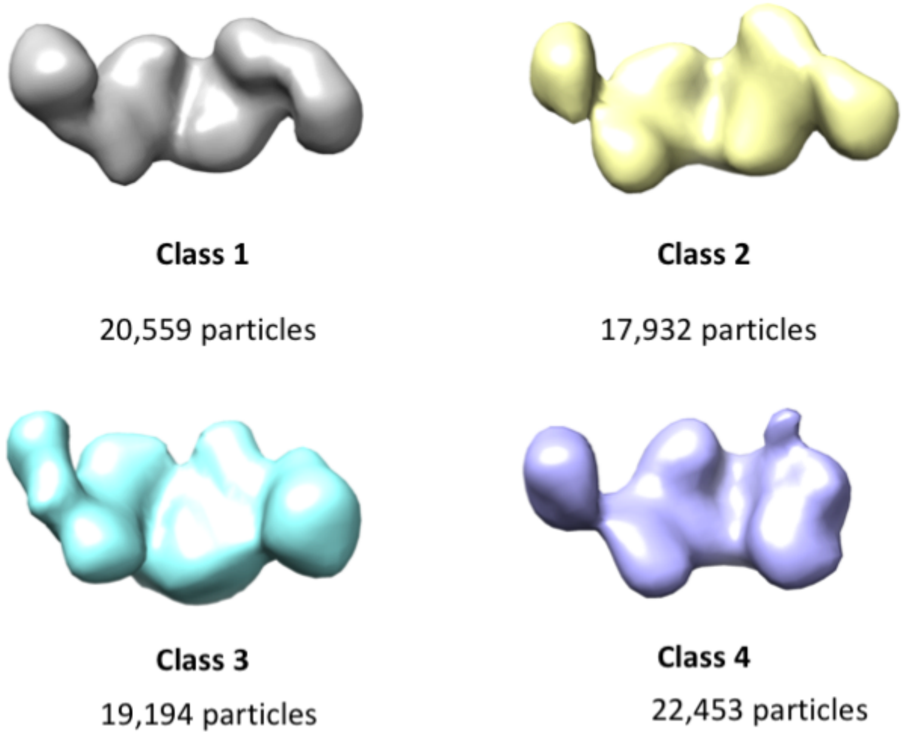


Figure 33: Negative stain EM 3D reconstruction of the ARISC-RAP80 complex

[A] Representative 2D classes selected for the generation of the Initial model and the initial model in C1 symmetry. [B] 3D maps obtained from 3D classification. The maps shared similar topology to one another. The fourth map (purple) was missing a MERIT40 subunit. Classes two (yellow) and three (blue) contained additional density on the left arm near the location of BRCC45. All four maps were asymmetrical. The maps were refined to resolutions of 25 (grey), 25 (yellow), 26 (cyan blue), 28 (purple) aÅ.



Class 1

20,559 particles

Class 2

17,932 particles

Class 3

19,194 particles

Class 4

22,453 particles

Figure 34: Top view of the ARISC-RAP80 complex 3D maps

The top view of the four ARISC-RAP80 3D maps show similarities between classes 1,2 and 3. The right arm of classes 1 and 2 and left arm of class 3 share a similar topology. Class 4 is the most distinct class as it is missing a second arm.

3.4.4.1 ARISC-RAP80 data processing in C2

Data processing of ARISC-RAP80 in C1 revealed a potential extra density indicating the presence of RAP80. Given that the complex, like the BRISC-SHMT2, may have C2 symmetry, the data were also processed with applied C2 symmetry. The same initial model presented in figure 33A and 13 2D classes containing 39, 390 particles used to obtain the C1 3D maps was used, and four classes were requested. Results from the 3D classification revealed models that mirrored results from the ARISC-AIR complex and C1 ARISC-RAP80 3D models (Figure 35). In addition to this, the data identified a potential new RAP80 region interacting with BRCC45, MERIT40 and Abraxas1. The 3D maps were refined individually against themselves (i.e. class 1 was refined against itself) to resolution of 22-25 Å.

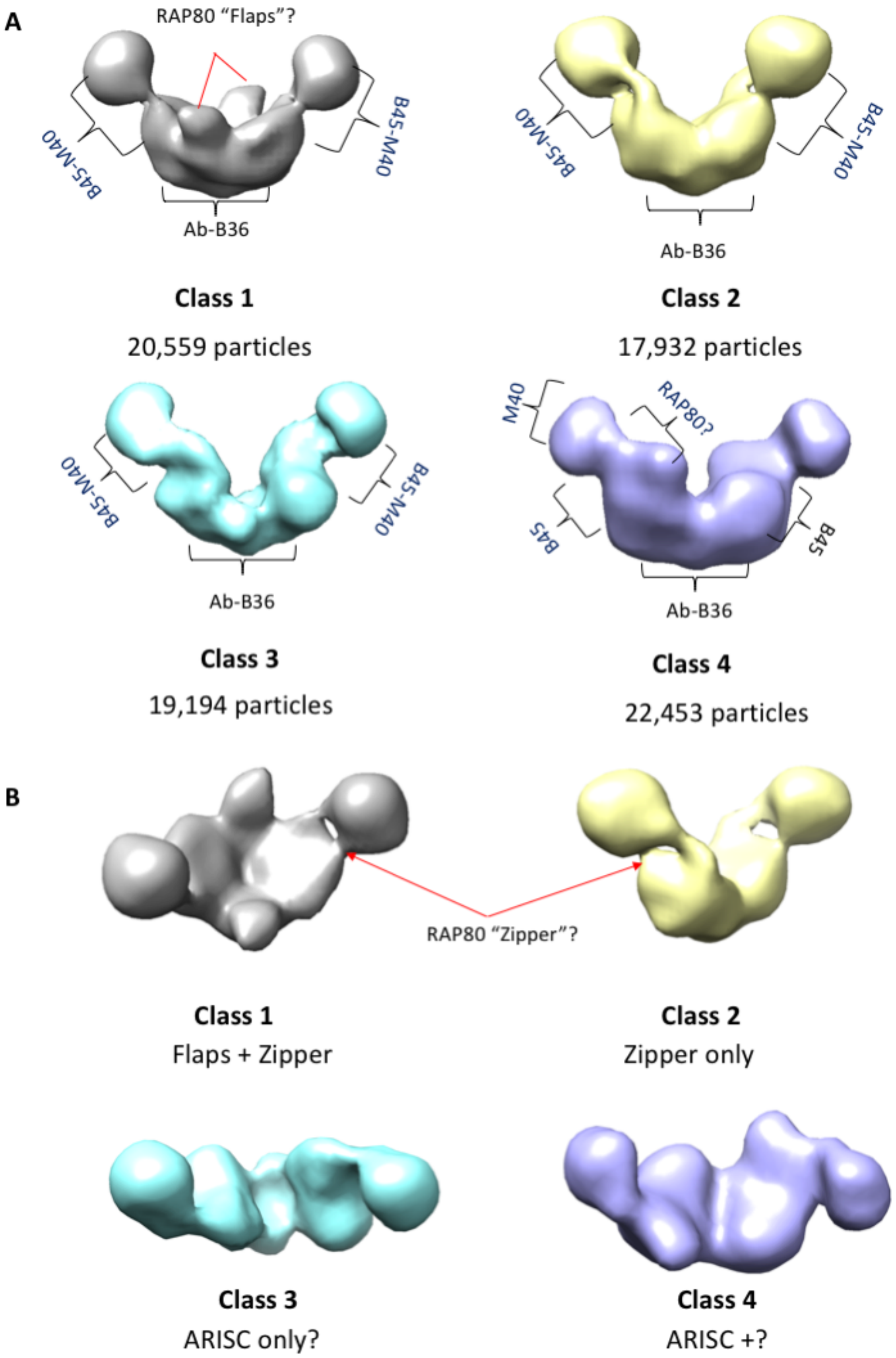
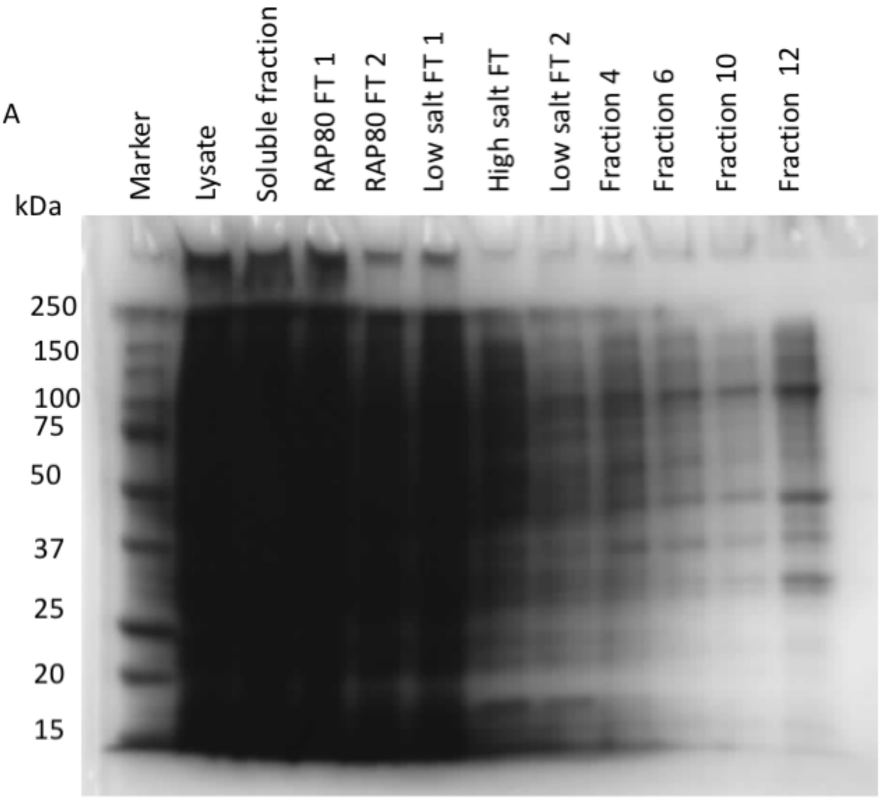


Figure 35: C2 symmetry 3D maps of the ARISC-RAP80 complex

A) Four 3D maps of the ARISC-RAP80 complex with applied C2 symmetry. Classes 1,2 and 4 contain additional densities and structural differences possibly relating to the presence of RAP80. B) Alternative view of the 3D maps. Class 1 contains two flaps protruding from the core of the complex. Both class 1 and 2 share two "zipper regions. The top view of class four is distinct from class three suggesting that they represent different classes

3.5 ARISC-RAP80 fraction 24

Negative stain EM results of the ARISC-RAP80 sample identified a potential ARISC-RAP80 3D map in which all five proteins may be present. Like the ARISC and ARISC-AIR datasets, the ARISC-RAP80 data displayed some structural heterogeneity. To help reduce sample heterogeneity, a new preparation of ARISC-RAP80 complex was purified from frozen Sf9 cell pellets. Taking advantage of a single 6-His-purification tag at the N-terminus of BRCC45, the 5-membered ARISCΔNΔC-RAP80 complex was purified using Ni-affinity chromatography (Figure 36). The 6-His purification tag was removed by addition of tobacco etch protease (TEV) and any noncleaved proteins were separated by Ni-subtraction (Figure 37). Following gel filtration chromatography, the fractions were kept separate, and samples were run on an SDS page gel. Results from the SDS PAGE gel showed that the lane corresponding to fractions 24-26 contained BRCC36, Abraxas, BRCC45, MERIT40 and RAP80. Fraction 24 corresponded to the beginning of the peak on the chromatography spectrum, this suggested that fraction 24 contained an intact ARISC-RAP80 complex (Figure 38).



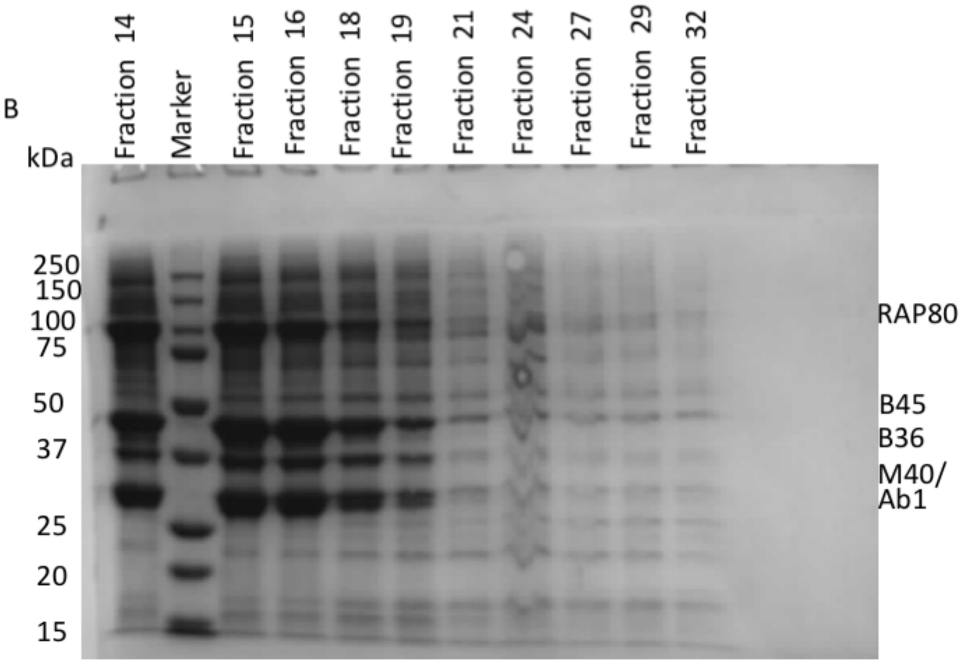


Figure 36: Nickel purification of the ARISC-RAP80 complex

[A] Nickel purification gel containing samples from the cell lysate, soluble fraction, protein flow through and washes. The fractions from washing with low salt and elution buffer was also included. The gel shows that the ARISC-RAP80 is present in fraction 12. [B] Fractions 14-32 of the ARISC-RAP80 nickel purification. The gels show that ARISC-RAP80 is present in fractions 12-19.

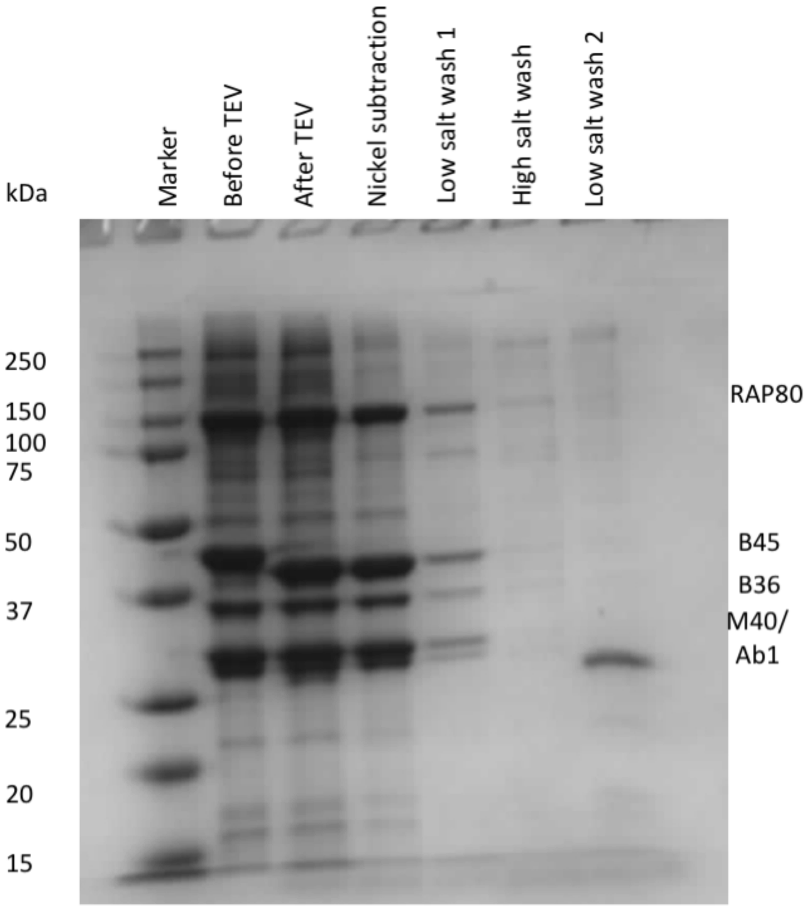


Figure 37: Nickel subtraction to remove TEV from protein sample

TEV was added to fraction 14-20 and the sample was dialysed overnight. Following dialysis, the sample underwent nickel subtraction to remove TEV and samples were run on an SDS PAGE gel. The gel shows that ARISC-RAP80 is present in the nickel subtraction flow through and low salt wash 1 flow through. There is also a clear shift in the molecular weight of the band corresponding to BRCC45 after TEV cleavage, indicating that TEV has cleaved the His tag off the protein.

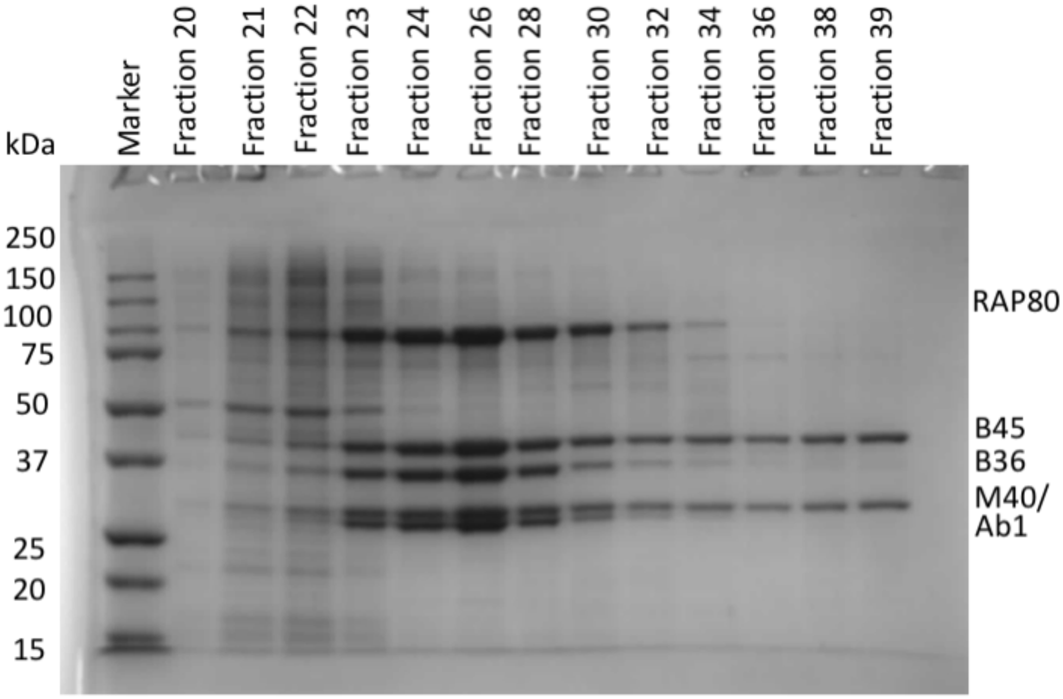


Figure 38: Gel filtration of the ARISC-RAP80 complex

The gel shows that Fractions 20-22 contain some contaminants in addition to Abraxas 1(30 kDa), MERIT40 (30 kDa), BRCC36 (36 kDa), BRCC45 (45 kDa) and RAP80 (80 kDa). Fraction 24-32 are the cleanest fractions. The BRCC45-MERIT40 heterodimer is overexpressed in insect cells and are present as a dimer in fractions 36-39.

3.5.1 Grid preparation and data collection

0.008 mg/ml of fraction 24 ARISC-RAP80 was deposited on glow discharged carbon coated TEM grids. The grids were imaged on the F20 and 214 micrographs were collected at a magnification of 29,000x, a KV of 200 and a defocus range of -1 to -5 μm. The dataset was processed using RELION 3.0 beta (Scheres, 2012; Zivanov et al., 2018).

3.5.3 Micrograph analysis and particle picking

Visual evaluation of the micrographs revealed that the particles displayed the distinct U shape characteristic of ARISC-RAP80 complexes (Figure 39A) and aggregation levels were visibly lower than datasets from concentrated and stored protein preparations. Approximately 1,000 particles were manually picked, extracted and aligned to create 2D templates (Figure 39B). Suitable 2D class averages were selected and used as templates to auto-pick 80,947 particles. The particles were extracted using a mask size of 300 Å and subject to reference free 2D classification.

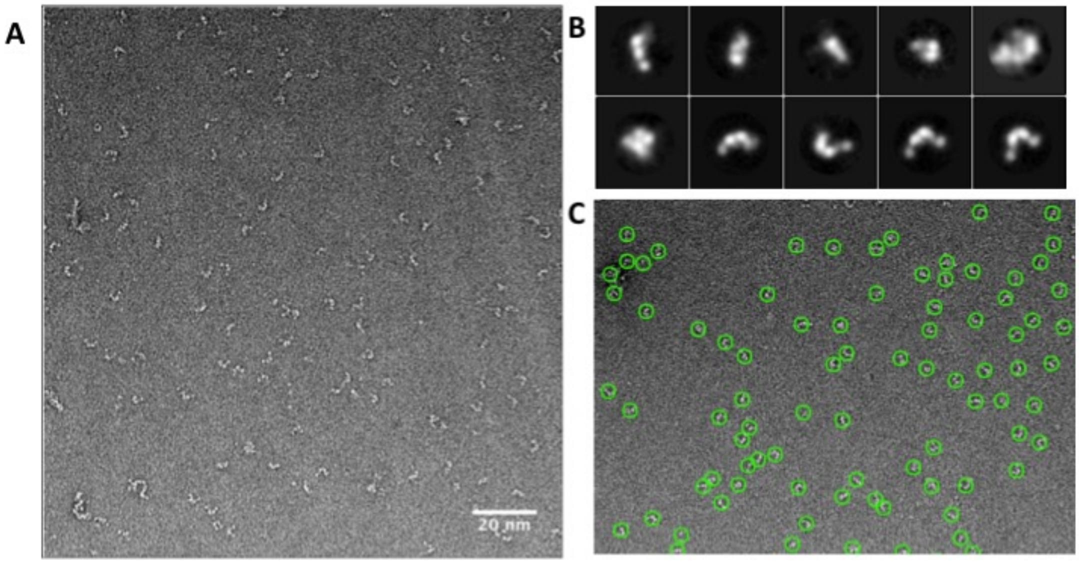


Figure 39: Pre-processing of ARISC-RAP80 fraction 24 negative stain EM data

[A] Representative micrograph of the ARISC-RAP80 fraction 24 dataset. [B] 2D templates used for auto-picking. [C] A representative auto-picked micrographs.

3.5.4 2D classification

2D classification revealed that the majority of the classes were well aligned and intact. Twenty five class averages displayed the U shape expected of ARISC. The remaining classes were either of ARISC-RAP80 subcomplexes or of the complex in a different view and orientation. The dataset also contained some fuzzy classes that likely contained mis-aligned particles, aggregates or non-particles (Figure 40).

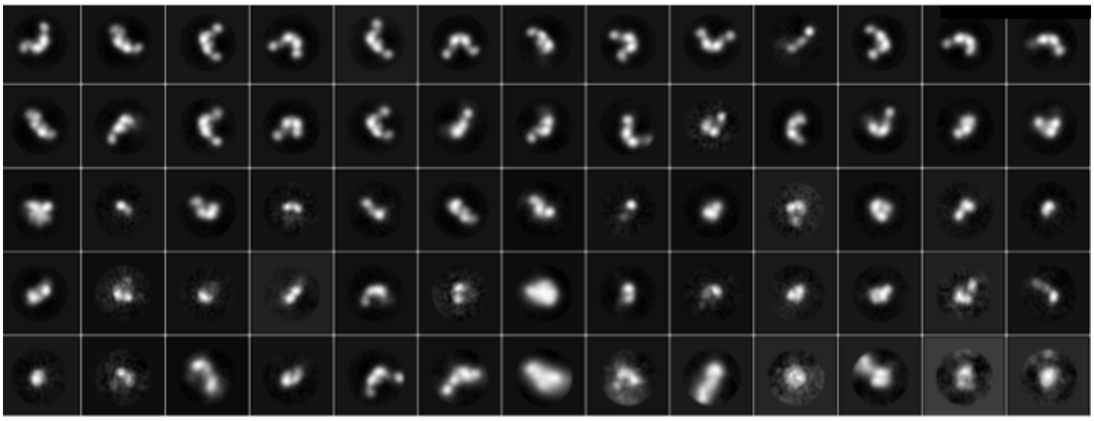


Figure 40: 2D classification of ARISC-RAP80 fraction 24 complex

Reference free 2D classification aligned the extracted particles into 100 class averages (65 represented here). Twenty five classes displayed shapes similar to the ARISC model. MERIT40 in some class averages had dissociated, and other classes appeared to contain only two subunits.

3.5.5 3D reconstruction and refinement

Seventeen classes containing 3,400 particles (particles per class were limited to 200) were selected to generate an initial model in C1 symmetry. The initial model contained two arms but was asymmetric, reflecting some of the 2D classes (Figure 41A). To obtain a more accurate 3D map, 31 suitable classes containing 53,217 particles were selected and subject to 3D classification using the initial model as reference. For the selection of classes for 3D classification, the number of particles per class was not restricted. Results from the 3D classification (Figure 41B) revealed that the data was more homogenous compared to the ARISC-RAP80 negative stain EM data as three out of the four classes were structurally similar. The first class was distinct from the other classes as it appeared to be missing an ‘arm’ (a BRCC45-MERIT40 heterodimer). As classes two and four were very similar to one another, particles contributing to those classes were combined and were refined using the initial model as reference to a resolution of 25 Å. Class one and three were refined separately to a resolution of 28 Å. The initial model was not used for the refinement of class one as they were structurally different to one another, instead the map was refined against itself.

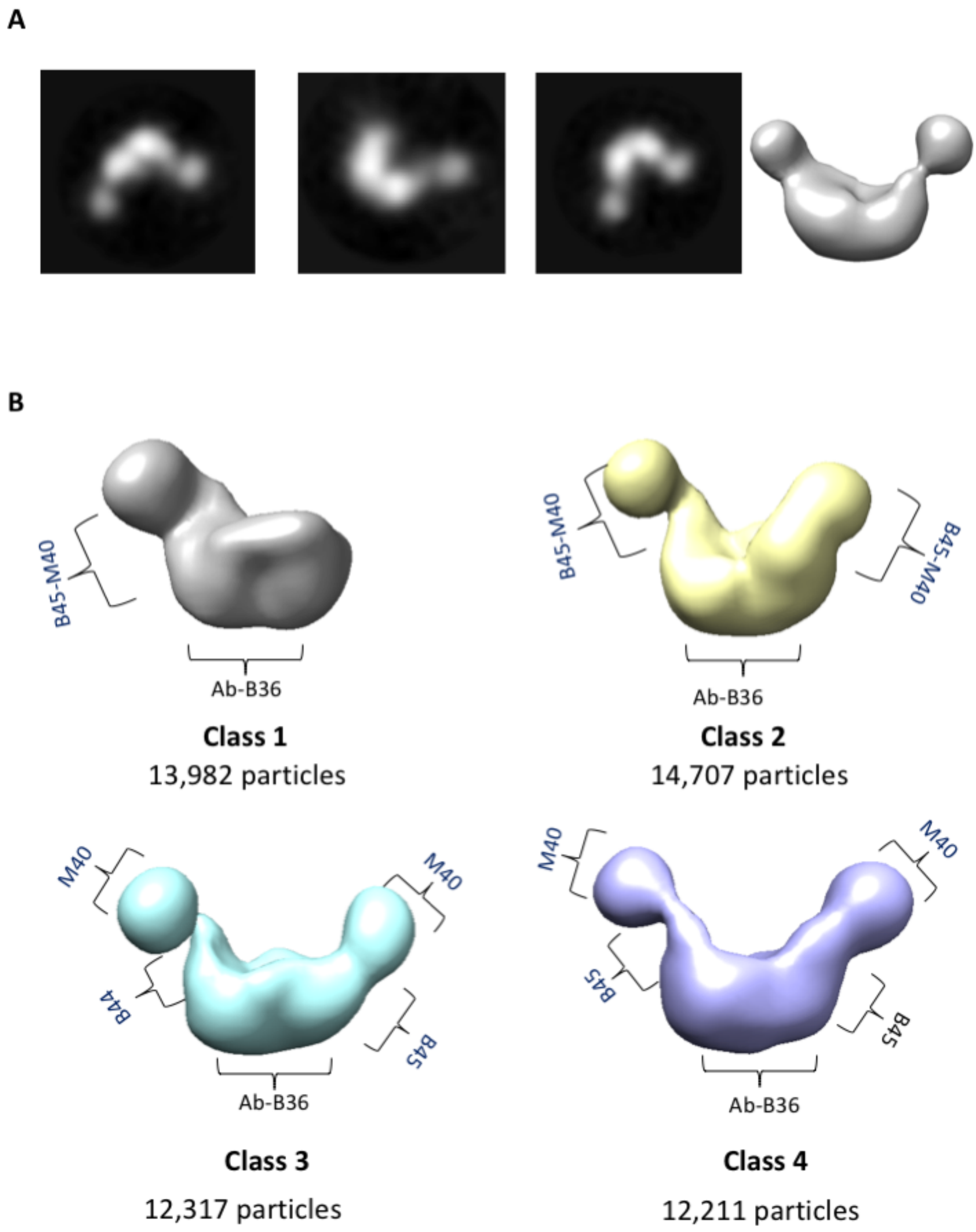


Figure 41: Negative stain EM 3D reconstruction of fraction 24 ARISC-RAP80 complex

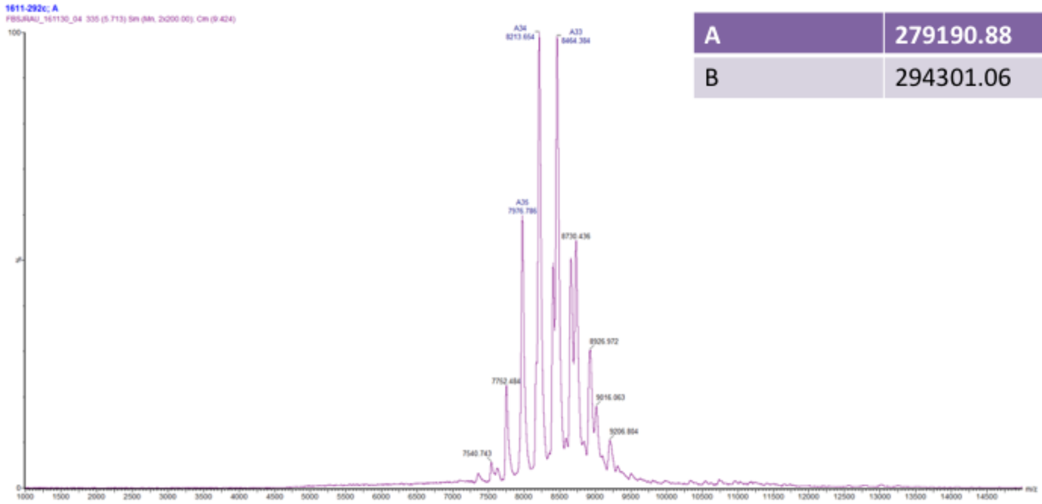
[A] Representative 2D classes selected for the generation of the Initial model and the initial model in C1 symmetry. [B] 3D maps obtained from 3D classification. Maps two, three and four shared similar topologies to one another and the first map was missing a BRCC45-MERIT40 heterodimer. Class one (grey map) was refined against itself to a resolution of 28 Å. Particles contributing to classes two and four were combined and refined to a resolution of 25 Å. Class three was refined to a resolution of 28 Å.

3.6 Mass spectrometry analysis of the ARISC complexes

Negative stain EM analysis of the ARISC complexes provided low resolution structural information of the complexes, however, due to the heterogenous nature of the complexes and the limited resolution of the technique, the stoichiometry of the ARISC-AIR and ARISC-RAP80 complex was difficult to determine. To this end, MS was utilised to identify the stoichiometry of both complexes. Prior to analysis, both samples were diluted to 2 mg/ml using gel filtration buffer and buffer exchanged. The reader is referred to chapter 2 for detailed breakdown of the method.

3.6.1 ARISC-AIR complex

Native MS analysis of the AIRSC-AIR complex produced a mass spectrum that revealed two main species at 279 and 294 kDa (Figure 42). The two molecular weights of the species reflect a stoichiometry of 2:2:2:2 and 2:2:2:2:1. These results suggest that ARISC and ARISC-AIR has been detected.



A	279190.88
B	294301.06

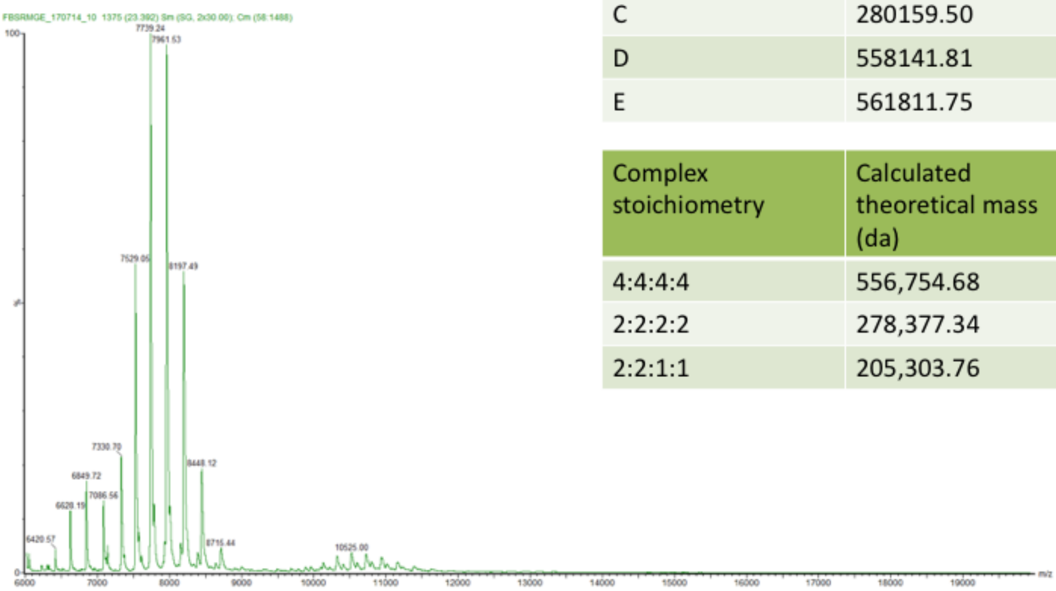
Complex stoichiometry	Calculated theoretical mass (Da)
2:2:2:2:2	308,362.80
2:2:2:2:1	293,352.06
2:2:2:2	278,377.34

Figure 42: Native mass spectrometry analysis of ARISC-AIR complex

The analysis of ARISC-AIR revealed two main species with a molecular weight of 279 and 294 kDa. The two molecular weights correspond to the 2:2:2:2 ARISC complex and 2:2:2:2:1 ARISC-AIR complex

3.6.2 ARISC-RAP80 complex

MS analysis of the ARISC-RAP80 complex revealed two main species at 278 and 280 kDa, indicating a stoichiometry of 2:2:2:2. Interestingly, two additional low abundance species were identified whose molecular weight were double that of the main species (558 and 561 kDa), and corresponded to the expected molecular weight of the ARISC dimer of octamer complex. A smaller species with a mass of 205.540 kDa was also detected possibly corresponding to ARISC one arm complex (Figure 43).



A	205540.39
B	278746.53
C	280159.50
D	558141.81
E	561811.75

Complex stoichiometry	Calculated theoretical mass (da)
4:4:4:4	556,754.68
2:2:2:2	278,377.34
2:2:1:1	205,303.76

Figure 43: Native mass spectrometry analysis of the ARISC-RAP80 complex

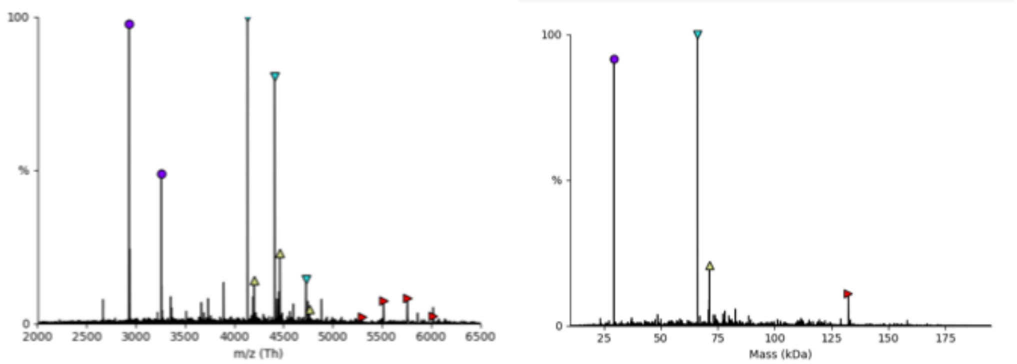
Two high abundance species (A and B) were identified with molecular weights of 278 and 280 kDa. The two species correspond to the stoichiometry of 2:2:2:2 indicating that the two species are the ARISC complex. A smaller species with a molecular weight of 205 kDa was detected. 205 kDa correspond to the ARISC 'one arm' complex. Two lower abundance species were identified (D and E), whose molecular weight corresponds to the expected molecular weight of the ARISC dimer of octamers (4:4:4:4).

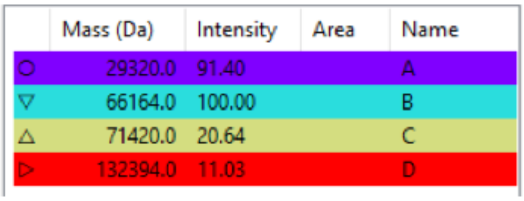
3.6.3 MS analysis of the ARISC-RAP80 complex using the Orbitrap

To try and obtain a more conclusive mass spectrum and to identify the RAP80 protein, the ARISC-RAP80 complex was further analysed using the Thermo Scientific Q Exactive UHMR Orbitrap mass spectrometer. The Orbitrap is an ion trap mass analyzer that consists of two outer electrodes and a central electrode, which enable it to act as both an analyzer and detector. Ions entering the Orbitrap are captured through "electrodynamic squeezing," after which they oscillate around the central electrode and in between the two outer electrodes. Different ions oscillate at different frequencies, resulting in their separation. By measuring the oscillation frequencies induced by ions on the outer electrodes, the mass spectra of the ions are acquired using image current detection.

ARISC-RAP80 was introduced into the instrument by direct infusion nanoESI through an in-house prepared gold/palladium-coated borosilicate glass capillary at 1 µM concentration in 1M ammonium acetate. Full scan data were acquired at a resolving power of 12500 in the m/z range 2,000-15,000. The mass spectra presented in the results section are comprised of approximately 150 scans. Data processing was performed with the use of the UniDec Software program.

The data was processed by analysing the low molecular weight and high molecular weight ranges. The processing of the low molecular range identified species corresponding to MERI40, Abraxas1-BRCC36 dimer, Abraxas1-BRCC36 dimer of dimers and an unknown species likely to be the BRCC45-MERIT40 dimer (Figure 44). The processing of the high molecular weight region revealed three mains species with molecular weights of 206 kDa, 279 kDA, and 360 kDa. These masses reflect the masses of MERI40-BRCC45-Abraxas1 dimer, MERIT40-BRCC45-Abraxas1-BRCC36 dimer and the MERIT40-BRCC45-Abraxas1-BRCC36 dimer and a single RAP80 protein (Figure 45).



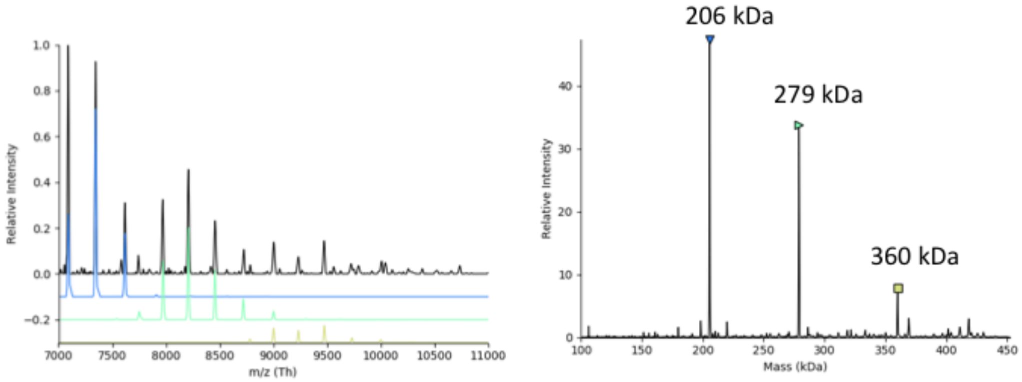


	Mass (Da)	Intensity	Area	Name
○	29320.0	91.40		A
▽	66164.0	100.00		B
△	71420.0	20.64		C
▷	132394.0	11.03		D

A = MERIT40
B = ABRAXAS1 /BRCC36
C = ?
D = (ABRAXAS1 /BRCC36)*2

Figure 44: Low molecular weight processing

Low molecular weight processing of the data revealed subcomplexes of the ARISC-RAP80 complex.



