Cryo-EM studies of Phosphatidylinositol-3-kinase related kinase (PIKK) structures from yeast

Transcription-associated protein 1 (Tra1) and Mitosis-entry checkpoint 1 (Mec1), and their role as conserved genomic regulators

Luis Miguel Díaz-Santín

Institute of Structural and Molecular Biology (ISMB)

- University College London (UCL) and Birkbeck College, University of London -

This thesis is submitted as partial fulfilment of the requirements

for the degree of Doctor of Philosophy in Structural and Molecular Biology

December 2019

Declaration

The work described in this thesis was carried out at the Institute of Structural and Molecular Biology at Birkbeck College, London, under the supervision of Dr. Alan C.M. Cheung.

I, Luis Miguel Díaz-Santín confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature

Date

To my parents

Acknowledgements

I thank my supervisor Alan C.M. Cheung for giving me the opportunity to do a PhD in his group and to create the ideal scientific environment to grow every day. I will always remember this exciting and intense scientific experience. I am very certain that the scientific evolution I have experienced since I started to the present moment could not be possible in many other places.

I thank Natasha Lukoyanova, an exceptionally skilled EM expert who has been essential for the generation of the presented cryo-EM datasets. In addition to have been an enormously supportive person to my scientific work, providing high-quality cryo-EM data, I would like to thank her for the support she provided while working at the EM facility, and her supervision to make it run smoothly every day.

I am enormously grateful to my sponsor the Wellcome Trust foundation for generously accepting to pay my PhD fees through their funding. I have tried to pay back every penny with unconditional passion and full dedication to my research. I also thank the UCL for admitting me into the PhD programme at the Division of Biosciences, because it has been an honour to be part as a PhD student of this world recognised institution.

Likewise, I am very grateful to Birkbeck College and the ISMB, my home research institution, and the Department of Crystallography, where I performed my research. I feel very lucky of being part of this historical institution, because its magnificent environment is largely responsible for the development of my skills is Structural Biology. The extraordinary Birkbeck community has always been open to friendly share their extensive knowledge on structural biology. As part of this community, I would like to thank Helen Saibil and Elena Orlova for the great scientific environment they have created at Birkbeck, and especially to Helen for being my thesis committee chair. I also thank Finn Werner for being a member of my thesis committee. I would like to thank the EM group at Birkbeck College, because it has been a wonderful experience to be part of the initially small but very powerful community, which has been largely increasing in the number of members since I started. I have learned very much from their expositions and discussions, and enjoyed its friendly environment. I would like to thank Giulia Zanetti, because she supported us with her expertise the first time we collected data at DLS. I especially want to thank Dave Houldershaw, who has been enormously efficient and supportive with any technical request I have had during the exhaustive processing time that I have spent during my stay at Birkbeck College.

I would like to thank my lab mates from the G. Waksman and C. Vaughan groups, because they have been always very helpful and supportive in the lab, and because the friendly atmosphere they created, many of them already left before I finished. I thank in particular to Aravindan Ilangovan, Andreas Busch and Tiago Costa for their support in the lab and engaging scientific discussions, and Adam Redzej because he has always been very attentive. I thank Wenjuan Zhang because she has been a fabulous lab and office mate, always enthusiastic for scientific discussion and a very supportive person. I also thank my other office mates for their interesting conversations and also because of the friendly environment they created. I would like to thank Alejandro Peña for his support in the EM lab, and for insightful and engaging scientific discussions.

Finally, I would like to thank Peter Cherepanov and Lori Passmore for accepting to be my PhD examiners, and especially for the valuable feedback and suggestions that they have provided.

Abstract

The regulation of the genome is an essential biological task. In eukaryotes, fundamental nuclear processes such as DNA replication, transcription and repair require a large variety of protein effectors to ensure the correct transmission of genetic information, for regulated gene expression and to maintain genome integrity.

The phosphatidylinositol 3-kinase-related kinase (PIKK) protein family represents a group of very large and highly conserved Ser/Thr-protein kinases, where most function in genome regulation processes. *S. cerevisiae* proteins Tra1 and Mec1 (known as TRRAP and ATR in humans) are two essential PIKKs found in the nucleus.

Tra1 is a 3,744-residue protein and a shared subunit between the chromatin modifying complexes SAGA and NuA4. Tra1 stimulates transcriptional activation by directly interacting with transcription activators and is also involved in DNA repair. Here, I present the near-atomic resolution cryo-EM structure of the full-length Tra1 polypeptide. A *"de novo"* atomic model was built from the map, revealing the unique arrangement of Tra1 alpha-solenoid, which is organised into an N-terminal Finger segment inserted into a circular Cradle region before the Kinase domain. Structural interpretation of previously published Tra1 mutants allowed to identify a Gal4 binding region.

Mec1 is a 2,368-residue protein that forms a complex with Ddc2 (ATRIP in humans), and participates in the DNA-Damage-Response, Replication-Stress-Response, in telomere maintenance and meiosis, and its kinase activity influences cell cycle progression. Here, I present the high-resolution cryo-EM structure of the heterodimeric Mec1-Ddc2 complex, providing the first complete atomic model of Mec1. The structure reveals how the complex is assembled, showing extensive interactions between Ddc2 solenoid region and Mec1. The map also shows the integration point of Ddc2 N-terminal region into the complex, and together with existing structural data for the interaction of Ddc2 coiled-coil with RPA, an integrative model shed light in the understanding of Mec1 recruitment to RPA-coated ssDNA.

Impact statement

Since many PIKK proteins have an essential role in genome regulation, targeting them has been shown to cause genome instability and cell lethality. Therefore, PIKKs participating in the DNA damage response (DDR) can be targeted for cancer treatment. However, the difficulty to purify PIKK proteins has traditionally represented a handicap to develop new drugs and also to obtain structural information.

In this respect, the first high-resolution structure of the conserved and essential Transcription-associated-protein 1 (Tra1) from *S. cerevisiae* is presented in this study, solved by single-particle cryo-EM. Tra1 has a fundamental role in transcription activation as a general coactivator protein, participating in the expression of most genes in the cell. Given its high-level of conservation in evolution, the atomic structure of the Tra1 polypeptide also provides a structural framework to understand transcription activation in humans. Tra1 human homolog TRRAP has in addition clinical implications, as is a coactivator targeted by oncoproteins. The oncogenic protein c-Myc has been shown to interact with TRRAP, as well as the transcription factor E2F1, reported to present increased levels in many human cancers. Disrupting the interaction between TRRAP and oncogenic factors such as c-Myc has the potential to control oncogenic processes. In this regard, Tra1 was incubated with the transcription activator Gal4 for cryo-EM analysis purpose, representing to my knowledge the first attempt to analyse an activator bound to its full-length coactivator target using direct detector technology.

In addition, the first complete atomic model of *S. cerevisiae* Mec1 polypeptide, homolog to human ATR, is also presented in this study. The reported high-resolution structure of Mec1 is solved in complex to its binding partner Ddc2, homolog to the human ATRIP protein. After more than 20 years since ataxia-telangiectasia mutated and Rad3-related (ATR) was cloned for the first time, still important questions related to the regulation of this essential and conserved protein kinase remain to be answered, partially due to the lack of structural information. The reported structure reveals the complex architecture and assembly, in addition to the integration of Ddc2 N-terminal region to the rest of the complex, which previously remained unsolved. In addition, the full-length of the activation-loop was built in the Kinase domain, allowing to compare the key catalytic residues present in the complex to other eukaryotic protein kinases and shedding light into its mechanism of action. ATR^{Mec1} is the only essential and catalytically active PIKK protein directly participating in the DDR, during homologous recombination and activating the Fanconi anemia pathway. ATR also has a central role in the stabilisation of the replication fork upon replication stress situations. Therefore, all together makes ATR^{Mec1} one of the most attractive targets for cancer therapy within the PIKK protein family.

The research described in this work represents a proof of concept validating the drug development possibilities in the PIKK field, both in terms of biochemistry and structural biology, where the high-resolution structures of the essential and conserved Tra1^{TRRAP} and Mec1^{ATR} polypeptides are presented.

Publications

Parts of this work have been published:

Diaz-Santin, Luis Miguel, et al. "Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution." *Elife* 6 (2017): e28384.

Cheung, Alan CM, and Luis Miguel Díaz-Santín. "Share and share alike: the role of Tra1 from the SAGA and NuA4 coactivator complexes." *Transcription* 10.1 (2019): 37-43.

Table of contents

List of figures

List of tables

1. Introduction

The experimental work described in this thesis was focused on the PIKK proteins Tra1 and Mec1. Both are highly conserved and essential PIKK proteins that traditionally have been described to participate in transcription regulation and DNA repair, respectively. Whereas Tra1 is a direct target of transcription activators (TA) and functions within the SAGA and NuA4 acetyltransferase complexes, Mec1 forms a dimeric complex and participates in homologous recombination, among other biological processes. Hence, the following sections describe the PIKK protein family (1.1) and detail the functional backgrounds to Tra1 and Mec1 activities (1.2, 1.3 and 1.4).

1.1 The Phosphatidylinositol 3-kinase (PI3K) - related kinases (PIKKs) protein family

Eukaryotic protein kinases (ePK) constitute a group of enzymes which typically represent the 2 % in most eukaryotic genomes, with more than 500 identified members in humans (Manning et al., 2002). The ePKs are subclassified in different subfamilies according to homology criteria. In this regard, the Phosphatidylinositol 3-kinase (PI3K)-related kinases (PIKKs) constitute a family of very large and essential protein kinases, conserved in the eukaryotic domain of life, and with a direct role in cell regulation and genome stability (Hiom, 2005; Smith and Jackson, 2010; Bareti and Williams, 2014; Rivera-Calzada et al., 2015; Blackford and Jackson, 2017).

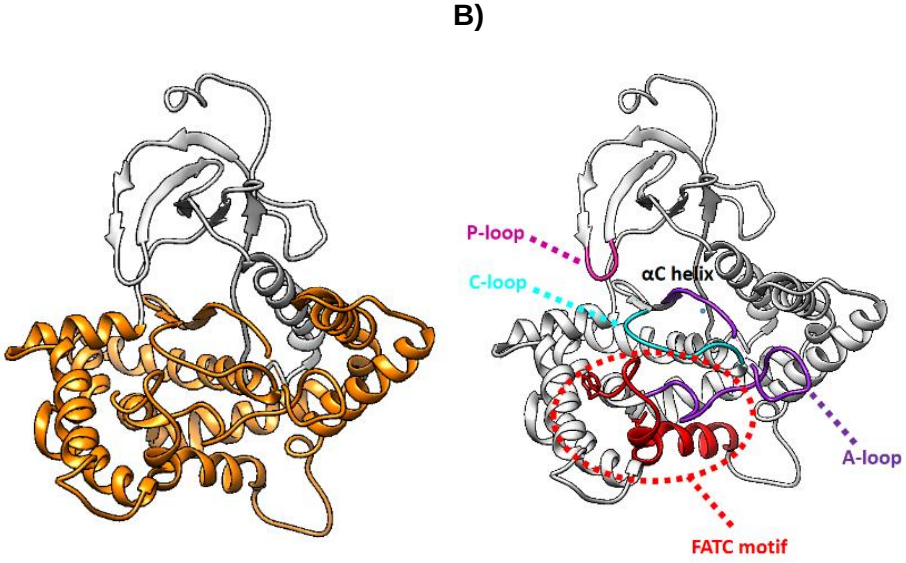


Figure 1.1 | The conserved architecture of eukaryotic protein kinases (ePKs). Tra1 atomic model (this study) is used to illustrate the general architecture in ePKs. **A)** The N-lobe within the kinase domain is shown in grey and the C-lobe in orange colour, both of them present in all ePKs. **B)** The different catalytic elements are outlined: P-loop (pink), activation loop (purple) and catalytic loop (blue). The structurally conserved αC helix in the N-lobe is also indicated. In addition, the conserved FATC motif at the C-terminal end of the protein, present in the PIKK protein family, is also shown (red colour).