206 kDa = (MERIT40/ABRAXAS1/BRCC45)*2

279 kDa = (MERIT40/ABRAXAS1/BRCC45/BRCC36)*2

360 kDa = (MERIT40/ABRAXAS1/BRCC45/BRCC36)*2 + RAP80*1

Figure 45: High molecular weight processing

High molecular processing of the data showed three main species whose molecular weight corresponded to the ARISC complex bound to a single RAP80 protein, the ARISC complex and the Abraxas-BRCC45-MERIT40 dimer.

3.6.4. Fluorescence activity assay

IQF-di-ubiquitin (LifeSensors) is a substrate used for continuous fluorescent measurement of isopeptidase activity in which the C-terminus of ubiquitin is conjugated via an isopeptide bond to K63 of a second ubiquitin moiety. The resultant di-ubiquitin forms an internally quenched fluorescent FRET pair (IQF). Each ubiquitin molecule is labelled with a single fluorescent dye or a highly efficient quenching dye. The cleavage of the IQF di-ubiquitin by deubiquitylating enzymes result in the separation of the fluorophore from the quencher and an increase in fluorescence. This section of the chapter compares the enzyme activity of the ARISC-AIR and ARISC complex. The aim was to determine whether AIR enhances the complex's activity.

3.6.4.1 ARISC-AIR titration

Three titration series of ARISC-AIR activity assays were carried out using enzyme concentrations of 2.5 nM, 5 nM, 10 nM, 20 nM and 40 nM to determine the optimal concentration for observing enzyme activity. These results show that within the first 20 minutes, the 20 and 40 nM reactions reached plateau and that the 2.5 and 5 nM reactions were significantly slower, suggesting that 10 nM was the optimum concentration for observing the progress of the assay (Figure 46).

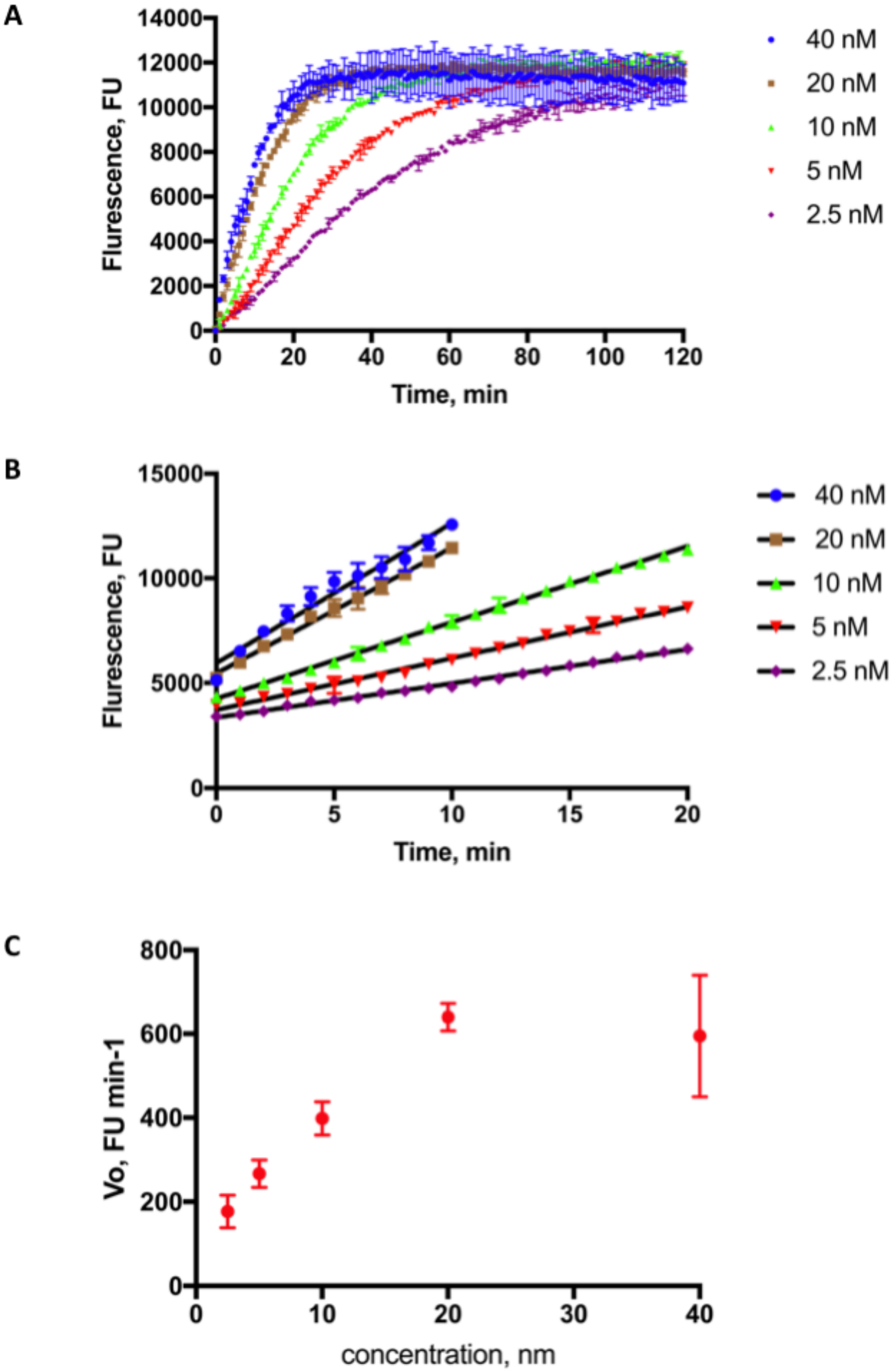


Figure 46: Titration activity assay of ARISC-AIR and IQF di-ubiquitin

[A] Representative titration assay results. [B] Representative graph of initial velocity. [C] The average initial velocity from the three repeats.

3.6.4.2 Comparison of ARISC and ARISC-AIR activity

After determining that 10 nM was the optimum concentration for observing the activity assay. The activity of the ARISC and ARISC-AIR complex was observed in parallel. Results of the fluorescence IQF di-ubiquitin activity assay revealed that ARISC-AIR displays a slightly higher activity than ARISC (Figure 47). The enzyme-substrate reaction of both complexes occurred between timepoint zero and approximately 50 minutes, after which the reaction plateaued.

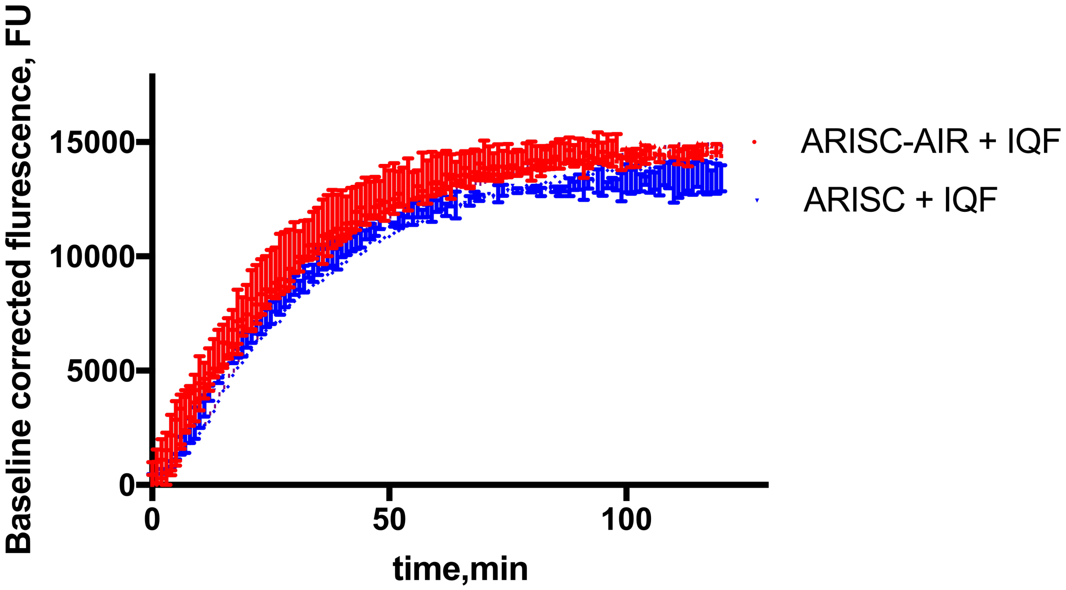


Figure 47: Fluorescence activity assay of the ARISC and ARISC-AIR complexes

Results from the activity assay demonstrates that the activity of the ARISC-AIR complex against the IQF di-ubiquitin substrate is slightly higher than that of the ARISC complex.

3.7 Discussion

The aim of this chapter was to obtain low resolution structures of the ARISC, ARISC-AIR and ARISC-RAP80 complexes by negative stain EM, to uncover the stoichiometry of the ARISC-AIR and ARISC-RAP80 complexes and investigate whether the targeting protein enhances ARISC's enzyme activity.

3.7.1 ARISC complex

The micrographs in this dataset contained particles with poor contrast, likely due to poor staining during grid preparation. Poor particle staining affects the quality of the dataset and will influence the final resolution of the 3D reconstructed maps. Despite the issues with particle contrast, approximately 1,000 particles were manually and randomly selected to generate 2D class averages that were used as templates for automated picking (auto-picking).

The negative stain ARISC 3D maps presented in this chapter showed some structural similarities with the published ARISC model (Kyrieleis et al., 2016). However, MERIT40 in the first 3D map appeared to point in a different direction. The differences between the map and the model can be attributed to the fact that in order to obtain the model from Kyrieleis et al.,2016 the ARISC complex was cross-linked with glutaraldehyde and the sample formed a symmetric dimer of octamers (two ARISC complexes; figure 20). Cross-linking proteins can result in the formation of non-native interactions. The differences observed between the ARISC 3D map presented in this study and the published model could be due to the fact that the later was obtained by crossing-linking. It is also possible that due to the complex being susceptible to dissociation, the ARISC "relaxed" conformation seen here was the result of complex flexibility as the molecules were dissociating. However, these multiple structural models showing flexibility of the ARISC arms and multiple conformations, may be the biologically relevant form of ARISC, and this could be stabilised in the presence of adaptor proteins such as RAP80 or substrate proteins like poly-ubiquitin chains or both.

Interestingly, class 4 obtained from 3D classification resembles the symmetrical ARISC dimer of dimer seen by Kyrieleis et al 2016. It is possible that a subpopulation of ARISC forms a dimer with itself and that the complex is further stabilising by cross-linking. However, in this dataset Class 4 contained the highest number of particles suggesting that a good majority of particles were that of the ARISC dimer.

Multiple attempts were made to obtain EM data of a stable ARISC complex. However, each attempt resulted in either a dataset in which the complex had completely dissociated as the result of over diluting the sample or in a dataset not suitable for single particle analysis as the micrographs contained too many protein particles, some of which were aggregated. This led to the assumption that the complex may exist at an equilibrium in which over diluting it destabilises the interactions between some of the subunits. Since these maps were at really low resolution (35-45 Å), it was difficult to determine which subunits had dissociated from the complex.

It is apparent from the dataset that ARISC is heterogeneous. Whether this heterogeneity is the result of sample preparation and image processing or a true representation of the sample in biology is difficult to conclude. Biophysical and biochemical analysis of the sample such as MS and native page gel would provide further insights onto the native stoichiometry of ARISC.

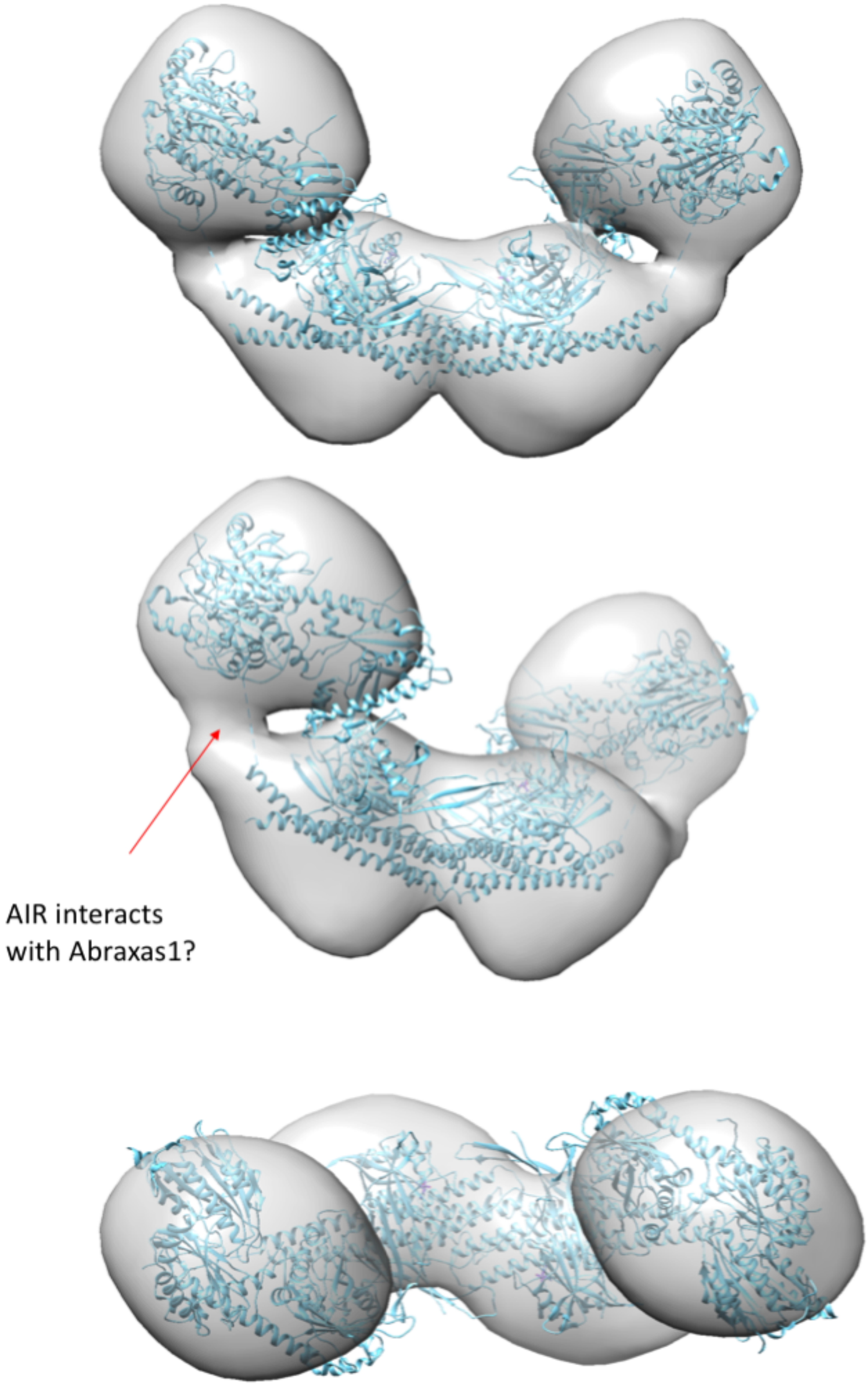
3.7.2 ARISC- AIR complex

Visual observation of the micrographs suggests that the ARISC-AIR complex is more stable than the ARISC complex as some particles exhibited a U-shape characteristic of ARISC complexes. Since similar grid preparation, protein concentrations and data collection methods were used to collect the ARISC and ARISC-AIR datasets, it is possible that the interaction of AIR with ARISC helps to stabilise the complex. Unlike the 2D classes of the ARISC dataset, the ARISC-AIR class averages contained non-dissociated particles, further supporting the hypothesis that the presence of AIR helps to stabilise the complex. However, due to the small nature of AIR, it is difficult to assess whether AIR remained bound to Abraxas1 or to precisely locate its position. Like the ARISC 2D classes, some of the ARISC-AIR class averages were missing a single BRCC45-MERIT40 arm. On the other hand, a subset of class averages contained both arms. Some of the two arm class averages had one arm that was fuzzy and less visible. This was likely the result of particle misalignment in the arm regions of the complex as a result of the way the particle is orientated on the grid. For example, if one arm is slightly more buried or embedded on the grid, and the other one is not, the more visible arm will be better coated with stain, will generate better signal, and will be easier to align.

The topology of the 3D classes reflected the characteristic U-shape of the published ARISC model. Class two and four were the most structurally similar 3D maps, differing only in small details at regions corresponding to the Abraxas1-BRCC36 dimer of heterodimers and possibly AIR. Both 3D maps contained a significant number of particles. As the differences were found in a similar region of both 3D maps, it is possible that RELION (Scheres, 2012; Zivanov et al., 2018) separated particles that otherwise would have aligned well in a single 3D map. Due to the limited resolution of negative stain EM, it is difficult to determine the significance of the difference observed between class two and four.

AIR is a 15 kDa protein. If the stoichiometry of the ARISC-AIR complex is 2:2:2:2:2, the density of AIR will should be visible on the complex. It is therefore possible that class three represents the ARISC-AIR complex. However, the size of the densities at the core of the complex is larger than what would be expected of a two 15 kDa proteins, this suggests that multiple AIR proteins could be present in the map. If the stoichiometry of the ARISC-AIR complex is 2:2:2:2:1, AIR would be too small to accurately locate on the complex. Map three could also be explained by assuming that the map is an artefact. However, this is unlikely because there are a large number of particles contributing to the map, meaning that is a representation of a significant subset of the data.

During the writing of this thesis, a crystal structure of the ARISC-AIR complex from mouse was published (Rabl et al., 2019). The structure suggests that the AIR interacts with MERIT40, BRCC45 and Abraxas1, inducing a conformational change in the arms of the complex. The topology of the published structure resembles class one from 3D classification as density is seen bridging MERIT40 and the core complex. Fitting class one and the ARISC-AIR structure revealed a generally good fit (Figure 48). The map obtained in this data was significantly larger than the published structure, however, this can be accounted for by the difference in resolution (3.75 Å vs 25Å and possibly due to the fact that human ARISC-AIR was used instead of mouse. In the published structure, AIR is shown to interact with Abraxas1, this is demonstrated in the structure as dotted connecting lines (Rabl et al., 2019). The map of class one surprisingly reflects this, suggesting that the map may contain the AIR protein.



AIR interacts with Abraxas1?

Figure 48: class one ARISC-AIR map fitted with published ARISC-AIR structure

The fit shows similarities between the ARISC-AIR C2 symmetry map presented in this chapter and the published structure of the mouse ARISC-AIR complex (Rabl et al., 2019) PBD: 6GVW.

To conclude, the overall topology of the 2D classes and 3D maps of ARISC-AIR was in accordance to the expected topology of ARISC (Kyrieleis et al., 2016), validating the data. The increased stability of the complex suggests that the binding of AIR to ARISC facilitates complex integrity. The data, however, does not confirm the stoichiometry of the ARISC-AIR complex, high resolution data will be required to correctly identify the location of AIR.

3.7.3 ARISC-RAP80 complex

The ARISC-RAP80 3D maps were structurally distinct from the ARISC-AIR and ARISC. The maps contained additional density at the core of the complex, possibly identifying RAP80. As the location of the additional densities differed in each of the ARISC-RAP80 maps, it is possible that the complex exists in different conformations, and that this dataset caught snap shots of each state.

The maps obtained from processing the data in C1 symmetry were similar to one another in that they contained additional densities predominantly in one region of the map- the left arm. This suggests that the maps likely contain a single RAP80 protein supporting the mass spectrometry data presented in 3.6.3. The application of symmetry (C2) surprisingly generated more structurally diverse maps. Class one contained two protruding densities at the centre of the complex, possibly identifying a flexible region of the RAP80 protein. Additionally, both class one and two contained two "zipper" regions in the arms of the complex in which the MERIT40 region interacts with the core complex. This zipper region was also identified in the ARISC-AIR data presented in this chapter. This raises the possibility that RAP80 binds to the complex in a way in which AIR interacts with MERIT40, BRCC45 and Abraxas1. This hypothesis is supported by the published ARISC-AIR mouse structure (Rabl et al., 2019).

Class three obtained from re-processing the ARISC-RAP80 data in C2 appears to represent the ARISC complex. MERIT40 in this map is positioned differently to MERIT40 class one of the ARISC 3D data. The differences in the positioning of the subunit could be due to complex flexibility. It's likely that if the complex is very flexible that different snapshots are captured during the staining and particle embedding process. Class four strongly resembled the 3D maps obtained from processing the data in C1 symmetry. These results imply that applying symmetry helped to further separate the particles into suitable classes. A caveat of applying symmetry is that it could result in the inaccurate description of a complex's stoichiometry.

3.7.4 ARISC-RAP80 fraction 24

The negative stain EM data of all three complexes (ARISC, ARISC-AIR, and ARISC-RAP80) displayed varying degrees of heterogeneity. It is uncertain whether this was due to the protein sample or was the result of the negative staining grid preparation, or both. To help reduce sample heterogeneity, additional ARISC-RAP80 complex was purified. Following size exclusion chromatography, the samples were separated based on the analysis of the chromatography spectra and SDS-PAGE gels. Fraction 24 was believed to contain intact ARISC-RAP80 complex as this fraction corresponded to the beginning of the gel filtration elution peak, suggesting that the protein in the sample had a higher molecular weight.

Visual observation of the micrographs revealed that they contained particles with good contrast to the background, and that unlike the ARISC-RAP80 data, the particles displayed distinct individual U shapes and less aggregation. 2D classification of the ARISC-RAP80 fraction 24 data contained more intact class averages than the ARISC-RAP80 dataset. However, there were still some visible conformational heterogeneity, and some of the particles displayed relaxed U shapes, whilst others were more compact. These differences could be due the complex being flexible and dynamic, and staining them with heavy metal solution resulted in snap shots of the different conformational states.

Freshly purifying the complex and separating the fractions successfully helped to reduce complex heterogeneity and aggregation. However, due to the maps being low resolution, it is difficult to accurately place RAP80 in the ARISC complex. What is clear though is that the ARISC-RAP80 complex when compared to the ARISC model is asymmetric, more relaxed and contains densities that are unaccounted for. A high resolution structure of the ARISC-RAP80 complex is required to accurately determine the location of RAP80 and the stoichiometry of the complex. The homogenous nature of ARISC-RAP80 fraction 24 made it a suitable contender for future cryo-EM experiments.

3.7.5 Mass spectrometry analysis of the ARISC complexes

MS analysis of the ARISC-AIR and ARISC-RAP80 complex was carried out to elucidate the stoichiometry of both complexes and to determine whether the complexes existed in various subcomplexes.

Analysis of the ARISC-AIR complex revealed two main species whose molecular weight corresponded to a stoichiometry of 2:2:2:2 and 2:2:2:2:1, indicating that the sample contained a mixture of the four membered ARISC complex and the ARISC complex bound with a single AIR protein (ARISC-AIR). These results reflect the ARISC-AIR data which generated 3D maps of the ARISC complex and ARISC + additional densities, possibly corresponding to the AIR protein. The 2:2:2:2:1 stoichiometry, however, doesn’t reflect the stoichiometry information of the complex obtained from a recent mouse ARSIC-AIR structure which showed two AIR protein bound to the ARISC complex (Rabl et al., 2019). It is possible that under the mass spectrometry conditions, AIR dissociates from the complex, explaining why the data generated masses the correlates to ARISC and ARISC + 1 AIR protein.

Native mass spectrometry analysis of the ARISC-RAP80 complex was unable to identify RAP80 bound to the complex. However, the analysis identified two main

ARISC species with a molecular weight of 278 and 280 kDa and two species with a lower abundance whose molecular weight was double that of the main species (558 and 561 kDa). The masses of the two lower abundance species suggest a complex stoichiometry of 4:4:4:4, indicating that the ARISC complex has formed a dimer with itself. The identification of the ARISC dimer supports the finding of Kyrieleis et al., 2016 where they reported that the ARISC complex self-dimerises and class four of the ARISC 3D classification maps presented in this chapter.

It is possible that RAP80-AIR and full length RAP80 dissociates from the complex under the native mass spectrometry conditions. The preservation of native structures of protein and non-covalent interaction during ionisation has been a challenge in biological MS. Prior to being subject to analysis, the samples were buffer exchanged into ammonium acetate, a volatile salt. Volatile salts are used in mass spectrometry because they prevent the formation of abundant salt clusters and non-specific protein adducts. However, a drawback of using volatile salts like ammonium acetate is that they have poor buffering capabilities compared to non-volatile salts such as tris, phosphate and phosphate-buffered saline. Due to the presence of species with different molecular weight, it is possible that two AIR and RAP80 proteins bind to ARISC and that under mass spectrometry conditions one is dissociating. Altering the buffer conditions may help to further stabilise the complex.

To try and detect the RAP80 protein in the sample, ARISC-RAP80 was re-analysed using the Orbitrap. The output of the analysis revealed that the proteins that make up ARISC-RAP80 also exist in subcomplexes. It also helps to paint a picture of how the complex assembles. For example, the Abraxas1-BRCC36 dimer was detected and so was the Abraxas1-BRCC36 dimer of dimers (super dimer). This suggests that the formation of the super dimer occurs in stages. In addition to this, Abraxas1 was found to form a hexamer with BRCC45 and MERIT40 independent of BRCC36, possibly identifying a novel structure with unknown, non-catalytic functions.

The orbitrap was able to identify a species whose molecular weight corresponded to the ARISC-RAP80 complex with a stoichiometry of 2:2:2:2:1, similar to the native mass spectrometry analysis of ARISC-AIR complex. Higher resolution structural data will be required to accurately determine the stoichiometry of the Human ARISC-AIR and ARISC-RAP80 complexes.

3.7.6 conformational and complex heterogeneity

The study of the ARISC complexes by negative stain EM and mass spectrometry gave rise to EM maps that displayed various conformations and mass spectrums with multiple high abundance species. With negative stain EM resolution, it is not possible to determine whether the various different complex conformations are due to sample dissociation or whether it relates to complex assembly and possible allosteric regulation of the complex. Previous structural studies have revealed significant conformational changes are mostly associated with substrate (ubiquitin)- induced fit, instead of allosteric (in)activation by a regulatory protein (Köhler et al., 2010; Avvakumov et al., 2006; Hu, Min et al., 2002). However, as the negative stain EM data of the ARISC complexes presented in this thesis are not in conjugation with its substrate, it is possible that the conformation changes captured as snap shot are related to the assembly and function of the complex. This is could particularly be the case when comparing the stability of the four membered ARISC complex and the five memebered ARISC-RAP80 complex. The negative stain EM maps of the complexes suggest that the presence of RAP80 facilitates in the stabilisation of the complex. The poor ARISC complex integrity seen in this project was also observed by Kyrieleis et al (2016). When studying the complex by negative stain EM, the group observed that ARISC was prone to dissociation. To overcome this problem the group used GRAFIX cross-linking and successfully stabilised the complex by forming a dimer of octamers. The study of the heterotetrameric SAGA Histone H2B Deubiquitinase Module revealed the structural basis of how a DUB is integrated into multisubunit protein machinery and becomes activated by its nonsubstrate binding partners (Köhler et al., 2010). In their paper, Kohler et al presented a model describing the assembly and activation of the complex, they showed that the formation of the complex can be dissected into two functional, but not necessarily sequential steps – assembly and activation.

Complex assembly of the ARISC and ARISC-RAP80 complex related to activation. We know that for minimal activity, BRCC36-Abraxas1-BRCC45 is required (Patterson-Fortin et al., 2010). This likely forms the more stable complex. The binding of MERIT40 is helps structural integrity, however, without RAP80 the arms of the four-membered complex is flexible and less stable. The binding of RAP80 to the complex helps to stabilise the arms of the complex. Rabl et al., 2019 revealed in their crystal structure of the mouse ARISC-RAP80 AIR structure that the RAP80 protein interacts with MERIT40 and Abraxas1, this explains why the negative stain EM data of the ARISC-RAP80 complex is more stable than the ARISC complex.

For future experiments, to stabilise the complex for structural determination, the complexes can undergo GRAFIX cross-linking. In comparison to previously published EM studies of SAGA (Wu, P.-Y.J. et al., 2004; Durand et al., 2014), the gradual crosslinking of the GraFix method by Setiaputra et al.,2015 greatly preserved the presence of SAGA’s extended tail (Setiaputra et al., 2015). The first SAGA EM study by Wu et al., 2004 showed that the standard tandem affinity purification (TAP) procedure resulted in 35% of the particles not displaying the tail density (Wu, P.-Y.J. et al., 2004) . Meanwhile, a recent study by Durand et al., 2014 applied glutaraldehyde cross-linking to TAP-purified SAGA, decreasing the percentage of tail-less SAGA to 25% (Durand et al., 2014). By carrying out GraFix purification, Setiaputra et al.,2015 were able to reduce the number of dissociated tails to 9%, demonstrating the effectiveness of this treatment. Using negative stain EM, the group investigated the degree of conformational changes that SAGA undergoes and generated a 3D reconstruction of each conformation and were able to effectively demonstrate the extensive rearrangement of between the three conformations (Setiaputra et al., 2015).

3.7.7. Fluorescence activity assay

The activity assay showed that ARISC-AIR displayed higher activity than the ARISC complex, despite BRCC45 and Abraxas1 and BRCC36 being the main contributor to the activity of the complex (Patterson-Fortin et al., 2010). It's possible that the difference in activity observed was due to ARISC-AIR forming a more stable complex than the four-membered ARISC complex. Negative stain EM analysis presented in this chapter suggests that the presence of AIR helps to stabilise the ARISC complex. If the ARISC-AIR complex was more stable in solution during this assay, it could account for the difference in activity.

For future experiments, activity assays of Abraxas1-BRCC36, BRCC45-MERIT40, ARISC, ARISC-AIR, and ARISC-RAP80 with IQF and non-cleavable di-ubiquitin substrates will be carried out. Conducting a comparative activity assay profile with the various ARISC subcomplexes, ARISC, ARISC-AIR and ARISC-RAP80 will help to paint a better picture of how complex assembly and stability influences the activity of the complex.

CHAPTER 4

Cryo-EM analysis and functional characterisation of the ARISC complexes

4.1 Introduction

Cryo-electron microscopy was employed to study the structure of the ARISC-RAP80 complex. ARISC-RAP80 fraction 24 was used to study the complex as it was a newly purified sample that had not been concentrated. Concentrated protein samples were avoided because previous cryo-EM analysis using concentrated ARISC-RAP80 samples resulted in extreme complex dissociation and sample heterogeneity. Unless otherwise stated, the KRIOS1 microscope equipped with a Falcon 3EC detector was used for screening and data collection. Grid preparation and data collection methods can be located in chapter 2.3.5.

4.2 QUANTIFOIL grids

ARISC-RAP80 cryo-EM grids were first made using QUANTIFOIL 2.1 300 mesh grids and a protein concentration of 0.08 mg/ml, ten times the concentration used to analyse the complex by negative stain EM. The grids were screened, and results from the screening revealed that whilst ice was clearly visible on the grid, the were no particles embedded in the grid holes (Figure 49A).

As there were no visible protein particles in the 0.08 mg/ml ARISC-RAP80 grids. The protein concentration was increased to 0.2 mg/ml and the grids were screened. Results of the screening generated grids that contained overcrowded protein particles (Figure 49B). Nevertheless, data was collected and processed in RELION 2.1 (Scheres, 2012).

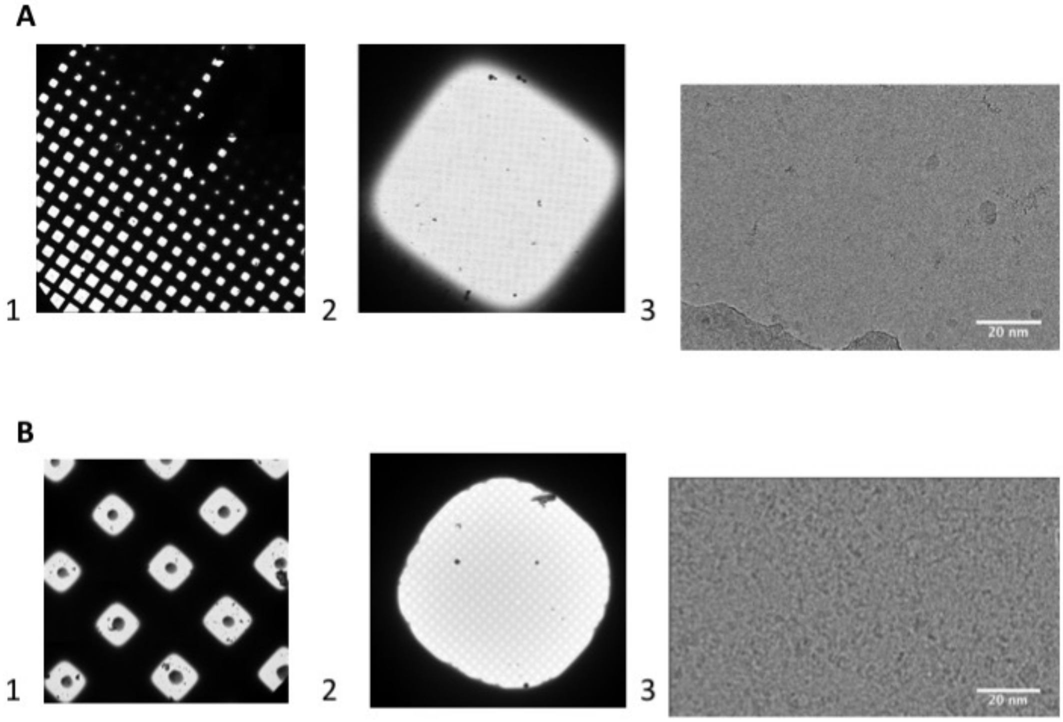


Figure 49: Comparison of the ARISC-RAP80 complex at 0.08 mg/ml and 0.2 mg/ml

Panel A shows representative images obtained from screening the complex at a concentration of 0.08 mg/ml. [1] An atlas image of the grid. The atlas indicated that there is a gradient of ice across the grid. The bottom left grid squares are larger than the grid squares at the centre and top right side of the grid. [2] An image of the grid square containing less ice. [3] A micrograph. The micrograph contained some contaminants but no ARISC-RAP80 protein particles. Panel B shows representative images obtained from screening the grid containing 0.2 mg/ml of ARISC-RAP80. [1] An atlas of the grid [2] A representative grid square. [3] A representative micrograph with visible protein particles.

Approximately 1,000 particles were manually and randomly selected from the micrograph, extracted and subjected to 2D classification. Out of the ten requested classes, eight contained class averages (Figure 50A). The first three class averages were the least noisy, however, they were small and spherical. The other classes were very noisy and either contained non-particles (for example class five) or subunits of the complex. Despite this, the first three classes were used as template to determine whether the program would be able to successfully identify protein particles. Results from the auto-picking job showed visible particles with minimal overlap (Figure 50B). Following auto-picking, 1,458,950 particles were extracted and subjected to 2D classification in which 100 classes were requested. A mask diameter of 300 Å was applied to each individual class averages.

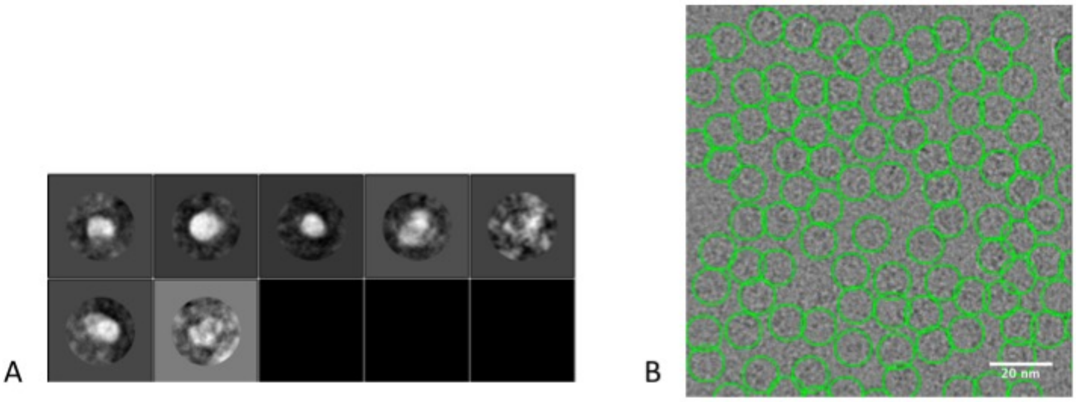


Figure 50: Particle auto-picking in RELION 2.1

[A] 2D templates generated from the manually selected particles. [B] A representative auto-picked micrograph. The particles were auto-picked using the class averages one, two, three, four, six and seven.

Results from the reference free 2D classification revealed class averages with high levels of background noise. The classes displayed topologies that were very different to that of the ARISC model (Figure 51).

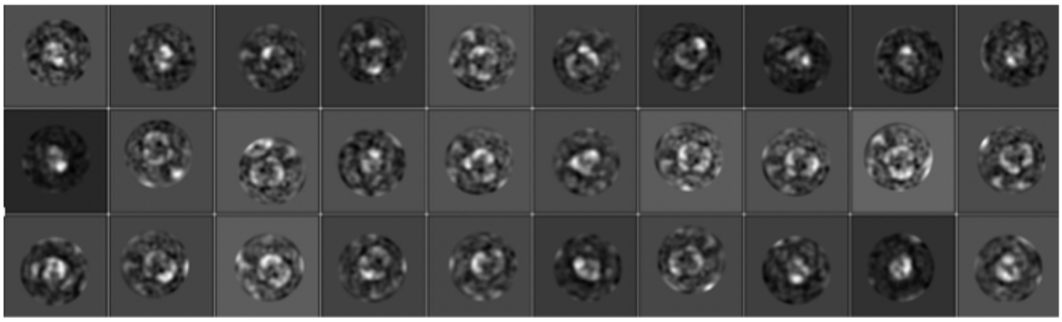


Figure 51: Representative 2D class averages of the 0.2 mg/ml ARISC-RAP80 complex

The class averages obtained from the classification job were extremely noisy. Some classes contained small dissociated proteins whilst others appeared to contain two subunits.

To determine the structure of the particles, a small subset of 2D class averages were selected as input for the generation of an initial model in C1 symmetry. The class averages selected all represented different views of the complex. Results from the initial model revealed a 3D map that loosely resembled the ARISC complex. The map contained two arms in which the region corresponding to MERIT40 were facing in opposite directions. Interestingly, the core of the map contained two protruding densities. Despite having obtained this low resolution initial model, the data was not further processed because it was too noisy and efforts were made to optimise grid preparation (Figure 52).

Figure 52: A C1 initial model of the 0.2 mg/ml ARISC-RAP80 complex

2D class averages representing different views of the data were selected to generate the initial model, and a C1 symmetry was applied. The model contained two arms. The regions possibly corresponding to MERIT40 were orientated in opposite directions. There were two densities at the core of the complex protruding out of the complex.

4.3 Optimisation of cryo grid preparation

4.3.1 Use of an in house vitrification machine

Attempts to obtain cryo-EM grids suitable for data collection using protein concentrations of 0.08 mg/ml and 0.2 mg/ml and the Vitrobot for preparing cryo-grids were futile. A novel approach was employed using an in-house vitrification machine. The machine was built in house by David Klebl and Dr Dimitrios Kontziampasi and the main principle investigators were Dr Stephen P Muench and Professor Howard D White. At the time of use, the machine was undergoing testing and constant development(Kontziampasis et al., 2019). The vitrification apparatus functions by spraying samples onto the grid as the grid descends into liquid ethane. Following vitrification, the grid is manually removed and stored in liquid nitrogen. To generate ARISC-RAP80 cryo-EM grids, a protein concentration of 0.1 mg/ml was used and the sample was deposited on Quantifoil r2/1 300 mesh grids.

Grid screening revealed grid squares with varying amounts of ice. Some grid squares contained grey regions which represented areas where the protein had been sprayed. Analysis of the grids showed that the majority of the areas in which the sample was deposited contained very thick ice. Grid areas containing a reasonable amount of ice were screened. However, despite the presence of ice, there were no visible protein particles (Figure 53).

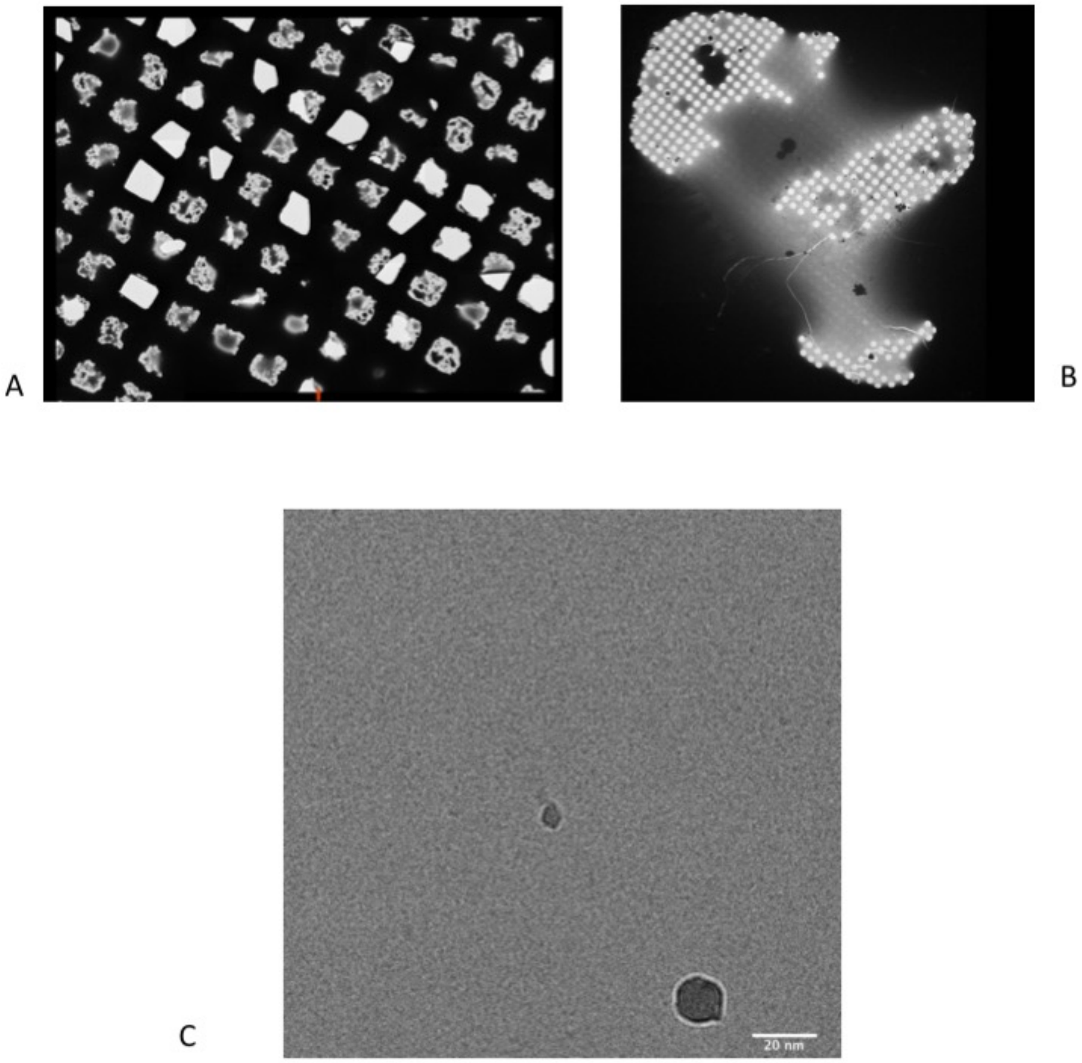


Figure 53: Screening results from the in-house spray vitrification machine

[A] An atlas of the grid. The atlas shows darker grey areas on the grid square where the sample was deposited by the machine. [B] A representative grid square. The dark grey areas are regions of the grid with thick ice where the sprayed droplets landed. [C] A representative micrograph obtained from screening the grid. The micrograph contains large and small ice contamination but no visible protein particles.

4.4 Testing of grid support materials

Protein samples have varying affinity to the material used to make the TEM grid (Thompson et al., 2016). To test whether the ARISC-RAP80 complex displayed better affinity to gold or carbon, UltrAUFoil and grids coated with a thin lacey carbon film were used. The grids were glow discharged using the Cressington glow discharge apparatus and a protein concentration of 0.95 mg/ml was used. The grids were vitrified using the Vitrobot, and the parameters of the Vitrobot were altered so that the grids were blotted using a higher force (a force value of five) and for a longer period of time (six seconds). Parameters regarding humidity and temperature were not changed.

The gold and lacey grids were screened on the KRIOS 1 microscope. Both grids contained overcrowded protein particles, and the particles in the lacey grids were more overcrowded than the particles in the UltrAUFoil grids (Figure 54).

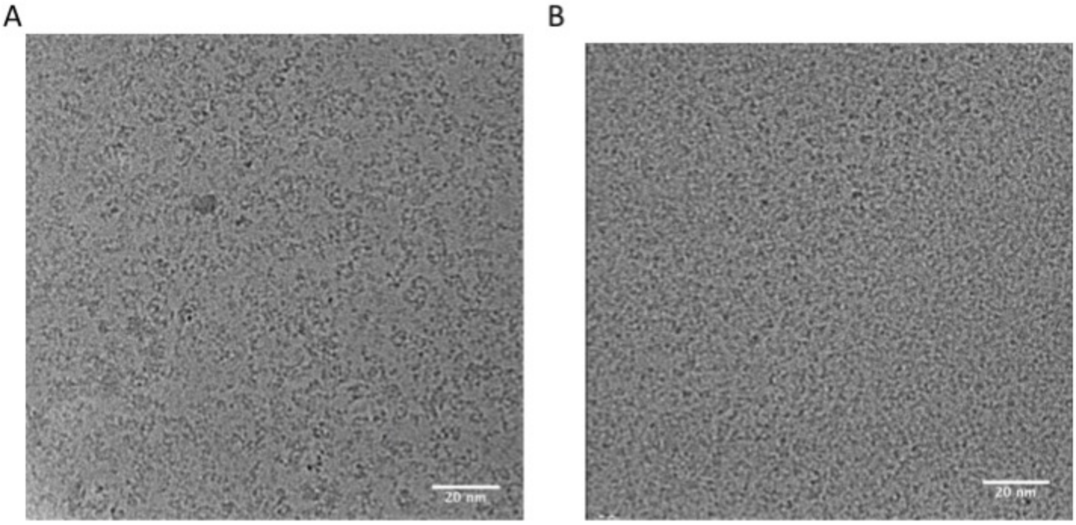


Figure 54: UltrAUFoil and lacey ARISC-RAP80 grids

[A] A representative UltrAUFoil ARISC-RAP80 micrograph. The micrograph contains dark grey ARISC-RAP80 particles that do not appear to be dissociated had have good contrast. [B] A representative lacy ARISC-RAP80 micrograph. The micrograph contains over crowded protein particles with slightly poorer contrast.

4.5 Cryo-EM data collection and processing of the ARISC-RAP80 complex

Following the successful initial screening of the ARISC-RAP80 complex using UltrAUFoil and lacey grids (Figure 54), additional UltrAUFoil grids were prepared using a protein concentration of 0.5 mg/ml. Screening of the grids revealed that the ARISC-RAP80 particles displayed a preference for thick ice. Four datasets were collected over a total of 168 hours on four separate occasions. The datasets were individually pre-processed using RELION 3.0 (Scheres, 2012; Zivanov et al., 2018) as described in 2.3.5 before being processed in cryoSPARC (Punjani, Ali et al., 2017). Following processing in cryoSPARC particles were imported back into RELION for further processing (Punjani, Ali et al., 2017; Zivanov et al., 2018). Figure 55 outlines the processing pipeline.

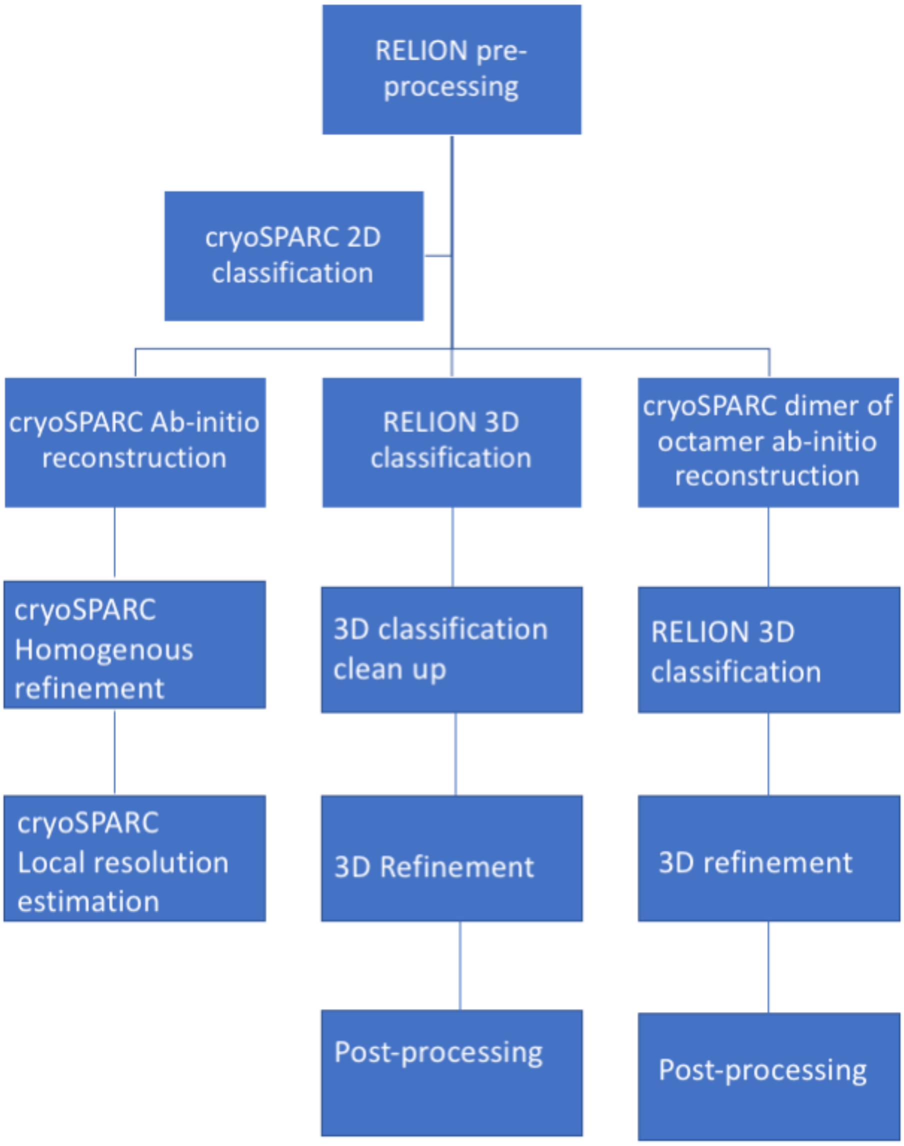


Figure 55: data processing overview

The flow chart outlines the approach taken to process all four datasets. The datasets were first pre-processed in RELION. Following pre-processing, extracted particles were imported into cryoSPARC and processed to 3D refinement. To compare the efficiency of processing in cryoSPARC with RELION, 2D class averages were selected and imported back to RELION for 3D processing (Punjani, Ali et al., 2017; Zivanov et al., 2018).

Visual inspection of the micrographs showed dark grey particles embedded against a grey background. Approximately 1,000 particles were manually selected at random from each dataset, extracted and subject to 2D classification. The small round of 2D classification, for the first three datasets, produced white class averages that displayed the same topology of the ARISC model (Figures 56-58). Surprisingly, the first three datasets contained class averages in which the ARISC complex appeared to dimerise with itself. This conformation had not been observed before and looked different from the cross-linked dimer of octamers described by Kyrieleis et al., 2016. The 2D class averages obtained from the fourth dataset were extremely noisy and not as detailed as the class averages from the first three dataset (Figure 59). In addition to this, RELION (Scheres, 2012; Zivanov et al., 2018) was only able to align the manually picked particles into six classes, two of which contained ice contamination (Figure 59C).

All ten of the 2D class averages from the first three data set were used as auto-picking template for their respective data (Figures 56-58). Classes 1-4 from the fourth dataset were used to auto-pick the fourth dataset (Figure 59). 155,417, 1,065,399, 597,235 and 745,569 particles were auto-picked from datasets one, two, three and four, respectively. The extracted particles from each dataset individually underwent a 2D classification. 2D classification using RELION (Scheres, 2012; Zivanov et al., 2018) generated extremely noisy class averages. Attempts were made to clean up the classes and obtain well align class averages, however after multiple rounds of 2D classifications the resulting class averages contained high levels of background noise (Figure 60).

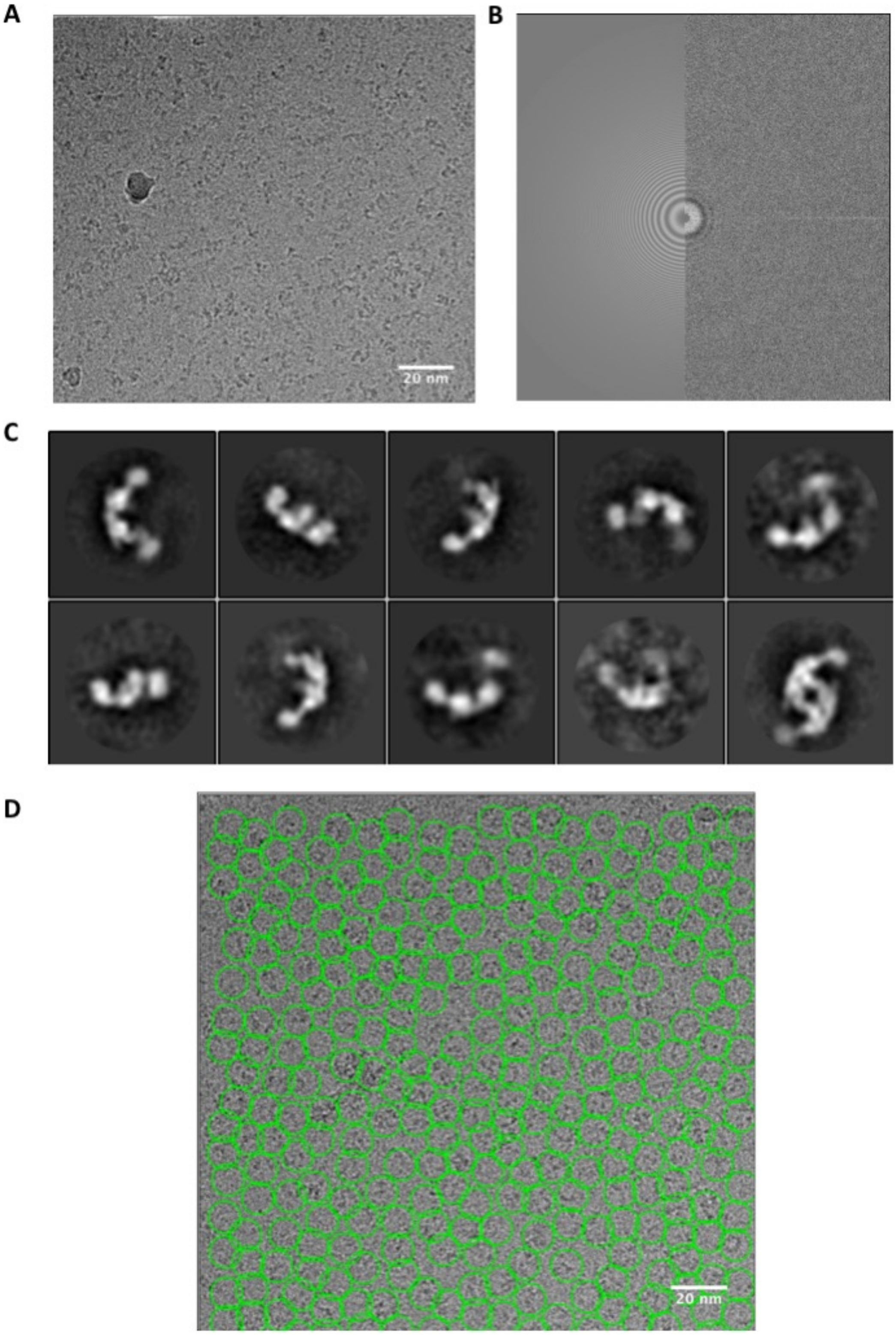


Figure 56: Cryo-EM pre-processing of the first ARISC-RAP80 complex dataset

[A] A representative micrograph showing clear protein particles. [B] A representative CTF corrected image. [C] 2D templates used for auto-picking. [D] A representative auto-picked micrograph.

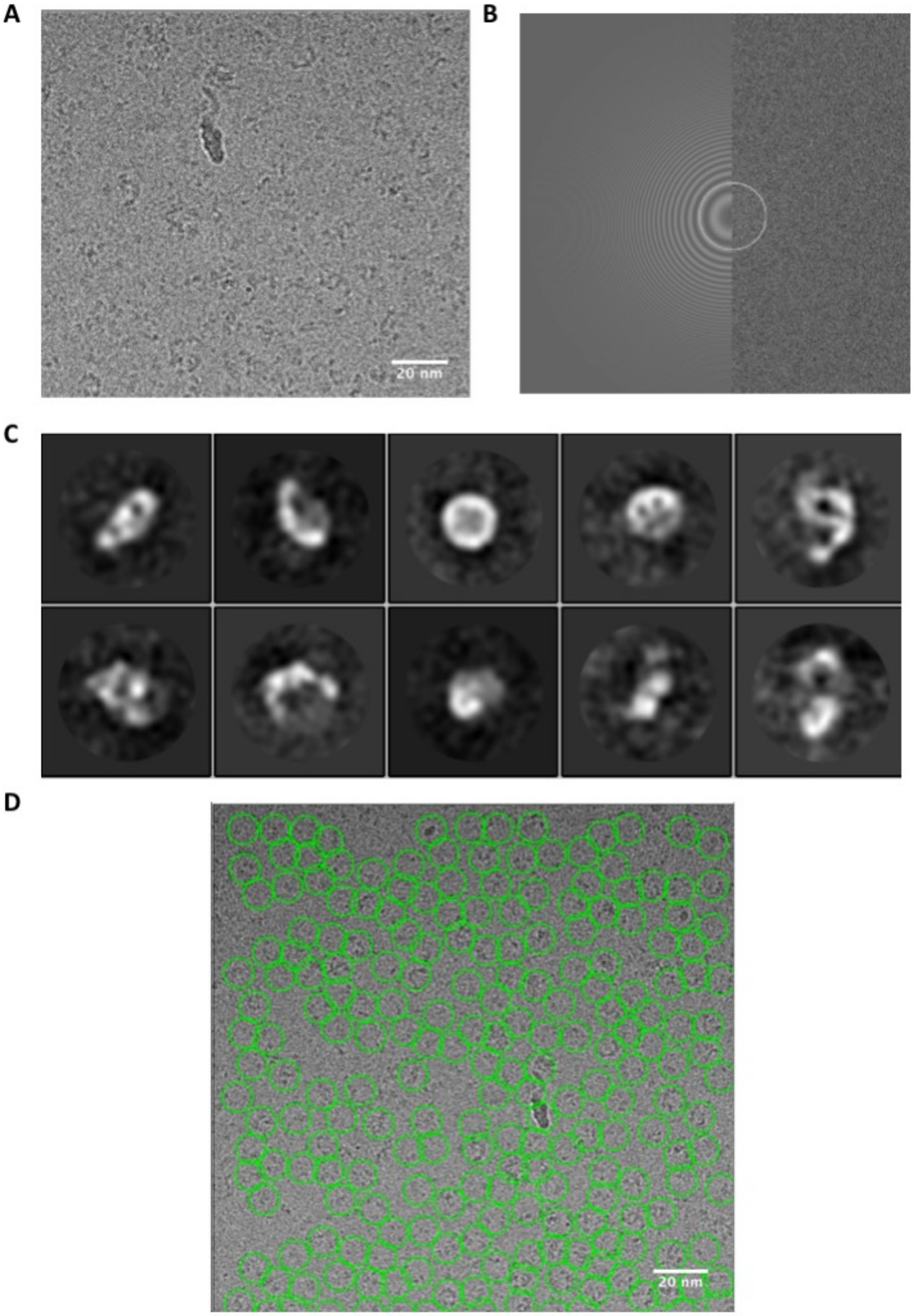


Figure 57: Cryo-EM pre-processing of the second ARISC-RAP80 complex dataset

[A] A representative micrograph showing clear protein particles. [B] A representative CTF corrected image. [C] 2D templates used for auto-picking. [D] A representative auto-picked micrograph.

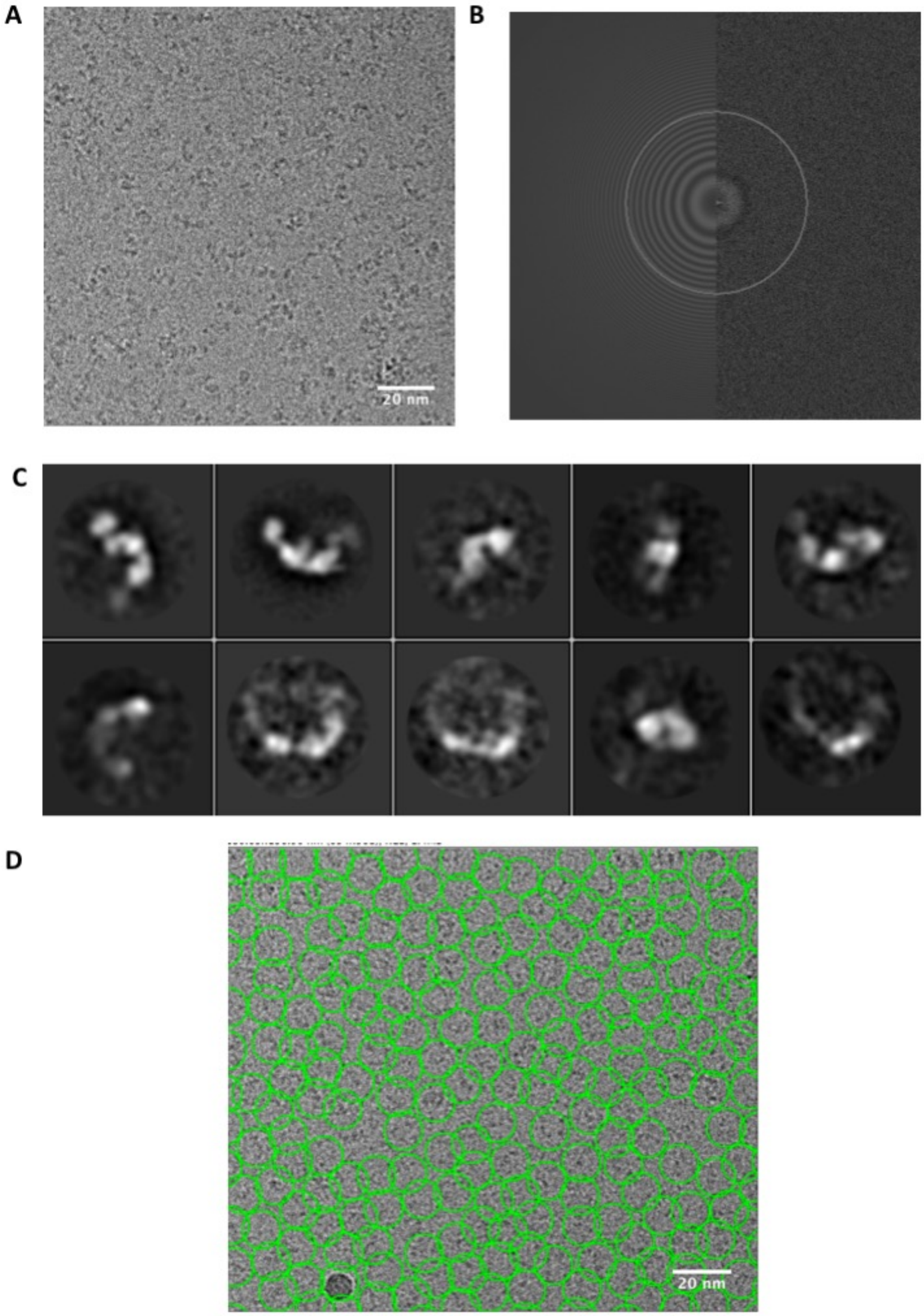


Figure 58: Cryo-EM pre-processing of the third ARISC-RAP80 complex data

[A] A representative micrograph showing clear protein particles. [B] A representative CTF corrected image. [C] 2D templates used for auto-picking. [D] A representative auto-picked micrograph.

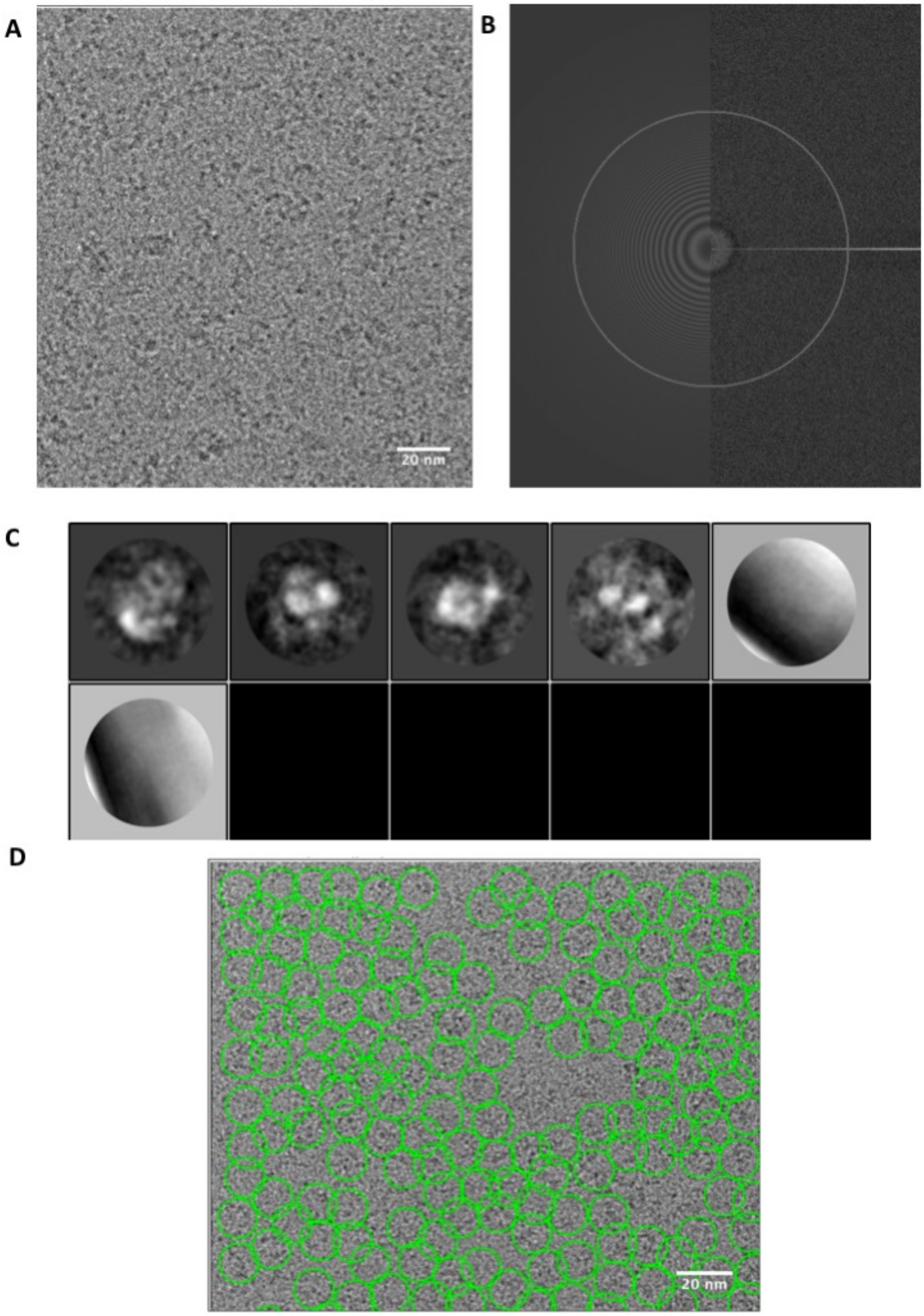


Figure 59: Cryo-EM pre-processing of the fourth ARISC-RAP80 complex dataset

[A] A representative micrograph showing clear protein particles. [B] A representative CTF corrected image. [C] 2D templates used for auto-picking. [D] A representative auto-picked micrograph.

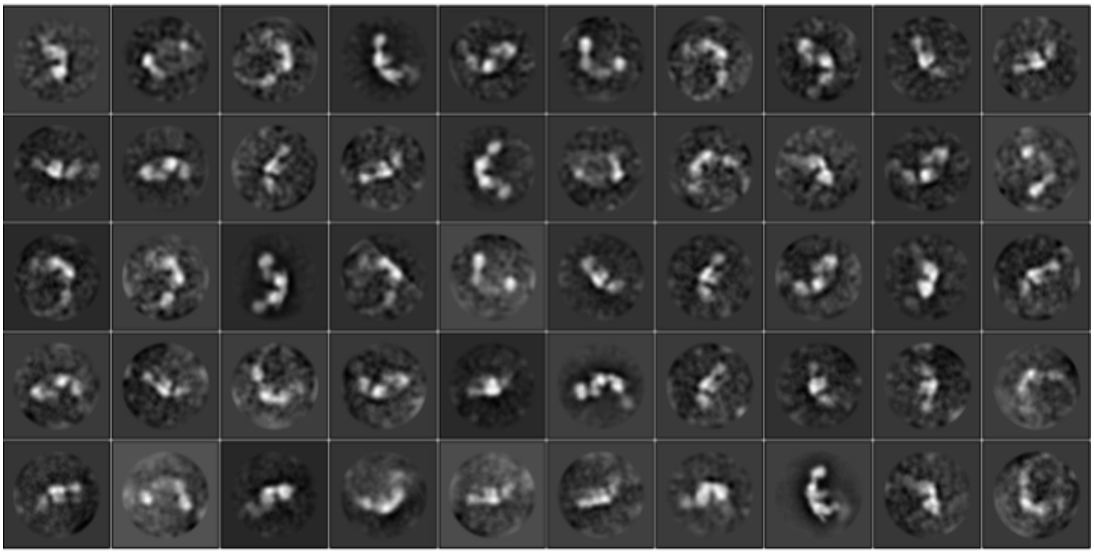


Figure 60: Representative class averages obtained from the fifth round of 2D classification of the second data set

2D classification generated seven well aligned white class averages with black background and some noise. The remaining classes contained high levels of background noise, however, the signal from the complex was slightly stronger, allowing them to be identified.

4.6 Data processing in CryoSPARC

Processing the four datasets to 2D classification in RELION (Scheres, 2012; Zivanov et al., 2018) resulted in extremely noisy class averages. Attempts to "clean up" the datasets and obtain well aligned and detailed class averages was unsuccessful. To overcome this problem, the extracted particles from all four datasets were combined, creating a single dataset and processed in CryoSPARC (Punjani, Ali et al., 2017). CryoSPARC (Punjani, Ali et al., 2017) was selected as the experience of other users was that it classified noisy particles better and faster.

4.6.1 2D classification in CRYOSPARC

2,563,620 particles were subject to 2D classification. 2D classification was repeated three times. After each round, class averages whose signal was stronger than the background noise by visual inspection were selected for further processing. Results from the 2D classification suggested that the complex displayed a high level of heterogeneity. Some of the 2D class averages appeared to correspond to the core of the complex (the Abraxas1-BRCCC36 super dimer), others appeared to represent the four membered ARISC complex. The majority of the ARISC 2D classes appeared to be missing MERIT40 or BRCC45-MERIT40 (so-called “one arm” classes) (Figure 61). CryoSPARC (Punjani, Ali et al., 2017), like RELION (Scheres, 2012; Zivanov et al., 2018), was able to identify and align the ARISC dimer of octamer particles. The 2D class averages were present in different orientations, and likely displayed different views of the complex such as the top, side, front and bottom view. Following 2D classification, 54 well-aligned class averages were selected for downstream processing, class averages pertaining to the ARISC dimer of octamer were omitted from this selection.

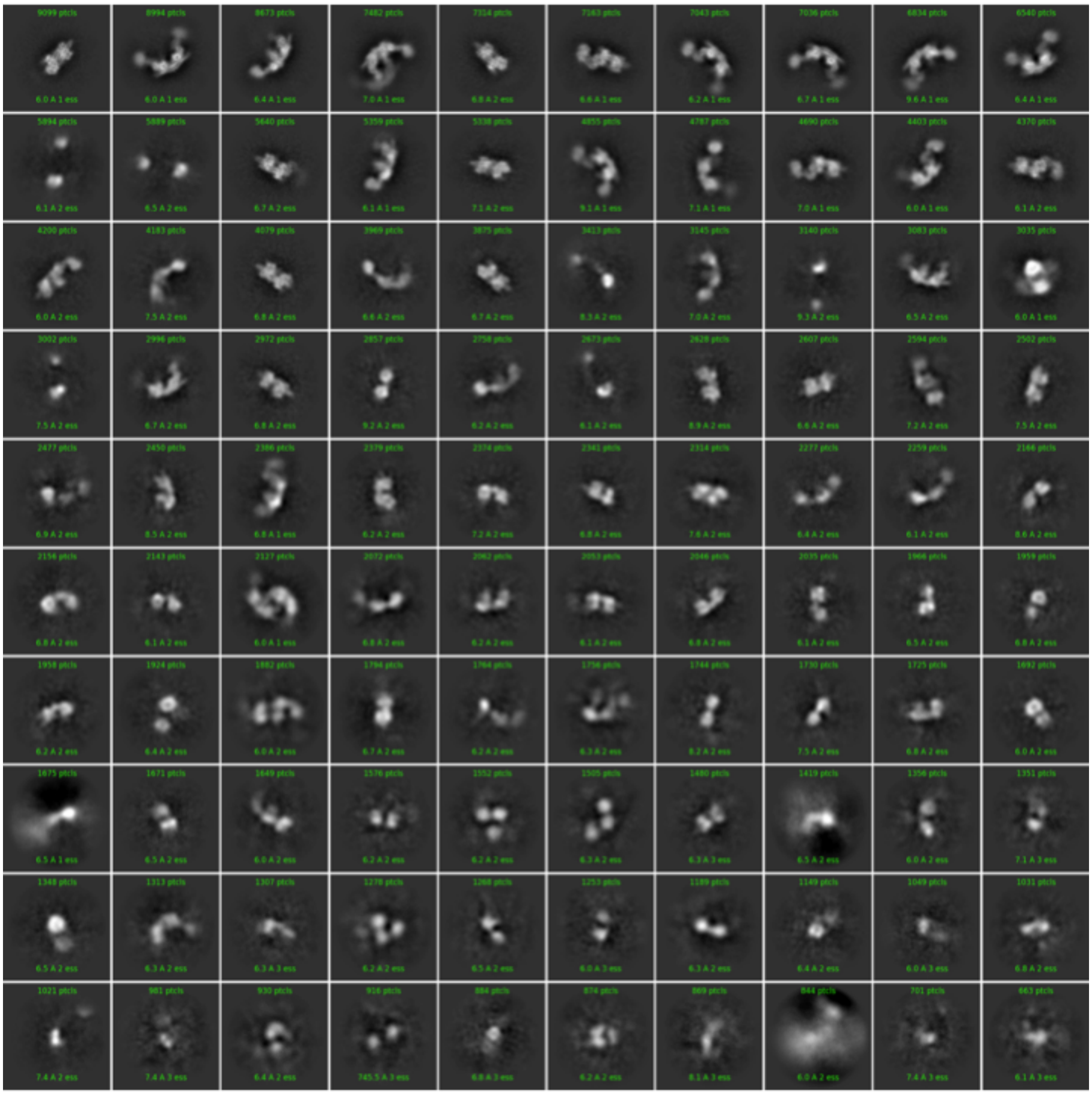


Figure 61: Final round of 2D classification in cryoSPARC

Results from 2D classification indicates sample and stoichiometry heterogeneity as some class averages appear to be subcomplexes of the ARISC-RAP80 complex. 2D classification was able to also accurately align the ARISC dimer of octamer particles.

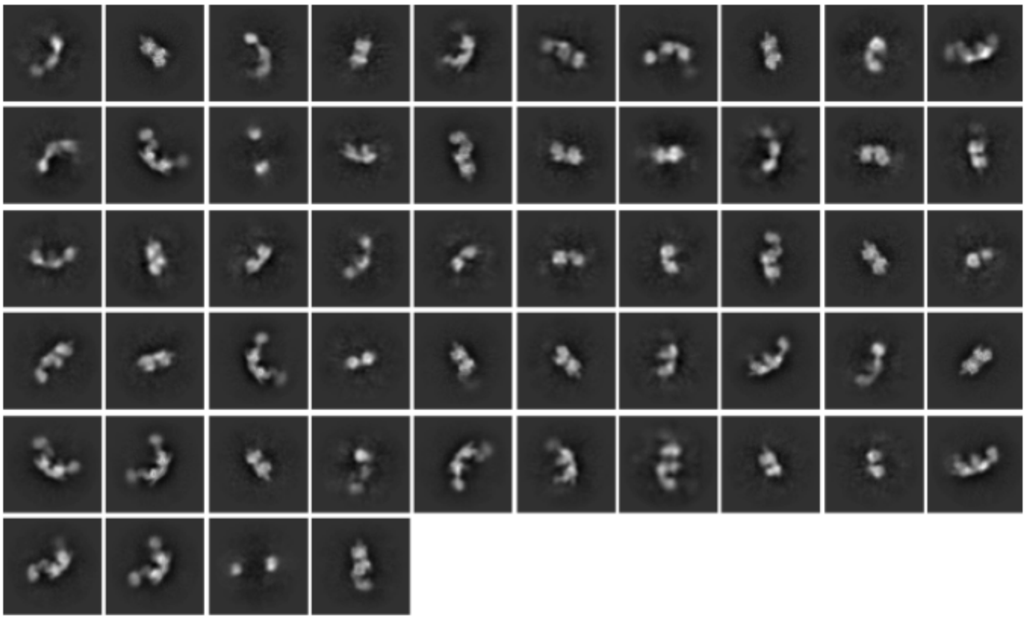


Figure 62: 2D class averages selected for downstream processing

Well aligned class averages with no to minimal background noise were selected for further processing. The ARISC dimer of octamer classes were omitted from the selection.

4.6.2 Ab-initio reconstruction

211,994 particles were selected for ab-initio reconstruction. For better classification, and to identify unique structures accurately representing the dataset, the program was asked to generate five initial models (Figure 63). This revealed two 3D maps that were missing densities corresponding to BRCC45-MERIT40 (one arm). Maps three and five appeared to represent the core complex (the Abraxas1-BRCC36 super dimers). Due to the high noise level and lack of discernible pattern it was not possible to assign the identity of the fourth map. The second map (yellow), contained protruding density at the position where alpha helical coil-coils were expected, which are typically visible at a resolution of approximately 10 Å. Particles contributing to the 3D maps with one BRCC45-MERIT40 arm were subjected to independent homogenous refinement.

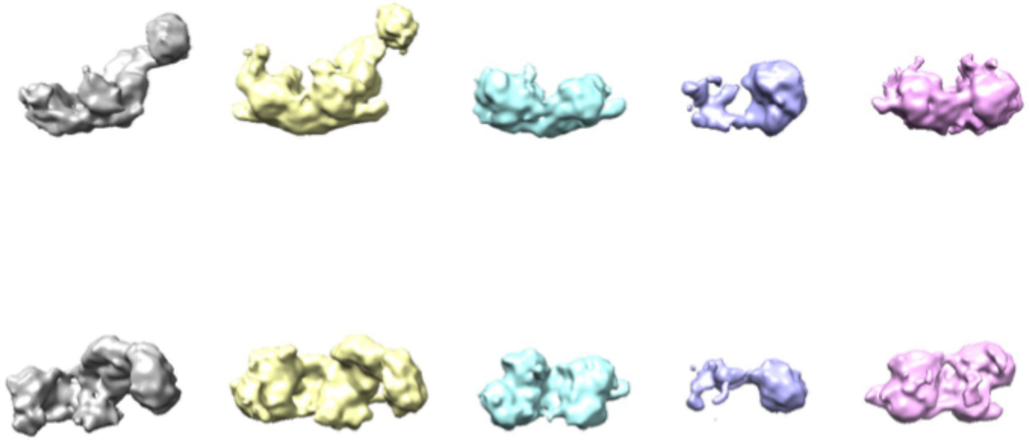


Figure 63: Five ab-initio reconstructed initial models

Classes one (grey), two (yellow), three (cyan blue), four (purple) and five (pink) contained 45,000, 65,000, 37,000, 26,000, 38,994 particles, respectively. Maps one and two were missing a BRCC45-MERIT40 heterodimer. Maps three and five represent the core complex, it is unclear what the fourth map represents.