The kinase domain in PIKKs shows a high level of homology when compared to the PI3K protein family, and therefore its name, but instead of phosphorylating lipids they can target a high number of protein substrates by adding a phosphate group to serine or threonine residues (Smith and Jackson, 2010). PIKKs were initially classified as atypical protein kinases (aPK) because of the lack in sequence similarity when compared to other ePKs. However, we can identify a very similar fold in the kinase domain of PIKKs comparing to other ePKs, as well as the presence of previously described catalytic elements in ePKs (Fig. 1.1).

1.1.1 PIKKs as genome and metabolism regulators

PIKKs have a central role in genome stability and are involved in fundamental biological processes such as the DNA damage response (DDR) (Blackford and Jackson, 2017), DNA replication stress response (RSR) (Friedel, Pike and Gasser, 2009), telomere maintenance (Mutiu et al., 2007), meiotic recombination (Cooper et al., 2014), transcriptional regulation (B. A. Knutson and Hahn, 2011), cell growth and metabolism control (Smith and Jackson, 2010), RNA quality control (Lovejoy and Cortez, 2009) and cell cycle progression and apoptosis (Smith and Jackson, 2010) (Table 1.1).

Table 1.1 The phosphatidylinositol 3-kinase-related kinase (PIKK) protein family

PIKK member			Biological function	Oligomeric state	Binding partners
S. cerevisiae	***S. pombe***	**Humans**			
Tra1	Tra1 and Tra2	TRRAP	Direct interaction with TAs	Monomer	$SAGA^{STAGA}$ and $NuA4^{Tip60}$ acetyltransferase complex subunits
-	-	DNA-PKcs	DDR (NHEJ)	Monomer	KU70/80
Mec1	Rad3	ATR	DDR (HR) and RSR	Dimer	$Ddc2^{ATRIP}$
Tel1	Tel1	ATM	DDR (HR)	Dimer	MRX^{MRN} complex
Tor1 and Tor2	Tor1 and Tor2	mTOR	Metabolism control	Dimer	LST8, Raptor, Rictor
-	-	SMG1	NMD	-	SMG8, SMG9

Table 1.1. In superscript the corresponding human homolog; in the case of NuA4, the complex fused in evolution together with the SWR1 complex to form the Tip60 complex found in humans (Fig. 1.16). NHEJ: non-homologus end joining. DDR: DNA damage response. HR: homologous recombination. RSR: replication stress response. NMD: non-sense mediated mRNA decay. The RSR is considered by some authors as part of the DDR (Forment and O'Connor, 2018); however, in the present text the DDR and the RSR will be described as different processes for convenience and explanation purposes.

Because of the essential nature of many of these proteins and the ability of several PIKKs to regulate a large number of proteins involved in cell cycle progression, some PIKK proteins have to be considered as essential cell master regulators in eukaryotes.

PIKKs function in many biological processes that require the binding of RNA and/or also DNA in different forms - single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), mRNA and

histone-bound DNA (Fig. 1.2). These interactions often regulate their kinase activities, and are usually not performed directly by PIKKs but instead they are mediated by adaptor proteins. In the case of DNA-PKcs, for example, the well-known KU proteins (KU70/80) mediate its interaction with nucleic acids. The MRX complex (MRN in humans) mediates the interaction in the case of Tel1^{ATM}, and the RPA trimeric complex in the case of Mec1^{ATR}. Tra1^{TRRAP}, as part of two different chromatin-modifying complexes, i.e. SAGA and NuA4, interacts indirectly with histone-bound DNA through the multiple histone binding domains present in other complex subunits.

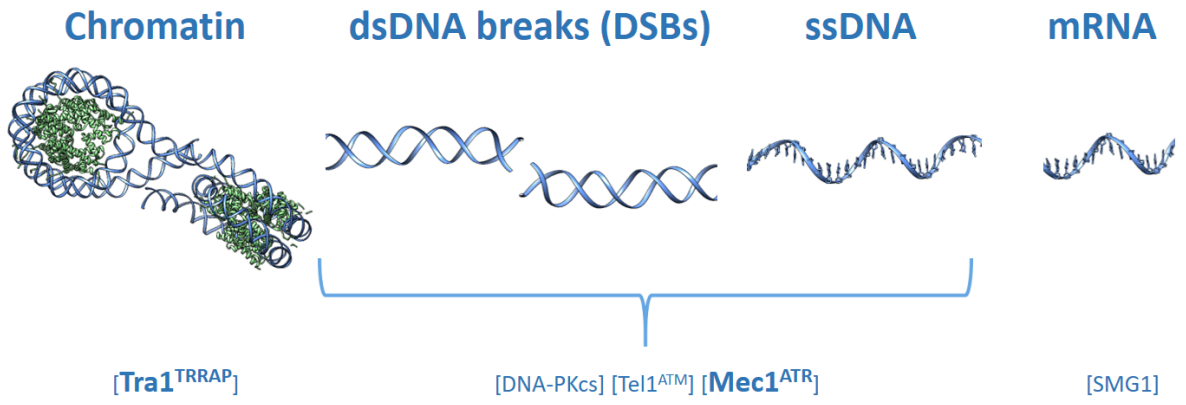


Figure 1.2 | PIKKs regulate nucleic acids in different forms. All PIKKs but mTOR have been shown to have a direct role in directly regulating nucleic acids through adaptor proteins. PIKKs are associate to the regulation of nucleic acids in different forms, i.e. histone-bound DNA, free dsDNA, ssDNA and mRNA. The participation of each PIKK is shown in the figure. Yeast names for each PIKK are given, with human naming in superscript, except DNA-PKcs and SMG1 which are absent in *S. cerevisiae.* (PDB 1ZBB is shown to represent chromatin and was used for illustration purpose in the other cases).

1.1.2 The relationship between different PIKK protein family members

Because of their very large size and the difficulty to purify enough protein for crystallisation assays, very limited structural information has been available for full-length PIKKs before the recent revolution in cryo-EM, with only two but very valuable X-ray crystallography works released before 2015 (Sibanda, Chirgadze and Blundell, 2010; Yang *et al.*, 2013), corresponding to partial regions of two different PIKK proteins. However, the cryo-EM revolution in single-particle analysis has deeply transformed our understanding of the PIKK protein family in a very short period of time, with multiple cryo-EM structures released since 2015, all of them solved by direct-detector technology at a resolution sufficient at least to identify secondary structure elements. In addition to sequence alignment and homology analysis tools (Fig. 1.3), the recent advancements in cryo-EM allows for the very first time to perform a structural comparison (Fig. 1.4) among the different PIKK members, revealing the existence of pairs of architecturally equivalent PIKKs where only one member is essential for the cell within each pair. This classification based on structural and sequence homology (Fig. 1.3 and 1.4) between the different PIKKs provides the following three pairs or groups of

paralog proteins[1]: TRRAP^{Tra1} and DNA-PKcs, ATR ^{Mec1} and ATM^{Tel1}, mTOR^{Tor} and SMG1. SMG1 has been very recently published at near-atomic resolution (Gat *et al.*, 2019; Zhu *et al.*, 2019), being the last member to be solved by a direct-detector device; homology modelling analysis also identified mTOR as the PIKK protein with the highest level of homology to SMG1 and therefore they have been grouped together (Fig. 1.3 and 1.4). This scheme for the PIKK protein family strongly suggests that a duplication event occurred in evolution, generating a total of six different members in humans, and where three of them are essential proteins in the cell. All PIKKs have been described to be nuclear proteins but mTOR. In the case of SMG1, very little literature there is available, but given the level of homology to mTOR and according to the Human Protein Atlas information for tissues (Uhlen *et al.*, 2010), it could be categorised as cytoplasmic protein as well. In the yeast *S. cerevisiae* we find two different homologs for Tor - named Tor1 and Tor2 - instead of only one, as is the case in humans. Finally, SMG1 and DNA-PKcs are found in humans and absent in *S. cerevisiae*.

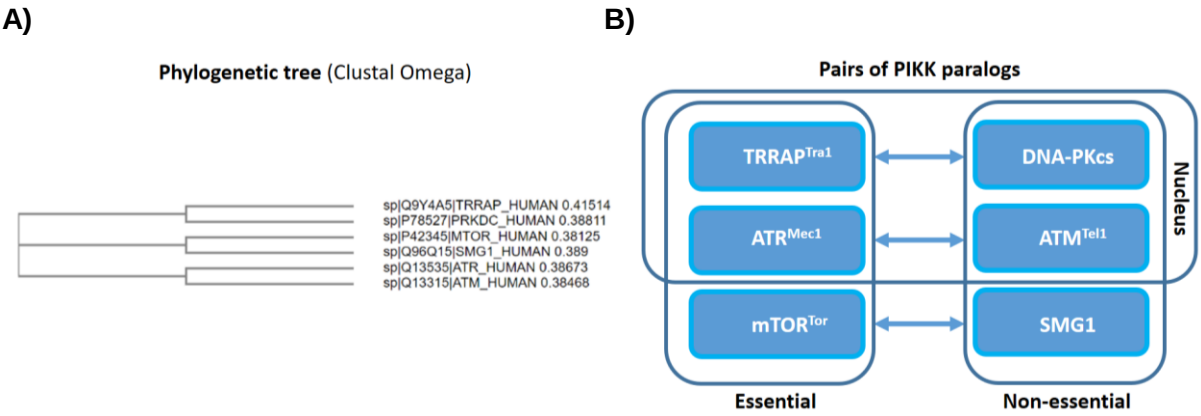


Figure 1.3 | Overview of the PIKK protein family: sequence and evolutionary analysis. A) Sequence alignment in Clustal Omega of the six human PIKK proteins, showing the phylogenetic tree analysis. **B)** Three pairs of paralog PIKK proteins are found in humans; within each pair, only one PIKK member is essential for the cell. Four PIKK proteins present a very well described nuclear localisation and function, as is shown in the figure. Human names for each PIKK are given, with yeast naming in superscript except for DNA-PKcs and SMG1, which are absent in yeast as described in the main text.

Increasing evidence in the literature suggests additional roles for different PIKKs, distinct to the primary biological processes where they were originally described (Table 1.1). In this regard, Tra1 may have roles in DNA repair and telomere integrity in addition to participate as a transcriptional regulator (Askree *et al.*, 2004; Robert *et al.*, 2006; Hoke *et al.*, 2008; Kvas, Gloor and Brandl, 2012). In addition, several PIKK phosphorylation targets have been described to be common to several members, indicating a clear cross-talk between the pathways controlled by the different kinases within the family (Maréchal and Zou, 2013). This

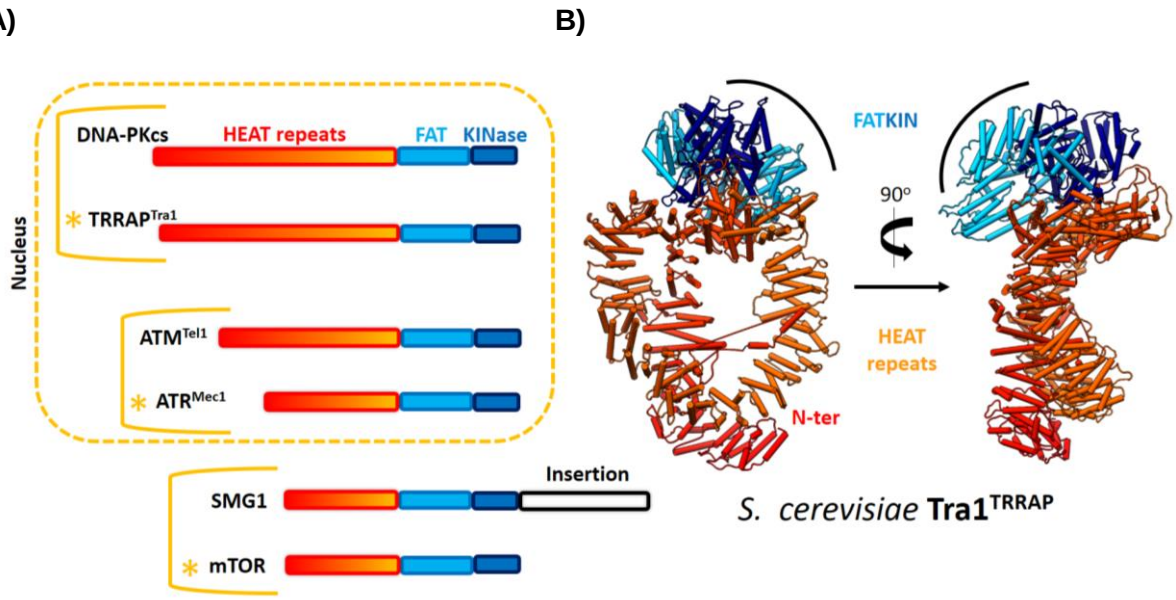
[1] The corresponding PIKK ortholog (either yeast or human) will be written in superscript to remind to the reader when considered appropriate, i.e. yeast (y) Tra1^{TRRAP} and human (h) TRRAP^{Tra1}, yMec1^{ATR} and hATR^{Mec1}, yTel1^{ATM} and hATM^{Tel1}.

model where different PIKKs overlap over similar biological functions may provide redundancy that ensures for the stability of the genome. Given the structural similarity between the PIKKs and the sometimes overlapping biological processes where they are involved, the understanding of the each individual PIKK within its particular context is essential to understand the others, since both similarities and differences can be exploited to gain a deeper understanding of the specific biological functions of each PIKK member.

1.1.3 General architecture of the PIKK protein family

This protein family of very large polypeptides – ranging from 2,549 to 4,128 residues in humans - present a general and common domain organisation, shared by the different family members. PIKK proteins exhibit an extensive α-solenoid region, varying from 1,938 residues as found in SMG1 (Gat et al., 2019) - the human PIKK with the shortest solenoid region - to 3,565 residues in DNA-PKcs (Sibanda et al., 2017) - the PIKK with the largest solenoid region. This α-solenoid region is composed of two different types of helical repeats, i.e. HEAT repeats (**H**untingtin, elongation factor 3 (**E**F3), protein phosphatase 2**A** (PP2A), yeast kinase **T**OR1 domain) and TPR repeats (**T**etratrico **p**eptide **r**epeat). The HEAT repeat region in PIKKs folds in the three dimensional space in a very particular manner depending on its length and composition, generating a singular topology which distinguishes each member within the family (Fig. 1.4). Whereas HEAT repeats are 37-47 amino acid long, TPR repeats have been defined to be 34 amino acids long, and both repeat types consist of pairs of antiparallel alpha helices that create a solenoid-like structure stabilised by a hydrophobic core. HEAT and TPR repeats mediate protein-protein interactions (Blatch and Lässle, 1999; Groves *et al.*, 1999) participating in many different biological processes (1.1.1), including protein transport and protein folding (Hunter *et al.*, 2008). When they are present in a tandem fashion, they generate a right-handed super-helical structure, and have been described as highly flexible super-secondary structure elements (Jínek *et al.*, 2004; Forwood *et al.*, 2010; Kappel *et al.*, 2010). TPR repeats associate in 3-16 tandem repeats to generate a TPR motif. In PIKKs, TPR motifs are found at the FAT domain (**F**RAP-**A**TM-**T**RRAP), which is located in between the long track of HEAT repeats present at the N-terminal region and the conserved Kinase domain present in all PIKK members at the C-terminal region of the polypeptide. As is the case for other eukaryotic protein kinases (ePK), the PIKK kinase domain is organised in two different lobes (Fig. 1.1). Although in PIKKs the catalytic residues are conserved comparing to other ePKs, the canonical catalytic motifs found in ePKs are poorly conserved in PIKKs in terms of sequence, and for that reason they are considered as atypical protein kinases (aPKs). Still, previously defined residues in ePKs important for catalysis can be identified in PIKKs in the active site, and the same structural motifs are also found, including the so-called phosphate

binding loop (P-loop), the activation loop (A-loop) and the catalytic loop (C-loop) (Fig. 1.1). The Kinase domain is clamped by the FAT domain in all PIKKs and the two domains together form the globular HEAD domain (Díaz-Santín *et al.*, 2017), also called FATKIN (Baretić *et al.*, 2017), which is structurally similar in all PIKK members (Fig. 1.4). In the case of SMG1, an insertion is found at the very C-terminal region after the Kinase domain, something unique within the family. The FAT domain of human mTOR (Yang et al., 2013) has been previously described to be composed at the same time of four discontinuous domains (Yang et al., 2013), i.e. three TPR repeat domains (TRD1, TRD2 and TRD3) and a single HEAT repeat domain (HRD) at the C-terminal region of the FAT domain. However, in the case of Tra1^{TRRAP}, the FAT domain has been described to be composed of 15 TPR repeats (Díaz-Santín *et al.*, 2017). Given that the repeat type or architecture is not clear for many pairs of α-helices because they do not present a canonical pattern or structure, for simplicity reasons in the present text the FAT region of Tra1^{TRRAP} and Mec1^{ATR} is described as essentially a TPR repeat region.



C)

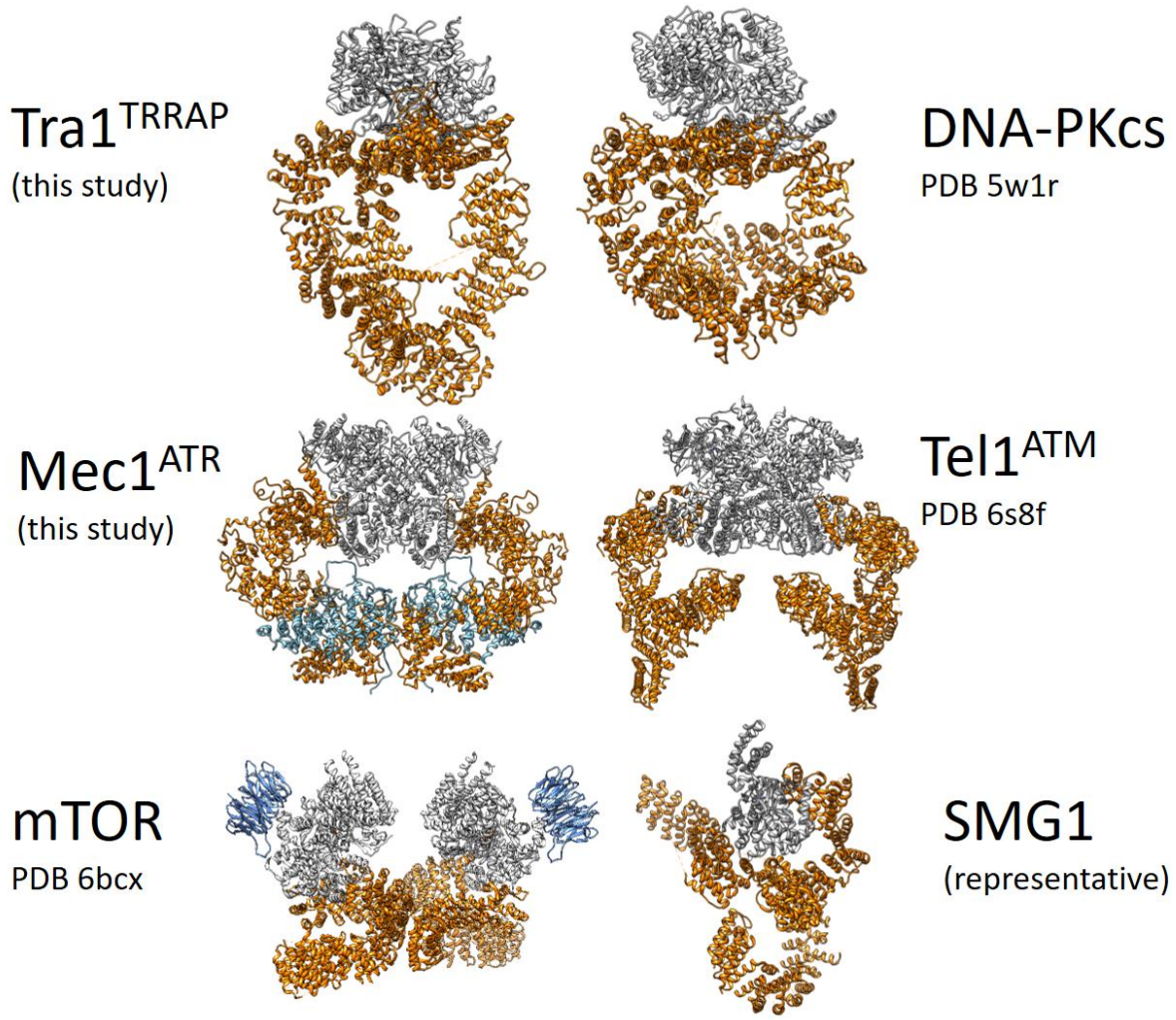


Figure 1.4 | General architecture of the PIKK protein family. The PIKK protein family and its different topological regions. **A)** The six different PIKK members found in humans are represented schematically, and the yeast ortholog for each PIKK is shown in superscript; DNA-PKcs and SMG1 are absent in *S. cerevisiae* and therefore no superscript is used. The HEAT region, variable in length among PIKKs, is shown in red to yellow colour gradient. The FAT domain is shown in light blue and the kinase region in dark blue, and together they form the globular FATKIN domain. The kinase domain is highly conserved among PIKKs. In the case of SMG1, there is a unique insertion at the C-terminal end of the protein after the Kinase domain. *The star indicates essential proteins for the cell. The brackets group pairs of paralog proteins together. **B)** Tra1 atomic structure (this study) represented in pikes and planks is used to illustrate in 3D the different PIKK topological regions shown in **A**. (**A** and **B** were inspired by Imseng and Aylett, 2018). **C)** The atomic models of available PIKKs together with the atomic models presented in this study are rendered following the same organisation as in Fig. 1.3. The HEAT repeat region is shown in orange and the FAT and the Kinase domains in grey, creating a globular domain together; in the case of Mec1, the Ddc2^{ATRIP} subunit is shown in light blue, and for mTOR, the mLST8 subunit is shown in dark blue, whereas the Raptor subunit has been removed for clarity. Although two high-resolution structures of SMG1 have been published so far, the corresponding atomic models are not available yet, so that a representative model of monomeric apo-mTOR is shown instead (PDB 6bcx; mLST8 and Raptor subunits were removed for rendering). Distinct to **A**, in the case of *S. cerevisiae* Tra1, Mec1 and Tel1, the corresponding human ortholog is shown in superscript.