4.6.3 3D refinement

45,000 particles in class one and 65,000 particles in class two were refined to 10.6 and 7.4 Å, respectively (Figure 64-65). Comparison of the two refined maps revealed that class one was poorly refined as the map was very smooth and contained less features than the initial model. The refined class two map displayed more structural features and contained density of the second MERIT40 subunit.

Analysis of Class two’s gold standard FSC curve generated by CryoSPARC (Punjani, Ali et al., 2017) showed that the lines corresponding to the corrected map fluctuated between 18 Å and 9 Å, dipping and rising before crossing the threshold of 0.143 (Punjani, Ali et al., 2017). Visual inspection of the map in CHIMERA (Pettersen, Eric F., et al. 2004), revealed a lot of noise. Removal of the noise using the ‘hide dust’ feature revealed a map containing two clear arms. The high level of noise was likely due to the map reaching its resolution limit. Adjusting the volume viewer contour levels in CHIMERA (Pettersen, Eric F., et al 2004) to remove the noise gave rise to a complex that was missing densities corresponding to BRCC45 and some parts of

MERIT40. This indicates that the resolution of that arm is lower than the resolution of the core and opposite arm (Figure 65).

As both class one and two maps were relatively similar to one another, they were combined, and particles were subjected to homogenous refinement using class two as a reference (Figure 66). The combined particles were refined to a resolution of 7.4 Å and one region of the map contained stronger BRCC45 and MERIT40 density. The corresponding gold-standard FSC graph also displayed significant fluctuations between 18 and 7 Å. Fluctuations in FSC curves are typically observed when working with samples that are flexible, and or contain disordered and unstructured regions.

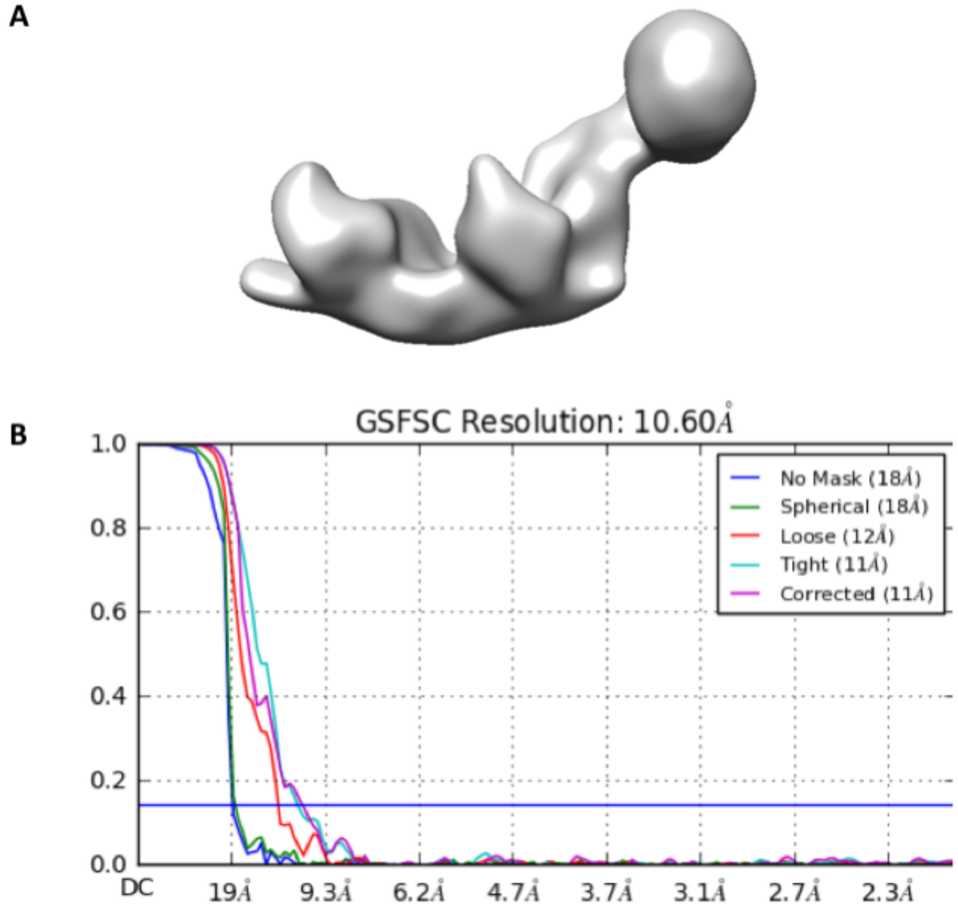


Figure 64: Homogenous refinement of class one

[A] The refined map. The map contained less structural details than the ab-initio model suggesting that the map was poorly refined. [B] The gold-standard FSC graph. The resolution of the map using a threshold of 0.143 was found to be 10.60 Å.

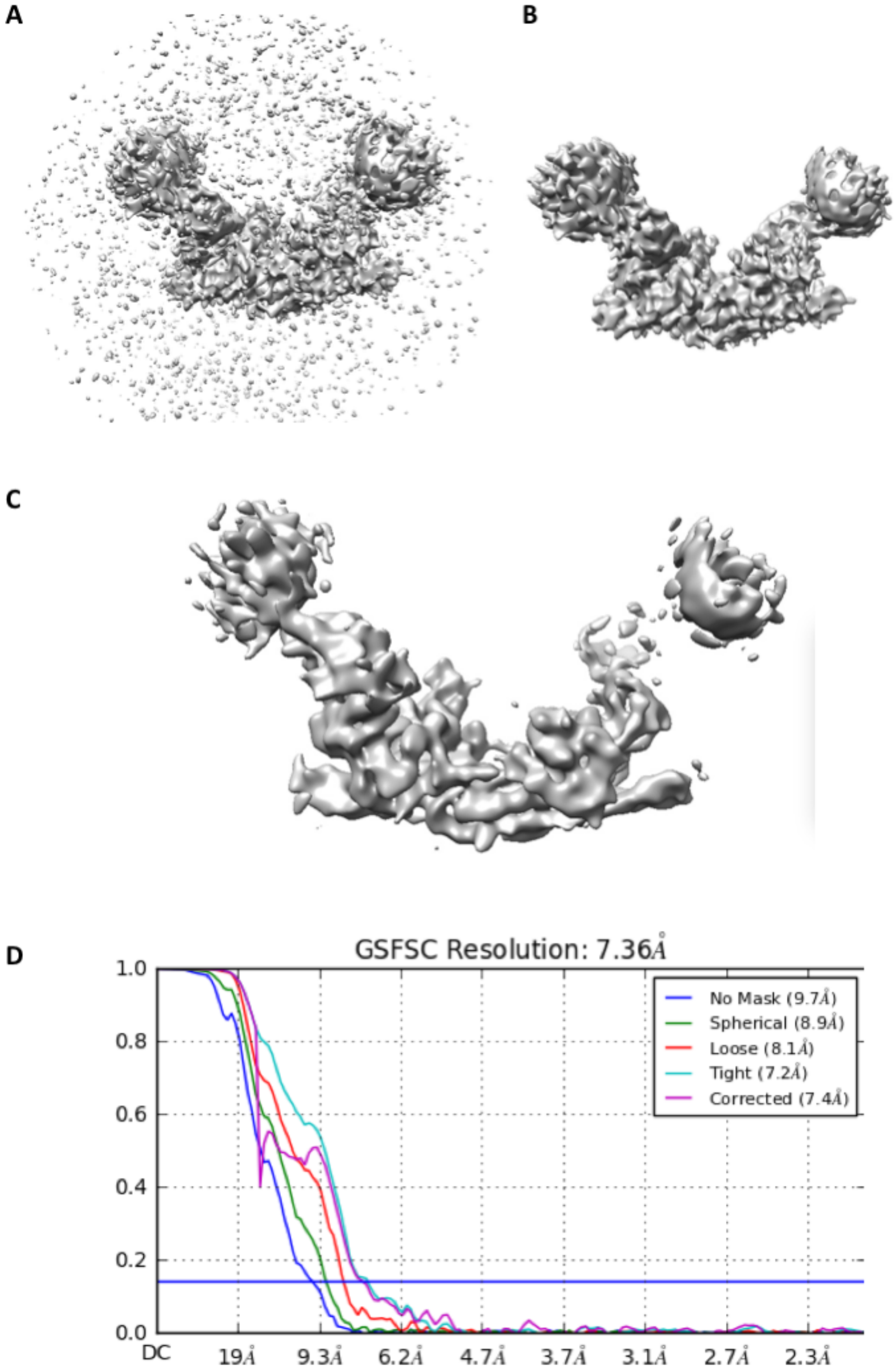


Figure 65: Homogenous refinement of class two

[A] The two arm refined map prior to noise removal. [B] The refined map after noise removal. The surrounding noise was hidden using the 'hide dust' feature on Chimera. [C] The contoured refined map. The map was contoured until the surrounding noise was no longer visible. This resulted in the loss of some MERIT40 and BRCC45 density. [D] The gold-standard FSC graph. Using a threshold of 0.143, the resolution of the map was found to be 7.36 Å.

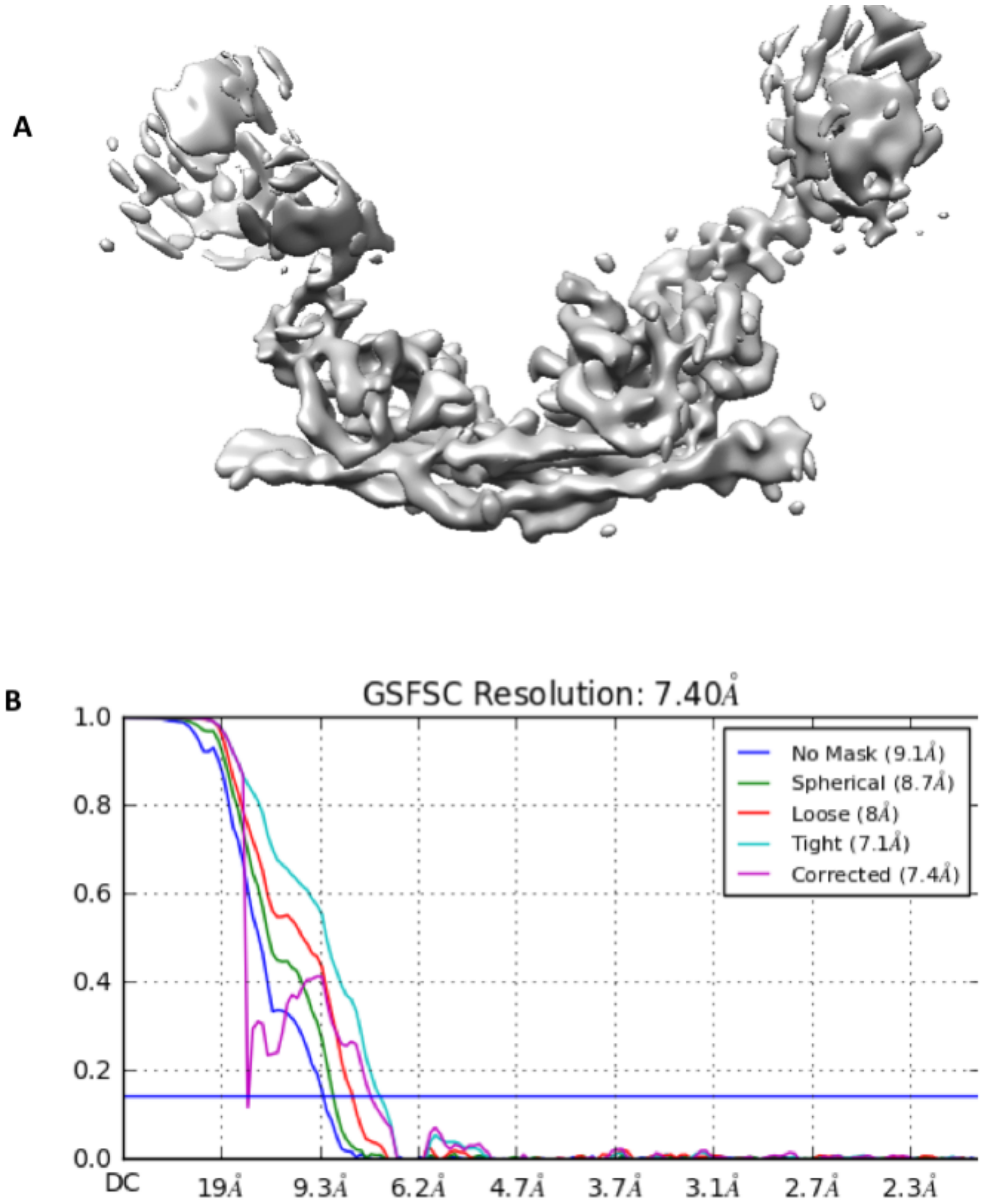


Figure 66: Homogenous refinement of particles from class one and two

[A] The refined map of the combined class one and two particles. The map contained more BRCC45 and MERIT40 than the individual classes. [B] The gold-standard FSC graph. Using a threshold of 0.143, the resolution of the map was determined to be 7.40 Å.

4.6.4 Local resolution estimation

The resolution provided during 3D refinement describes the global resolution. To obtain a more accurate description of the resolution across the map, the local resolution of the refined class two map was calculated. Results from the estimation showed that the majority of the map had a resolution of approximately seven Å. The histogram also revealed that some regions of the map were of higher resolutions (4-6 Å), whereas other regions had lower resolutions (8-18 Å). The structure of the BRISC complex was obtained by subtracting the SHMT2 dimer from the BRISC-SHMT2 structure (protein data bank code 6r8f) (Walden et al., 2019). The BRISC structure was fitted to the map and the model and the cryo-EM map agreed well with one another. The alpha helices of Abraxas2 and BRCC36 subunits fitted well into the core density of the map confirming that this region is Abraxas 1 and BRCC36. The fit of BRISC to the map revealed additional densities in the BRCC45 region of the right arm of the map and MERIT40 on the left arm of the map (Figure 67).

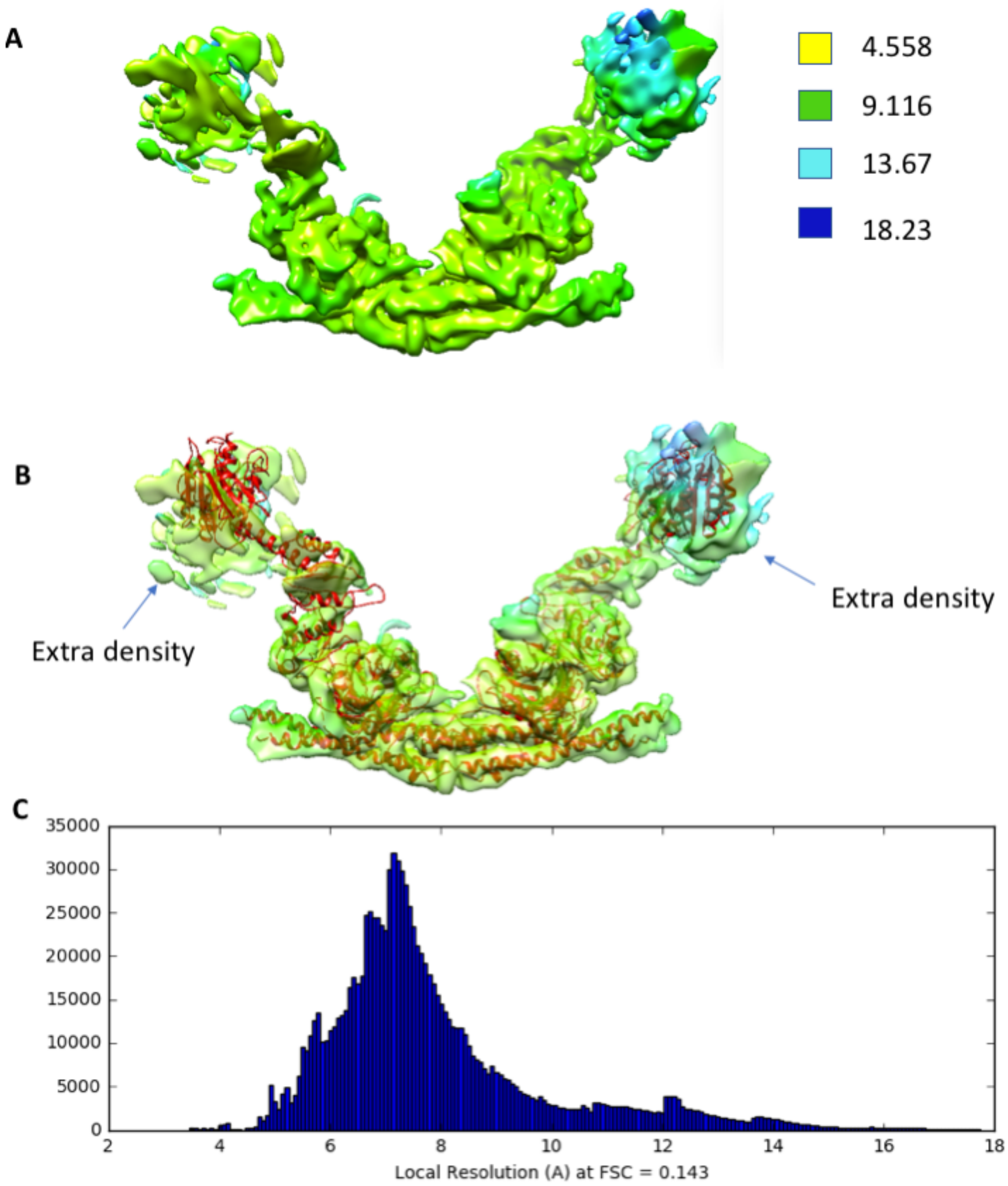


Figure 67: Local resolution estimation of class two

[A] Local resolution estimated map. The map was coloured according to the estimated resolution of each region. The majority of the map were lime green and green suggesting that the resolution in those regions were around 5-9 Å. [B] The map fitted with the BRISC structure obtained from subtracting the SHMT2 dimer from the BRISC-SHMT2 structure (PDB:6r8f). The fit revealed that the structure occupied the majority of the map, confirming that the map is of the ARISC complex. [C] The histogram of the local resolution estimation. The histogram reveals that the majority of the map is at around 7 Å and some regions had a lower or high resolution.

4.7 Data processing in RELION 3.0

Data processing in cryoSPARC (Punjani, Ali et al., 2017) generated excellent 2D class averages, however, 3D processing of the particles produced inconsistent (e.g. poor refinement of class one: Figure 64) and unreliable results. It was then decided that processing the ‘cleaned-up’ data in RELION (Scheres, 2012; Zivanov et al., 2018) may generate more reliable 3D maps. Particles selected from CryoSPARC (Punjani, Ali et al., 2017) 2D classification were converted into star files and imported into RELION 3.0 (Scheres, 2012; Zivanov et al., 2018) for 3D classification. A python script published on Github by UCSF PyEM was run to convert the particles from a CryoSPARC (Punjani, Ali et al., 2017) format (.CS) into a RELION (Scheres, 2012; Zivanov et al., 2018) compatible format (.star) files. However, the conversion resulted in loss of CTF information. A script written by Dr Matthew Iadanza was then used to obtain the corresponding CTF information for each particle from the CTF output files. This allowed successful import of particles into RELION (Scheres, 2012; Zivanov et al., 2018) and further data processing. To determine the quality of the particles, each dataset underwent 3D classification prior to being processed as a single large dataset.

4.7.1 Validating the quality of selected cryoSPARC particles

To assess the quality of each dataset and determine whether the particles selected from cryoSPRAC (Punjani, Ali et al., 2017) processing would be suitable for processing in RELION (Scheres, 2012; Zivanov et al., 2018), the individual datasets underwent 3D classification independently (Figure 68). 3D classification of the first dataset was run in C1 symmetry using map two from the cryoSPARC (Punjani, Ali et al., 2017) ab-initio reconstruction as reference (Figure 63). The output of the job revealed a single map that strongly resembled the ARISC model (class five), three maps potentially representing the ARISC-RAP80 complex (class1-3) and an undefined map (class four). Following this, class five was used as the initial model to classify the second, third and fourth dataset. 3D classification of the second dataset generated two maps that resembled the two arm ARISC complex (class three and four), one arm ARISC complex (class one) and two ARISC-like maps (class two and five). 3D classification of the third dataset generated four maps that mirrored the ARISC model (class, one, two, four and five) and a map that contained additional density at the core of the complex (class three). Six classes were requested to better separate and classify the particles of the fourth dataset. Results from the classification revealed two maps that resembled the ARISC complex (class three and four). The remaining maps contained features that could be argued to resemble the ARISC complex but were more distorted. Overall, the result of this analysis suggests that the particles imported following 2D classification in cryoSPARC (Punjani, Ali et al., 2017) were of good quality and can be further processed in RELION (Scheres, 2012; Zivanov et al., 2018).

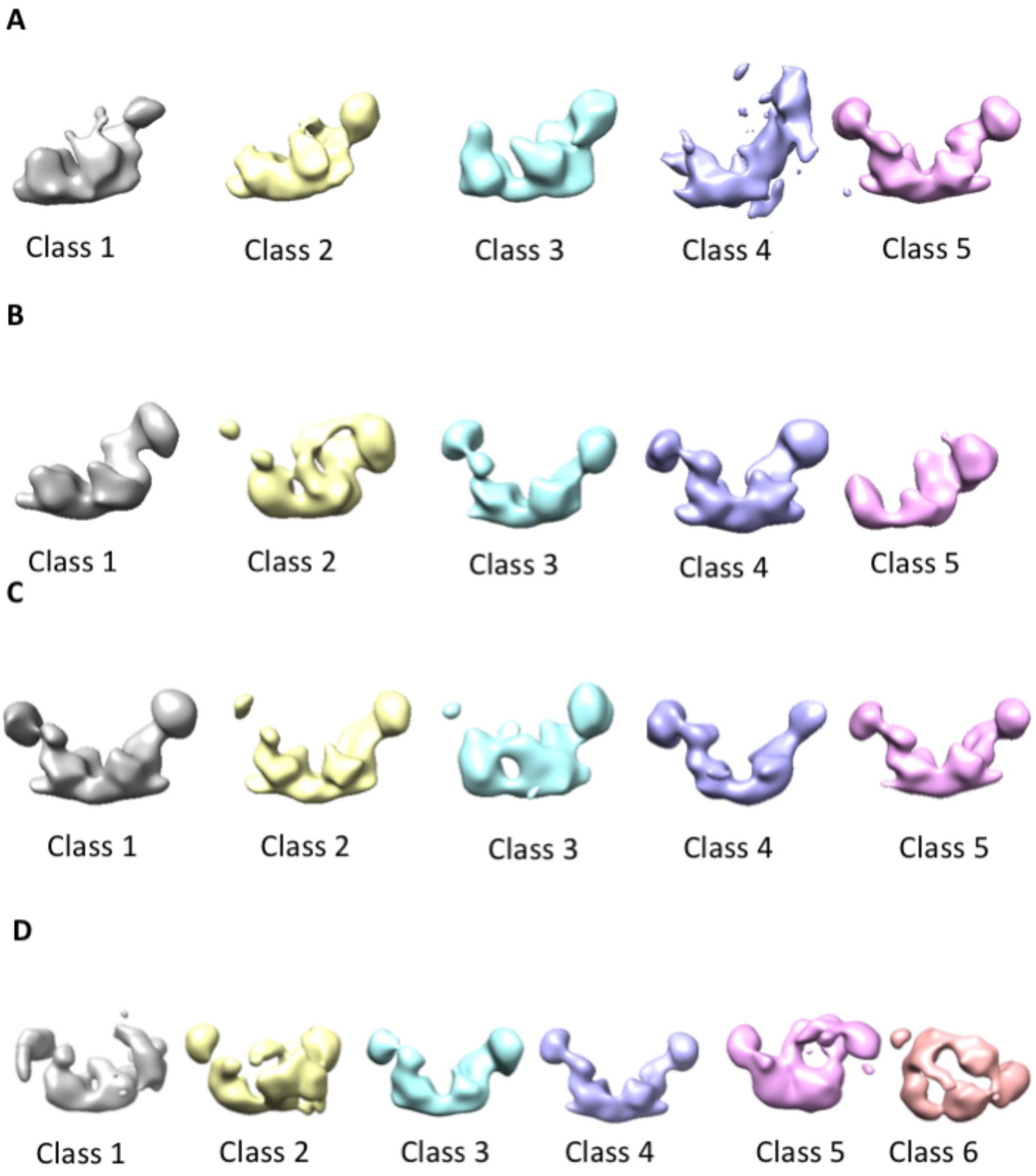


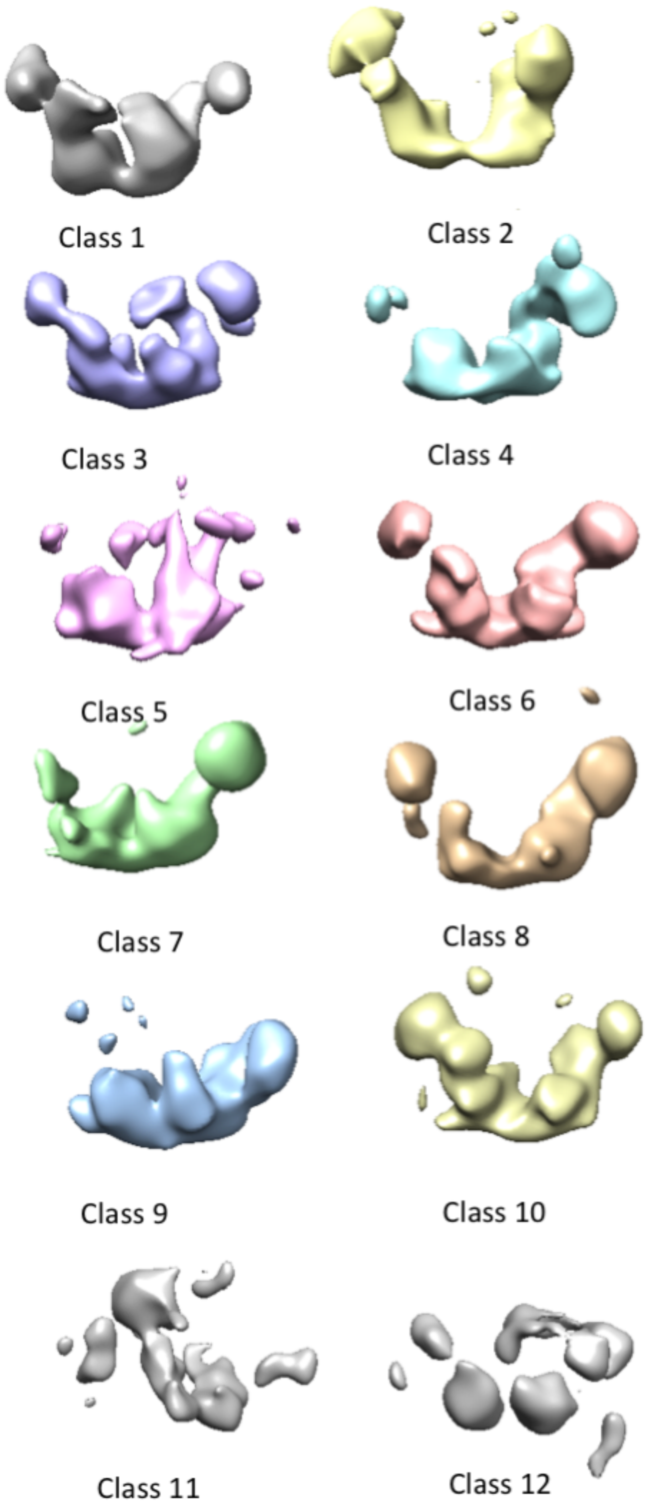
Figure 68: 3D classification of the four datasets in RELION

[A-D] 128,000, 13,578, 58, 798 and 65, 613 particles from datasets 1(A), 2 (B), 3 (C) and 4 (D) respectively underwent 3D classification. The results suggest that 2D classification in cryoSPARC (Punjani, Ali et al., 2017) generated particles suitable for 3D processing in RELION (Scheres, 2012; Zivanov et al., 2018).

4.7.2 3D classification of all combined particles

Individual 3D classification of each dataset revealed that RELION (Scheres, 2012; Zivanov et al., 2018) was able to align the particles and generate 3D maps whose structure resemble the ARISC model and the cryoSPARC (Punjani, Ali et al., 2017) maps. Re-evaluation of the cryoSPARC (Punjani, Ali et al., 2017)2D class averages omitted from the downstream processing resulted in the identification of class averages that may represent different views or additional subcomplex structures. To prevent losing crucial structural information, all the particles from the final round of 2D classification were selected, converted to .star files, and processed in RELION (Scheres, 2012; Zivanov et al., 2018). As there were a large number of particles (380,000), 12 classes were requested.

Out of the 12 3D maps generated, 10 looked promising and displayed recognisable structural features resembling the characteristic ARISC U-shape (Figure 69). Maps 11 and 12 did not display any distinct topology and likely contained junk particles. Maps six and ten contained two arms or 1.5 arms (i.e. they were missing regions of BRCC45). Other maps that resembled the ARISC complex were slightly distorted such as maps two, four, seven, and eight. Map three had one stable BRCC45-MERIT40 arm but also had density that curved out from the back of the structure. It is possible that this density represents an ARISC dimer of octamers conformation. Maps five and nine appear to be two ARISC subcomplexes as both BRCC45-MERIT40 arms were missing. Map one had a more compact U shape. It also contained density spanning out from the BRRC45 region on the left arm of the complex. The Abraxas1-BRCC36 core density also appears to be re-arranged.



Class 1 Class 2
Class 3 Class 4
Class 5 Class 6
Class 7 Class 8
Class 9 Class 10
Class 11 Class 12

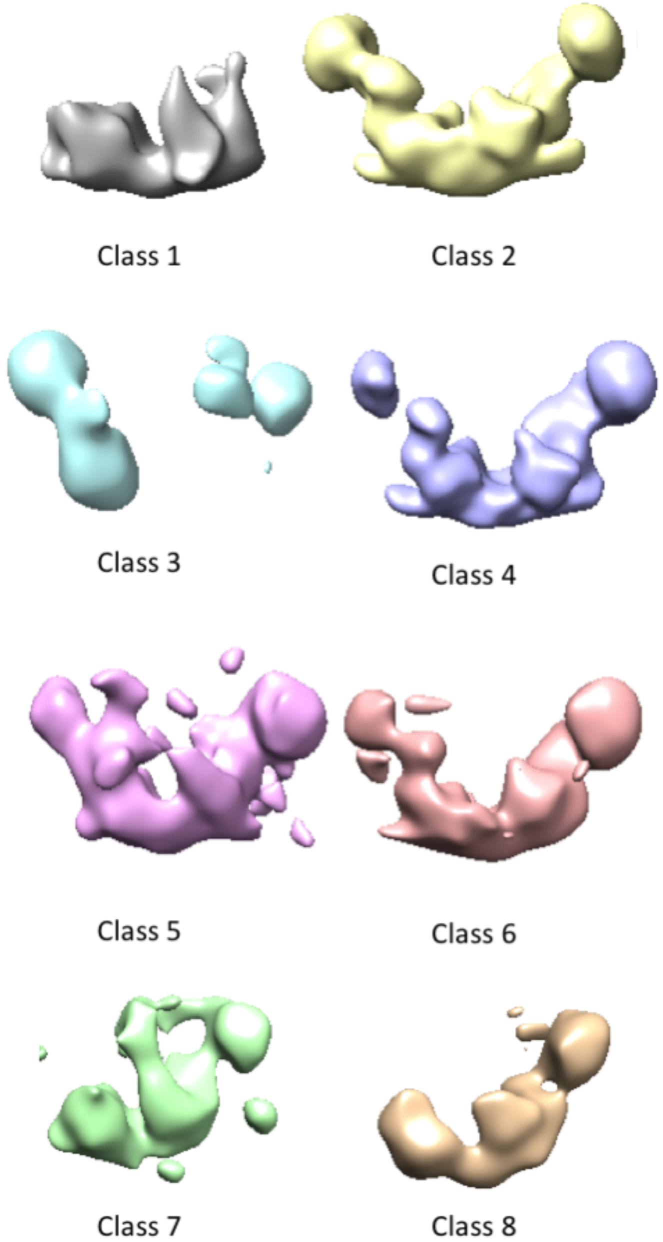
Figure 69: 3D classification of all particles from the combined dataset of the ARISC-RAP80 complex

3D classification generated 12 maps. The maps displayed stoichiometry heterogeneity as some maps represented the one arm ARISC complex, two arm ARISC complex and the core complex.

4.7.2.1 3D Second round of 3D classification of the two arm maps

A second round of 3D classification was carried out to better separate the particles in the dataset and better align them. Class one was omitted as it looked different from the other maps and was processed separately. Classes eleven and twelve were also omitted from this round of 3D classification as these classes likely included junk particles.

Results from the second round of 3D classification (Figure 70) showed a map that corresponded to the core complex (class one), and three maps that resembled the ARISC model (class two, four and six). Classes two and six had two arms whereas class four was missing some density in the BRCC45-MERIT40 region on the left arm of the complex. The fifth map also resembled the ARISC model, but it contained density protruding out of the arms of the complex. Maps three seven and eight were all incomplete maps.



Class 1

Class 2

Class 3

Class 4

Class 5

Class 6

Class 7

Class 8

Figure 70: Second round of 3D classification for better alignment of particles

3D classification produced eight classes. The first class (grey) represents the core complex. The second (yellow), fourth (purple), and sixth (coral) classes resemble the ARISC model. Classes three (cyan blue), seven (green) and eight (orange) are all of dissociated complexes. The fifth class (pink) has a similar structure to the ARISC model but also contains additional densities protruding from the arms of the complex.

4.7.2.2 3D refinement and post-processing of the core complex

Class one from the second round of 3D classification (section 4.7.2.1) was selected for 3D refinement. The class contained 52,375 particles, which were refined against the map. Results from the refinement gave rise to a map whose resolution was estimated at 16.9 Å. Following refinement, a mask was created using a binarization value of 0.0176. Analysis of the map showed that the map was large enough to encapsulate the structure. The mask and the refined half maps were then used as inputs for post-processing, which generated an unsharpened map with an estimated resolution of 18 Å. Analysis of the FSC curve shows that the mask was not used to calculate the resolution of the map.

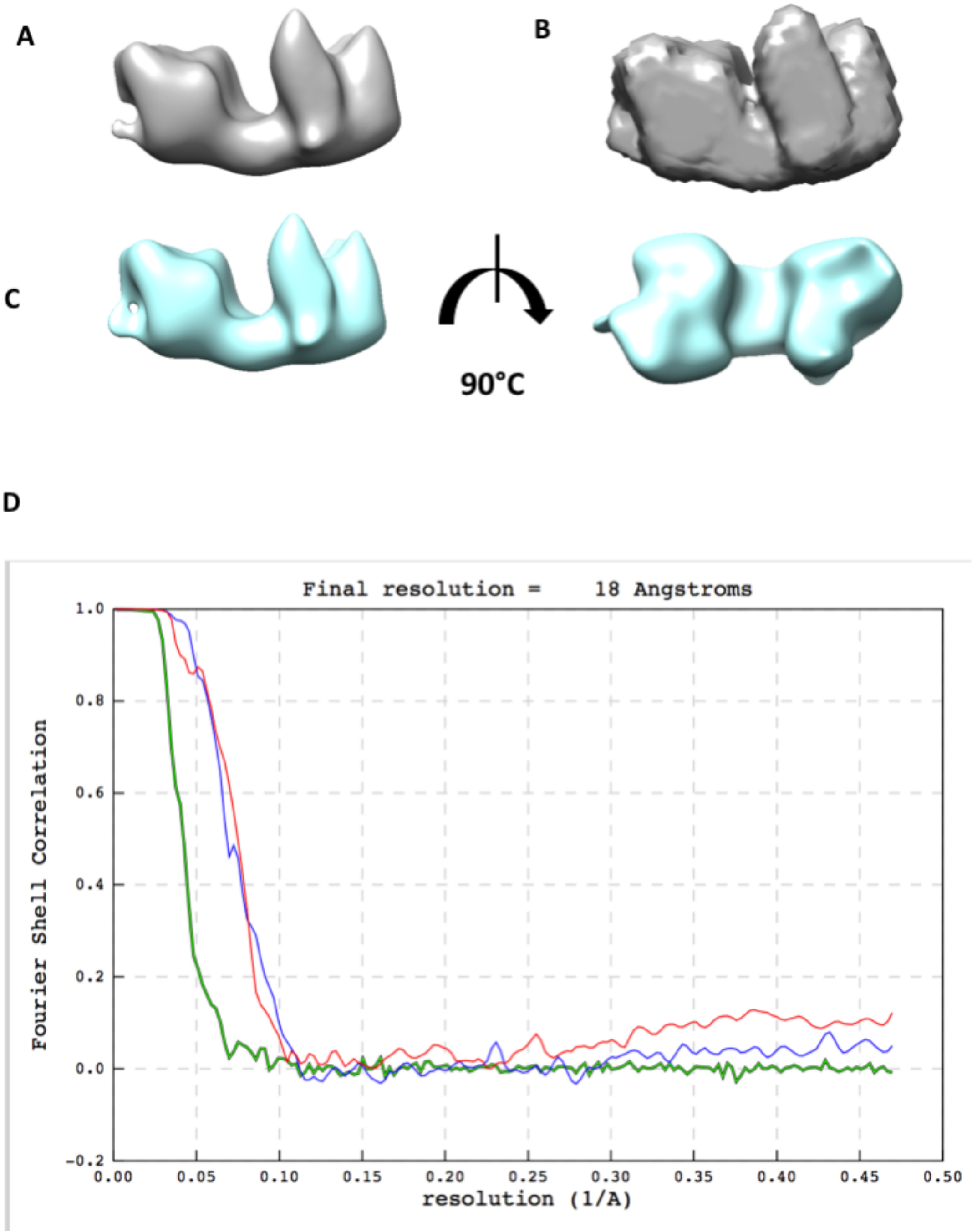


Figure 71: 3D refinement and post-processing of the core complex

[A] The refined map of the core complex. [B] The mask created using a binarization value of 0.0176, low pass filtered to 15 Å. [C] The post-processed map in different orientations. [D] The FSC curve generated after post-processing. Using a threshold value of 0.143, the resolution of the map was calculated at 18 Å.

4.7.2.3 3D refinement and post-processing of classes two and four

As classes two and four contained maps that closely resembled the ARISC model (chapter 4.7.2.1: Figure 70), they were selected for refinement. The maps were first refined individually. The refinement of class four generated a map that was missing BRCC45 and some MERIT40 density on the left arm of the complex. The output of the refinement gave an estimated resolution of 13 Å. The second map was also refined to an estimated resolution of 13 Å. This map was not missing any BRCC45 or MERIT40 density. Maps from classes two and four closely resembled each other and the particles contributing to those maps were combined and refined using map two as a reference. The refinement of the combined particles gave rise to a 3D map with two arms. The estimates resolution of this map was 11 Å. Masks of class two refined map and the combined refined map were then generated using a binarization threshold of 0.052. The masks were then used to mask the maps during post-processing. Post-processing of map two and the combined particle map generated maps with resolutions of 10 Å and 9.1 Å, respectively.

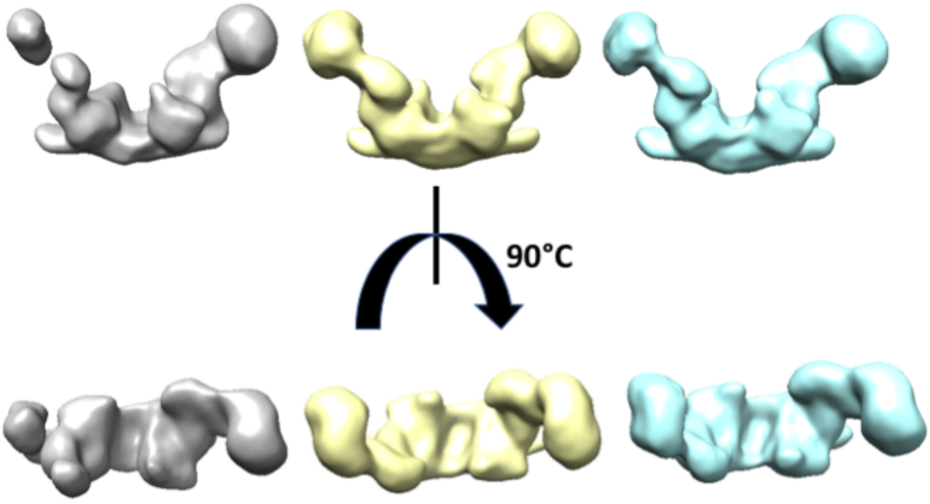


Figure 72: 3D refinement of class two and four particles

The particles contributing to class four (grey map) and two (yellow map) were refined to 13 Å. The particles were then combined and refined into a single map (cyan blue) using class two as reference. The resolution for the combined data was estimated at 11 Å.