In addition, in PIKKs at least other three different regulatory elements or secondary structure motifs can be found within the kinase domain; from N to the C-terminal region: the FRB motif - made of four α-helices, and absent in ATR and ATM -, the LBE motif - made of two α-helices- and the FATC motif, also composed of two α-helices, and located at the very C-terminal end of the protein. In addition to the previously mentioned, a PRD motif in the C-lobe has been also described (Lovejoy and Cortez, 2009), located before the FATC motif. Although the sequence of the PRD domain does not seem to be very well conserved among PIKKs, the position of the PRD motif is structurally conserved among PIKK members as has been revealed by the comparison of the recently published structures (4.3.3). In SMG1, the unique insertion that is found after the Kinase domain consists of 1,094 residues in humans when SMG1 is aligned to the other PIKKs. Although two high-resolution cryo-EM structures of SMG1 have been recently released (Gat *et al.*, 2019; Zhu *et al.*, 2019), the density for the long insertion at the C-terminal region seems to be missing, possibly indicating a high degree of conformational flexibility for this region of the polypeptide. Interestingly, secondary structure prediction analysis by the PSIPRED server (Buchan and Jones, 2019) shows that the insertion is mostly populated by alpha-helices, where the unstructured region as predicted by the Disopred3 tool corresponds to the last ~ 142 residues, accounting for only ~ 13% of the insertion. Furthermore, homology modelling analysis of the 1,094-residue insertion by Phyre2 (Kelley *et al.*, 2015) predicts that the most similar protein structure, according to sequence homology, is mTOR (PDB 6SB2), and secondary structure analysis in Phyre2 predicts that the 75% of the insertion presents alpha-helix fold. These analyses suggest that the folding of this region might consist in a long α-solenoid segment as is the case for the N-terminal region in all PIKKs, and followed in this particular case by a ~ 142 residue unstructured C-terminal region at the end of the insertion. Although the functional and mechanistic reasons to explain the observed experimental results still need to be addressed, the recently released high-resolution cryo-EM structures of SMG1 provide very valuable information regarding to the interaction with its binding partners, SMG8 and SMG9, and allows the direct comparison to the mTOR structure bound to its G-protein partner RHEB.

Finally, and related to the overall architecture of PIKKs, it has been found that (m)TOR, ATR^{Mec1} and ATM^{Tel1} form dimeric complexes (Fig. 1.4), whereas $TRRAP^{Tra1}$ and DNAP-PKcs have been solved at near-atomic resolution as a monomer, including the recently released structures of SMG1 (Gat *et al.*, 2019; Zhu *et al.*, 2019).

1.2 Transcriptional activation in Eukaryotes and the involvement of Transcription-Associated Protein 1 (Tra1)

The synthesis of messenger RNA (mRNA) from a DNA template molecule is an essential biological process for life and is tightly regulated in all living organisms. In eukaryotes, most of genes are transcribed by RNA polymerase II (RNA Pol II from hereafter), a multi-protein complex with enzymatic activity that is recruited to promoter DNA regions (1.2.1) where transcription initiation occurs. The RNA Pol II works in concertation with many other transcription factors (TFs), which together have been defined as the general transcription machinery. In this regard, the promoter DNA represents a platform for many TFs to assemble and form together with the RNA Pol II the so-called pre-initiation complex (PIC) (1.2.2), a requirement previous to transcription initiation. The recruitment of RNA Pol II to promoter DNA regions is not performed directly but instead requires the interaction of transcription activators with coactivator proteins (1.2.1 and 1.2.3). TAs are TFs bind to specific DNA regions named Upstream Activating Sequence (UAS) in yeast, and "enhancer" in metazoans, to activate transcription. The direct interaction of TAs with transcription coactivator complexes causes the recruitment of these complexes to promoter DNA regions. Finally, coactivator complexes are able to recruit the basal transcription machinery with the RNA Pol II at its core, stimulating the formation of the PIC and transcription initiation (Fig. 1.5). Coactivator examples include complexes such as the SAGA complex (1.2.3.3) that catalyses histone acetylation, the TFIID complex (1.2.3.1) that loads TATA-binding protein (TBP) into the promoter DNA previous to the assembly of the PIC, and Mediator complex (1.2.3.2), that stimulates the assembly of the PIC. Additional complexes such as NuA4 (Tip60 in humans) (1.2.3.4) have also been described as transcription coactivators.

In this regard, the SAGA essential subunit Tra1 (1.2.4) in yeast physically interacts with several transcription activators such as Gal4, Gcn4 and Hap4, and it is conserved in evolution from yeast to humans. Because the mechanism of activator binding and specificity to the specific target is not yet understood, and because Tra1 has been largely described as a target of several transcription activators, the Tra1 protein represents an exceptional model for the study of transcription activation.

1.2.1 The binding of transcription factors to the promoter DNA in eukaryotes

As coactivators lack intrinsic DNA binding activity, TAs are needed for their recruitment to the promoter DNA; they are proteins typically composed of a DNA binding region or domain (DBD) that binds specific DNA sequences, and a transactivation domain (TAD) that binds coactivators (Fig.1.5). The DBD and the TAD are usually connected by a flexible or unstructured region. This modularity makes TAs difficult targets for full-length structural

analysis, distinct to the case of globular, single domain proteins. Some activators such as Gal4 and Gcn4 present a high content in acidic residues, and therefore their TAD is also called acidic activation domain (AAD). Therefore, transcription coactivators are direct targets of TADs, and very little is known regarding to the mechanism of this interaction.

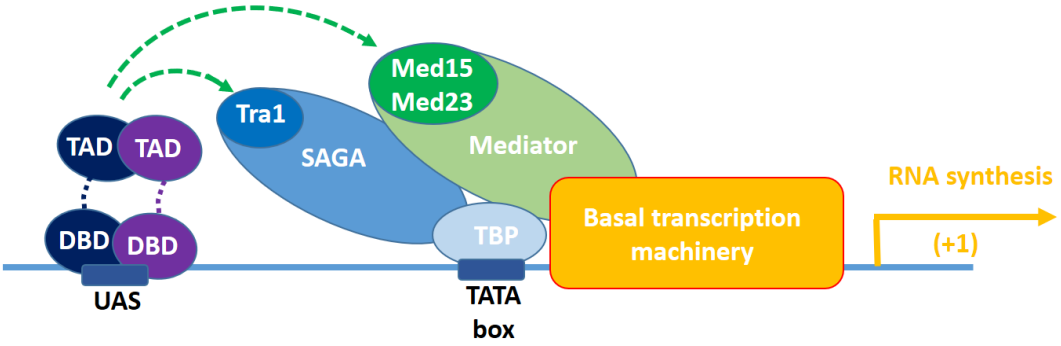


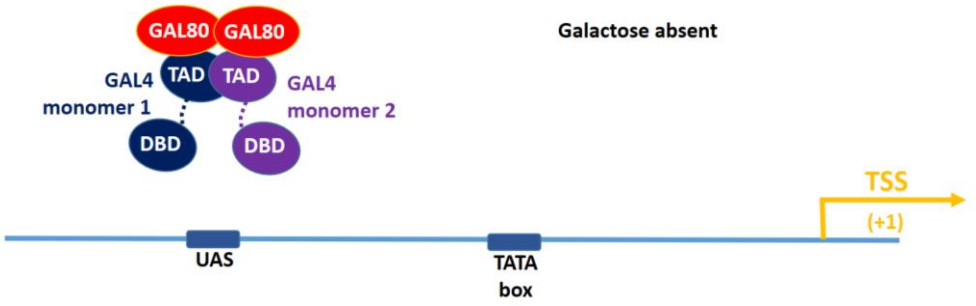
Figure 1.5 | Transcriptional activation in eukaryotes by transcription coactivators. A schematic representation of a typical promoter DNA in *S. cerevisiae* is shown, together with a model of transcription activation by activator proteins. Transcription initiates at the core promoter DNA, where the transcription start site (TSS) is found, represented by the +1 position. A transcription activator (TA) is represented as one DNA binding domain (DBD) and a trans-activating, trans-activation or transcription activation domain (TAD) connected by a flexible linker (dotted line). After the dimerisation of the two TAs and binding to the upstream activation sequence (UAS), TAs recruit coactivator complexes to the promoter DNA (such as SAGA and Mediator). The SAGA subunit Tra1 and Mediator subunits Med15 and Med23 directly interact with TAs. After the recruitment of coactivators to the promoter DNA, the basal transcription machinery is also recruited for transcription to start. (Inspired by Traven, Jelicic and Sopta, 2006).

The regulation of transcription in eukaryotes involves the direct binding of TFs to specific DNA sequences. In this regard, TAs target UAS regions (Fig. 1.5), and protein coding genes can be controlled by either single or multiple UASs and TAs. In yeast, the majority of UASs are located 5' of the promoter, and the level of transcription varies with the position and number of UASs (Hahn and Young, 2011). In higher eukaryotes, "enhancer" DNA elements can work upstream or downstream of genes and at vast distances from their regulated target gene (Pennacchio *et al.*, 2013; Shlyueva, Stampfel and Stark, 2014). A common feature for most UASs and active core promoters is the location in a nucleosome free region defined as nucleosome depleted region (NDR) (Haberle and Stark, 2018), allowing for the interaction with TAs such as the Gal4 activator in yeast and the recruitment of the basal transcription machinery . In humans, we find more than 1,600 genes encoding transcription factors (Lambert *et al.*, 2018; Cramer, 2019), and many of them present a DNA binding motif. It is possible to classify TAs in three different classes depending on the architecture of the DBD: zinc-finger, helix-turn-helix and the leucine-zipper motifs (Struhl, 1989; Chen and Rajewsky, 2007). The specificity of TFs for the binding to a given DNA sequence is central to understand TFs biology, and in this regard, chromatin immunoprecipitation followed by high-throughput

DNA sequencing has provided near single-nucleotide resolution of DNA binding sites when the ChIP-exo technique is used (Lis, 2019).

Some mechanisms of TA regulation involve the control of their expression levels or cell localisation. In the case of the Gal4 activator, its AD is blocked by the inhibitor Gal80 in the lack of galactose, and its presence releases Gal4 TADs to activate transcription (Traven, Jelicic and Sopta, 2006; Diep *et al.*, 2008) (Fig. 1.6).

A)



B)

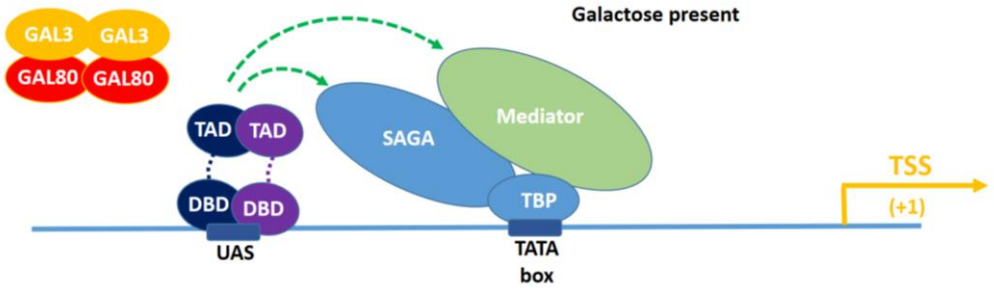


Figure 1.6 | Transcription regulation by the Gal4/Gal80 system. A) In the absence of galactose, Gal80 is bound to Gal4, impeding its binding to the UAS region in the promoter DNA. **B)** When galactose is present in the culture medium, Gal80 is bound to Gal3, and Gal4 is free to bind to the promoter. After dimerisation and binding to the UAS, Gal4 directly interacts with coactivator complexes (such as SAGA and Mediator) to activate transcription.

As described before, transcription activation involves the assembly of RNA pol II and general transcription factors at the core promoter DNA – containing the regulatory UAS elements – to form the PIC. The core promoter DNA can be defined as the minimal DNA element necessary for basal transcription (Fig. 1.7). In this context, it has been shown that at least 51 bp of promoter DNA is in contact with the core of the PIC, including the RNA pol II and the participating TFs (Forget *et al.*, 1997).