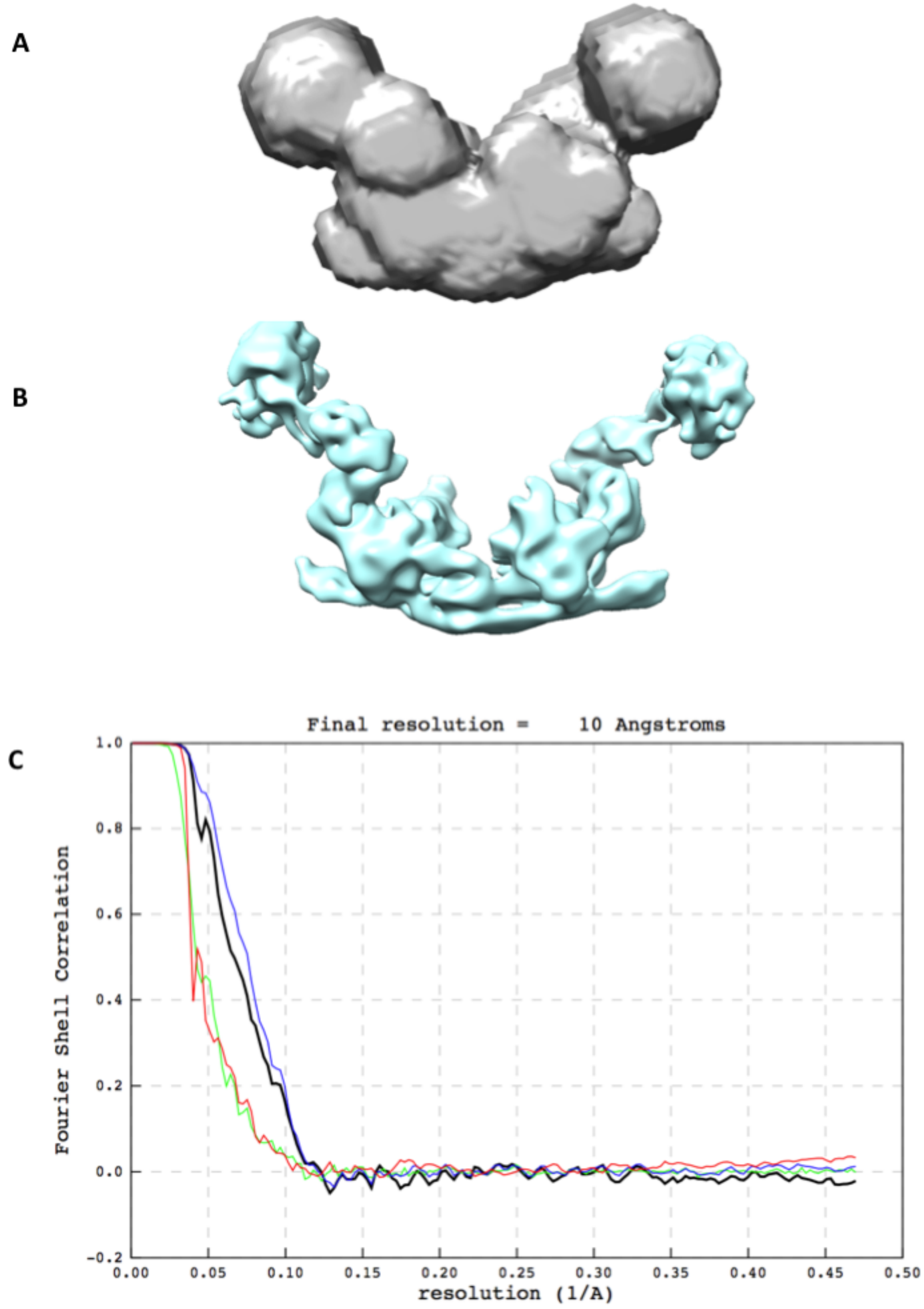


Figure 73: Post-processing of class two refined map

[A] The mask created using a binarization value of 0.052, low pass filtered to 15 Å. [B] The post-processed map. [C] The FSC curve generated after post-processing. Using a threshold value of 0.143, the resolution of the map was calculated to be 10 Å.

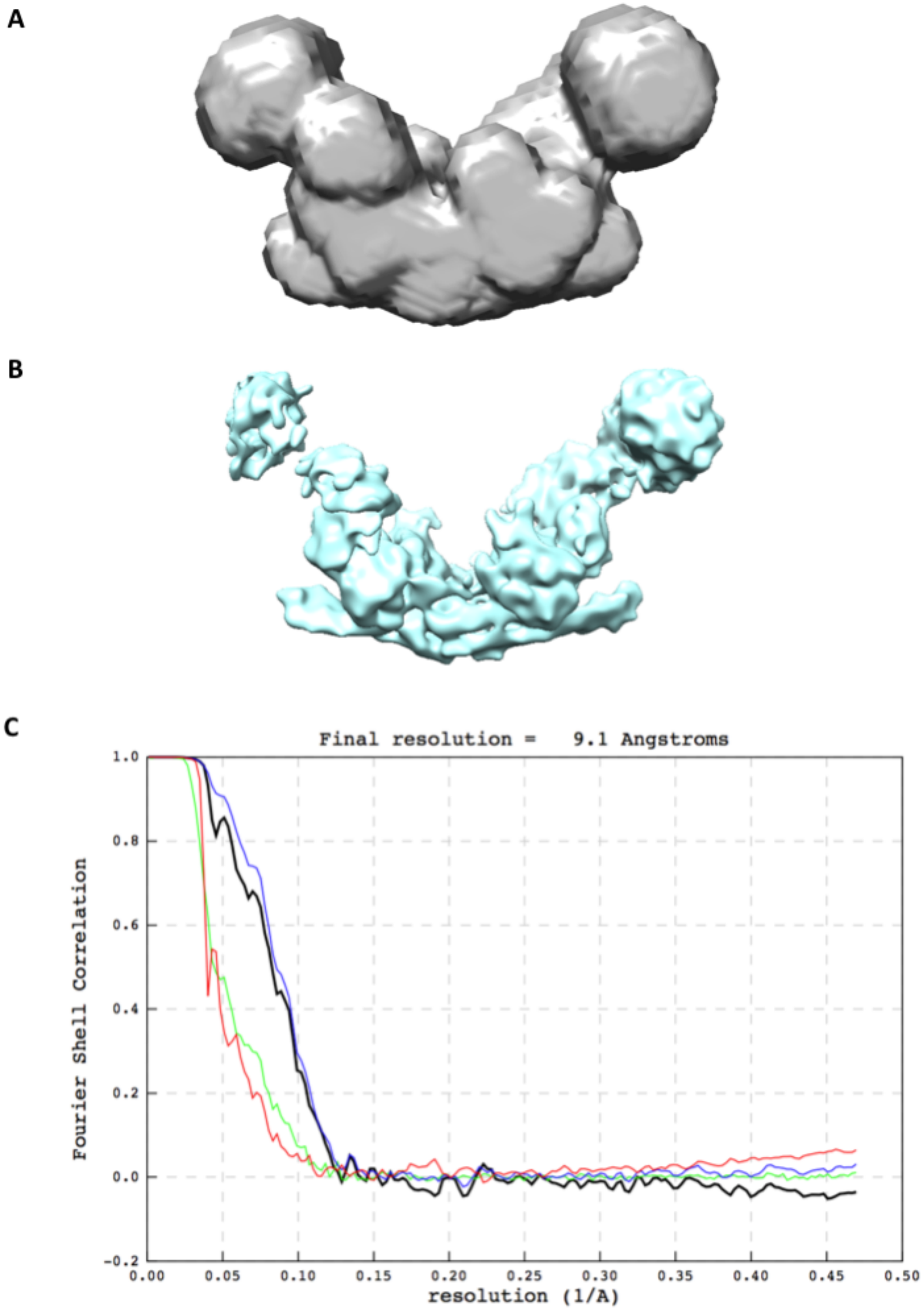


Figure 74: Post-processing of combined one and two arm refined map

[A] The mask created using a binarization value of 0.052, low pass filtered to 15 Å. [B] The post-processed map. [C] The FSC curve generated after post-processing. Using a threshold value of 0.143, the resolution of the map was calculated to be 9.1 Å.

4.7.2.4 3D classification of the ARISC dimer of octamer

To aid in the identification of the ARISC dimer of octamers, the ARISC dimer of octamer 2D class averages from cryoSPARC (Punjani, Ali et al., 2017) were selected and processed to ab-initio reconstruction in cryoSPARC (Punjani, Ali et al., 2017). Results from the ab-initio reconstruction produced a single map that showed one MERIT40 subunit of each complex interacts with the core Abraxas1-BRCC36 region of the other complex.

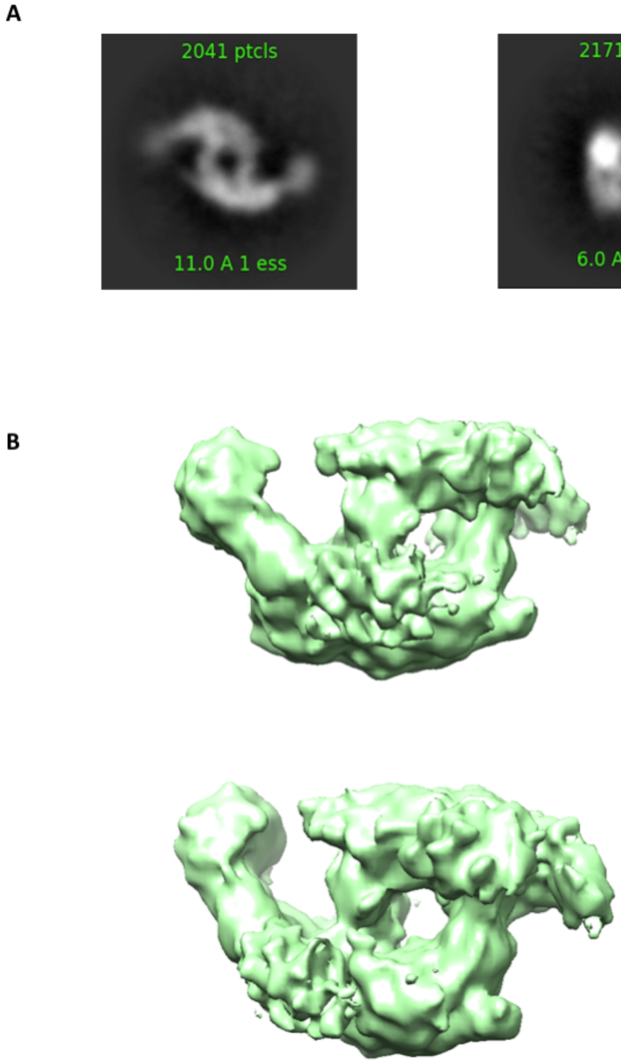


Figure 75: Ab-initio reconstruction of the ARISC dimer of octamer complex

[A] Representative 2D class averages used to generate the initial model. [B] Ab-initio reconstructed initial model of the ARISC dimer of octamers.

The ab-initio reconstructed dimer of the octamer model was lowpass filtered to 15 Å, clipped and scaled using the relion_image_handler. The filtered, clipped and re-scaled map was then used as reference for 3D processing in RELION (Scheres, 2012; Zivanov et al., 2018). Classes two, three, four, five, seven, eight, eleven and twelve from the first round of 3D classification (chapter 4.7.2: Figure 69) were selected for 3D classification to uncover the ARISC dimer of octamers. Results from the 3D classification produced four 3D maps (Figure 76B). The first map showed two complexes interacting with the MERIT40 and Abraxas-BRCC36 core complex. The second map resembled the first map but the second octameric complex was missing MERIT40 density. Maps three and four loosely resembled the first two maps but were disordered.

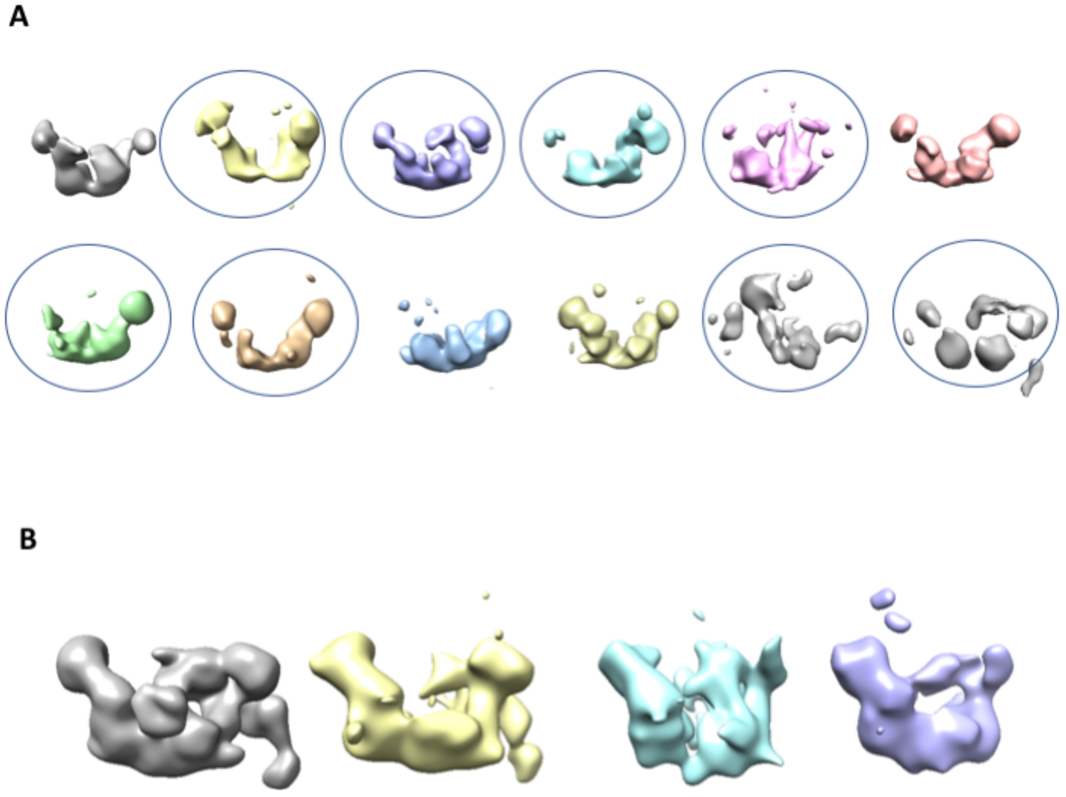


Figure 76: 3D classification to identify the dimer of octamer

[A] The first RELION 3D classification of the combined dataset. The maps circled were selected for a separate round of 3D classification using the dimer of octamer cryoSPARC (Punjani, Ali et al., 2017) initial model as reference. [B] The four 3D maps obtained from 3D classification. The first map (grey) resembles the initial model. The second map (yellow) also resembles the initial model but is missing density corresponding to the MERIT40 subunit. Maps three and four (cyan and purple) likely contained misaligned and junk particles.

4.7.2.5 3D refinement and post-processing of the dimer of octamers

Class one was selected and refined on its own to an estimated resolution of 16 Å. A mask was then generated using a binarization value of 0.123. The mask was then used to sharpen the map and exclude noise in the solvent region. Results from post-processing revealed a sharpened map with a calculated resolution of 16 Å. Analysis of the FSC curve shows that the black line corresponded to the masks map reaches the 0.143 threshold at 16 Å and again at around 10 Å. This suggests that the map (Figure 77) contains regions with lower and higher resolution.

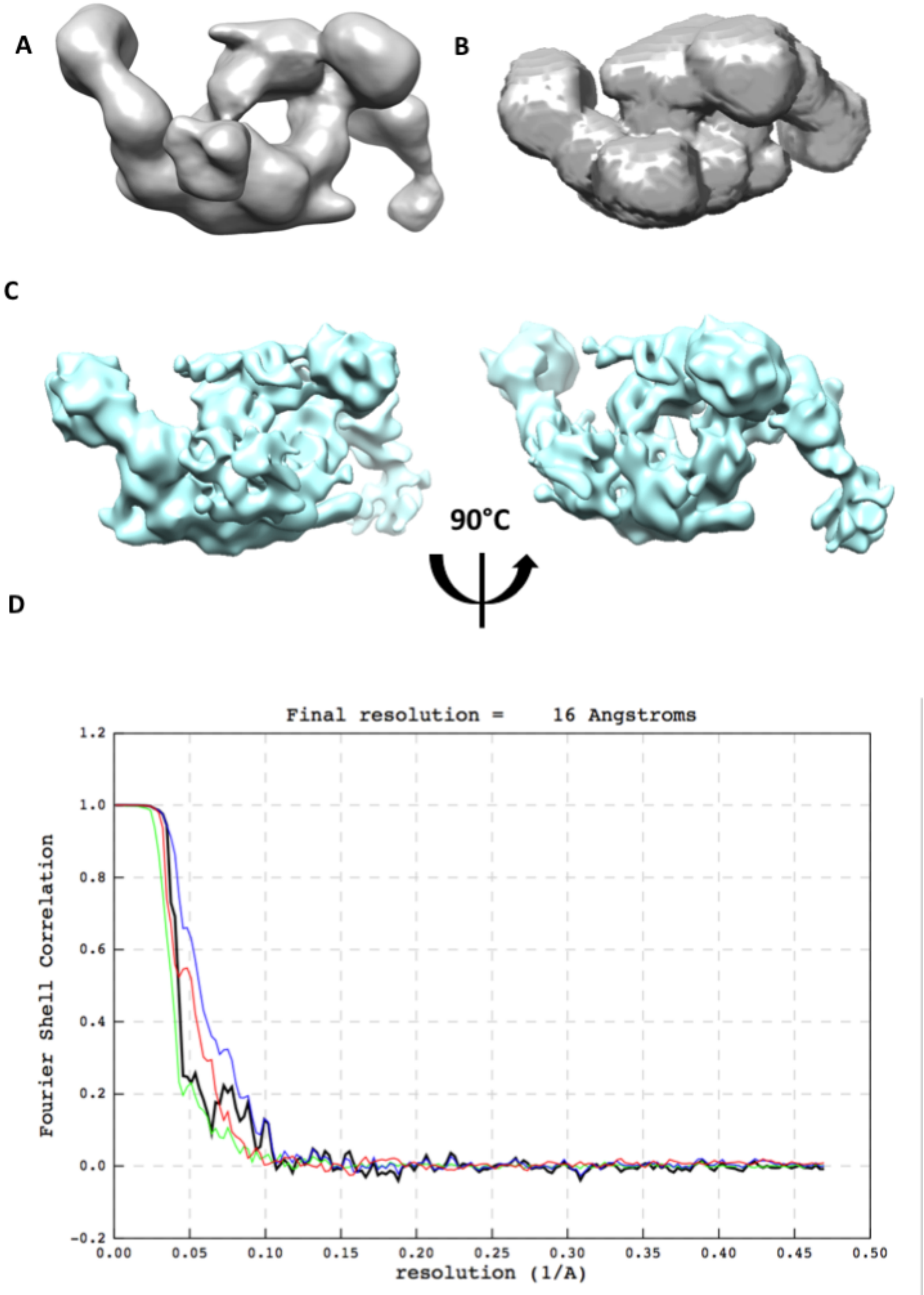


Figure 77: 3D refinement and post-processing of the dimer of octamer complex

[A] The mask created using a binarization value of 0.0123, low pass filtered to 15 Å. [B] The post-processed map in different views [C] The FSC curve generated after post-processing. Using a threshold value of 0.143, the resolution of the map was calculated at 16 Å.

4.7.2.6 3D refinement and post-processing of the unknown complex

Class one from the first 3D classification results of the combined dataset in chapter 4.7.2. was selected for further processing. The map was composed of 19,990 particles, which were refined against the map to a resolution of 26.6 Å. Following refinement, a mask of the complex was created using a binarization value of 0.0118. The mask was then used as an input in post-processing to sharpen the map and mask out solvent noise. The results from post-processing revealed that the map had not been sharpened and that the resolution was calculated using a threshold of 0.143 to be 27 Å

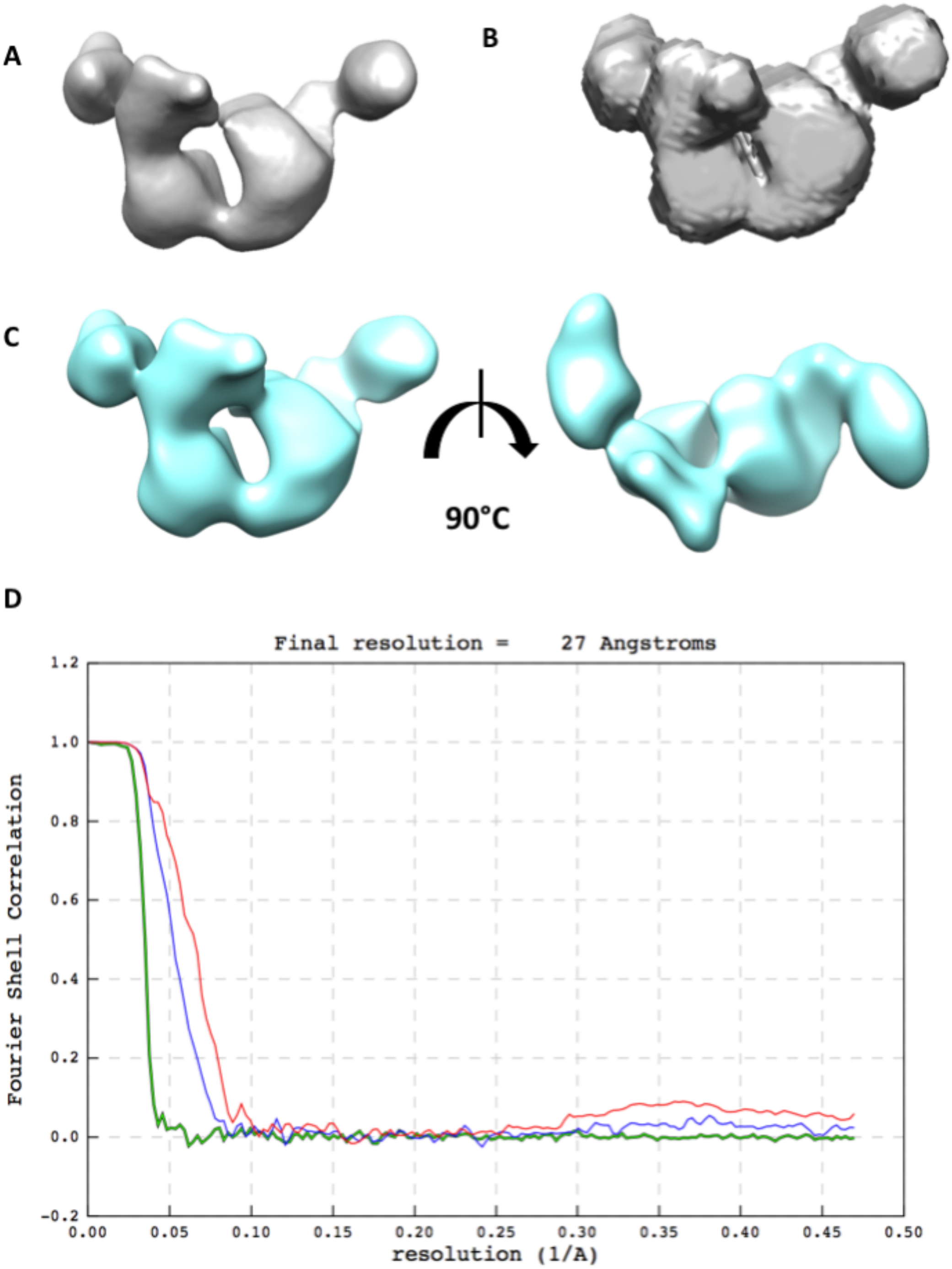


Figure 78: 3D refinement and post-processing of the unknown complex

[A] The refined map of the complex. [B] The mask created using a binarization value of 0.0118, low pass filtered to 15 Å. [C] The post-processed map showing front and top view. [D] The FSC curve generated after post-processing. Using a threshold value of 0.143, the resolution of the map was calculated at 27 Å.

4.8 Map interpretation and model fitting

4.8.1 ARISC

Five different structures were obtained from the cryo-EM analysis of the ARISC-RAP80 fraction 24 protein sample. The maps were all mid to low resolution, and therefore de novo model building was not possible. To validate the maps a structure of the BRISC complex obtained by removing SHMT2 from the published BRISC-SHMT2 structure (Walden et al., 2019) was fitted into the maps to try and identify densities not accounted for, and to identify any possible differences in the way the two complexes assembly. During the write up of this thesis, a 3.75 Å ARISC-RAP80 crystal structure was also published (Rabl et al., 2019). The crystal structure was also compared with the 3D maps presented in this thesis.

Fitting the BRISC structure into the class two map showed that the structure occupied the map well. However, there were some unoccupied densities in the MERIT40 and BRCC36 regions (Figure 79). The newly published ARISC-RAP80 AIR structure revealed that RAP80 AIR binds in the arm region of the complex, near MERIT40 (Rabl et al., 2019). To analyse whether the class two map contained RAP80, the map was fitted with the structure. Results of the fit shows that, while “the base” BRCC36-Abraxas1 complex was positioned in a similar way, the arm regions containing BRCC45 and MERIT40 were different (Figure 80). The arm regions in the map presented in this project are more elongated when compared to the published ARISC-RAP80 AIR structure (Figure 80). These differences made it difficult to accurately fit the published structure into the map. Analysis of the fit from the bottom view showed that Abraxas1 and BRCC36 fit in the core region of the map. A side view comparison of the map fitted with BRISC and the map fitted with the published ARISC-RAP80 AIR structure showed that MERIT40 from the BRISC complex accurately occupied the MERI40 density, however some areas remained unaccounted for (Figure 81). Conversely, MERIT40 from the published ARISC-RAP80 AIR structure was positioned lower on the map and did not accurately occupy the experimentally determined map from my cryo-EM studies. Together these results suggest that the cryo-EM class two map is of the ARISC complex and not the ARISC-

RAP80 complex. However, it is important to note that some of the differences observed could be due to the Rabl et al., 2019 structure being of the full length ARISC complex and the ARISC-RAP80 sample utilised for this project was ARISCΔNΔC.

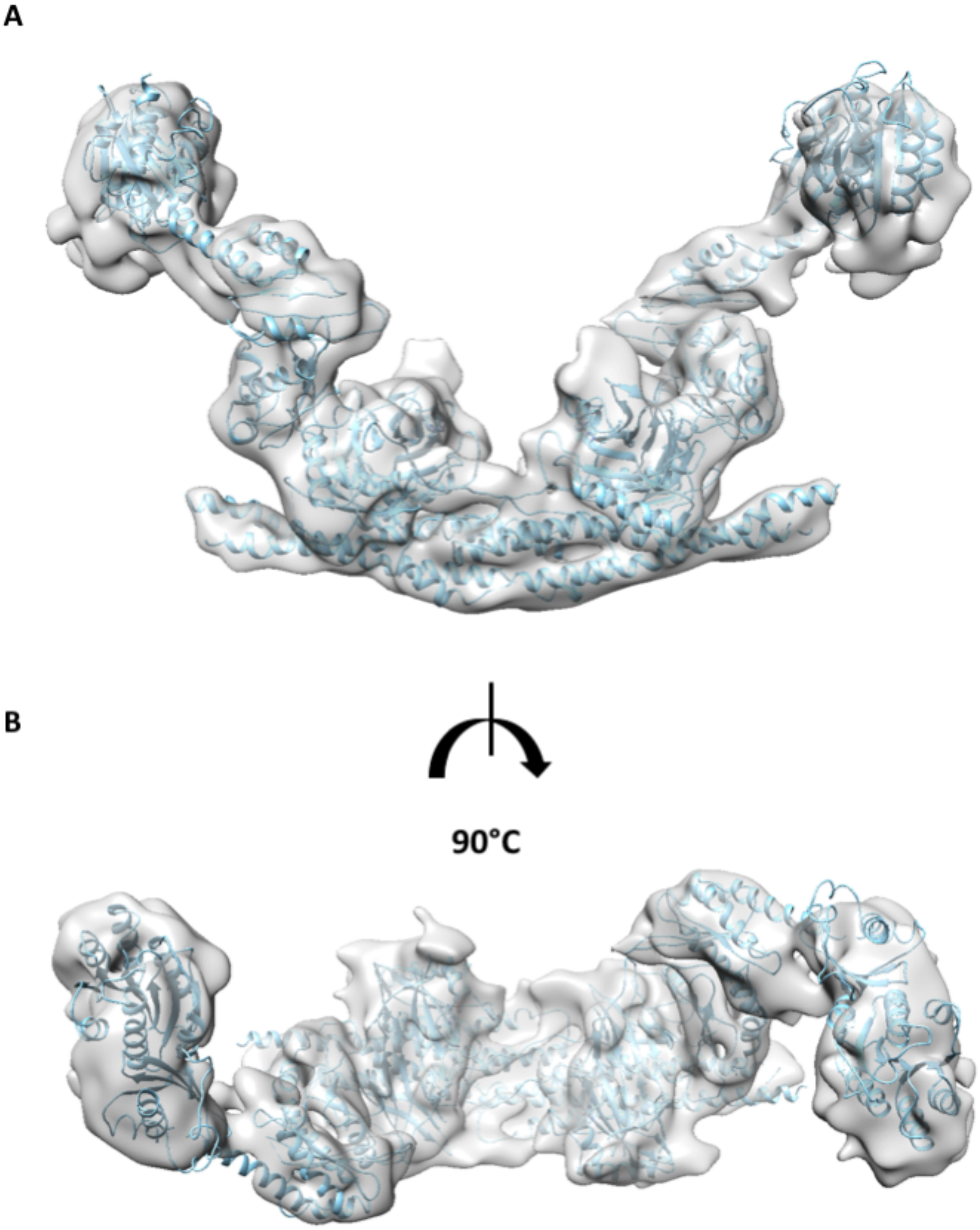


Figure 79: Fitting the BRISC structure into the class two map

[A] The post-processed class two map was fitted with the BRISC complex. The BRISC structure was obtained by subtracting the SHMT2 dimer from the BRISC-SHMT2 structure (Walden et al., 2019) (PDB:6R8F). The front view shows that the structure accurately occupies the majority of the map. [B] A top view of the fitted map. This view shows that MERIT40 of the BRISC structure correctly occupies MERIT40 of the map, and that the other subunits of BRISC (BRCC45, BRCC36, Abraxas2) also fit well within the map.

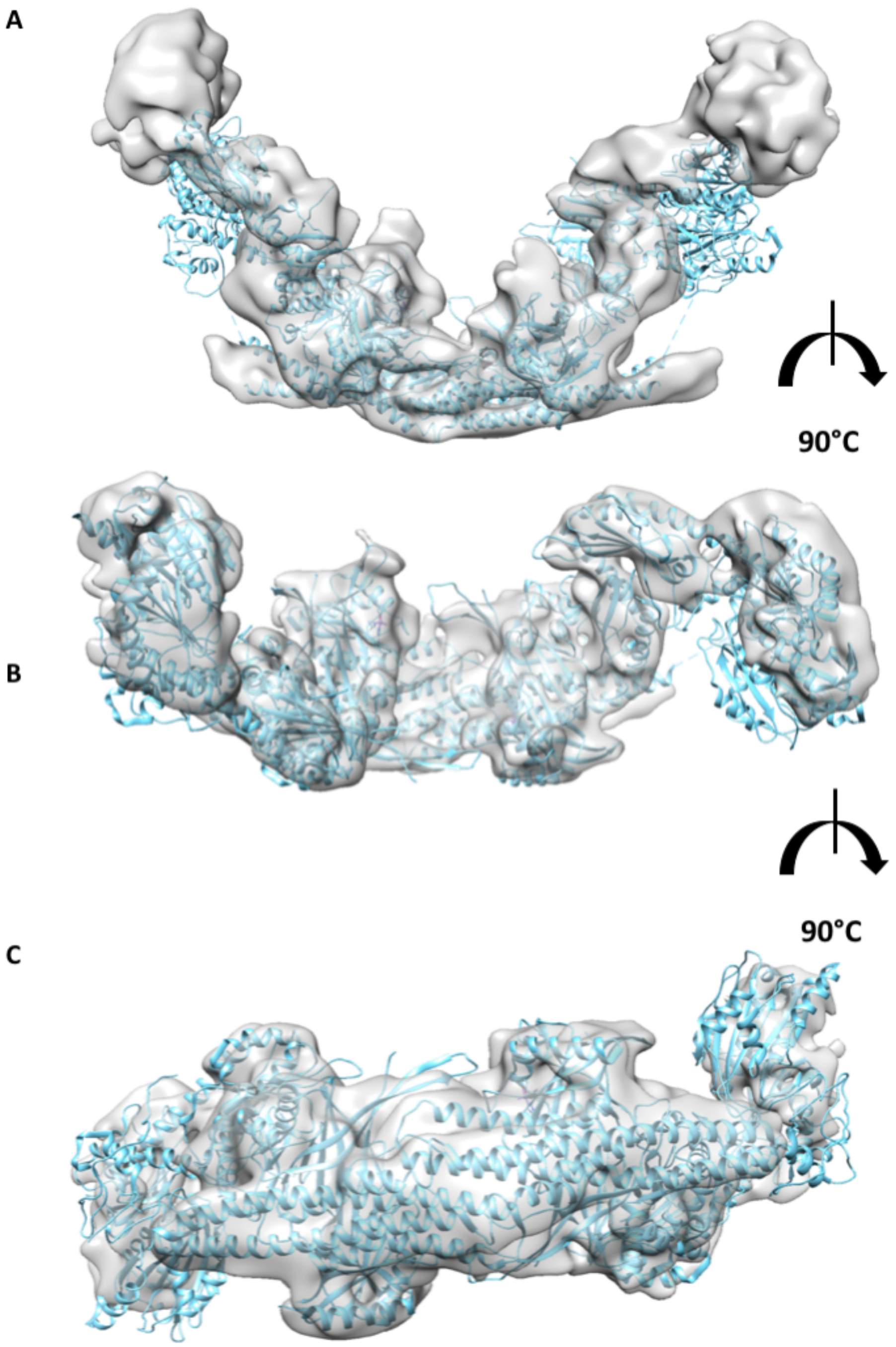


Figure 80: Fitting the structure of the mouse ARISC-AIR complex with class two

[A] A front view of the post-processed class two map fitted with the mouse ARISC-AIR complex (Rabl et al., 2019) (PDB:6gvw). [B] Top view of the ARISC-AIR complex fitted with the class two post-processed map. [C] A rotated bottom view showing the base of the map fitted with the mouse ARISC-AIR complex.

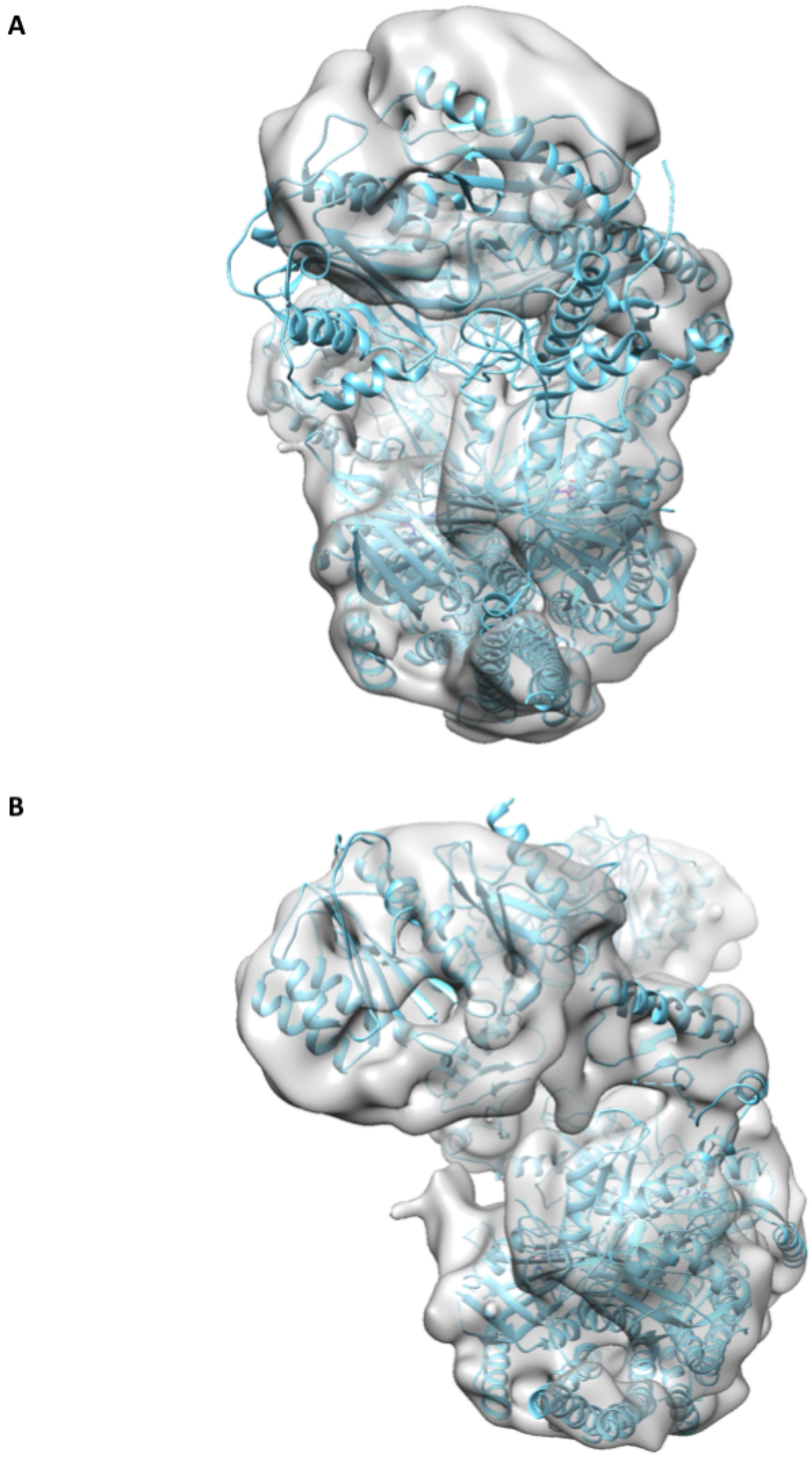


Figure 81: Side comparison of class two fitted with BRISC and ARISC-AIR

[A] A side view of the post-processed map of class two was fitted with the mouse ARISC-AIR crystal structure (Rabl et al., 2019) (PDB:6gvw). The fit revealed differences in the positioning of the BRCC45-MERIT40 heterodimer. [B] A side view of the post-processed map of class two fitted with the human BRISC structure.

4.8.2 ARISC dimer of octamers

Two copies of the BRISC structure were fitted into the post-processed dimer of octamer map. Fitting the two into the map showed that the BRISC model fitted well into both complexes. There were however some unoccupied densities, for example on MERIT40 on the right arm of the lower complex. Evaluation of the fitted map from another angle also identified more unoccupied densities around the BRCC45-MERIT40 area. The discrepancies between the structure and the map may be due to the differences in resolution and/or structural rearrangements between ARISC octamer and ARISC dimer of octamers. It is also possible that the extra density corresponds to structural differences between ARISC and BRISC (Figure 82). The map revealed that the self-dimerisation of the ARISC complex results in the occlusion of the BRCC36 active site, thus hindering the complex’s ability to deubiquitylate K63 chains.

4.8.3 Abraxas1-BRCC36 core complex

For the validation of the core complex map, the ant BRCC36-Abraxas2 crystal structure was fitted into the map (Zeqiraj et al., 2015). The fit revealed that the structure was accommodated within the map, suggesting that the map represents the BRCC36-Abraxas1 super dimer. There were, however, some unoccupied densities on the right side of the map near BRCC36 and the alpha helices of Abraxas2 (Figure 83).