In metazoans, different elements have been identified as important core promoter sequences. In addition to the **TATA box,** these elements might include the TFIIB recognition element (**BRE**) upstream of the transcription start site (TSS) and initiator (**Inr**), whereas the motif 10 element (**MTE**) and downstream promoter element (**DPE**) are located downstream of the TSS

(Fig. 1.7) (Juven-Gershon and Kadonaga, 2010; Haberle and Stark, 2018). In addition, other regulatory elements can be found such as the downstream core elements (**DCE**) and the DNA replication related element (**DRE),** among others. The TATA binding protein recognises the TATA element whereas MTE, DPE and Inr are specifically recognised by TBP-associated factors (TAFs) present in the coactivator complex TFIID (Nogales, Patel and Louder, 2017) (1.2.3.1). BRE instead is recognised by the general transcription factor TFIIB. The previously described elements are combined into a given core promoter to regulate activator specificity (Fig. 1.7).

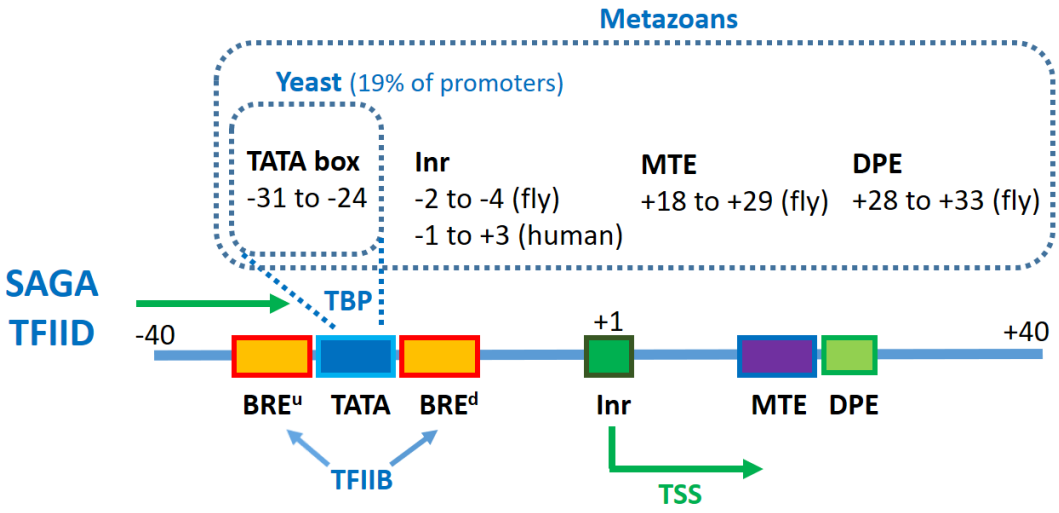


Figure 1.7 | The core promoter DNA in metazoans versus yeast. Different elements can be found at the promoter DNA in mammalian cells, including the TATA box, Inr, MTE and DPE elements. In *S. cerevisiae*, only the TATA box is found, which is present in only the 19% of promoters. Both the SAGA and TFIID coactivators are recruited by transcription activators and load the TATA binding protein (TBP) into the core promoter DNA to stimulate transcription initiation (indicated by the green arrow) (Cramer, 2019). BREu: BRE upstream. BREd: BRE downstream. TSS: transcription start site. (Inspired by Juven-Gershon and Kadonaga, 2010)**.**

The **TATA** element is the only metazoan motif that is conserved in yeast (Basehoar, Zanton and Pugh, 2004), and although TFIID participates in the expression of most genes in the cell, the elements necessary for TFIID recognition have not been identified yet.

Two different mechanisms of PIC assembly were initially described, dependent on the Transcription Factor II D (TFIID) or the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex. These complexes directly interact with activators at UAS and recruit TBP to the promoter DNA region. It was initially believed from genome wide experiments that promoters presenting a TATA box depend on the SAGA complex, considered to be highly regulated and involved in the stress response (Basehoar, Zanton and Pugh, 2004; Huisinga and Pugh, 2004). Approximately 19% of promoters in *S. cerevisiae* contain a TATA box, and the prevailing paradigm in the field was that 10% of all yeast promoters were controlled by the SAGA complex, whereas the remaining 90% of yeast promoters was believed to be activated by TFIID, typically constitutively active (often with a role as housekeeping genes) and generally

TATA-less genes. However, a recently published work (Baptista *et al.*, 2017) demonstrates that SAGA is rather a general transcription cofactor of the RNA Pol II, as existing studies suggested before (Lee *et al.*, 2000). The SAGA coactivator complex maps to the UAS of most yeast genes and shows overlap with Mediator binding. Furthermore, SAGA is responsible for TBP recruitment at promoter DNA regions in a stable manner, and the disruption of SAGA decreases the global levels of RNA pol II transcription as measured by newly synthesised mRNA rather than steady state mRNA levels. In a parallel study, it was shown that the general coactivator TFIID is responsible for the expression of almost all RNA Pol II dependent genes in *S. cerevisiae* (Warfield *et al.*, 2017), with similar TFIID dependency for both TATA and TATA-less promoters. The resultant hypothesis is that SAGA and TFIID associate together at most RNA pol II dependent genes, contrary to the previous believe that the transcription of most of yeast genes is dominated either by one complex or the other acting as alternate transcriptional factors. This suggests that the SAGA complex is a central regulator of most genes. Therefore, its study is necessary for the understanding of eukaryotic transcriptional regulation and highlights the importance of obtaining knowledge on the structure and function of the SAGA complex to understand transcription initiation in yeast. In this regard, Tra1 plays a key role as an essential subunit of the complex and direct target of transcription activators.

1.2.2 The pre-initiation complex (PIC) assembly

Transcription initiation of protein-coding genes starts with the assembly of a PIC on promoter DNA (Roeder, 1996). The PIC is made of RNA pol II and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Fig. 1.8) (Nogales, Patel and Louder, 2017). The Mediator complex is a transcriptional coactivator that functions as a transcriptional chaperone (from a functional point of view exclusively) by stabilising the PIC and can be considered as a general factor for transcription initiation. Mediator is recruited to the promoter DNA by the direct interaction with transcription activators, as is the case for the Tra1 subunit present in the SAGA and NuA4 complexes (1.2.4) that also promotes the assembly of the PIC. There are existing structures for the PIC lacking TFIIH (termed core PIC or cPIC) from yeast and human at near-atomic resolution (He *et al.*, 2016; Plaschka *et al.*, 2016), as well as crystal structures of core Mediator (cMed) complex from yeast (Nozawa, Schneider and Cramer, 2017) and cryo-EM structures of full-length free Mediator from yeast (Tsai *et al.*, 2017) and mouse (El Khattabi *et al.*, 2019). There are also existing structures for TFIIH as part of the PIC together with RNA Pol II and Mediator from yeast (Schilbach *et al.*, 2017) (Fig. 1.8) and humans (Louder *et al.*, 2016). The yeast structure of PIC-cMed complex containing the TFIIH factor is a 2 MDa structure consisting of 46 polypeptides essential for transcription initiation in yeast, with more than 70% of the mass modelled at near-atomic level (Fig. 1.8). Although in

the published structure of the yeast cPIC-cMed-TFIIH complex the Mediator tail module is not present, its position on the PIC has been already described in yeast (Robinson *et al.*, 2016). Free TFIIH structures have been already reported at high-resolution by single-particle cryo-EM (Greber *et al.*, 2017, 2019; Kokic *et al.*, 2019), extending the model from transcription regulation to nucleotide excision repair (NER). Finally, a composite model of the TFIID-based PIC (complete PIC) in humans has been published (Nogales, Patel and Louder, 2017), containing all the TFs described before in addition to TFIIS, and generated by the superimposition of three different existing cryo-EM structures (He *et al.*, 2016; Louder *et al.*, 2016; Tsai *et al.*, 2017).

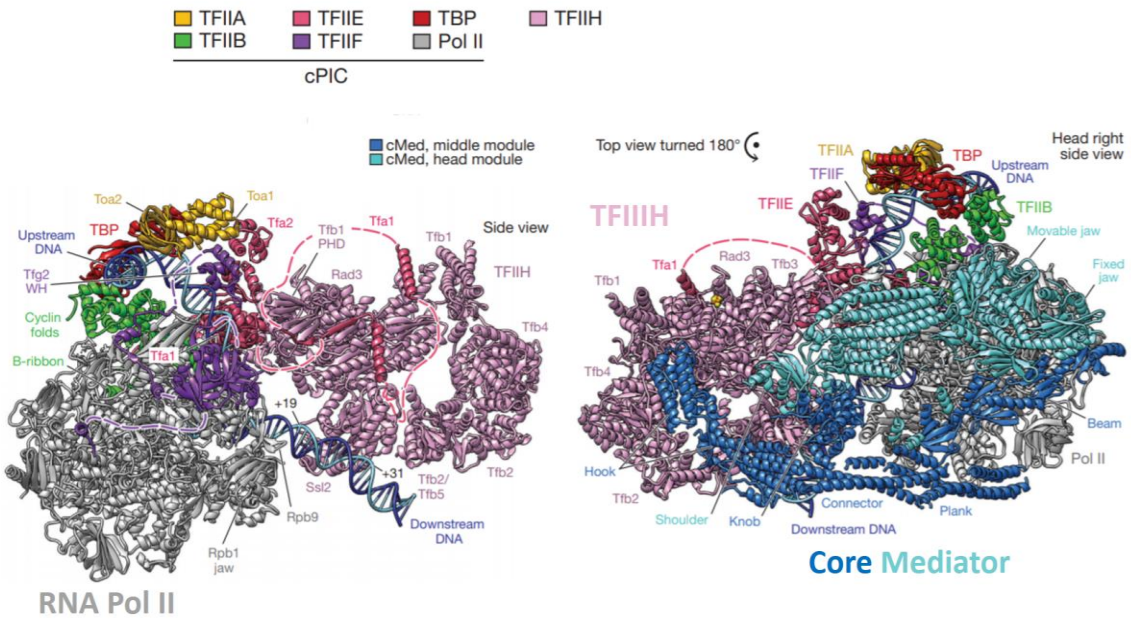


Figure 1.8 | cPIC bound to core Mediator (Head and Middle modules) and TFIIH. The atomic model for the core PIC in complex with core Mediator and TFIIH is rendered in ribbon. Mediator performs extensive contacts with both the RNA Pol II and TFIIH. Mediator Tail module, with a role in the direct interaction with transcription activators, is not present in the structure, as is the case for the kinase module. (Adapted from Schilbach et al. 2017).

Upon PIC assembly there is a transition to the open complex that involves a strong conformational change. In this process where promoter DNA opens, DNA is positioned over the RNA pol II cleft through interactions between TBP and transcription factors TFIIA, TFIIB, TFIIE and TFIIF (Schilbach *et al.*, 2017).

1.2.3 General Transcription Coactivators

In addition to TFIID and Mediator, another conserved general transcription coactivator complex in eukaryotes is the SAGA complex. These three transcription coactivators are essential and necessary for the transcription of nearly all protein-coding genes transcribed by RNA pol II in *S. cerevisiae* (1.2.1 and 1.2.2). Although they can be recruited to the promoter

DNA region through similar mechanisms, i.e. interaction with transcriptional activators, they have different roles in transcription initiation. In the case of TFIID and SAGA, both bind to the promoter DNA and are able to recruit the basal transcription machinery, but only SAGA presents enzymatic activity on nucleosomes (Fig. 1.9) (Koutelou, Hirsch and Dent, 2010). Therefore, although both complexes are able to load TBP to the promoter DNA region (Cramer, 2019), SAGA can also modify the chromatin structure at the promoter region by both nucleosome acetylation and deubiquitination, activities well-defined into different modules within the complex.

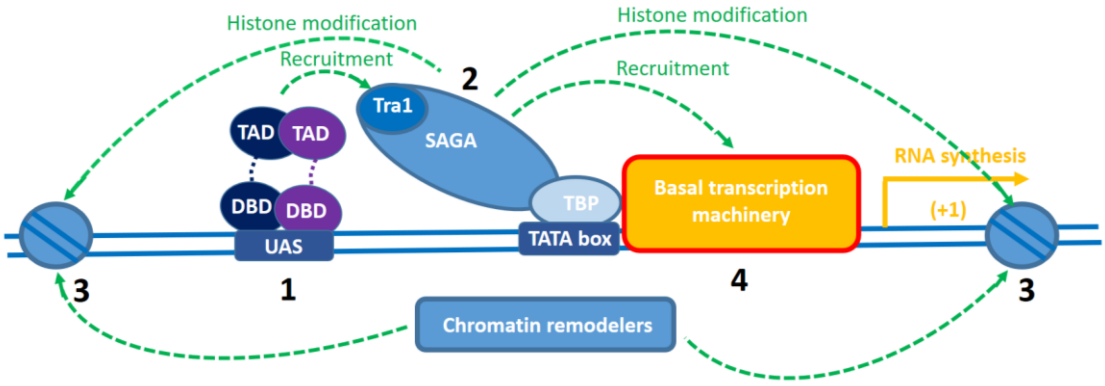


Figure 1.9 | SAGA is a coactivator complex with histone modification activities. Upon binding to an UAS sequence at the promoter DNA, a transcription activator (1) is able to recruit a co-activator complex, exemplified by the SAGA coactivator in the figure (2). The Tra1 subunit within the SAGA complex is responsible for the direct interaction with transcription activators. Upon recruitment, SAGA modifies chromatin (3) and interacts with the basal transcription machinery via TBP (4) to initiate transcription. After histone acetylation by the SAGA complex, ATP-dependent chromatin remodeling complexes can further modify chromatin by directly interacting with acetylated nucleosomes.

Histone acetylation might be followed by nucleosome depletion after the recruitment of chromatin remodeling complexes (Fig. 1.9). According to this model, a nucleosome-free promoter DNA region would facilitate the access to the Mediator coactivator complex for the formation of the PIC. Mediator can interact with transcription activators and mediates PIC assembly prior to transcription initiation by RNA pol II, so that its function can be understood as a necessary “transcriptional chaperone” if interpreted from a functional point of view only.

1.2.3.1 TFIID

The transcriptional coactivator TFIID factor is a megadalton multiprotein complex composed of TBP and 13 different TBP associated factors, the so-called TAF proteins (Papai, Weil and Schultz, 2011), which are proteins presenting histone fold domains. TFIID has been described as the first general transcription factor that binds to the promoter region and displays a high level of conformational flexibility, necessary for promoter binding, and presenting distinct

conformational states (Cianfrocco *et al.*, 2013) (Fig. 1.10). Three different lobes can be identified in TFIID, named A, B and C.

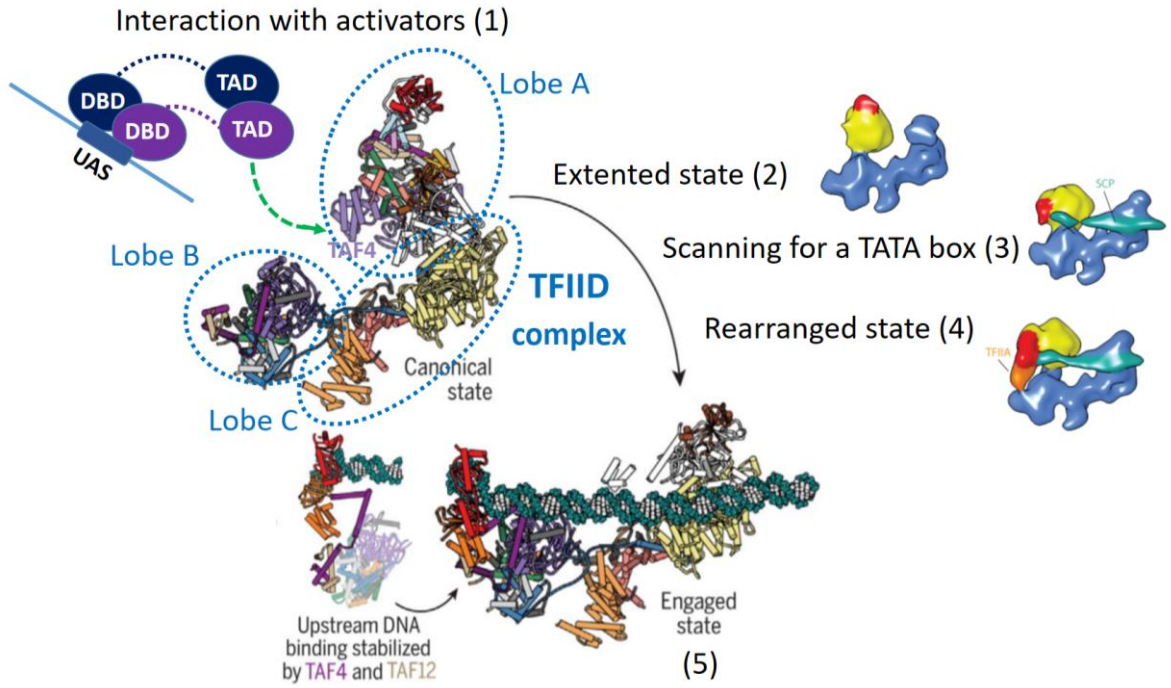


Figure 1.10 | TFIID recruitment by transcription activators and binding to the promoter DNA. The figure shows a schematic representation of the transition from the canonical state to the rearranged state. TFIID is recruited to the promoter DNA by the direct interaction with transcription activators. In metazoans, activators bind to enhancer elements instead of UAS sequences as occurs in yeast. Large conformational changes occur in TFIID before binding to the promoter. The interaction with TFIIA and the promoter DNA are necessary for the transition to occur. Although activators interact with several TAF proteins, the conserved glutamine rich and TAFH domains of TAF4 within its flexible N-terminal region have been shown to play a very important role in this interaction. For the electron density map, The "Lobe A" is shown in yellow, TBP in red, super core promoter (SCP) DNA in green and TFIIA in orange. (Adapted from Patel *et al.*, 2018).

The lobe A (Fig. 1.10) is able to translocate through a distance of 100 Å between these conformational states. The overall structure presents two main states termed canonical and rearranged. In the canonical state, the lobe A contacts lobe C whereas in the rearranged state the lobe A is positioned on the other side of the BC core, close to lobe B. There is a clear biological relevance for these different states given that TFIIA and promoter DNA promote the rearranged state, and TFIID is only able to bind to the promoter DNA in the rearranged state (Cianfrocco *et al.*, 2013). TFIID has a central role in the nucleation of the PIC upon core promoter DNA recognition, which is essential for transcription initiation (Louder *et al.*, 2016).

1.2.3.2 Mediator

Mediator is a multiprotein complex and transcription coactivator that binds to the transcription PIC and regulates RNA pol II. It is made of 25 subunits, 15 of them essential for function and defining the previously mentioned core Mediator complex (cMed) (Fig. 1.11). Mediator

architecture can be divided into four different modules: Head, Middle, Tail and Kinase. cMed is formed by the head and middle modules (Nozawa, Schneider and Cramer, 2017) (Fig. 1.8), whereas the tail and kinase domains have regulatory roles.

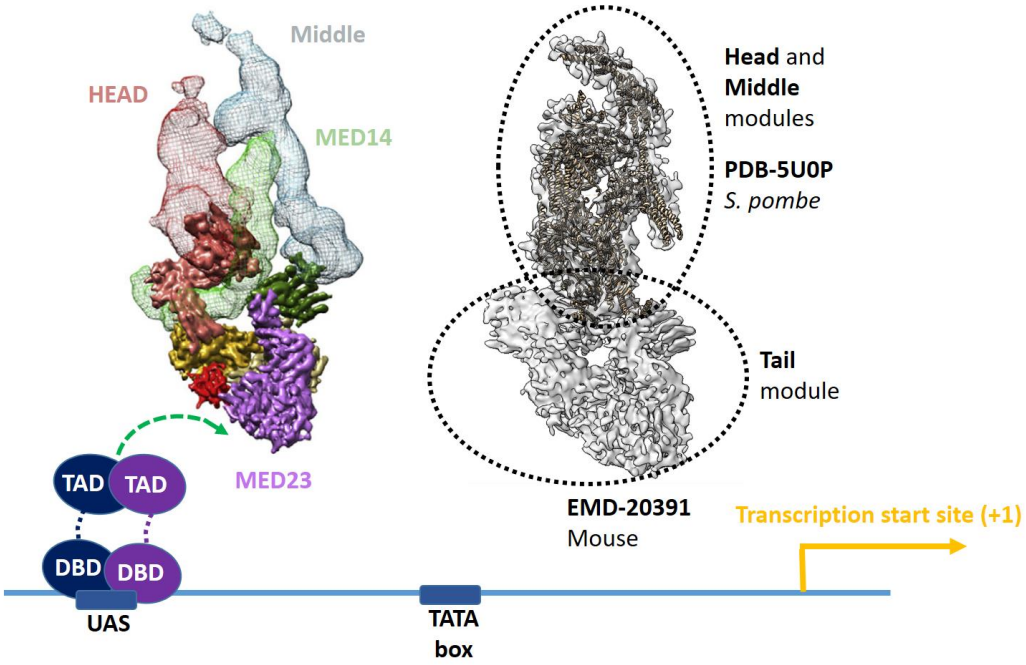


Figure 1.11 | A Mediator complex structure (Head, Middle and Tail modules) and its recruitment to promoter DNA by transcription activators. A mediator structure by cryo-EM is shown in the figure (EMD-20391). Upon UAS binding in yeast (or enhancer elements in metazoans), TAs directly interact with Mediator Med23 subunit, which is present in the Tail module. Interestingly, the Med23 is mostly composed of α-solenoid, similar to the SAGA subunit Tra1. The atomic model of *S. pombe* Mediator Head and Middle modules (PDB 5U0P) is fitted into a medium resolution cryo-EM map of mouse Mediator (**right**), showing the presence of the Tail module, and corresponding to the opaque coloured regions of the map (**left**). (Adapted from Tsai *et al.*, 2017; El Khattabi *et al.*, 2019).

The binding of cMed to the RNA pol II is made in proximity to Rpb4-Rpb7 subunits – defined as the RNA pol II stalk – close to the CTD of the polymerase. The Head module of Mediator interacts with the dock region of RNA Pol II and the TFIIB ribbon, stabilising the PIC (Fig. 1.8). Transcription activators positioned at the upstream DNA are predicted to directly interact with the Tail module of Mediator, and RNA Pol II phosphorylation of the CTD by TFIIH might be stimulated by the “Arm” and “Hook” domains of Mediator by positioning the two complexes together (Plaschka *et al.*, 2015; Schilbach *et al.*, 2017).