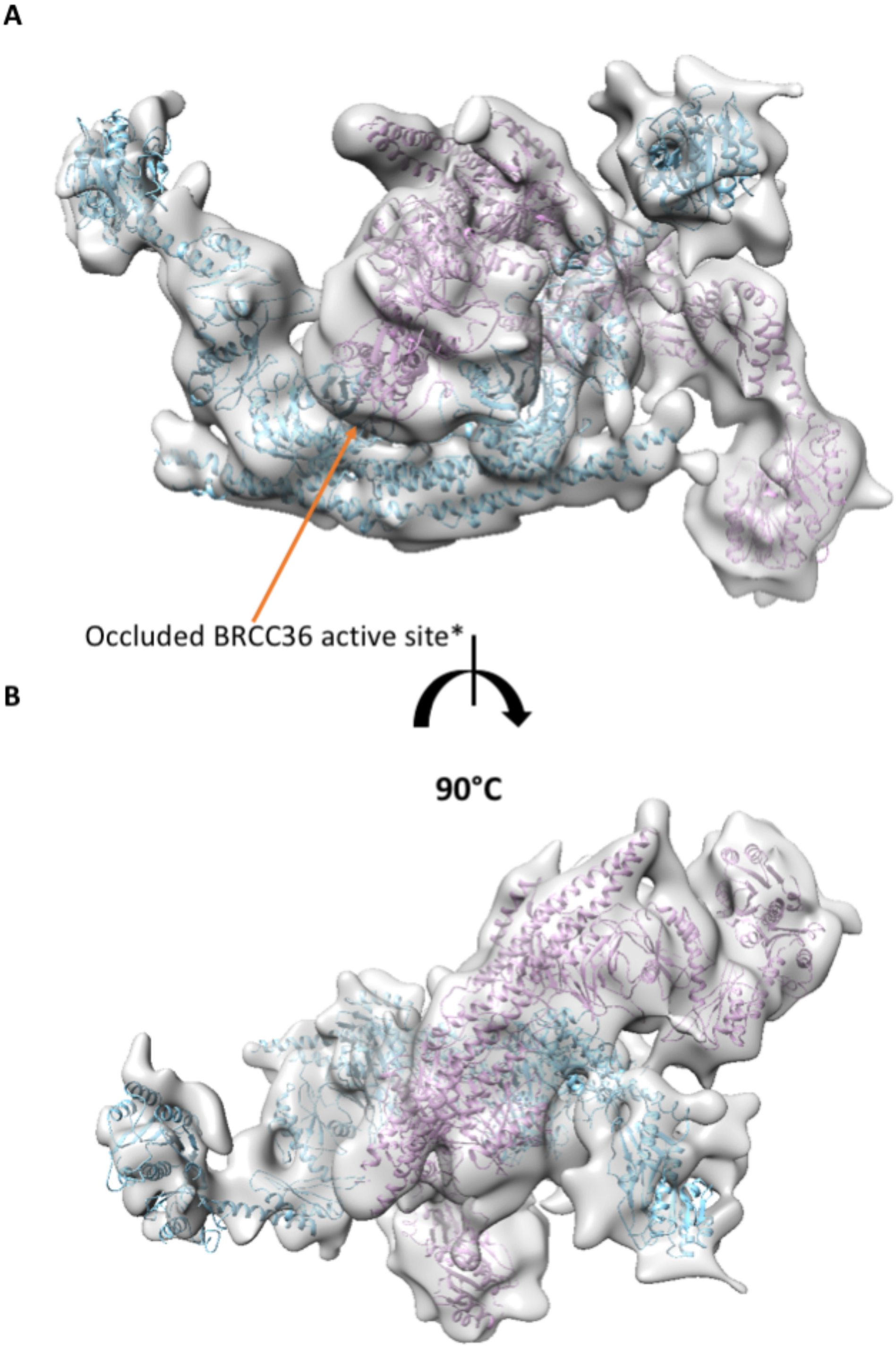


Figure 82: Fitting the structure of the BRISC complex with the dimer of octamer

[A] The post-processed 3D map of the dimer of octamer was fitted with two BRISC structures (PDB:6r8f). [B] A top view of the map. The fit showed that the BRISC structure accurately occupied both ARISC densities.

* An approximate location of the BRCC36 active site

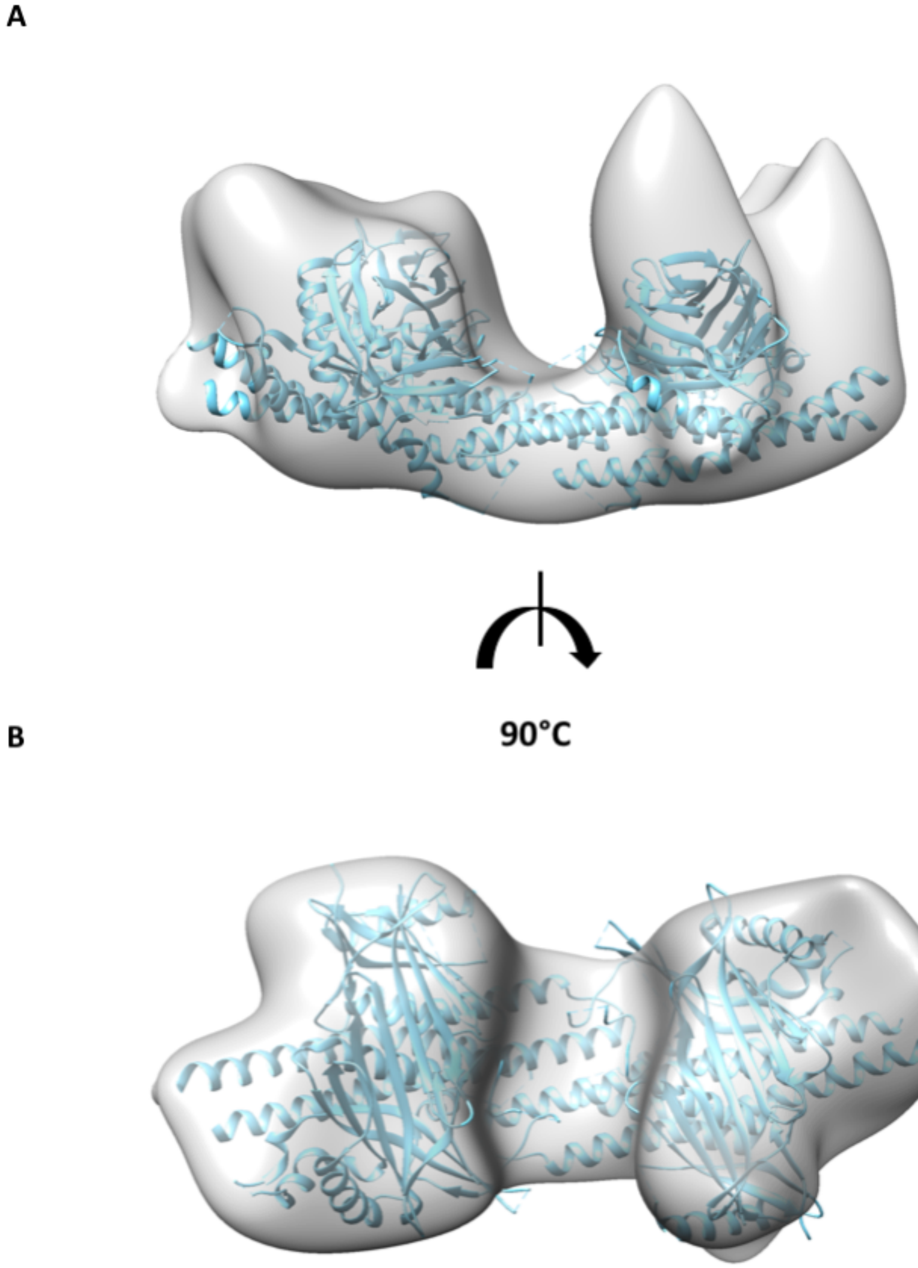


Figure 83: Fitting the structure BRCC36-Abraxas2 with the core complex

[A] The front view of the core complex fitted with the ant BRCC36-Abraxas2 complex (Zeqiraj et al., 2015) (PDB:5cw3). [B] A top view of the core complex fitted with the ant BRCC36-Abraxas2 complex.

4.8.4 ARISC-RAP80 potential structure

The class 1 map from the first round of 3D classification in chapter 4.72 is depicted as the ‘unknown structure’ due to the fact that despite its resemblance with the ARISC structure, it contains additional density regions that are unaccounted for. The fitting of the ‘unknown structure’ map with the structure of BRISC revealed that the BRISC MERIT40 subunits fits into the MERIT40 region of the map. However, the Abraxas2-BRCC36 core does not. This suggests that the complex has undergone significant conformational changes. In addition to this, there are large densities unaccounted for near BRCC45 of the left and right arm. The view of the fit from the back shows that the unknown density spans across one BRCC45 subunit to the other, but only on one side of the complex, leaving BRCC36-Abraxas1 exposed (Figure 84). To evaluate whether the unaccounted density is the RAP80 protein, the map was also fitted with the published ARISC-AIR structure from mouse (Rabl et al., 2019) (Figure 85). Results of the fit showed that only the BRCC45-MERIT40 subunits fitted in the map. Like with the BRISC fit, the positioning of the Abraxas1-BRCC36 super dimer was not well aligned in the map. Additionally, there were densities near the arm and core of the complex that were unaccounted for. It is possible that the RAP80 binds to the complex in a manner that elicits a conformational change at the core of the complex.

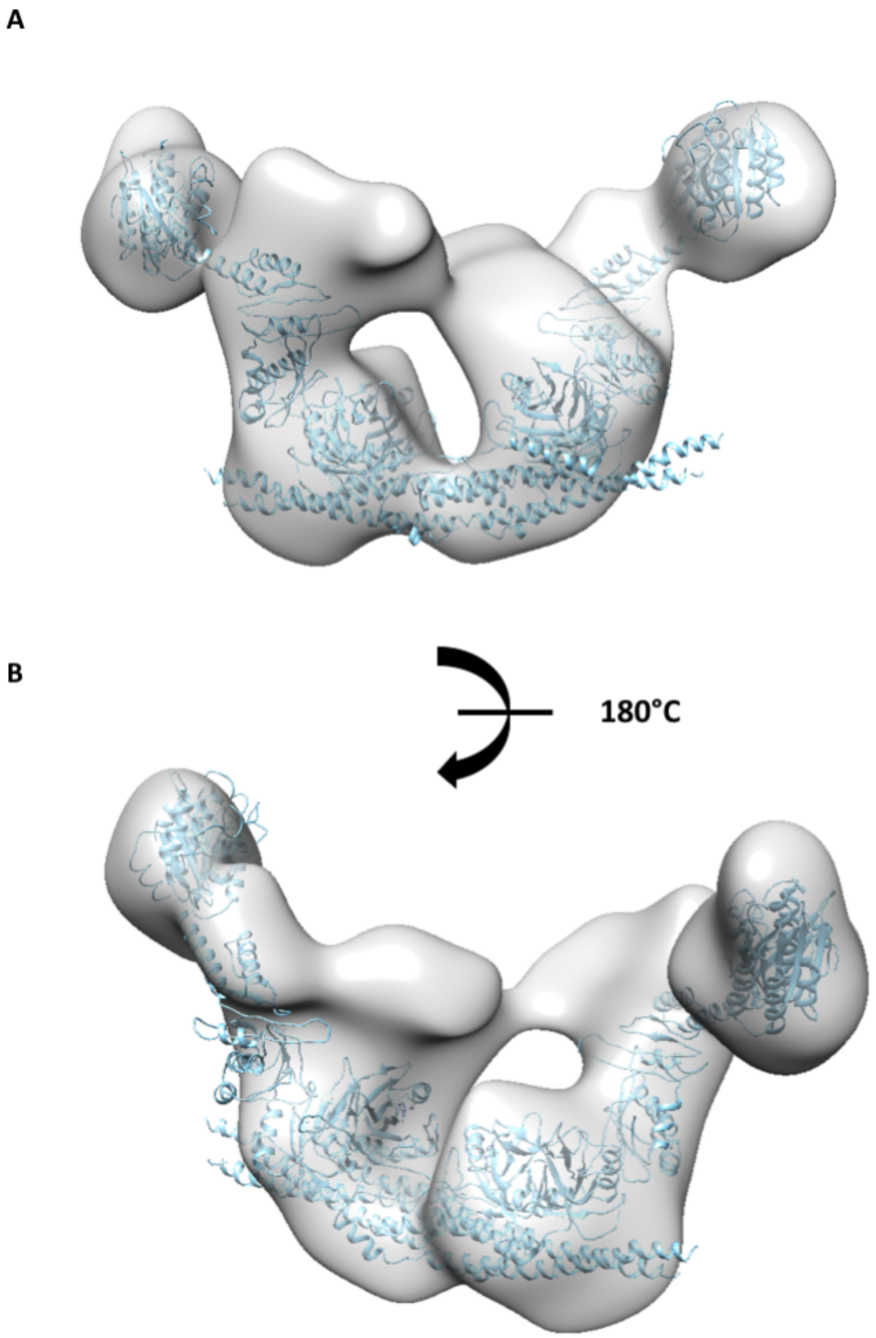


Figure 84: Fitting the structure of the BRISC complex with the unknown complex

[A]A front view of the unknown complex fitted with the cryo-EM BRISC structure (PDB:6r8f). [B] The back view of the unknown complex fitted with BRISC structure.

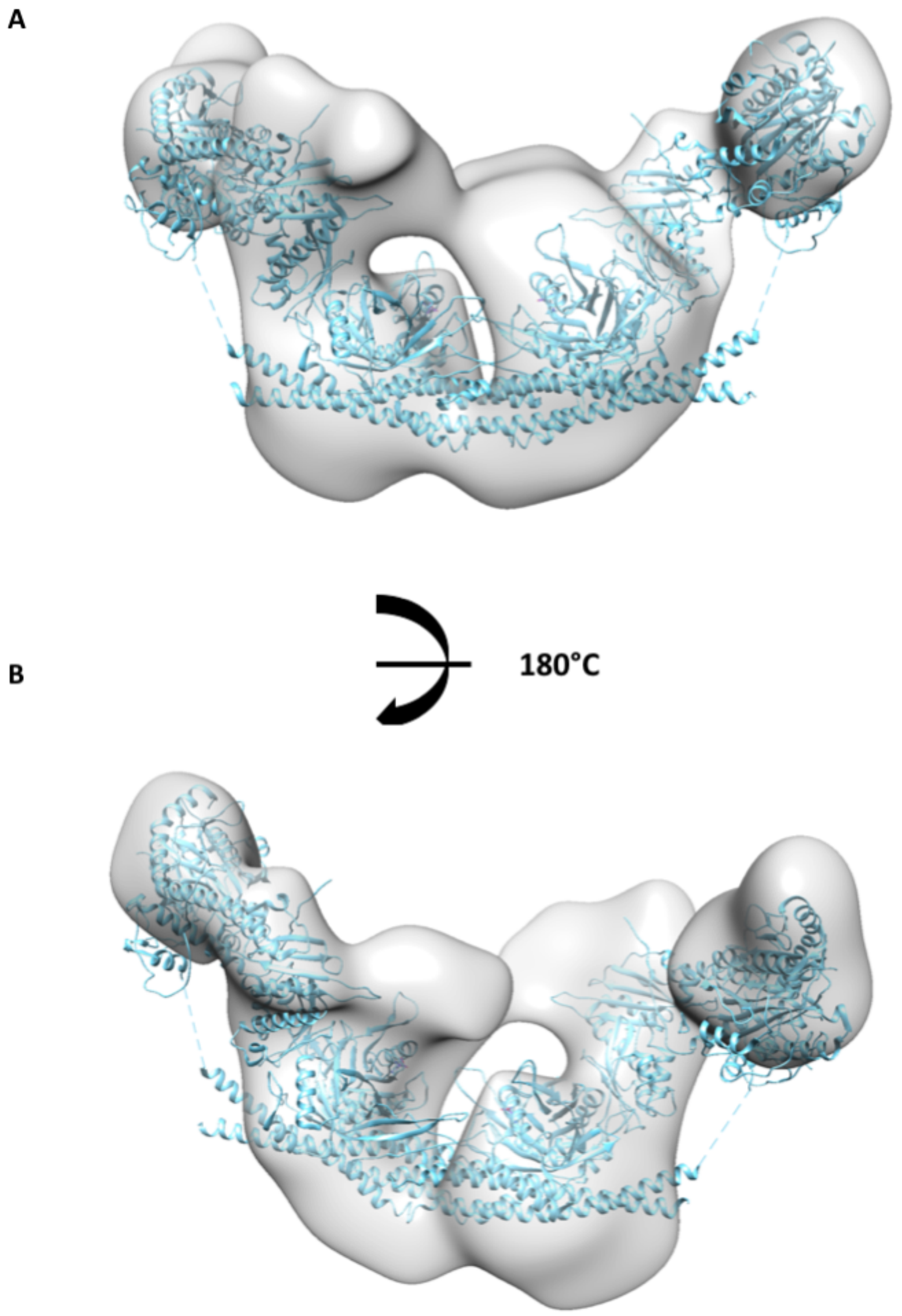


Figure 85: Fitting the structure of the mouse ARISC-AIR complex with the unknown complex

[A] A front view of the unknown complex fitted with the mouse ARISC-AIR crystal structure (Rabl et al., 2019) (PDB:6gvw). [B] The back view of the unknown complex fitted with the ARISC-AIR complex.

4.9 ARISC-RAP80 substrates and complex composition

Chapter 4.9 aims to determine whether ARISC-RAP80 is capable of binding DNA via the zinc finger domains of the RAP80 protein, whether BRCC36, Abraxas1, BRCC45 and MERIT40 have DNA binding capabilities and if it's ability to bind DNA is dependent on the length of DNA. Preliminary negative stain and cryo-EM data were also collected in parallel (data not shown) to obtain structural information of how the complex binds DNA. The negative stain EM data were processed to 2D classification, however, the data revealed that in the presence of DNA the complex dissociated. Cryo-EM analysis of the ARISC-RAP80 complex allowed for the identification of DNA particles but not the complex, confirming that ARISC-RAP80 dissociated. Future experiments are needed to try and stabilise the interaction of the complex with DNA for structural studies. This chapter also explores the complex's ability to bind ubiquitin. Complex cross-linking with di-ubiquitin was carried out to determine whether the binding of the substrate would stabilise the complex. Preliminary negative stain EM analysis of the ARISC-RAP80 complex with non-cleavable di-ubiquitin was carried out, however, in the presence of the substrate, the complex dissociated. This data is not presented in this thesis. The section then concludes with the analysis of the ARISC complexes on a Native PAGE gel. Analysis by Native PAGE was carried out to understand whether the heterogeneity observed during negative stain EM, cryo-EM and mass spectrometry analysis was reflective of the composition of the protein sample.

4.9.1 ARISC-RAP80 and 300 bp DNA

Gel shift assay using 300 bp DNA revealed that at 10, 5 and 2.5 µM ARISC-RAP80 was capable of strongly binding DNA. In addition to this, it showed that the ARISC-AIR gel lanes were comparable to the DNA control lane, suggesting that the ARISC-AIR complex did not bind DNA. The gel also revealed that 10 µM ARISC was capable of weakly binding DNA, and that the complex's ability to bind DNA correlated with a higher protein concentration (Figure 86). Collectively, these results suggest that RAP80 is the main protein subunit responsible for facilitating the interaction between the ARISC-RAP80 complex and DNA, and that Abraxas' contribution to the interaction is weak.

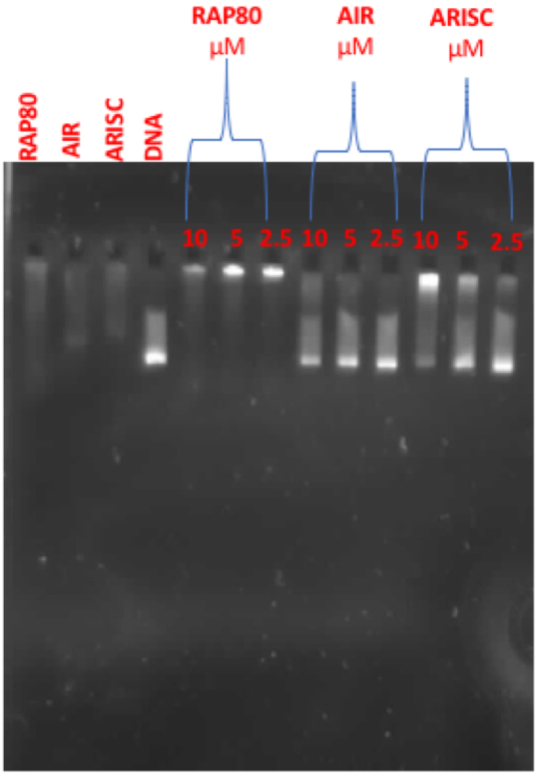


Figure 86: Gel shift assay of the ARISC complexes with 300 bp DNA

ARISC, ARISC-AIR and ARISC-RAP80 were incubated with 300 bp DNA for 2 hours. Following incubation, samples were run on a 2.5% agarose gel. Results from the assay show that at 10, 5 and 2.5 µM ARISC-RAP80 is able to strongly bind DNA. The gel also shows that at a protein concentration of 10 µM, the ARISC complex binds DNA, and that its ability to bind DNA correlates with protein concentration. The ARISC-AIR complex did not bind DNA.

4.9. 2 ARISC-RAP80 and 50 bp DNA

Gel shift assay using 50 bp DNA revealed that ARISC-RAP80 retained its DNA binding ability and that ARISC did not. The BRCC45-MERIT40 heterodimer was also tested for its DNA binding capabilities. Results from the assay indicates that the BRCC45-MERIT40 heterodimer does not bind DNA. Glycogen synthase was used as a control as it is known to not interact with DNA. The protein only controls in the 300 bp gel shift assay (Figure 86) all contained minimal background signals, whereas protein only controls in the 50 bp gel shift assay (Figure 87) did not.

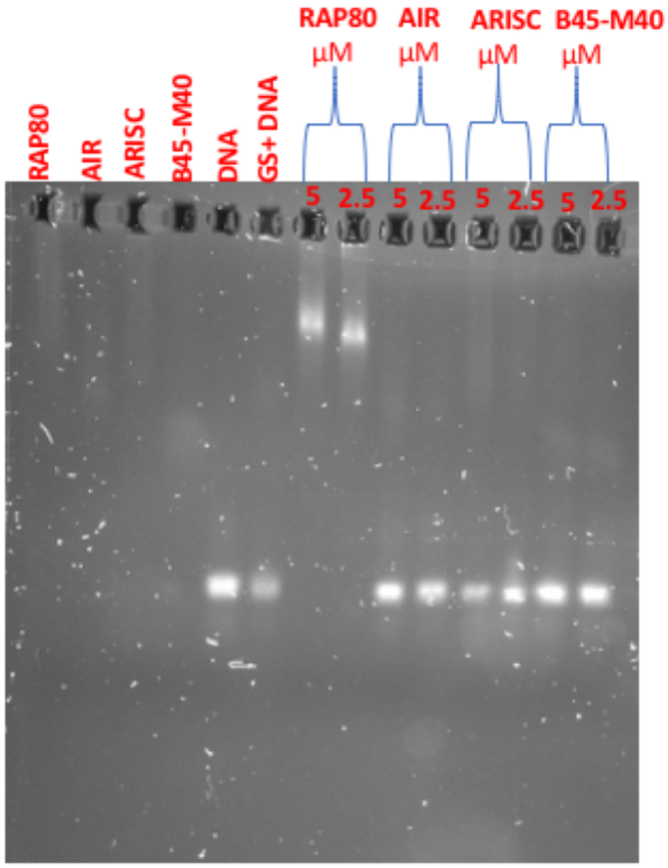


Figure 87: Gel shift assay of ARISC complexes with 50 bp DNA product

Six controls consisting of protein only, DNA only and glycogen synthase and DNA. The protein only controls showed that there was no DNA present in the sample. The glycogen synthase and DNA control showed that glycogen synthase does not bind DNA. Results from the ARISC-RAP80 DNA reaction showed that at 5 and 2.5 µM, the complex was able to bind DNA. ARISC-AIR, ARISC, and BRCC45-MERIT40 were all unable to bind 50 bp DNA.

4.9.3 Cross-linking photoreactive Ubiquitin reveals ubiquitin binds to RAP80

Cross-linking experiments were carried out using the UbiQ photoreactive K63 mono-ubiquitin and the ARISC-RAP80 complex to determine which ARISC-RAP80 subunit is capable of binding mono-ubiquitin. As the RAP80 protein contains two ubiquitin interacting motifs and sumo interacting motifs which have been shown to be ubiquitylated, it is possible that the protein can bind up to four ubiquitin moieties. Additionally, the BRCC36 active subunit is capable of recognising and cleaving ubiquitin and the BRCC45 subunit contains ubiquitin E2 variants (UEV) domains. Together, it’s possible that the ARISC-RAP80 complex can bind up to 10 ubiquitin molecules with up to four binding to each RAP80 protein, two binding to BRCC45 and a single ubiquitin molecule binding to BRCC36.

The protein sample was incubated with ubiquitin for 30 minutes at 30 °C in the dark and then placed in a light source and UV-irradiated for 30 minutes at 355 nm. Samples were taken at a 30 minute interval for 90 minutes, and run on a 4-20% gradient gel. Results from the gel shows that RAP80 is the only subunit that bound the photoreactive monoubiquitin (Figure 88). At 30 minutes, there is a clear shift in the experimental RAP80 band compared to the control. At the 60 minute mark, multiple RAP80 bands are formed, suggesting that RAP80 binds multiple ubiquitin moieties. At 90 minutes, there are no additional bands formed. The amount of free mono-ubiquitin in the gel decreases over time as more protein binds to RAP80. The experimental protein bands corresponding to Abraxas1, BRCC36, BRCC45 and MERIT40 were the same as the respective controls at each time stamp indicating that they did not bind ubiquitin.

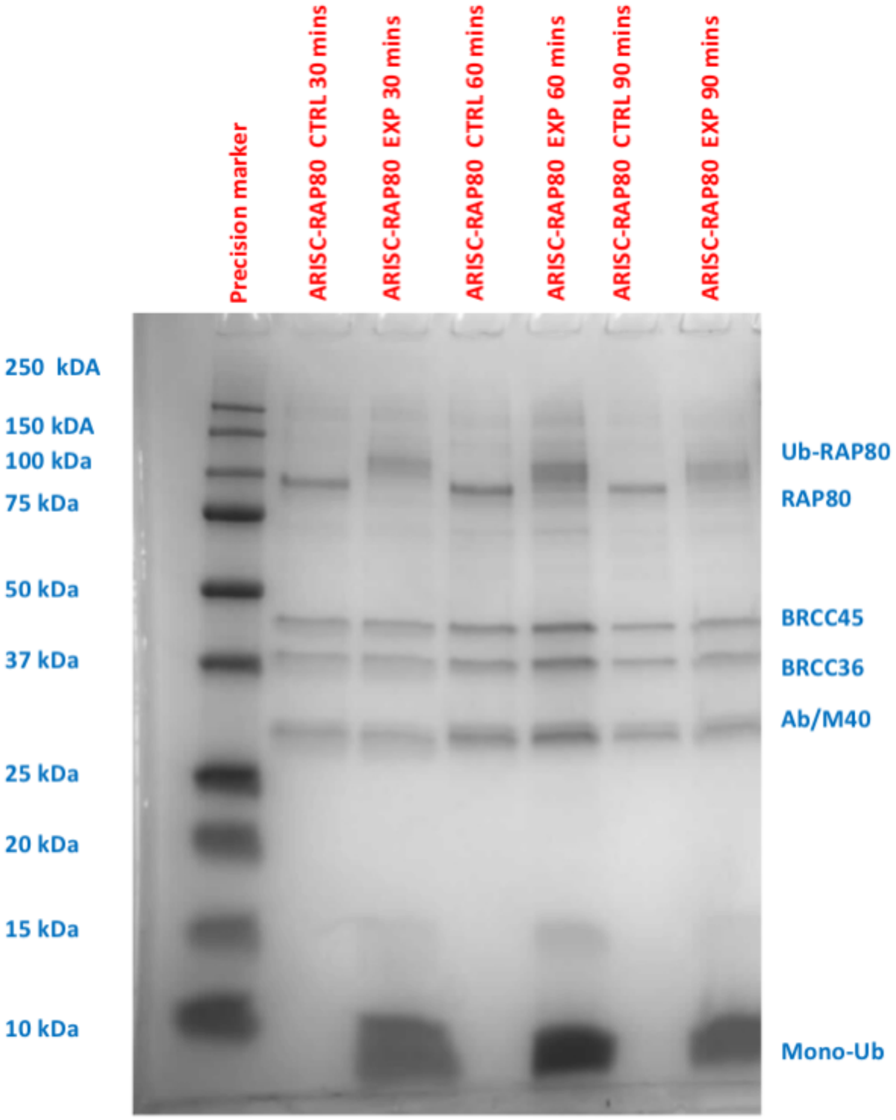


Figure 88: RAP80 is the only subunit capable of binding K63 photoreactive ubiquitin

The cross-linking results of the ARISC-RAP80 complex and photoreactive k63 ubiquitin shows that RAP80 was the only subunit bound to ubiquitin. At the 30 minute timestamp RAP80 been successfully cross-linked to ubiquitin. With increased exposure to UV, more ubiquitin was cross-linking to RAP80 and there were less free mono-ubiquitin present on the gel. The intensity of the RAP80-ubiquitin band was stronger at the 60 minute time stamp and multiple bands were present, suggesting that multiple ubiquitin molecules were bound to RAP80. The experimental lane at 90 minutes also shows that ubiquitin is bound to RAP80, however the intensity of the bands are weaker. The Abraxas1, BRCC36, BRCC45 and MERIT40 bands are the same in both the control and the experimental lanes.

4.9.4 Analysis of the ARISC complexes by Native PAGE

SDS-PAGE gels were employed as a means to confirm the presence of each protein subunit following protein purification by gel filtration. Results from the SDS-PAGE gel showed that all the desired subunits were present in the sample. SDS-PAGE gels are run under reducing conditions, it is possible that whilst the proteins are present in the sample, they do not form the same complex. Native mass spectrometry and EM analysis of the complexes revealed heterogenous subunit composition. Both methods identified subcomplexes and the ARISC dimer of octamer. The ARISC dimer of octamers was identified during mass spectrometry through the identification of species with the molecular weight of 558 and 561 kDa (chapter 3.6.2). Cryo-EM 2D and 3D reconstruction of the ARISC-RAP80 complex revealed two interacting ARISC complexes, confirming the presence of the ARISC dimer of octamers (chapter 4.7.2.5 and 4.8.2). A 4-16% Native PAGE gel was run with aims to try and determine whether the subunits exist in various structures or form a single complex in solution.

The complexes were analysed under two different concentrations: 0.5 and 0.008 mg/ml. These concentrations reflected the concentrations used to for cryo-EM and negative stain EM analysis. Results from the Native PAGE gel reveals that at 0.5 mg/ml the ARISC complex has three bands, the lowest band at approximately 200 kDa, the second at around 480 kDa and the highest band situated at approximately 600 kDa. When diluted to 0.008 mg/ml only the band at approximately 480 kDa is visible (Figure 89). The ARISC-AIR bands at 0.5 and 0.008 mg/ml corresponded to the ARISC bands at the same concentrations. The bands at 200 kDa likely represent the Abraxas1-BRCC45-MERIT40 dimer shown in chapter 3.6.3 and the band at approximately 600 kDa the ARISC dimer of octamer. Interestingly the band at 480 kDa is present in all the protein samples.

During the course of this project, multiple protein preps were utilised. For simplicity, all protein preps with the exception of the ARISC-RAP80 fraction 24 prep is referred to as “old”. “Old” ARISC-RAP80 complex protein that was used to obtain negative stain EM results was also analysed. The gel showed that the complex was extremely heterogenous with distinct bands at approximately 100 kDa, 200 kDa, 480 kDa, and approximately 500 kDa. In addition to these bands, the sample generated a large smear that span the gel from 66 kDa to 1,048 kDa, suggesting high levels of protein aggregation. Freshly purified ARISC-RAP80 samples discussed in chapter 3.5 were also analysed. Fraction 25 and 26, although not used for analysis in this project were also run to determine whether the sample contained the same protein complexes as fraction 24. Results from running the sample suggests that the two fractions contained different ARISC-RAP80 protein composition. At 0.5 mg/ml fraction 25-26 contained a high intensity band at around 480 kDa, and smears below it at 66 kDa, 200 kDa. When diluted to 0.008 mg/ml a single distinct band was visible at 480 kDa. At 0.5 mg/ml fraction 24 had two main protein bands at approximately 480 kDa and 720 kDa. The band at 480 kDa possibly corresponds to the ARISC-RAP80 complex and the 720 kDa band the ARISC dimer of octamers. At 0.008 mg/ml, the ARISC-RAP80 fraction 24 bands were fainter due to lower protein concentration.

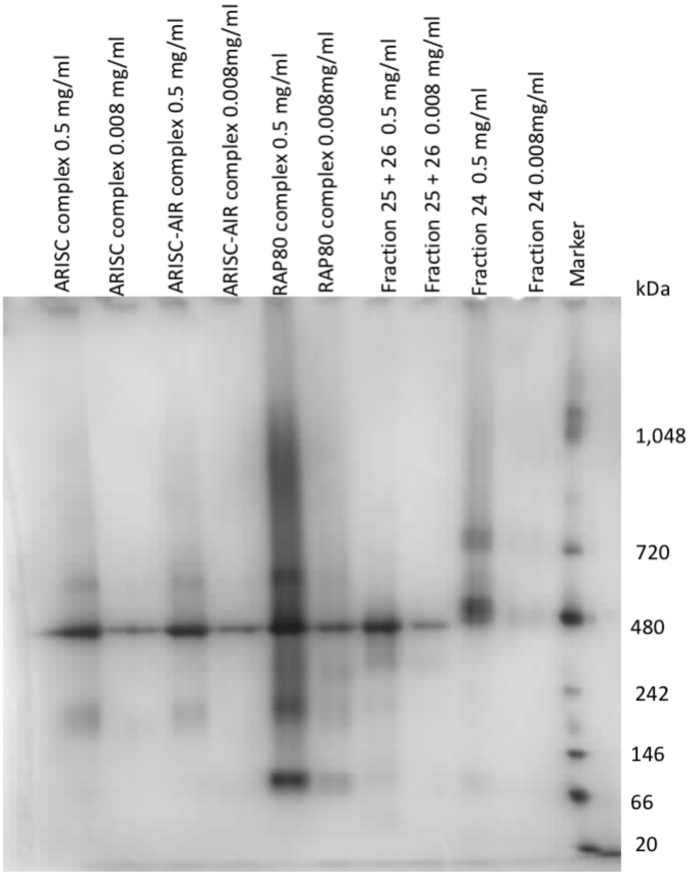


Figure 89: Native gel analysis of the ARISC complexes

The coomassive stained Native PAGE gel (4-16%) revealed that all the samples were heterogenous. The ARISC and ARISC-AIR complex contained the same amount of heterogeneity with protein bands at approximately 200, 480 and 500 kDa. The ARISC-RAP80 complex was the most heterogeneous as it had protein band smear that ranged from approximately 100 to 1,048 kDa. ARISC-RAP80 fractions 25-26 and 24 had protein bands that corresponded to approximately 100 kDa, 450 kDa, 480 kDa, and 480 kDa and 720 kDa respectively.

4.10 Discussion

At a protein concentration of 0.08 mg/ml ARISC-RAP80 particles were not visible on the micrographs. The inability to see particles at that concentration may be due to complex dissociation. Negative stain EM data presented in chapter 3 indicate that the ARISC complexes are susceptible to dissociation. It is possible that the ARISC-RAP80 complex more readily dissociates under cryo-EM because unlike negative stain EM, the particles were not fixed in position on the grid. Increasing sample concentration to 0.2 mg/ml resulted in the identification of protein particles. However, at that protein concentration the particles were overcrowded, making single particle analysis difficult. Nevertheless, data was collected and processed to 3D initial model. Data processing in RELION 3.0 (Zivanov et al., 2018) generated class averages that contained high levels of background noise and neighbouring particles further confirming that the protein concentration was too high for analysing the sample. Interestingly the initial model obtained from the dataset displayed the general U shape architecture of the ARISC complex, despite this the model was still unreliable as the 2D class averages were difficult to interpret.

Various researchers have reported success using TEM grids made with other materials such as gold (Russo and Passmore, 2014). To this end, ARISC-RAP80 grids were made using protein concentrations of 0.95 mg/ml, ultrAUFoil, and lacey carbon grids. A high protein concentration of 0.95 mg/ml was used to see protein particles and to help determine by how much the samples should be diluted for data collection. As expected, results from screening both the ultrAUFoil (gold) and lacey carbon grids showed overcrowded protein particles. Particles on the ultrAUFoil grids had slightly better contrast, which could be due to the fact that the grid was less crowded than the lacey grid. Despite the same protein concentration being used on both grids, gold grids contained fewer protein particles than the lacey grids, since unlike the lacey grids, gold grids likely have less of an attraction to particles than carbon.

Following screening, gold grids were selected as the grid of choice for downstream screening and data collection. Successful data collection was achieved using protein samples diluted to 0.5 mg/ml. A total of four datasets were collected over a five month period. During screening and data collections, particles were only visible and collected in regions with thicker ice. It is possible that thicker ice helped to better support the complex. Three different glow discharge machines were used to introduce negative charge to the grid surface, making them hydrophilic. Out of the three machines, the QUORUM generated the most consistent results.

4.10.1 Data processing in cryoSPARC and RELION

Processing the datasets using two different programs enabled the 3D reconstruction of the complexes presented in this thesis. Both programs had their advantages and disadvantages with regards to how they processed data. For example, 2D classification in RELION (Scheres, 2012; Zivanov et al., 2018) generated extremely noisy class averages that were difficult to “clean up” even after multiple rounds of 2D classifications. CryoSPARC (Punjani, Ali et al., 2017), on the other hand, produced well aligned 2D class averages that contained visible secondary structure features. Another strength of processing the data to 2D classification in cryoSPARC (Punjani, Ali et al., 2017) was efficiency. 2D classification in RELION (Scheres, 2012; Zivanov et al., 2018) took weeks due to the large particle number whereas class averages were obtained using cryoSPARC (Punjani, Ali et al., 2017) in a matter of days. A major drawback, however, of cryoSPARC (Punjani, Ali et al., 2017) though is the inability to continue a job should it crash.

Despite its ability to generate excellent 2D class averages, the 3D maps obtained from cryoSPARC (Punjani, Ali et al., 2017) were either too noisy, did not visually align with the attributed resolution or were inconsistent. The 3D maps generated using RELION (Scheres, 2012; Zivanov et al., 2018), on the other hand, were reliable and reflected the estimated resolution. To conclude, integrating cryoSPARC (Punjani, Ali et al., 2017) into the data processing pipeline, allowed for the identification and alignment of good protein particles, allowing them to be reconstructed into 3D maps.

4.10.2 Fitting the BRISC and ARISC-AIR structures into cryo-EM maps

To determine whether the class two map was of the ARISC complex, the human BRISC and mouse ARISC-AIR structures were fitted with the map. The fit showed that the BRISC complex accurately occupied the EM map, however there were some unoccupied regions in the MERIT40 and Abraxas 1 densities. This could be due to differences in resolution. The fit of the mouse ARISC-AIR structure with the class two map revealed significant differences in the BRCC45-MERIT40 region. In their paper, Rabl et al., 2019 determined that the binding of the truncated RAP80 AIR protein resulted in a significant conformational shift in BRCC45-MERIT40 arm region. Obtaining a 10 Å map of the ARISC complex allows for the comparison of the ARISC and ARISC-RAP8 AIR complex and provides additional insight on complex assembly and most likely also how the complex may carry out its function within the nucleus. Patterson-Fortin et al.,2010 showed that the binding of RAP80 to ARISC does not increase or abolish the complex's activity. Therefore, it is possible that the conformation shift observed by Rabl et al., 2019 relates to the function of the enzyme – unrelated to its activity such as localisation to sites of double stranded DNA breaks.

Two BRISC structures were fitted with the map of the dimer of octamers. The fit suggests that the dimer of octamers is formed by the interaction of two ARISC complexes and not two ARISC-RAP80 complexes. The identification of this novel complex and the revelation that the binding of the AIR protein induces a conformation shift within the 'arms' of ARISC complex eludes to the possibility that the dimerization of the ARISC complex regulates the activity of the complex in vivo as complex dimerization likely occludes the BRCC36 active site. It is important to note that the asymmetric dimer of octamers uncovered here is a new ARISC structure and differs from the symmetric dimer of octamers structure described previously by Rabl et al., 2019 and Kyrieleis at al., 2016. Both of these studies used glutaraldehyde cross-linked samples for their EM analysis, which may be responsible for the observed structural differences (Figure 90).