Med23, the largest subunit of Mediator Tail module, physically associates with TAs (Fig. 1.11). Interestingly, the structure is made of 25 HEAT repeats organised into five α-solenoid super-secondary structure elements (Monté *et al.*, 2018). Therefore, the direct interaction with TAs is carried out through a α-solenoid interface, as is the case for Tra1^{TRRAP} (1.2.4). In addition, Med15 subunit has been reported to participate in the direct interaction with TAs (Brzovic *et*

al., 2011). NMR analysis shows that Gcn4 activation domain (AD) interacts with Med15 activator binding domains (ABDs) contributing to activate transcription (Tuttle *et al.*, 2018).

1.2.3.3 The SAGA complex

The SAGA complex is a chromatin modifying, 1.8 MDa modular multiprotein complex consisting of 19 different subunits – plus the TBP protein – (Fig. 1.12).

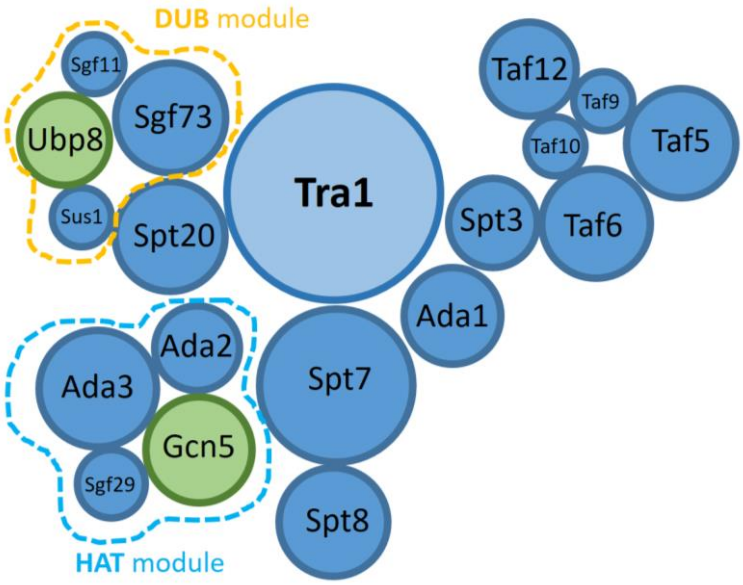


Figure 1.12 | Schematic representation of the SAGA complex showing the different modules. The size of each subunit is represented proportionally to its corresponding molecular size. The catalytic subunits are depicted in green. DUB: deubiquitinase module. HAT: histone acetyltransferase module.

SAGA – named STAGA in humans - contains five separate activities that combine to activate gene expression: activator binding, nucleosome binding, histone acetyltransferase (HAT), recruitment of the basal transcriptional machinery (Doyon *et al.*, 2004; Y. Han *et al.*, 2014) and histone deubiquitinase (DUB) activity. Its catalytic subunits are Gcn5 and Ubp8, responsible for the histone acetyltransferase and deubiquitinase enzymatic activities, respectively, where the HAT subunit belongs to the GNAT family of HATs (Fig. 1.13). The activator binding activity is mediated by the essential Transcription-associated factor protein 1 (Tra1 protein), a shared subunit with the histone acetyltransferase complex NuA4 (1.2.3.4), which is a member of the MYST family of HATs and the only essential HAT complex in *S. cerevisiae* (apart from the NuA4 subcomplex Piccolo). Tra1 is a direct activator target in both SAGA and NuA4 complexes.

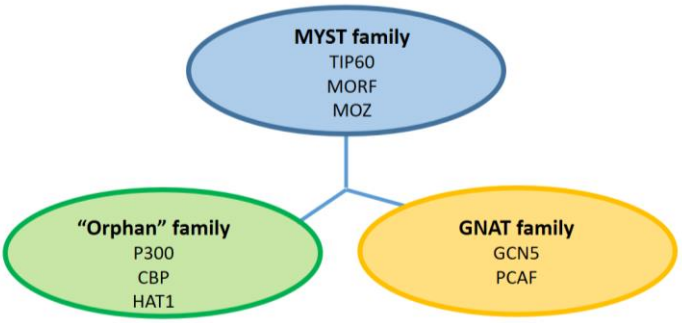
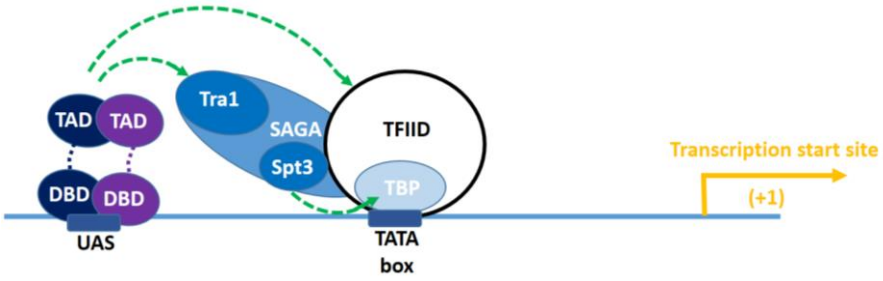


Figure 1.13 | Schematic representation of the HAT families in eukaryotes. The catalytic activity in NuA4 corresponds to the essential Esa1 subunit (Tip60 in humans), and to Gcn5 in the SAGA complex. Whereas SAGA preferably acetylates histone H3, NuA4 acetylates histone H4. As is shown in the figure, the HAT module of the SAGA complex belongs to the GNAT family, whereas the NuA4 HAT module belongs to the MYST family.

A recently published and previously mentioned work (Baptista *et al.*, 2017) (1.2.1) redefines the importance of the SAGA complex for transcriptional activation. Contrary to the belief that SAGA is responsible for the activation of only ~ 10% of *S. cerevisiae* genes, this work shows that SAGA maps to the UAS of most genes (Fig. 1.14) so that the transcription of most genes is affected when essential SAGA subunits are depleted.

A)



B)

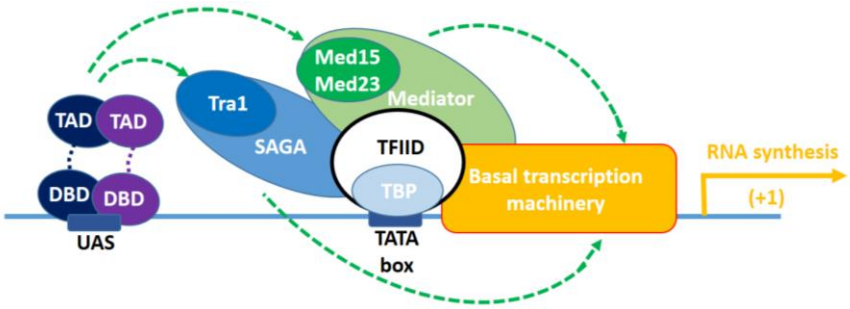
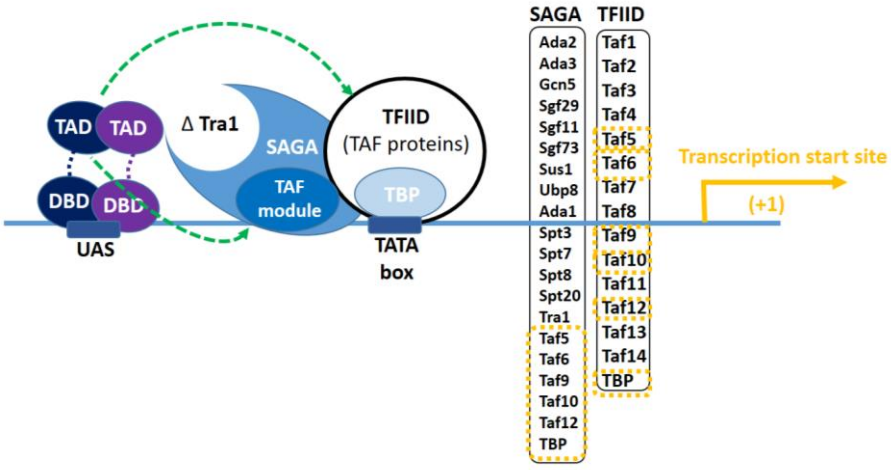


Figure 1.14 | The transcriptional co-activativors SAGA, TFIID and Mediator physically interact with transcription activators. A) Schematic representation of SAGA and TFIID coactivator complexes as general transcription factors. SAGA is recruited to the majority of *S. cerevisiae* genes and is necessary for TBP recruitment at promoter regions. **B)** After SAGA and TFIID binding, Mediator is also recruited to the promoter DNA for transcription activation. Like SAGA and TFIID, Mediator also interacts directly with TAs. (Inspired by the graphical abstract of Baptista *et al.* 2017).

Several TAF proteins are shared between the SAGA and TFIID complexes (Fig. 1.15). One of the remaining questions for the SAGA complex is the subunit stoichiometry of the TAFs, given that the TAFs in the TFIID complex are present in more than one copy (Bieniossek *et al.*, 2013; Patel *et al.*, 2018). Interestingly, in *S. pombe* we find two Tra1 paralogs: Tra1 and Tra2, which integrate differentially into the SAGA and NuA4 complexes. Tra1 is present in the SAGA complex and is a non-essential gene, whereas Tra2 is integrated into the NuA4 complex and is essential for the cell (Helmlinger *et al.*, 2011) (Fig. 1.15). The complete deletion of Tra1 in *S. pombe* is partially compensated by other subunits such as Ada1, Gcn5, Taf6, and Taf12 (Klein *et al.*, 2003; Fishburn, Mohibullah and Hahn, 2005; Reeves and Hahn, 2005; Zhang *et al.*, 2014), so that SAGA is still recruited to some promoter DNA regions.

Similar to Tra1, the α-solenoid Mediator subunit Med23 has been shown to directly interact with transcription activators (Fig. 1.15), in addition to the Med15 subunit.

A)



B)

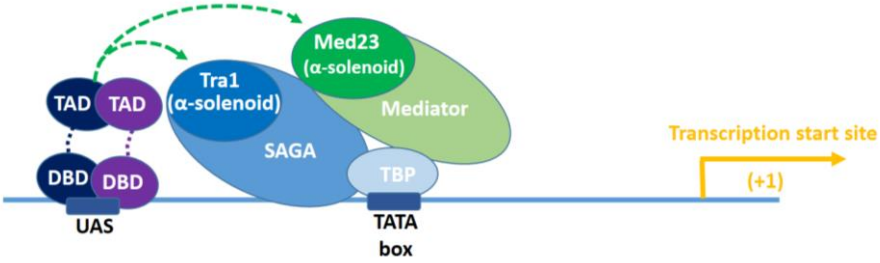


Figure 1.15 | Common mechanisms of interaction with transcription activators between SAGA, TFIID and Mediator coactivator complexes. A) In *S. pombe*, the complete deletion of Tra1 is partially compensated by the TAF module present in the SAGA complex, so that *S.p.* Δ Tra1 SAGA is still recruited to some promoter DNA regions through the direct interaction with TAs, a process that is mediated by the TAF module. The different SAGA subunits that are shared with the TFIID complex are also outlined (dotted lines in orange). **B)** Mediator α-solenoid Med23 subunit directly interacts with transcription activators. Tra1, like Med23, presents an extensive α-solenoid region that physically interacts with TAs.

Whether Tra1 has additional roles within the SAGA and NuA4 complexes in addition to the recruitment to UASs remains to be discovered yet, as new emerging roles are being

discovered in the case of other PIKK proteins. In this regard, evidence for the role of the SAGA and NuA4 complexes in DNA repair are increasing, as different research works describe their role in both DNA double-strand break (DSB) repair and nucleotide excision repair (NER) (1.3).