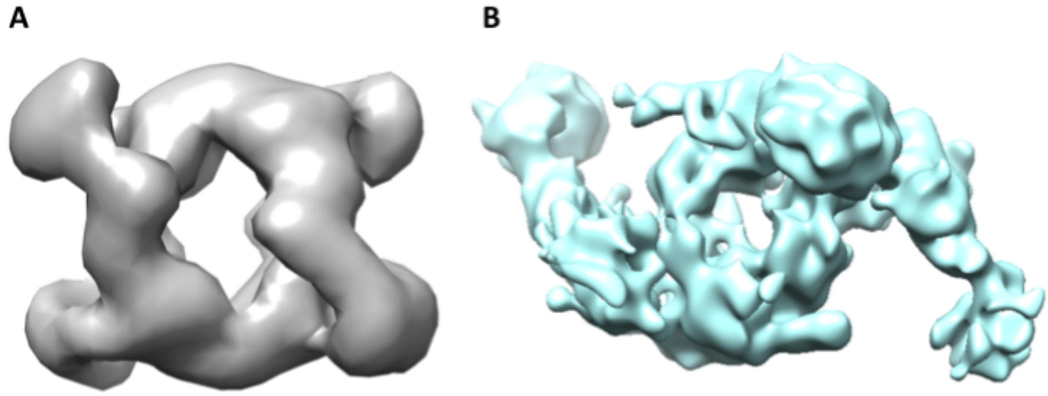


Figure 90: Symmetric and asymmetric ARISC dimer of octamer

[A] The Kyrieleis at al 2016 symmetric ARISC dimer of octamer obtained as a result of cross-linking with glutaraldehyde. The map shows MERIT40 and BRCC45 of one complex interacting with the Abraxas1 tail. [B] The asymmetric ARISC dimer of octamer uncovered as part of this research project. The map shows MERIT40-BRCC45 interacting with BRCC36.

Cryo-EM analysis of the ARISC-RAP80 complex resulted in the identification of five complex, as opposed to just two (ARISC-RAP80 and ARISC dimer of octamer). It is possible that in addition to targeting the complex to sites of double stranded breaks, RAP80 inhibits complex dimerization and therefore enzyme inactivation and stabilises the complex. However, in the absence of RAP80 and prior to the formation of the dimer of octamer which stabilises the complex, members of the complex dissociate, resulting in the ARISC one arm map and the core Abraxas1-BRCC36 super dimer. This is supported by the fact that a substantial number of particles only contained the dissociated Abraxas1-BRCC36 super dimer (chapter 4.7.2.2). What is not known is the affinity of the subunits to one another. Understanding the affinity of each subunit may help to understand which ARISC complex is the more stable complex and may help to elucidate whether the subcomplexes, ARISC, ARISC-RAP80 and the dimer of octamers exist in a delicate equilibrium. A novel way to study this would be to utilise a method termed Metabolically-labelled Affinity-tagged Subunit Exchange (MASE) (Pan, 2011). MASE can be used to determine the dissociation rate constant k_{off} of subunits within a complex. MASE works by coupling metabolic stable isotope labelling, affinity purification and mass spectrometry and is based on a subunit exchange process between an unlabelled affinity-tagged variant and a metabolically-labelled untagged variant of a complex (Pan 2011).This relatively new method is a more suitable approach to traditional methods for measuring k_{off} as the more traditional methods require high-quality purification and specific modifications of protein complexes (Pan 2011). It is for this reason that the dissociation kinetics have only been studied for a small set of model complexes (Pan 2011).

The crystal structure of BRCC36-KIAA0157 (Abraxas2) was also fitted into the map of the Abraxas1-BRCC36 core complex presented in chapter 4.8.3. The fit showed that the structure occupies the majority of the map, however, there were some unoccupied densities. As the map is of low resolution, it is difficult to draw further conclusions.

4.10.3 Localising the RAP80 protein

The two main research questions driving this chapter were: 1) How does the ARISC-RAP80 complex assemble, and 2) where and how does RAP80 interact with members of the complex. Cryo-EM analysis of the ARISC-RAP80 complex resulted in the identification of five complexes. Fitting the BRISC, ARISC-AIR and Abraxas2-BRCC36 published structures into the maps described in chapter 4.8.1, 4.8.2 and 4.8.3 suggests that those maps do not contain RAP80.

The map termed ‘the unknown complex‘remained the only map that could contain RAP80. To understand what the unknown complex represents the BRISC structure was fitted into the map. The fit suggests that the core Abraxas1-BRCC36 structure at the heart of the unknown complex has undergone a conformational change. The positioning of the structure into the map also revealed large unaccounted for densities that interacted with MERIT40, BRCC45 and Abraxas1. This finding is consistent with the findings of Rabl et al., 2019 in which they showed that the AIR protein interacts with MERIT40, BRCC45 and Abraxas1. In their structure, they showed that when bound to ARISC, AIR induced a conformational change in the arm region of MERIT40 and BRCC45. It is possible that the full RAP80 protein interacts with other regions of the complex inducing a more global conformational change.

The published ARISC-RAP80 AIR structure was also fitted into the unknown map. The fit showed considerable differences both in the core and arm of the complex. These differences could be accounted for by the fact that the Abraxas protein studied in this project is truncated (ΔC 1-265). The ARISC-AIR crystal structure indicates that the C-terminus of Abraxas interacts with MERIT40 and AIR. As the C-terminus of Abraxas1 has been truncated in this study, the conformational changes at the arms of the complex is likely to be inhibited. It is plausible that full length ARISC-RAP80 complex assembles in a manner that induces conformational changes both at the arms and core of the complex generating a structure that does not resemble an arc. As the map is of low resolution, it is difficult to accurately position each protein subunit, especially since large parts of the RAP80 protein are structurally unknown. Higher resolution data will be required to determine the exact nature of the observed conformational changes and structural differences.

To conclude and reflect on the research questions posed in this chapter, the data presented here suggests that the ARISC-RAP80 complex is prone to dissociation, resulting in pronounced heterogeneity with the formation of various subcomplexes and that during the dissociation process or prior to the binding of RAP80, ARISC forms a stable dimer with itself. In addition to this, analysis of the various four-membered complexes identified in chapter 4.7.1 suggests that the complex may also display degrees of flexibility and that the maps captured various snapshots of the complex.

Due to the heterogenous nature of the protein, it was impossible to obtain high resolution structural information despite the considerable size of the combined datasets. Without high resolution maps, it is difficult to accurately locate RAP80 within the complex. However, the unknown complex identified here may represent a snapshot of the ARISC-RAP80 complex. Despite not being able to definitively identify RAP80, the data here suggests more questions regarding the ARISC-RAP80 complex and its assembly *in vivo*. Is the ARISC dimer of octamers an autoinhibited conformation of the complex? If so, what drives complex dissociation in vivo? The data presented in this chapter essentially captured various complex assembly stages of the ARISC-RAP80 complex, does this reflect the nature of the complex in biology, and how is the equilibrium maintained or established? These are all very important research question; further work is required to understand the structure and function of the human full length ARISC-RAP80 complex.

4.10.4 Gel shift assay (EMSA)

The gel shift assay results showed that ARISC-RAP80 and, to a much lesser extent, ARISC were capable of binding 300 bp DNA and that ARISC-AIR was not. This suggests that RAP80 and Abraxas1 are the subunits responsible for facilitating the complex's interaction with DNA. As AIR is the truncated domain of RAP80 that interacts with Abraxas1, it's possible that the region of Abraxas1 that interacts with AIR is responsible for contributing to ARISC's DNA binding ability and that the binding of AIR to Abraxas1 inhibits the DNA-protein interaction. The binding of DNA to ARISC was shown to be concentration dependent: as the concentration of ARISC decreased from 10 to 5 μM so did the binding capabilities of the complex as more free DNA was visible on the lanes. At a protein concentration of 2.5 μM, ARISC no longer efficiently bound DNA and the amount of free DNA on the lane was comparable with the control. The binding capabilities of ARISC-RAP80 did not appear to be affected by reduction in protein concentration and lower concentrations will need to be tested to see this, possibly reflecting a high ARISC-RAP80 binding affinity. This was apparent from the lack of free DNA on the lanes.

To determine whether the DNA binding capabilities of ARISC-RAP80 and ARISC were length specific, the gel shift assay was repeated using 50 bp DNA. Gel shift assay using 300 bp DNA found that Abraxas1 contributed to ARISC DNA binding abilities. To decipher whether other components of the complex also facilitated in ARISC and ARISC-RAP80 DNA interaction, BRCC45-MERIT40 heterodimer was also tested for its DNA binding ability. Glycogen synthase, a metabolic enzyme known to not bind DNA was used as a control. Results from the assay shows that only ARISC-RAP80 retained its ability to interact with DNA and that BRCC45-MERIT40 did not bind DNA. This suggests that Abraxas1 may only interact with longer DNA and that BRCC45-MERIT40 do not contribute to ARISC ability to bind DNA. The ARISC 300 bp DNA gel shift assay showed that ARISC bound DNA tightly at 10 μM protein concentration and less efficiently at 5 μM. As the assay conducted with 50 bp DNA only used protein concentrations of 5 and 2.5 μM it's expected that ARISC would not bind DNA efficiently. These results show that the binding of DNA to ARISC is both protein concentration dependent and DNA length specific. Similar with the 300 bp gel shift assay, ARISC-AIR did not interact with 50 bp DNA.

ARISC-RAP80 was able to strongly bind 300 and 50 bp DNA, suggesting that the complex may not display length specificity. To test this, the assay would need to be repeated using a wider variety of DNA products such as 600 bp and 20 bp. RAP80 is composed of SIM, UIM, AIR and zinc finger domains (Sato et al., 2009). Zinc finger domains are known to facilitate protein-protein interaction and bind DNA (Pavletich and Pabo, 1991; Fairall et al., 1993; Matthews et al., 2000).. It is plausible that RAP80's zinc finger domains were responsible for binding DNA in the assays. To test this, the individual domains would have to be cloned, expressed and tested for their DNA binding abilities.

The gel was run for 40 minutes as opposed to 20, this allowed for the bands to migrate further down the gel. To improve the reliability of the 300 bp gel shift results, the gel needs to be repeated and run for longer to allow for the bands to migrate further down the gel and the experiment will have to include BRCC45-MERIT40 to determine whether the heterodimers interact with longer length DNA.

4.10.5 ARISC-RAP80 photoreactive ubiquitin

The ARISC-RAP80 complex was reacted with photoreactive monoubiquitin to determine the components capable of binding ubiquitin. The results revealed that RAP80 was the only subunit capable of binding the substrate. Analysis of the gel also showed that the RAP80-monoubiquitin bands wear smeared, suggesting that the protein was binding varying amounts of ubiquitin. This is especially clear at the 60 minute time point. RAP80's ability to bind multiple ubiquitin molecules is due to the protein containing two UIM domains. Ubiquitin are known to form mixed chains composed of SUMO and ubiquitin moieties. For example, the SUMO-targeted ubiquitin ligases (STUBLs) add ubiquitin chains to proteins that have already been modified with SUMO. *In vitro*, STUBLs have been shown to directly ubiquitylate SUMO tags. These enzymes have also been implicated in several signalling pathways including DNA damage (Yau and Rape, 2016). As the ubiquitin code has been shown to be increasingly more complex, it is possible that UIM and SIM domains can recognise ubiquitin-SUMO heterotypic chains, thus making it plausible that the SIM domains of RAP80 recognised monoubiquitin with lower affinity but as the experiments involved cross-linking the SIM-ubiquitin interaction was stabilised and detected.

Surprisingly, BRCC36 was incapable of binding monoubiquitin, despite its ability identify and deubiquitylate K63-linked ubiquitin chains. BRCC36's inability to capture monoubiquitin is likely due to the possibility that its normal role involves interaction with longer ubiquitin chains. To this end, I propose that RAP80 is responsible for locating and binding ubiquitin chains, and after binding ubiquitin, RAP80 feeds the ubiquitin chain to BRCC36 with the support for BRCC45 and MERIT40 for catalyses and that BRCC36 may only recognise and bind longer length ubiquitin chains. This hypothesis is supported by experiments carried out by Kyrieleis et al., 2016 and Rabl *et al*., 2019. In their paper Kyrieleis et al using the purified BRCC45-MERIT40 complex showed that BRCC45-MERIT40 have an affinity for binding to K63-linked di-ubiquitin of around 17 mM using biolayer interferometry.

In the Rabl et al., 2019 study, the group reacted di-ubiquitin, tri-ubiquitin and tetra-ubiquitin with the BRCA1-A complex and found cleavage of tetra-ubiquitin but no cleavage of di- and tri-ubiquitin. In light of this, they also conducted a cleavage experiment using fluorescent labelled proximal or distal end of tetra-ubiquitin chains to determine the orientation in which the chains are cleaved. They found that following incubation of the complex with proximally labelled ubiquitin generated (Ub)2 and (Ub)3 cleavage products, and that distally labelled samples resulted in a different cutting pattern generating predominantly Ub cleavage products. These findings suggest that (Ub)4 chains are fed to BRCC36 with the distal ubiquitin pointing to the protein and being cleaved first. Together these result along with the data presented in this chapter suggests that RAP80 is the sole subunit capable of binding monoubiquitin and that BRCC36 recognises and likely deubiquitylates K63 polyubiquitin chains composed of at least four ubiquitin moieties.

4.10.6 Native PAGE analysis

Native page analysis of the ARISC complexes is consistent with the sample heterogeneity observed during negative stain EM, cryo-EM and mass spectrometry analysis. Together these results suggest the subunits do not form a single complex but multiple complexes including the ARISC dimer of octamer. The smear observed in the gel points to protein aggregation and indicate that over time the protein aggregates. The cleanest bands were of the ARISC-RAP80 fractions 25 and 26, and 24. This suggests that experimental analysis of the complex will benefit from use of freshly purified proteins.

CHAPTER 5

Data collection of the BRISC-SHMT2 complex using a higher protein concentration

5.1 Introduction

During this research project, the structure of the BRISC-SHMT2 complex solved by Dr Miriam Walden. Over time it became apparent that both the ARISC-RAP80 complex and BRISC-SHMT2 complex displayed heterogeneity in the form of the one-arm and two-arm structures. It was also noted that the ARISC-RAP80 complex EM datasets were collected at a higher protein concentration than the BRISC-SHMT2 complex and showed a higher proportion of 2-arm particles. Thus, using the information gained from my studies of ARISC, we decided to collect a dataset of the BRISC-SHMT2 complex at a higher concentration than usual. Data was collected at a higher concentration of 0.5 mg/ml instead of 0.05 mg/ml and pre-processed to generate 2D templates. Downstream processing to 2D classification was then carried out by Dr Miriam Walden. Following this, suitable 2D class averages were combined with the data previously collected Dr Miriam Walden (at lower concentration) and processed to a final overall resolution of 3.8 Å. The aim of this chapter is to determine whether the BRISC-SHMT2 complex, like the ARISC-RAP80 complex, forms a more stable complex at higher protein concentration, and to collect suitable data to improve the resolution of the BRISC-SHMT2 map.

5.2 Pre-processing and 2D classification of the BRISC-SHMT2 complex

QUANTIFOIL R1.2/1.3 grids were glow discharged using the Pelco easiglow glow discharge apparatus and BRISC-SHMT2 protein sample (0.5 mg/ml) was applied. The grids were vitrified using the Vitrobot using blotting force and time (seconds) of 3 and 4, respectively. The humidity was set to 100% and temperature to 4°C. The grid was subsequently screened on the KRISO1 microscope and 2,615 micrographs were collected over a period of 24 hours.

Following data collection, the grids were Motion and CTF corrected in RELION (Scheres, 2012; Zivanov et al., 2018) using the same procedure outlined during the processing of the ARISC-RAP80 complex (Chapter 4). Visual observation of the micrographs revealed some ice contamination and BRISC-SHMT2 particles in areas of thick ice (Figure 91A). Despite this, attempts were made to process the data. Due to higher protein concentration, more particles were manually selected compared to the ARISC-RAP80 complex data (approximately 2,500 particles). The selected particles were then extracted using a particle box size of 350 and processed to 2D classification to obtain 2D templates. After iterative rounds of manual picking and 2D classifications, 2D templates clearly displaying the BRISC-SHMT2 complex were obtained (Figure 91B) and these were selected as templates for autopicking.

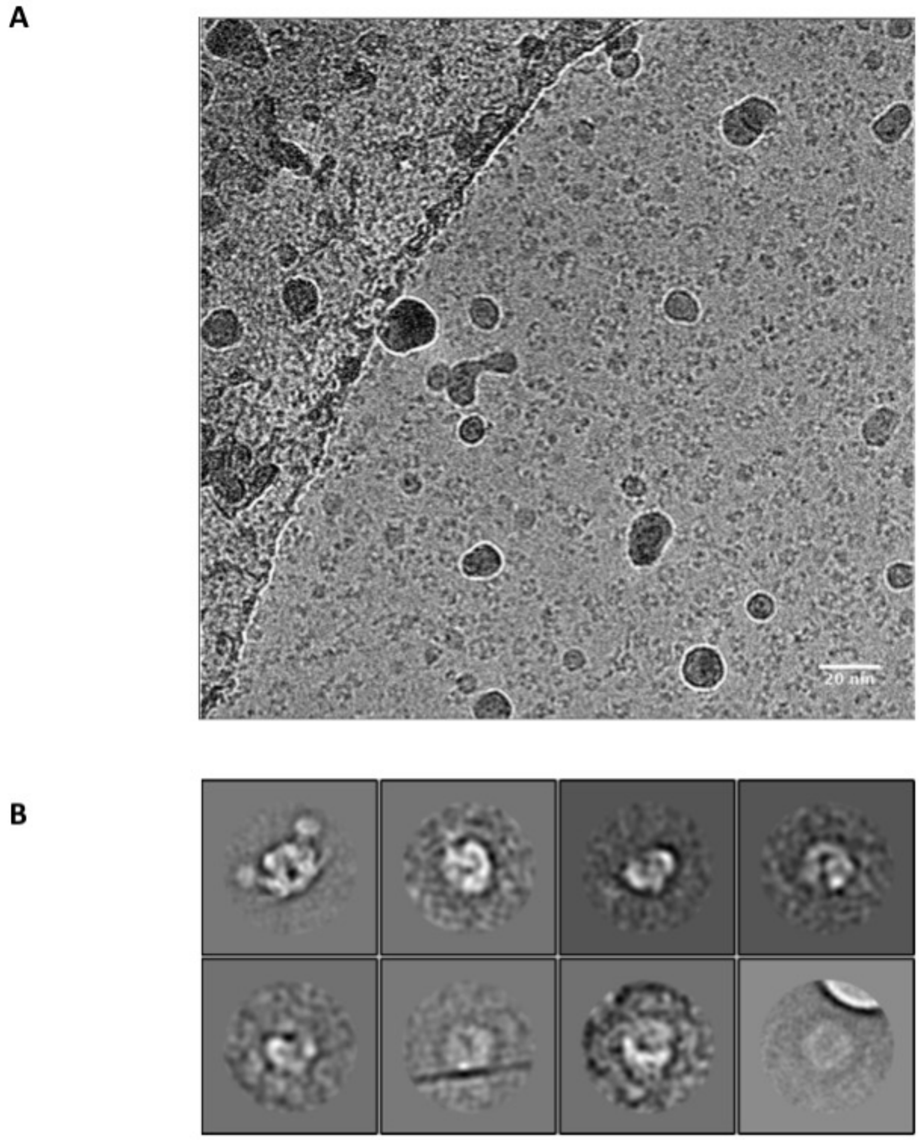


Figure 91: Pre-processing of the BRISC-SHMT2 complex

[A] A representative micrograph of the BRISC-SHMT2 complex dataset. [B] 2D class averages obtained following iteratively re-classifying selected class averages.

444,752 particles were autopicked and extracted using a box size of 350 pixels. The extracted particles were then subjected to three rounds of 2D classification. Results from the final round of 2D classification revealed one- and two-arm class averages in different orientations (Figure 92A).

The 2D classes were sorted so that the class averages containing the most particles appeared at the top. Visual comparison of the three BRISC-SHMT2 2D classes suggests that a higher protein concentration generated more stable protein particles. For example, the first two 2D class averages from the 0.05 mg/ml datasets were of the BRISC-SHMT2 subcomplexes (Figure 92B-C). This suggests that the majority of the well-aligned particles were of the subcomplexes. Conversely, the first two 2D class averages obtained using a higher protein concentration were of the two and one-arm complex (Figure 92A). This suggests that a higher protein concentration stabilises the complex, resulting in more two-arm BRISC-SHMT2 particles.

Figure 92: Comparison of the BRISCS-SHMT2 2D class averages obtained from the low and high concentration datasets

[A] Top 20 class averages from the 0.5 mg/ml dataset. The first five class averages contained more stable one and two-arm particles. Eight class averages contained a significant level of background noise. [B-C] Top 20 class averages from the two 0.05 mg/ml datasets. The first five class averages contained subcomplexes of the BRISC-SHMT2 complex, all class averages were well aligned and contained minimal background noise. 2D class averages with two arms are highlighted in red.

5.3 3D processing of the BRISC-SHMT2 complex

14 classes (132,081 particles) were then selected and combined with the two 0.05 mg/ml BRISC-SHMT2 datasets collected and processed by Dr Miriam Walden (2D classes shown in Figure 92B and 92C). The first low concentration dataset was composed of 1,921 movies. Approximately 1,000 particles were manually selected to generate 2D templates for auto-picking. Autopicking resulted in the selection of 486,307 particles. The extracted particles were then subjected to three rounds of 2D classification, of which 53 classes containing 202,834 particles were selected for 3D processing. The second BRISC-SHMT2 dataset was composed of 2,958 movies. During pre-processing, 655,383 particles were autopicked and extracted. The particles were subjected to two rounds of 2D classification, following which 41 class averages containing 307, 641 particles were selected for 3D processing.

A combined total of 642,557 particles (including suitable selected particles from the high concentration dataset) were processed to 3D classification using RELION 3.0 (Scheres, 2012; Zivanov et al., 2018). 3D reconstruction of the particles was first conducted using C1 symmetry. Two rounds of 3D classification were carried out to better classify the particles. Four classes containing a total of 403,516 particles were selected, refined and post-processed to 3.8 Å (Figure 93A). 3D reconstruction was then repeated using C2 symmetry. In this instance, two classes (64,403) depicting the intact two-arm complex were selected for 3D refinement and post-processing. The map was post-processed to a resolution of 4.2 Å (Figure 93B). The lower resolution could be attributed to lower particle number and the fact that resolution describes the global resolution of the map. The arm region of the map was of significantly lower resolution, this is evident from the lack of structural detail.

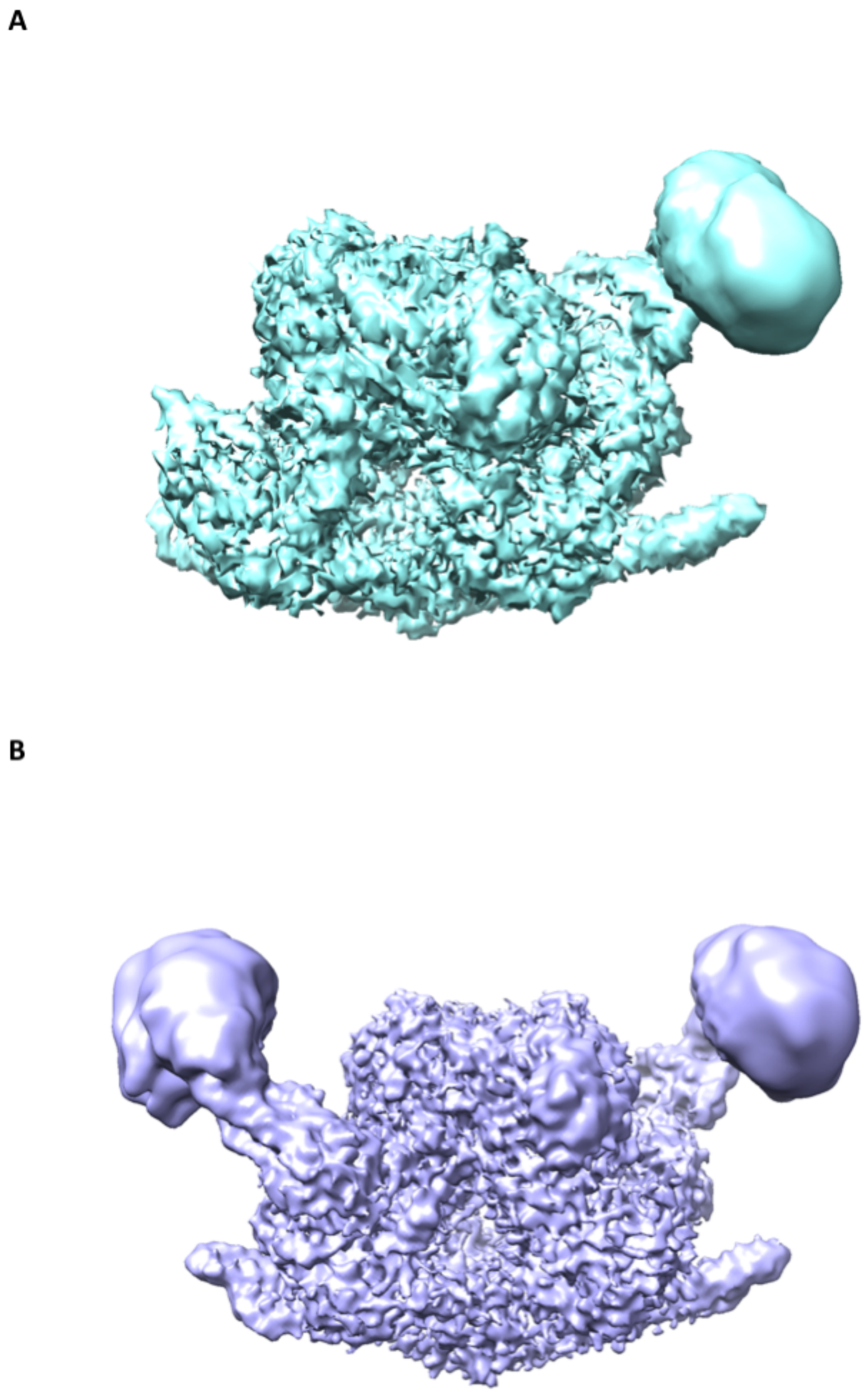


Figure 93: 3D post-processed maps of the combined BRISC-SHMT2 datasets

[A] 403, 516 particles were processed in C1 symmetry, refined and post-processed to 3.8 Å. The resulting map was missing an arm and contained high resolution information at the core of the complex and lower resolution at the arm. [B] Two classes (64,403 particles) depicting the intact two-arm complex were processed using C2 symmetry to a resolution of 4.2 Å. The resulting map was of the intact two-arm complex. The majority of the high resolution information was contained at the core of the complex.

Prior to the inclusion of the particles from the 0.5 mg/ml dataset, the 0.05 mg/ml datasets were 3D classified, refined and post-processed independently. 3D processing of the second 0.05 mg/ml dataset resulted in a map with an estimates global resolution of 4.7 Å (Figure 94A). The first dataset was only processed to 2D classification, the 53 selected classes (202,834 particles) were combined with the 307,641 particles from the second dataset. 3D processing of the combined particles gave rise to a map with an estimated global resolution of 4.43 Å (Figure 94B). The improved resolution is likely due to the increase in particle number. Despite the improvement in resolution, the combining of particles resulted in the averaging out of one arm. This is likely due to data containing an overwhelming number of one-arm BRISCH-SHMT2 particles.

Visual observation and comparison of the 3D post-processed maps suggests that the resolution differences across the structure obtained from the low concentration datasets were not as large as those found in the structure resulting from combining all three datasets. This becomes apparent when observing the lack of structural detail in the arm regions of figure 93 and comparing it to figure 94. The differences in detail and resolution in the arm region of the map, despite the map having a higher global resolution could be due to the fact that the core of the structure (Abraxas2-BRC336-SHMT2) were better aligned then the arm region of the complex (BRCC45-MERIT40). Poor alignment of the arms is likely due to the arms being more flexible and therefore being orientated at slightly different angles relative to one another. The stability and therefore better alignment of the core region of the complex is likely due to SHMT2 locking the proteins in place.

Despite discrepancies in the structural details of the arms of the complex, the incorporation of the high concentration dataset particles resulted in a significant increase in resolution and a map of the two-arm complex.

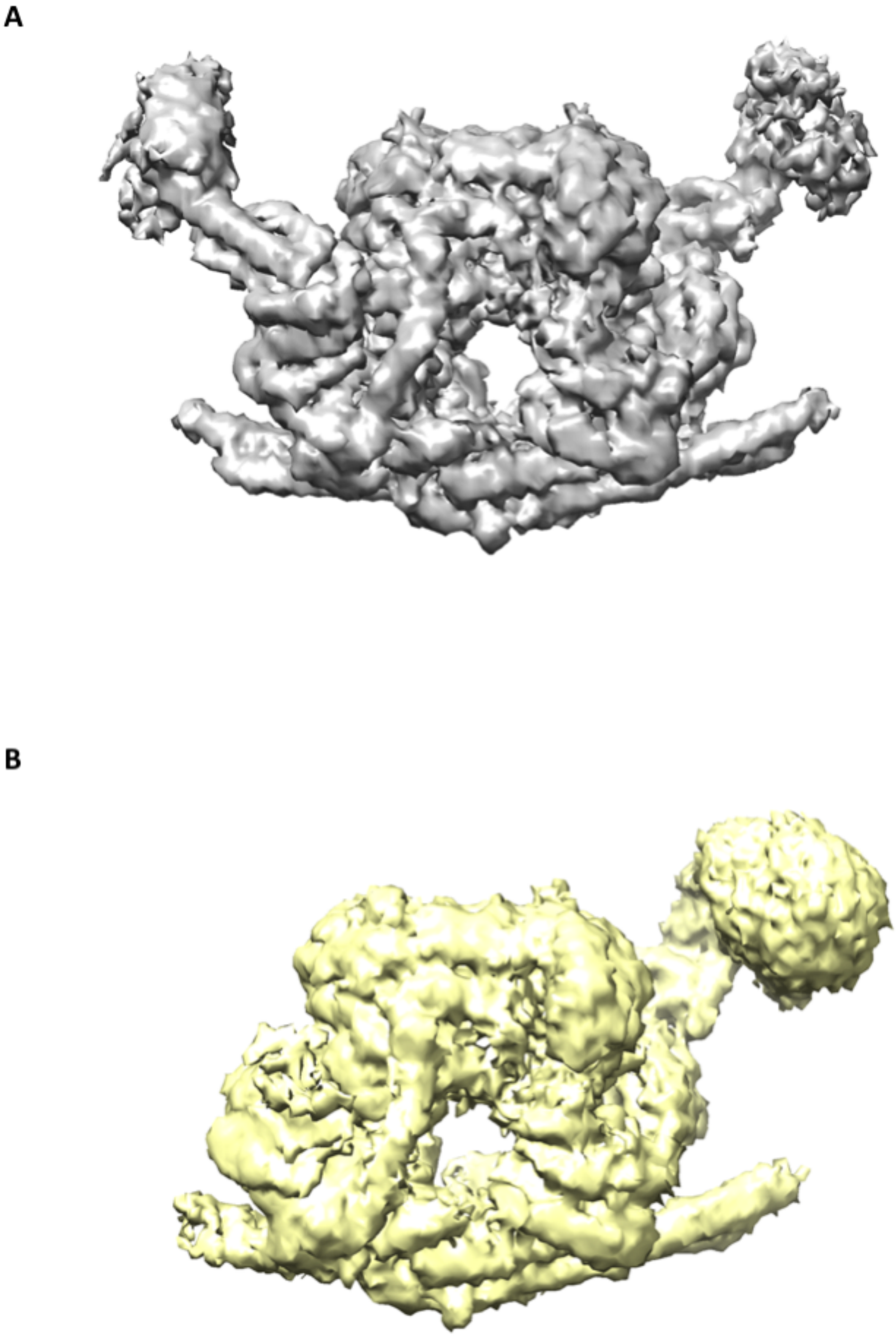


Figure 94: 3D post-processed maps of the low concentration BRISC-SHMT2 datasets

[A]. The second 0.05 mg/ml dataset was processed to 4.7 Å. The resulting map was missing densities corresponding to regions of the BRCC45-MERIT40 heterodimer. [B] 202, 837 particles from the first dataset was combined with selected particles from the second dataset and underwent 3D processing. The resulting post-processed map was of the one-arm BRISC-SHMT2 complex and had an estimated global resolution of 4.43 Å.

5.4 Discussion

This chapter investigates the effect concentration has on the stability of the BRISC-SHMT2 complex. Previous cryo-EM data of the complex collected at a lower concentration resulted in a more heterogenous datasets in which subcomplexes were identifiable during 2D classification. Data collected at 10 times the concentration resulted in overcrowded particles, which made data pre-processing more complex. Despite the difficulty encountered at the pre-processing stage, suitable 2D templates were obtained, allowing the program to effectively select, extract and align particles during 2D classification. Results from 2D classification indicated that the dataset contained more stable protein particles than the data collected at a lower concentration. Most protein particles prepared in vitreous ice for single-particle cryo-EM are adsorbed to air–water or substrate–water interfaces (Noble et al., 2018). Absorption to the air-water interface has the potential to cause protein particles to adopt preferred orientations as well as damage or degrade protein. This phenomenon could explain why at lower protein concentrations the BRISC-SHMT2 and ARISC-RAP80 complex appeared dissociated. Increasing protein concentration allowed the air-water interface to become saturated with protein thereby preventing the diffusion of more particles to the interface. This allowed data of a more intact BRISC-SHMT2 and ARISC-RAP80 complex to be collected and processed. To test this and avoid future protein contact with the air-water interface, surfactants could be added to the sample which could act as a barrier or the particles could be sequestered to a support film (Noble et al., 2018). Additionally, reducing the time between depositing the sample on the cryo grid and the placement of the grid in the cryogen could help limit particle diffusion to the air-water interface (Noble et al., 2018). There are currently multiple vitrification devices available that allow for a more rapid plunge-freezing of the grid such as a microfluidic spray-plunging machine developed by the Frank group (Feng, X. et al., 2017). A surface-acoustic-wave-based microfluidic dispenser from the de Marco group (Ashtiani et al., 2018), Spotiton, a robotic device that dispenses picoliter volumes of sample onto a self-blotting nanowire grid as it quickly moves past en route to vitrification (Jain et al., 2012; Dandey et al., 2018; Razinkov et al., 2016) (Razinkov et al., 2016)(and a sample spraying technology (Klebl et al., 2020).

This observation in both the ARISC-RAP80 and BRISC-SHMT2 complex suggests that despite differing in two protein subunits (targeting protein and MPN$^-$ protein), the complex assembly may occur in a similar pattern. Comparison of the data processing results suggests that the ARISC-RAP80 complex dissociates more readily than the BRISC-SHMT2 complex. This is likely due to the positioning of the target proteins SHMT2 and RAP80 on their respective complexes. The cryo-EM structure of the BRISC-SHMT2 complex shows that SHMT2 sits directly above the Abraxas2-BRCC36 dimer bridging the two arms of the BRSIC complex. SHMT2's positioning enables it to interact with the MPN domain of BRCC36 and Abraxas2, and the UEV domains of BRCC45. Additionally, the structure revealed that MERIT40 does not directly interact with SHMT2. The binding of SHMT2 to BRISC in this manner likely locks the complex in a more stable position. In contrast, it is possible that RAP80 binds to ARISC in a manner that still allows the complex to display high levels of flexibility. This flexibility could also explain the conformation heterogeneity observed in the negative stain EM dataset. Alternatively, it's possible that RAP80 and SHMT2 bind to their respective complexes with differing affinity and that RAP80 more readily dissociates from ARISC thus encouraging complex dissociation or stabilising via the formation of the ARISC dimer of dimer. These hypotheses are based solely on the observation of the low resolution structural data and biochemical experiments will be required to test this theory.

To conclude, using a higher protein concentration resulted in a more stable complex as at a higher concentration some particles diffused to the air-water interface saturating preventing further diffusion of particles to the interface. This in turn gave rise to more two-arm BRISC-SHMT2 particles and allowed to data be processed to a higher resolution. The Kd of the complex is at the cusp of the cut-off point for cryoEM, had it been 10-fold weaker, obtaining a high resolution structure would have required cross-linking and further optimisations.

CHAPTER 6
Final thesis discussion

6.1 Final discussion

Negative stain EM was conducted primarily to obtained preliminary structural information on the ARISC complexes. The data obtained from the ARISC complex revealed that the complex was prone to dissociations, however, the 3D maps were promising as they displayed a similar U-like shape to the published negative stain EM model of the ARISC complex (Kyrieleis et al., 2016). The group also noted the susceptibility of the complex to dissociate on the EM grid, a phenomenon also experienced during this project. The stability of the ARISC complex was increased with the presence of the 15 kDa RAP80 truncated protein- AIR. Negative stain EM analysis of the ARISC-RAP80 (AIR) complex, like the ARISC complex, produced distinct 3D maps. However, out of the four maps produced, class one displayed high level of similarities with the recently published structure of the mouse ARISC-AIR complex (Rabl et al.,). This data is promising at it suggests that both human and mouse ARISC-AIR complexes assemble in a similar manner. The data also eludes to the complex existing in various conformation.

The negative stain EM ARISC-RAP80 3D maps presented in chapter 3.4 all contained extra density located at one arm of the complex. It's possible that the density represented full length RAP80. If so, then the maps agree with the Orbitrap mass spectrometry data presented in chapter 3.6.3 which suggests that RAP80 binds to ARISC with a stoichiometry of 2:2:2:2:1. When the ARISC-RAP80 negative stain EM data processed using C2 symmetry, additional features such as the 'flaps' and 'zipper' became apparent (chapter 3.4.4: Figure 35). The zipper seen in the data was also observed in class one of the ARISC-AIR map, suggesting that the zipper is the AIR region of RAP80. The flaps could represent a snapshot of the protein. It is unlikely that the two flaps represent two RAP80 proteins as the density is too small to be RAP80. It is possible that flaps represent a more flexible region of RAP80 captured during the analysis and extended by the C2 symmetry operators. Interestingly, the fourth class resembled the 3D classes obtained from processing the data in C1 symmetry. Together this suggests, that applying symmetry enabled RELION (Scheres, 2012; Zivanov et al., 2018) to better separate the particles into different maps, thus identifying the different snapshots of the complex.

Cryo-EM analysis of the ARISC-RAP80 complex generated two ARISC maps (one arm and two arm), the Abraxas1-BRCC36 super dimer, the ARISC dimer of octamers, and an unknown complex potentially containing RAP80. Together, the data suggests that Abraxas1, BRCC36, BRCC45, MERIT40 and RAP80 can form different complexes with varying degrees of flexibility and conformations.

In their 2016 paper, Kyrieleis et al., 2016 noted that they observed that the ARISC complex forms a dimer with itself, to stabilise this interaction and obtain a model of the ARISC complex, the group cross-linked their protein sample. This resulted in an ARISC dimer of octamer that was symmetry. The ARISC dimer of octamers presented in this thesis interact in a different matter producing a more asymmetric complex. Differences in complex interaction are likely to be the result of the induced covalent interaction facilitated by cross-linking. The function of the ARISC dimer of octamers has yet to be elucidated, as this is a novel complex. However, I hypothesise that ARISC dimerization has a regulatory role in cells, potentially preventing the complex from deubiquitylating substrates indiscriminately. Visual inspection of the maps would suggest that ARISC dimer of octamers is an autoinhibited and catalytically inactive complex since the BRCC36 active sites would be inaccessible for ubiquitin chains. Further experiments are required to test this hypothesis.

The main aim of this project was to uncover the structure of the ARISC-RAP80 complex. I propose that the unknown structure is likely to represent the ARISC-RAP80 complex. This argument is supported by the fact that the map contains additional density on the left arm of the complex that mirrors the negative stain EM 3D maps of the complex and it is the only map that does not reflect the published ARISC model, ARISC-AIR structure or BRISC structure. Visual analysis of the map suggests that the binding of RAP80 induces a global conformational change and that RAP80 binds in a manner that results in a bridge formation across the two arms. The map presented as the ARISC-RAP80 complex shows a novel conformation for this complex and further optimisation such as GRAFIX cross-linking or the addition of a substrate will be required to stabilise and obtain a higher resolution map of this ARISC-RAP80 complex.

The cryo-EM map of the human ARISC complex presented in this thesis is the highest resolution map currently available of the complex. Although the map displays the U-like shape described by Kyrieleis et al., 2016, it differs in the positioning of the BRCC45-MERIT40 heterodimer. This suggest that the ARISC map presented in this thesis is of a novel conformation of the complex. The ARISC map is also distinctively different to the crystal structure of the mouse ARISC-AIR complex (Rabl et al., 2019). Fitting the ARISC-AIR structure into the cryo-EM map of the ARISC complex revealed that the binding of the AIR protein to the complex induces conformational changes in the arms of the complex (BRCC45-MERIT40). It is possible that the conformation changes observed by Rabl et al., 2019 accounts for the conformational changes seen in the structure depicted as the ‘unknown structure’. If the unknown structure is in fact a low-resolution map of the ARISC-RAP80 complex, this would make it the first topological information of the complex available to date. The possible location of the RAP80 protein on the complex would also provide insight as to why the complex remains active when bound to RAP80.

Cryo-EM data of the ARISC-RAP80 and BRISC-SHMT2 complex obtained using higher protein concentrations (0.5 mg/m) resulted in more two-arm particles. This suggests an interplay between protein concentration and complex stability. However, it is also possible that during grid preparation, protein particles were diffusing to the air-water interface resulting in their degradation, explaining why during the screening of the grids the protein particles appeared to have dissociated (Noble et al., 2018). Interestingly, despite both the ARISC-RAP80 and BRISC-SHMT2 data presented in this thesis being collected at comparable protein concentration, the ARISC-RAP80 dataset displayed higher levels of heterogeneity. This suggests variability in complex assembly and stability between the ARISC-RAP80 and BRISC-SHMT2 complex.

The data here points to the stoichiometry of the ARISC, ARISC-AIR and ARISC-RAP80 complex being 2:2:2:2 and 2:2:2:2:1, respectively. It could be argued that two AIR and RAP80 proteins may bind to ARISC in vivo, however, the mass spectrometry data presented here suggests that there is only a single RAP80 bound. Initially we hypothesised that the second RAP80 dissociated from the complex and however, it's possible that the subcomplexes identified in mass spectrometry are not dissociated but reflect the nature of the sample (that the subunits form different complexes in solution), as suggested by the native PAGE gel in chapter 4.9.4. Further experiments are needed to determine the in vivo stoichiometry and the potential biological role and relevance of these subcomplexes.

The stoichiometry of the ARISC-RAP80 complex being 2:2:2:2:1 explains the differences in activity between the BRISC-SHMT2 and ARISC-RAP80 complex. The binding of SHMT2 to BRISC renders to complex inactive as it interacts with the complex in a way that blocks the active site of BRCC36 (Walden et al., 2019). Having a single RAP80 protein bound to ARISC likely frees up the active site of a single or both BRCC36 subunits depending on how RAP80 'sits' on the complex. One of the aims of this project was to understand how the structure of ARISC-RAP80 allows the complex to interact with its substrates. The preliminary photoreactive ubiquitin experiments presented in chapter 4.9.3 suggests that RAP80 is the only subunit capable of binding monoubiquitin. These results, together with the negative stain EM and mass spectrometry analysis of the ARISC-RAP8O complex leads to the suggestion that RAP80 binds a ubiquitin moiety on the K63-polyubiquitin chain fand subsequently feeds the chain to BRCC36 for deubiquitylation. This hypothesis is supported by and supports data published by Rabl et al., 2019 in which the group observed that BRCC36 displays a specificity for longer chained K63-polyubuqitin chains. Together these results shed further light on substrate recognition by DUBS. The preliminary gel shift assay experiments shown in chapter 4.9.1 and 4.9.2 showed that the ARISC-RAP80 and to a lower extent the ARISC complex are capable of binding DNA. The data also indicates that RAP80 is the main contributor to DNA binding. Within the context of the BRCA1-A complex, RAP80 is described as the subunit responsible for targeting the complex to K63 polyubiquitin chains at sites of DNA damage (Wu, J. et al., 2012). It is not known whether the protein remains bound to the complex as the polyubiquitin signal is deubiquitylated. The preliminary data here suggests that even when in the presence of the ARISC complex, RAP80 is capable of binding DNA. To this end, I propose that following DSBs, the BRCA1-A complex localises to site of damage and that RAP80 binds DNA via its zinc finger domains anchoring the complex. After binding DNA, the complex then carries out its DUB activity. To confirm that RAP80 helps tether the complex on histone H2A as BRCC36 carries out its deubiquitylating activity, structural information of the ARISC-RAP80 in complex with histone H2A or DNA and non-cleavable polyubiquitin chain would be required.

To conclude, the research presented here has made significant progress in understanding the structure and function of the ARISC complexes, in particular its conformational dynamics and flexibility. Obtaining higher resolution information was limited due to the heterogeneous native of the complexes, although it is important to realise that this conformational heterogeneity may be reflective of "snapshots" of functional complex structures. The stoichiometry information gained from mass spectrometry analysis all point to the complex containing a single RAP80 protein, this however, is not in line with the newly published ARISC-AIR mouse structure. Higher resolution maps of the human ARISC-RAP80 complex will be required to accurately determine the real stoichiometry of the complexes.

To fully comprehend how ARISC-RAP80 carries out its activity within the nucleus, it will be important to study the protein in complex with polyubiquitin chains and bound to DNA. It would also be interesting to observe structural and activity differences when the complex is bound to the BRCA1-BARD1 heterodimer, which will contribute to further elucidation of the full BRCA1-A complex assembly and functions in DNA damage and repair.

6.2 Final conclusion

At the start of this project (October 2016) there was no information on the structure of the ARISC-RAP80 complex. Work performed for this thesis in the past 3.5 years has elucidated new and important structural and functional information. Firstly, ARISC, ARISC-RAP80 and BRISC-SHMT2 complexes sample different conformations and this appears to be dependent on protein concentration and subunit composition. For the first time, we have new low resolution structures that may be relevant to how the RAP80 complexes assemble in cells. Secondly, ARISC attains an autoinhibited dimer of octamers and displays a new asymmetric structure. Thirdly, RAP80 contains the highest affinity ubiquitin and DNA binding domains within the ARISC-RAP80 complex. This may explain the RAP80 crucial functions within ARISC and BRCA1-A complex and its ability to participate in detecting and aiding repair of DNA double strand breaks. Work presented in this thesis provides the foundations for obtaining higher resolution structures of ARISC-RAP80 complex with and without functionally relevant binding partners such as poly-ubiquitin chains, DNA, BRCA1 and BARD1.

CHAPTER 7
References

Abraham, R.T. 2001. Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes & Development.* **15**(17), pp.2177-2196.

Aebi, U. and Pollard, T.D. 1987. A glow discharge unit to render electron microscope grids and other surfaces hydrophilic. *Journal of electron microscopy technique.* **7**(1), pp.29-33.

Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D. and Urzhumtsev, A. 2018. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D: Structural Biology.* **74**(9), pp.814-840.

Akutsu, M., Dikic, I. and Bremm, A. 2016. Ubiquitin chain diversity at a glance. *J Cell Sci.* **129**(5), pp.875-880.

Ali, S.O., Khan, F.A., Galindo-Campos, M.A. and Yelamos, J. 2016. Understanding specific functions of PARP-2: new lessons for cancer therapy. *American Journal of Cancer Research.* **6**(9), pp.1842-1863.

Ambroggio, X.I., Rees, D.C. and Deshaies, R.J. 2003. JAMM: a metalloprotease-like zinc site in the proteasome and signalosome. *PLoS Biol.* **2**(1), pe2.

Annunziatio, A. 2015. DNA packing: Nucleosomes and chromatin. *Nat. Educ.* **1**, pp.1-26.
Ashtiani, D., Venugopal, H., Belousoff, M., Spicer, B., Mak, J., Neild, A. and de Marco, A. 2018. Delivery of femtolitre droplets using surface acoustic wave based atomisation for cryo-EM grid preparation. *Journal of structural biology.* **203**(2), pp.94-101.

Avvakumov, G.V., Walker, J.R., Xue, S., Finerty, P.J., Mackenzie, F., Newman, E.M. and Dhe-Paganon, S. 2006. Amino-terminal dimerization, NRDP1-rhodanese interaction, and inhibited catalytic domain conformation of the ubiquitin-specific protease 8 (USP8). *Journal of Biological Chemistry.* **281**(49), pp.38061-38070.

Bai, X.C., McMullan, G. and Scheres, S.H. 2015. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci.* **40**(1), pp.49-57.

Bailly, V., Lauder, S., Prakash, S. and Prakash, L. 1997. Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. *Journal of Biological Chemistry.* **272**(37), pp.23360-23365.

Ball, H.L., Myers, J.S. and Cortez, D. 2005. ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. *Molecular Biology of the Cell.* **16**(5), pp.2372-2381.

Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E. and Baumeister, W. 2012. Near-atomic resolution structural model of the yeast 26S proteasome. *Proceedings of the National Academy of Sciences.* **109**(37), pp.14870-14875.

Berger, I., Garzoni, F., Chaillet, M., Haffke, M., Gupta, K. and Aubert, A. 2013. The multiBac protein complex production platform at the EMBL. *JoVE (Journal of Visualized Experiments).* (77), pe50159.

Bertoli, C., Skotheim, J.M. and de Bruin, R.A.M. 2013. Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology.* **14**(8), pp.518-528.

Birol, M., Enchev, R.I., Padilla, A., Stengel, F., Aebersold, R., Betzi, S., Yang, Y., Hoh, F., Peter, M. and Dumas, C. 2014. Structural and biochemical characterization of the Cop9 signalosome CSN5/CSN6 heterodimer. *PloS one.* **9**(8), pe105688.

Bose, M., Sachsenweger, J., Laurila, N., Parplys, A.C., Willmann, J., Jungwirth, J., Groth, M., Rapakko, K., Nieminen, P. and Friedl, T.W. 2019. BRCA1 mislocalization leads to aberrant DNA damage response in heterozygous ABRAXAS1 mutation carrier cells. *Human molecular genetics.*

Bowman, G.D. and Poirier, M.G. 2015. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chemical Reviews.* **115**(6), pp.2274-2295.

Bray, F., McCarron, P. and Parkin, D.M. 2004. The changing global patterns of female breast cancer incidence and mortality. *Breast cancer research.* **6**(6), p229.

Burgess, R.J. and Zhang, Z.G. 2013. Histone chaperones in nucleosome assembly and human disease. *Nature Structural & Molecular Biology.* **20**(1), pp.14-22.

Burma, S., Chen, B.P., Murphy, M., Kurimasa, A. and Chen, D.J. 2001. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *Journal of Biological Chemistry.* **276**(45), pp.42462-42467.

Carroni, M. and Saibil, H.R. 2016. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods.* **95**, pp.78-85.

Castillo, A., Paul, A., Sun, B., Huang, T.H., Wang, Y., Yazinski, S.A., Tyler, J., Li, L., You, M.J. and Zou, L. 2014. The BRCA1-interacting protein Abraxas is required for genomic stability and tumor suppression. *Cell reports.* **8**(3), pp.807-817.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H. and Henderson, R. 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy.* **135**, pp.24-35.

Ciccia, A. and Elledge, S.J. 2010. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell.* **40**(2), pp.179-204.

Clapier, C.R. and Cairns, B.R. 2009. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry.* **78**, pp.273-304.

Clapier, C.R., Iwasa, J., Cairns, B.R. and Peterson, C.L. 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology.* **18**(7), pp.407-422.

Cook, W.J., Jeffrey, L.C., Carson, M., Chen, Z.J. and Pickart, C.M. 1992. Structure of a Diubiquitin Conjugate and a Model for Interaction with Ubiquitin Conjugating Enzyme (E2). *Journal of Biological Chemistry.* **267**(23), pp.16467-16471.

Cooper, E.M., Cutcliffe, C., Kristiansen, T.Z., Pandey, A., Pickart, C.M. and Cohen, R.E. 2009. K63-specific deubiquitination by two JAMM/MPN+ complexes: BRISC-associated Brcc36 and proteasomal Poh1. *The EMBO journal.* **28**(6), pp.621-631.

Cotto-Rios, X.M., Békés, M., Chapman, J., Ueberheide, B. and Huang, T.T. 2012. Deubiquitinases as a signaling target of oxidative stress. *Cell reports.* **2**(6), pp.1475-1484.

Cutter, A.R. and Hayes, J.J. 2015. A brief review of nucleosome structure. *Febs Letters.* **589**(20), pp.2914-2922.

Dandey, V.P., Wei, H., Zhang, Z., Tan, Y.Z., Acharya, P., Eng, E.T., Rice, W.J., Kahn, P.A., Potter, C.S. and Carragher, B. 2018. Spotiton: New features and applications. *Journal of Structural Biology.* **202**(2), pp.161-169.

Davies, C.W., Paul, L.N. and Das, C. 2013. Mechanism of recruitment and activation of the endosome-associated deubiquitinase AMSH. *Biochemistry.* **52**(44), pp.7818-7829.

Des Georges, A., Dhote, V., Kuhn, L., Hellen, C.U., Pestova, T.V., Frank, J. and Hashem, Y. 2015. Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature.* **525**(7570), pp.491-495.

Dikic, I., Wakatsuki, S. and Walters, K.J. 2009. Ubiquitin-binding domains - from structures to functions. *Nature Reviews Molecular Cell Biology.* **10**(10), pp.659-671.

Dove, K.K. and Klevit, R.E. 2017. RING-Between-RING E3 Ligases: Emerging Themes amid the Variations. *Journal of Molecular Biology.* **429**(22), pp.3363-3375.

Drulyte, I., Johnson, R.M., Hesketh, E.L., Hurdiss, D.L., Scarff, C.A., Porav, S.A., Ranson, N.A., Muench, S.P. and Thompson, R.F. 2018. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D: Structural Biology.* **74**(6).

Durand, A., Bonnet, J., Fournier, M., Chavant, V. and Schultz, P. 2014. Mapping the deubiquitination module within the SAGA complex. *Structure.* **22**(11), pp.1553-1559.

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

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature.* **366**(6454), pp.483-487.

Fan, J.Y., Gordon, F., Luger, K., Hansen, J.C. and Tremethick, D.J. 2002. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nature Structural Biology.* **9**(3), pp.172-176.

Feng, L., Wang, J. and Chen, J. 2010. The Lys63-specific deubiquitinating enzyme BRCC36 is regulated by two scaffold proteins localizing in different subcellular compartments. *Journal of Biological Chemistry.* **285**(40), pp.30982-30988.

Feng, X., Fu, Z., Kaledhonkar, S., Jia, Y., Shah, B., Jin, A., Liu, Z., Sun, M., Chen, B. and Grassucci, R.A. 2017. A fast and effective microfluidic spraying-plunging method for high-resolution single-particle cryo-EM. *Structure.* **25**(4), pp.663-670. e663.

Fradet-Turcotte, A., Sitz, J., Grapton, D. and Orthwein, A. 2016. BRCA2 functions: from DNA repair to replication fork stabilization. *Endocr Relat Cancer.* **23**(10), pp.T1-T17.

Galanty, Y., Belotserkovskaya, R., Coates, J., Polo, S., Miller, K.M. and Jackson, S.P. 2009. Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature.* **462**(7275), pp.935-U132.

Giglia-Mari, G., Zotter, A. and Vermeulen, W. 2011. DNA Damage Response. *Cold Spring Harbor Perspectives in Biology.* **3**(1).

Gonzalez, V.M., Fuertes, M.A., Alonso, C. and Perez, J.M. 2001. Is cisplatin-induced cell death always produced by apoptosis? *Molecular pharmacology.* **59**(4), pp.657-663.

Grainger, R.J. and Beggs, J.D. 2005. Prp8 protein: at the heart of the spliceosome. *Rna.* **11**(5), pp.533-557.

Green, M.R. and Sambrook, J. 2016. Precipitation of DNA with ethanol. *Cold Spring Harbor Protocols.* **2016**(12), ppdb. prot093377.

Hakem, R. 2008. DNA-damage repair; the good, the bad, and the ugly. *The EMBO journal.* **27**(4), pp.589-605.

Harper, J.W. and Elledge, S.J. 2007. The DNA damage response: Ten years after. *Molecular Cell.* **28**(5), pp.739-745.

Hrdinka, M. and Gyrd-Hansen, M. 2017. The Met1-linked ubiquitin machinery: emerging themes of (De) regulation. *Molecular cell.* **68**(2), pp.265-280.

Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J.-W., Gu, W., Cohen, R.E. and Shi, Y. 2002. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell.* **111**(7), pp.1041-1054.

Hu, M., Li, P.W., Li, M.Y., Li, W.Y., Yao, T.T., Wu, J.W., Gu, W., Cohen, R.E. and Shi, Y.G. 2002. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell.* **111**(7), pp.1041-1054.

Hu, X., Kim, J.A., Castillo, A., Huang, M., Liu, J. and Wang, B. 2011. NBA1/MERIT40 and BRE interaction is required for the integrity of two distinct deubiquitinating enzyme BRCC36-containing complexes. *Journal of Biological Chemistry.* **286**(13), pp.11734-11745.

Hu, X., Paul, A. and Wang, B. 2012. Rap80 protein recruitment to DNA double-strand breaks requires binding to both small ubiquitin-like modifier (SUMO) and ubiquitin conjugates. *Journal of Biological Chemistry.* **287**(30), pp.25510-25519.

Hu, Y., Scully, R., Sobhian, B., Xie, A., Shestakova, E. and Livingston, D.M. 2011. RAP80-directed tuning of BRCA1 homologous recombination function at ionizing radiation-induced nuclear foci. *Genes & development.* **25**(7), pp.685-700.

Ikeda, F. and Dikic, I. 2008. Atypical ubiquitin chains: new molecular signals. *EMBO reports.* **9**(6), pp.536-542.

Iyer, D.R. and Rhind, N. 2017. The Intra-S Checkpoint Responses to DNA Damage. *Genes.* **8**(2).

Jackson, S.P. and Bartek, J. 2009. The DNA-damage response in human biology and disease. *Nature.* **461**(7267), pp.1071-1078.

Jain, T., Sheehan, P., Crum, J., Carragher, B. and Potter, C.S. 2012. Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *Journal of structural biology.* **179**(1), pp.68-75.

Jasin, M. and Rothstein, R. 2013. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology.* **5**(11).

Jasin, M. and Rothstein, R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology.* **5**(11), pa012740.

Jiang, Q. 2018. Molecular Recognition at DNA Damage Sites.
Jiang, X.-X., Nguyen, Q., Chou, Y., Wang, T., Nandakumar, V., Yates, P., Jones, L., Wang, L., Won, H. and Lee, H.-R. 2011. Control of B cell development by the histone H2A deubiquitinase MYSM1. *Immunity.* **35**(6), pp.883-896.

Jin, L.Y., Williamson, A., Banerjee, S., Philipp, I. and Rape, M. 2008. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell.* **133**(4), pp.653-665.

Keusekotten, K., Elliott, P.R., Glockner, L., Fiil, B.K., Damgaard, R.B., Kulathu, Y., Wauer, T., Hospenthal, M.K., Gyrd-Hansen, M. and Krappmann, D. 2013. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell.* **153**(6), pp.1312-1326.

Kikuchi, K., Ishii, N., Asao, H. and Sugamura, K. 2003. Identification of AMSH-LP containing a Jab1/MPN domain metalloenzyme motif. *Biochemical and biophysical research communications.* **306**(3), pp.637-643.

Kirisako, T., Kamei, K., Murata, S., Kato, M., Fukumoto, H., Kanie, M., Sano, S., Tokunaga, F., Tanaka, K. and Iwai, K. 2006. A ubiquitin ligase complex assembles linear polyubiquitin chains. *The EMBO journal.* **25**(20), pp.4877-4887.

Klebl, D.P., Monteiro, D.C., Kontziampasis, D., Kopf, F., Sobott, F., White, H.D., Trebbin, M. and Muench, S.P. 2020. Sample deposition onto cryo-EM grids: from sprays to jets and back. *Acta Crystallographica Section D: Structural Biology.* **76**(4).

Köhler, A., Zimmerman, E., Schneider, M., Hurt, E. and Zheng, N. 2010. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell.* **141**(4), pp.606-617.

Komander, D., Clague, M.J. and Urbe, S. 2009. Breaking the chains: structure and function of the deubiquitinases. *Nature Reviews Molecular Cell Biology.* **10**(8), pp.550-563.

Komander, D. and Rape, M. 2012. The Ubiquitin Code. *Annual Review of Biochemistry, Vol 81.* **81**, pp.203-229.

Kontziampasis, D., Klebl, D.P., Iadanza, M.G., Scarff, C.A., Kopf, F., Sobott, F., Monteiro, D.C., Trebbin, M., Muench, S.P. and White, H.D. 2019. A cryo-EM grid preparation device for time-resolved structural studies. *IUCrJ.* **6**(6).

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell.* **128**(4), pp.693-705.
Kulathu, Y., Garcia, F.J., Mevissen, T.E., Busch, M., Arnaudo, N., Carroll, K.S., Barford, D. and Komander, D. 2013. Regulation of A20 and other OTU deubiquitinases by reversible oxidation. *Nature communications.* **4**, p1569.

Kumar, S., Kao, W.H. and Howley, P.M. 1997. Physical interaction between specific E2 and Hect E3 enzymes determines functional cooperativity. *Journal of Biological Chemistry.* **272**(21), pp.13548-13554.

Kwasna, D., Rehman, S.A.A., Natarajan, J., Matthews, S., Madden, R., De Cesare, V., Weidlich, S., Virdee, S., Ahel, I. and Gibbs-Seymour, I. 2018. Discovery and characterization of ZUFSP/ZUP1, a distinct deubiquitinase class important for genome stability. *Molecular cell.* **70**(1), pp.150-164. e156.

Kyrieleis, O.J., McIntosh, P.B., Webb, S.R., Calder, L.J., Lloyd, J., Patel, N.A., Martin, S.R., Robinson, C.V., Rosenthal, P.B. and Smerdon, S.J. 2016. Three-dimensional architecture of the human BRCA1-A histone deubiquitinase core complex. *Cell reports.* **17**(12), pp.3099-3106.

Lamond, A.I. 2002. Molecular biology of the cell, 4th edition. *Nature.* **417**(6887), pp.383-383.

Lange, O.F., Lakomek, N.A., Fares, C., Schroder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C. and de Groot, B.L. 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science.* **320**(5882), pp.1471-1475.

Lee, J.-G., Baek, K., Soetandyo, N. and Ye, Y. 2013. Reversible inactivation of deubiquitinases by reactive oxygen species in vitro and in cells. *Nature communications.* **4**, p1568.

Li, X. and Heyer, W.D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research.* **18**(1), pp.99-113.

Li, Y., Luo, K., Yin, Y., Wu, C., Deng, M., Li, L., Chen, Y., Nowsheen, S., Lou, Z. and Yuan, J. 2017. USP13 regulates the RAP80-BRCA1 complex dependent DNA damage response. *Nature communications.* **8**, p15752.

Liao, H.Y. and Frank, J. 2010. Definition and estimation of resolution in single-particle reconstructions. *Structure.* **18**(7), pp.768-775.

Liu, B.H., Yip, R.K.H. and Zhou, Z.J. 2012. Chromatin Remodeling, DNA Damage Repair and Aging. *Current Genomics.* **13**(7), pp.533-547.

Liu, P., Gan, W., Su, S., Hauenstein, A.V., Fu, T.-m., Brasher, B., Schwerdtfeger, C., Liang, A.C., Xu, M. and Wei, W. 2018. K63-linked polyubiquitin chains bind to DNA to facilitate DNA damage repair. *Sci. Signal.* **11**(533), peaar8133.

Lombardi, P.M., Matunis, M.J. and Wolberger, C. 2017. RAP80, ubiquitin and SUMO in the DNA damage response. *Journal of Molecular Medicine.* **95**(8), pp.799-807.

Mahajan, K.N., McElhinny, S.A.N., Mitchell, B.S. and Ramsden, D.A. 2002. Association of DNA polymerase mu (pol mu) with Ku and ligase IV: Role for pol mu in end-joining double-strand break repair. *Molecular and Cellular Biology.* **22**(14), pp.5194-5202.

Mallette, F.A. and Richard, S. 2012. K48-linked ubiquitination and protein degradation regulate 53BP1 recruitment at DNA damage sites. *Cell research.* **22**(8), p1221.

Mandemaker, I., van Cuijk, L., Janssens, R.C., Lans, H., Bezstarosti, K., Hoeijmakers, J., Demmers, J., Vermeulen, W. and Marteijn, J. 2017. DNA damage-induced histone H1 ubiquitylation is mediated by HUWE1 and stimulates the RNF8-RNF168 pathway. *Scientific Reports.* **7**(1), p15353.

Marchione, R., Leibovitch, S.A. and Lenormand, J.-L. 2013. The translational factor eIF3f: the ambivalent eIF3 subunit. *Cellular and Molecular Life Sciences.* **70**(19), pp.3603-3616.

Marechal, A. and Zou, L. 2013. DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harbor Perspectives in Biology.* **5**(9).

Marín, I., Lucas, J.I., Gradilla, A.-C. and Ferrús, A. 2004. Parkin and relatives: the RBR family of ubiquitin ligases. *Physiological genomics.* **17**(3), pp.253-263.

Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K., Benesch, J.L. and Robinson, C.V. 2015. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical chemistry.* **87**(8), pp.4370-4376.

Matthews, J.M., Kowalski, K., Liew, C.K., Sharpe, B.K., Fox, A.H., Crossley, M. and Mackay, J.P. 2000. A class of zinc fingers involved in protein–protein interactions: Biophysical characterization of CCHC fingers from Fog and U-shaped. *European journal of biochemistry.* **267**(4), pp.1030-1038.

Maynard, S., Schurman, S.H., Harboe, C., de Souza-Pinto, N.C. and Bohr, V.A. 2009. Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis.* **30**(1), pp.2-10.

Metzger, M.B., Hristova, V.A. and Weissman, A.M. 2012. HECT and RING finger families of E3 ubiquitin ligases at a glance. *Journal of Cell Science.* **125**(3), pp.531-537.

Mevissen, T.E. and Komander, D. 2017. Mechanisms of deubiquitinase specificity and regulation. *Annual review of biochemistry.* **86**, pp.159-192.

Michel, M.A., Swatek, K.N., Hospenthal, M.K. and Komander, D. 2017. Ubiquitin linkage-specific affimers reveal insights into K6-linked ubiquitin signaling. *Molecular cell.* **68**(1), pp.233-246. e235.

Milne, J.L., Borgnia, M.J., Bartesaghi, A., Tran, E.E., Earl, L.A., Schauder, D.M., Lengyel, J., Pierson, J., Patwardhan, A. and Subramaniam, S. 2013. Cryo-electron microscopy–a primer for the non-microscopist. *The FEBS journal.* **280**(1), pp.28-45.

Moretti, J., Chastagner, P., Gastaldello, S., Heuss, S.F., Dirac, A.M., Bernards, R., Masucci, M.G., Israel, A. and Brou, C. 2010. The translation initiation factor 3f (eIF3f) exhibits a deubiquitinase activity regulating Notch activation. *PLoS Biol.* **8**(11), pe1000545.

Moynahan, M.E., Chiu, J.W., Koller, B.H. and Jasin, M. 1999. Brca1 controls homology-directed DNA repair. *Molecular cell.* **4**(4), pp.511-518.

Nijnik, A., Clare, S., Hale, C., Raisen, C., McIntyre, R.E., Yusa, K., Everitt, A.R., Mottram, L., Podrini, C. and Lucas, M. 2012. The critical role of histone H2A-deubiquitinase Mysm1 in hematopoiesis and lymphocyte differentiation. *Blood.* **119**(6), pp.1370-1379.

Nikkilä, J., Coleman, K.A., Morrissey, D., Pylkäs, K., Erkko, H., Messick, T.E., Karppinen, S.-M., Amelina, A., Winqvist, R. and Greenberg, R.A. 2009. Familial breast cancer screening reveals an alteration in the RAP80 UIM domain that impairs DNA damage response function. *Oncogene.* **28**(16), p1843.

Noble, A.J., Wei, H., Dandey, V.P., Zhang, Z., Tan, Y.Z., Potter, C.S. and Carragher, B. 2018. Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nature methods.* **15**(10), pp.793-795.

Norbury, C.J. and Hickson, I.D. 2001. Cellular responses to DNA damage. *Annual Review of Pharmacology and Toxicology.* **41**, pp.367-401.

Norbury, C.J. and Zhivotovsky, B. 2004. DNA damage-induced apoptosis. *Oncogene.* **23**(16), pp.2797-2808.

Oberle, C. and Blattner, C. 2010. Regulation of the DNA Damage Response to DSBs by Post-Translational Modifications. *Current Genomics.* **11**(3), pp.184-198.

Ohtake, F., Saeki, Y., Ishido, S., Kanno, J. and Tanaka, K. 2016. The K48-K63 branched ubiquitin chain regulates NF-κB signaling. *Molecular cell.* **64**(2), pp.251-266.

Orlova, E.V. and Saibil, H.R. 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev.* **111**(12), pp.7710-7748.

Pan, C. 2011. Measuring dissociation rate constants of protein complexes through subunit exchange: Experimental design and theoretical modeling. *PloS one.* **6**(12), pe28827.

Pao, K.-C., Wood, N.T., Knebel, A., Rafie, K., Stanley, M., Mabbitt, P.D., Sundaramoorthy, R., Hofmann, K., van Aalten, D.M. and Virdee, S. 2018. Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature.* **556**(7701), pp.381-385.

Parrilla-Castellar, E.R., Arlander, S.J. and Karnitz, L. 2004. Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair (Amst).* **3**(8-9), pp.1009-1014.

Pascual-Montano, A., Donate, L., Valle, M., Barcena, M., Pascual-Marqui, R. and Carazo, J. 2001. A novel neural network technique for analysis and classification of EM single-particle images. *Journal of structural biology.* **133**(2-3), pp.233-245.

Pathare, G.R., Nagy, I., Śledź, P., Anderson, D.J., Zhou, H.-J., Pardon, E., Steyaert, J., Förster, F., Bracher, A. and Baumeister, W. 2014. Crystal structure of the proteasomal deubiquitylation module Rpn8-Rpn11. *Proceedings of the National Academy of Sciences.* **111**(8), pp.2984-2989.

Patterson-Fortin, J., Shao, G., Bretscher, H., Messick, T.E. and Greenberg, R.A. 2010. Differential regulation of JAMM domain deubiquitinating enzyme activity within the RAP80 complex. *Journal of Biological Chemistry.* **285**(40), pp.30971-30981.

Pavletich, N.P. and Pabo, C.O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science.* **252**(5007), pp.809-817.

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

PHENIX. *Cryo-EM validation tools in PHENIX.* [Online]. [Accessed 22 November]. Available from: https://www.phenix-online.org/documentation/reference/validation_cryo_em.html
Polo, S.E. and Jackson, S.P. 2011. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Development.* **25**(5), pp.409-433.

Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods.* **14**(3), pp.290-296.

Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* **14**(3), pp.290-296.

Razinkov, I., Dandey, V.P., Wei, H., Zhang, Z., Melnekoff, D., Rice, W.J., Wigge, C., Potter, C.S. and Carragher, B. 2016. A new method for vitrifying samples for cryoEM. *Journal of structural biology.* **195**(2), pp.190-198.

Reyes-Turcu, F.E., Ventii, K.H. and Wilkinson, K.D. 2009. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annual review of biochemistry.* **78**, pp.363-397.

Roos, W.P. and Kaina, B. 2006. DNA damage-induced cell death by apoptosis. *Trends in Molecular Medicine.* **12**(9), pp.440-450.

Rosen, E.M. 2013. BRCA1 in the DNA damage response and at telomeres. *Frontiers in genetics.* **4**, p85.

Russo, C.J. and Passmore, L.A. 2014. Ultrastable gold substrates for electron cryomicroscopy. *Science.* **346**(6215), pp.1377-1380.

Sancar, A., Lindsey-Boltz, L.A., Ünsal-Kaçmaz, K. and Linn, S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry.* **73**(1), pp.39-85.

Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A. and Fukai, S. 2009. Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *The EMBO journal.* **28**(16), pp.2461-2468.

Sawant, U.U. and Varma, A.K. 2014. Role of MERIT40 in stabilization of BRCA1 complex: A protein–protein interaction study. *Biochemical and biophysical research communications.* **446**(4), pp.1139-1144.

Scarff, C.A., Fuller, M.J., Thompson, R.F. and Iadaza, M.G. 2018. Variations on negative stain electron microscopy methods: tools for tackling challenging systems. *JoVE (Journal of Visualized Experiments).* (132), pe57199.

Scheres, S.H. 2010. Maximum-likelihood methods in cryo-EM. Part II: Application to experimental data. *Methods in enzymology.* **482**, p295.

Scheres, S.H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology.* **180**(3), pp.519-530.

Scheres, S.H. 2014. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife.* **3**, pe03665.

Scheres, S.H. 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol.* **579**, pp.125-157.

Scheres, S.H., Valle, M., Nunez, R., Sorzano, C.O., Marabini, R., Herman, G.T. and Carazo, J.M. 2005. Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol.* **348**(1), pp.139-149.

Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.-Q. and Yip, C.K. 2015. Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. *Journal of Biological Chemistry.* **290**(16), pp.10057-10070.

Shao, G., Lilli, D.R., Patterson-Fortin, J., Coleman, K.A., Morrissey, D.E. and Greenberg, R.A. 2009. The Rap80-BRCC36 de-ubiquitinating enzyme complex antagonizes RNF8-Ubc13-dependent ubiquitination events at DNA double strand breaks. *Proceedings of the National Academy of Sciences.* **106**(9), pp.3166-3171.

Shi, W., Tang, M.K., Yao, Y., Tang, C., Chui, Y.L. and Lee, K.K.H. 2016. BRE plays an essential role in preventing replicative and DNA damage-induced premature senescence. *Scientific reports.* **6**, p23506.

Shiloh, Y. 2003. ATM and related protein kinases: Safeguarding genome integrity. *Nature Reviews Cancer.* **3**(3), pp.155-168.

Shrestha, R.K., Ronau, J.A., Davies, C.W., Guenette, R.G., Strieter, E.R., Paul, L.N. and Das, C. 2014. Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. *Biochemistry.* **53**(19), pp.3199-3217.

Sigworth, F.J. 2016. Principles of cryo-EM single-particle image processing. *Microscopy (Oxf).* **65**(1), pp.57-67.

Simpson, R.T. 1978. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry.* **17**(25), pp.5524-5531.

Sinha, R.P. and Hader, D.P. 2002. UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences.* **1**(4), pp.225-236.

Smalle, J. and Vierstra, R.D. 2004. The ubiquitin 26S proteasome proteolytic pathway. *Annual Review of Plant Biology.* **55**, pp.555-590.

Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M. and Greenberg, R.A. 2007. RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science.* **316**(5828), pp.1198-1202.

Solyom, S., Aressy, B., Pylkäs, K., Patterson-Fortin, J., Hartikainen, J.M., Kallioniemi, A., Kauppila, S., Nikkilä, J., Kosma, V.-M. and Mannermaa, A. 2012. Breast Cancer–Associated Abraxas Mutation Disrupts Nuclear Localization and DNA Damage Response Functions. *Science translational medicine.* **4**(122), pp.122ra123-122ra123.

Talbert, P.B. and Henikoff, S. 2010. Histone variants - ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology.* **11**(4), pp.264-275.

Tenno, T., Fujiwara, K., Tochio, H., Iwai, K., Morita, E.H., Hayashi, H., Murata, S., Hiroaki, H., Sato, M. and Tanaka, K. 2004. Structural basis for distinct roles of Lys63-and Lys48-linked polyubiquitin chains. *Genes to Cells.* **9**(10), pp.865-875.

Tessarz, P. and Kouzarides, T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology.* **15**(11), pp.703-708.

Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P. and Ranson, N.A. 2016. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods.* **100**, pp.3-15.

Tran, H.J., Allen, M.D., Löwe, J. and Bycroft, M. 2003. Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry.* **42**(39), pp.11460-11465.

Travers, A. and Muskhelishvili, G. 2015. DNA structure and function. *Febs Journal.* **282**(12), pp.2279-2295.

van Attikum, H. and Gasser, S.M. 2009. Crosstalk between histone modifications during the DNA damage response. *Trends in Cell Biology.* **19**(5), pp.207-217.

Van Heel, M. and Frank, J. 1981. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy.* **6**(1), pp.187-194.

van Tilburg, G.B., Elhebieshy, A.F. and Ovaa, H. 2016. Synthetic and semi-synthetic strategies to study ubiquitin signaling. *Current opinion in structural biology.* **38**, pp.92-101.

Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C. and Fushman, D. 2004. Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of biological chemistry.* **279**(8), pp.7055-7063.

Varadan, R., Walker, O., Pickart, C. and Fushman, D. 2002. Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology.* **324**(4), pp.637-647.

Venkatesh, S. and Workman, J.L. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology.* **16**(3), pp.178-189.

Walden, M., Masandi, S.K., Pawłowski, K. and Zeqiraj, E. 2018. Pseudo-DUBs as allosteric activators and molecular scaffolds of protein complexes. *Biochemical Society Transactions.* **46**(2), pp.453-466.

Walden, M., Tian, L., Ross, R.L., Sykora, U.M., Byrne, D.P., Hesketh, E.L., Masandi, S.K., Cassel, J., George, R. and Ault, J.R. 2019. Metabolic control of BRISC–SHMT2 assembly regulates immune signalling. *Nature.* p1.

Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P. and Elledge, S.J. 2007. Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science.* **316**(5828), pp.1194-1198.

Weber, C.M. and Henikoff, S. 2014. Histone variants: dynamic punctuation in transcription. *Genes & development.* **28**(7), pp.672-682.

Weterings, E. and van Gent, D.C. 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair.* **3**(11), pp.1425-1435.

Wickliffe, K.E., Williamson, A., Meyer, H.-J., Kelly, A. and Rape, M. 2011. K11-linked ubiquitin chains as novel regulators of cell division. *Trends in cell biology.* **21**(11), pp.656-663.

Worcel, A., Strogatz, S. and Riley, D. 1981. Structure of Chromatin and the Linking Number of DNA. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences.* **78**(3), pp.1461-1465.

Worden, E.J., Padovani, C. and Martin, A. 2014. Structure of the Rpn11–Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nature structural & molecular biology.* **21**(3), p220.

Workman, J.L. and Kingston, R.E. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual Review of Biochemistry.* **67**, pp.545-579.

Wu, J., Huen, M.S., Lu, L.-Y., Ye, L., Dou, Y., Ljungman, M., Chen, J. and Yu, X. 2009. Histone ubiquitination associates with BRCA1-dependent DNA damage response. *Molecular and cellular biology.* **29**(3), pp.849-860.

Wu, J., Liu, C., Chen, J. and Yu, X. 2012. RAP80 protein is important for genomic stability and is required for stabilizing BRCA1-A complex at DNA damage sites in vivo. *Journal of Biological Chemistry.* **287**(27), pp.22919-22926.

Wu, P.-Y.J., Ruhlmann, C., Winston, F. and Schultz, P. 2004. Molecular architecture of the S. cerevisiae SAGA complex. *Molecular cell.* **15**(2), pp.199-208.

Wu, Q., Paul, A., Su, D., Mehmood, S., Foo, T.K., Ochi, T., Bunting, E.L., Xia, B., Robinson, C.V. and Wang, B. 2016. Structure of BRCA1-BRCT/Abraxas complex reveals phosphorylation-dependent BRCT dimerization at DNA damage sites. *Molecular cell.* **61**(3), pp.434-448.

Yau, R. and Rape, M. 2016. The increasing complexity of the ubiquitin code. *Nature cell biology.* **18**(6), pp.579-586.

Ye, Y. and Rape, M. 2009. Building ubiquitin chains: E2 enzymes at work. *Nature reviews Molecular cell biology.* **10**(11), p755.

Yoshida, A., Yoneda-Kato, N. and Kato, J.-y. 2013. CSN5 specifically interacts with CDK2 and controls senescence in a cytoplasmic cyclin E-mediated manner. *Scientific reports.* **3**, p1054.

Zeqiraj, E., Tian, L., Piggott, C.A., Pillon, M.C., Duffy, N.M., Ceccarelli, D.F., Keszei, A.F., Lorenzen, K., Kurinov, I. and Orlicky, S. 2015. Higher-order assembly of BRCC36–KIAA0157 is required for DUB activity and biological function. *Molecular cell.* **59**(6), pp.970-983.

Zhang, J., Cao, M., Dong, J., Li, C., Xu, W., Zhan, Y., Wang, X., Yu, M., Ge, C. and Ge, Z. 2014. ABRO1 suppresses tumourigenesis and regulates the DNA damage response by stabilizing p53. *Nature communications.* **5**, p5059.

Zhang, K. 2016. Gctf: Real-time CTF determination and correction. *Journal of structural biology.* **193**(1), pp.1-12.

Zheng, H., Gupta, V., Patterson-Fortin, J., Bhattacharya, S., Katlinski, K., Wu, J., Varghese, B., Carbone, C.J., Aressy, B. and Fuchs, S.Y. 2013. A BRISC-SHMT complex deubiquitinates IFNAR1 and regulates interferon responses. *Cell reports.* **5**(1), pp.180-193.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y. and Agard, D.A. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods.* **14**(4), pp.331-332.

Zhou, B.B.S. and Elledge, S.J. 2001. The DNA damage response: putting checkpoints in perspective (Reprinted from Nature, vol 408, pg 433-439, 2001). *Nature Reviews Molecular Cell Biology.* pp.S25-S31.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E. and Scheres, S.H. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife.* **7**, pe42166.

Zou, L. and Elledge, S.J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science.* **300**(5625), pp.1542-1548.