1.2.3.4 The NuA4 and Tip60 complexes

Although the SAGA complex has been a main research focus in comparison to other acetyltransferase complexes, the nucleosome acetyltransferase of H4 (NuA4) complex is the only essential acetyltransferase complex in yeast (Fig. 1.16) together with Piccolo, a NuA4 subcomplex as mentioned before. Its catalytic subunit is the essential Esa1 protein, homologous to the human Tip60 protein. Interestingly, in evolution, the NuA4 complex fused to the SWR1 chromatin remodeling complex to form the Tip60 complex found in humans (Fig. 1.16), combining therefore the acetyltransferase and remodeling activities together.

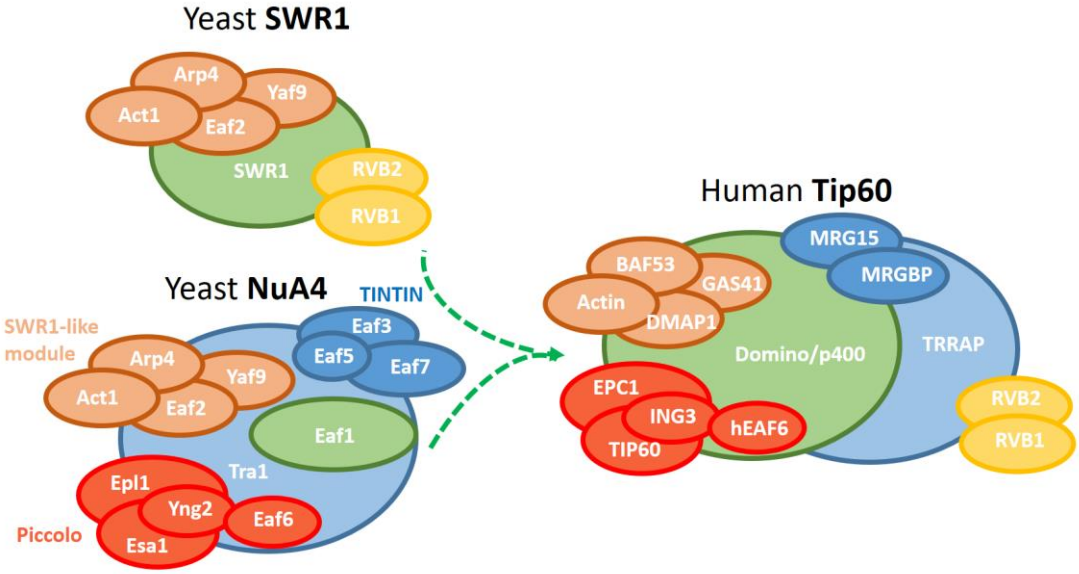


Figure 1.16 | NuA4, SWR1 and Tip60 chromatin remodeling and modifying complexes in *S. cerevisiae* and humans. In *S. cerevisiae*, we find the chromatin remodeling complexes SWR1 and INO80, as well as the chromatin modifier NuA4 complex. In metazoans, the SWR1 and NuA4 complexes fused together to form a super-complex called Tip60 in humans, represented in the figure by merging green arrows. The size of the essential Tra1^{TRRAP} protein is represented proportionally to the other complex subunits. NuA4 Piccolo and TINTIN subcomplexes, and the SWR1-like module are also illustrated. (Inspired by van Attikum and Gasser, 2005).

The NuA4 complex, containing the essential Tra1 subunit like the SAGA complex, has also been implicated in transcription activation (Allard, 1999), but its role is less studied in comparison to SAGA. In addition, the NuA4 complex also participates in the DNA damage response (DDR) (Doyon and Côté, 2004), as is the case for the SAGA complex (Ramachandran, Haddad, Li, Michael X. Le, *et al.*, 2016).

1.2.4 The Essential Transcription-Associated protein 1 (Tra1): a shared subunit of the SAGA and NuA4 complexes

Initially described as a component of the SAGA complex (Grant *et al.*, 1998; Saleh *et al.*, 1998), the yeast Tra1 is as a direct target of transcription activators and participates in yeast cells as a shared and essential subunit of both the NuA4 and SAGA complexes (Cheung and Díaz-Santín, 2019), as previously mentioned. Its human homolog TRRAP is present in the PCAF, TFTC, STAGA and Tip60 HAT complexes, and TRRAP deletion has been shown to be lethal in mouse embryos (Herceg *et al.*, 2001). Tra1 is a colossal ~ 433 kDa polypeptide made of 3,744 residues and the third biggest protein in *S. cerevisiae.* Therefore, the difficulty to purify and work with this very large polypeptide has extensively limited the knowledge we have so far in relation to this essential and conserved protein in evolution.

Tra1 belongs to the PIKK protein family (Fig. 1.1-1.4 and Fig. 1.17), presenting the same general architecture and domain organisation as found in other PIKKs; however, it is predicted to be the only pseudo-kinase within the family due to the lack of key catalytic residues. The N-terminal region of Tra1 comprises a large α-helical region made of 49 HEAT repeats - named as HEAT domain, residues ~ 1-2,600 - followed by a shorter 15 TPR repeat region - named as FAT domain, residues ~ 2,600-3,220. The TPR repeats surround and stabilise a very well conserved kinase domain located at the C-terminal region of the protein (residues 3,220-3,744) (Fig. 1.17).

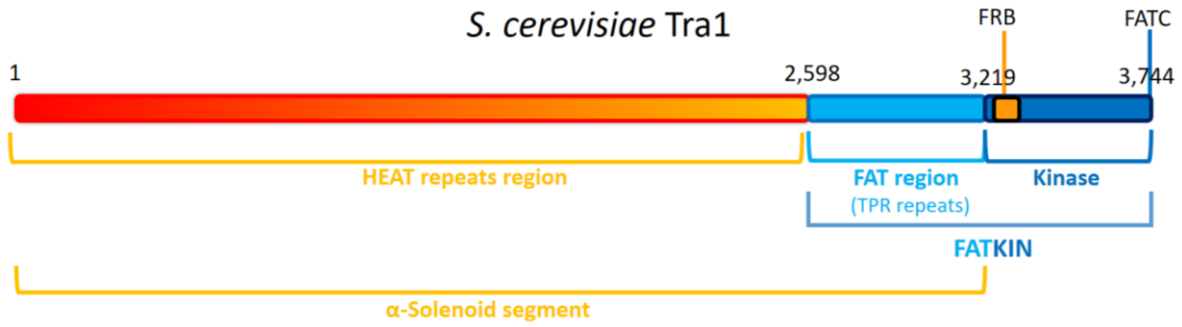


Figure 1.17 | Domain organisation of *S. cerevisiae* Tra1^{TRRAP} polypeptide. A large alpha-solenoid region composed of HEAT repeats is represented in red to orange gradient. A second solenoid segment composed of TPR repeats is shown in light blue. The conserved kinase domain at the C-terminal region is in dark blue, and together with the FAT region forms the globular FATKIN or HEAD domain.

A conserved FRB motif consisting of a bundle of four α-helices is found at the N-terminal region of the conserved Kinase domain, in addition to the also conserved FATC motif at the very C-terminal region, present in all PIKKs and composed of two α-helices.

Previously existing low resolution EM studies of the SAGA and NuA4 complexes that were available before the beginning of the present work are shown in Fig. 1.18 (Wu *et al.*, 2004; Chittuluru *et al.*, 2011; Durand *et al.*, 2014; Setiaputra *et al.*, 2015). In some of the mentioned

studies, the identification of Tra1 was uncertain, whereas in others, Tra1 was located with low confidence within a region of the low resolution EM map.

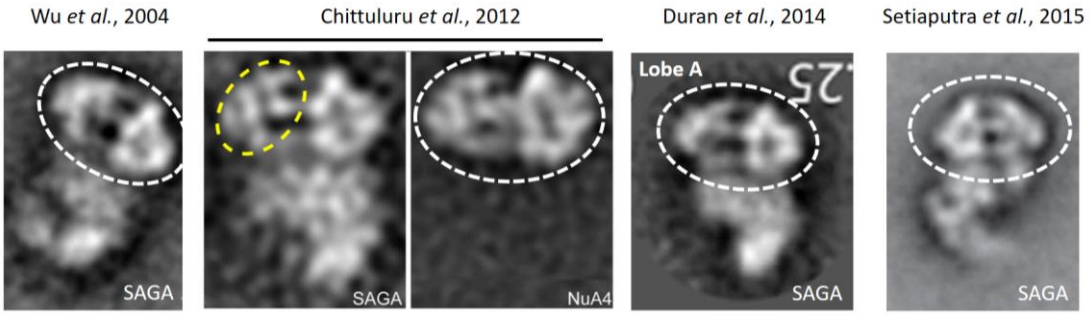


Figure 1.18 | Previously available electron-microscopy (EM) studies for SAGA and NuA4, and the location of Tra1 within the complex. A very well defined and constant density can be observed in previously available EM images. This density is outlined within a white dotted line drawn over the original 2D images to indicate the location of Tra1. The yellow coloured dotted line encloses a previously proposed location for Tra1 within the SAGA complex (Chittuluru e*t al* 2012). Note that the so-called "Lobe A" by Duran *et al.* (2014) corresponds to the density described as NuA4 by Chittuluru e*t al.* (2012).

In addition to the previously mentioned EM works (Fig. 1.18), data of cross-linking followed by mass spectrometry analysis is also available for the SAGA complex (Yan Han *et al.*, 2014), showing that Tra1 makes limited contacts with the rest of the SAGA subunits. *S.c.* Tra1 deletion studies (Knutson and Hahn, 2011) shows that removing between one to three HEAT repeats along the structure and the equivalent protein sequence length in the FATKIN domain leads in many cases to lethality, outlining the critical role of most of Tra1 regions to maintain structural integrity and function. In addition, secondary structure prediction analysis was performed in the mentioned work, showing that at the secondary structure level the identity of *S.c.* Tra1 to its human homolog TRRAP is 98%, demonstrating the high-degree of conservation in evolution from yeast to humans.

The role of Tra1 in the direct interaction with transcriptional activators has been previously studied by confocal microscopy combined with mutational analysis and also biochemical analysis. In this regard, pull-down experiments using yeast TAs show that the transcription activation domains of VP16, Gcn4, Gal4 and Hap4 fused to GST are able to specifically interact with Tra1 *in vitro* (Brown *et al.*, 2001). In addition, *in vivo* studies show that TAs fused to fluorophores directly interact with Tra1 (Bhaumik *et al.*, 2004), and that this interaction stimulates transcription. TAs are modular polypeptides, composed of a DBD connected to a TAD by a flexible linker (Fig. 1.5), and its activation *in vivo* requires previous binding to the DNA and dimerisation. In addition, TADs have been described as intrinsically disordered proteins, so that the interaction with a coactivator has been described as "fuzzy" (Tuttle *et al.*, 2019). An example of this "fuzzy" interaction or protein-protein fuzzy complex model is the described for the interaction of Gal4 acidic activation domain (AAD) with Mediator subunit

Med15, where different acidic binding domains (ABDs) have been defined within Med15 (Tuttle *et al.*, 2018). If functionally equivalent or defined ABDs also exists in Tra1, remains to be discovered yet.

Tra1 human homolog TRRAP has clinical relevance since it is a coactivator protein, target of oncoproteins (McMahon *et al.*, 1998; Murr *et al.*, 2007). The oncogenic protein c-Myc has been shown to interact with TRRAP, as well as the transcription factor E2F1, reported to present increased levels in many human cancers (Dang, Le and Gao, 2009).

Due to the essential role of different PIKKs in the cell and their influence in genome stability and cell cycle progression (Smits, Reaper and Jackson, 2006), we must consider some PIKK proteins as essential genome regulators in the eukaryotic domain of life. Therefore, understanding Tra1 structure and function will not only help to improve our present understanding of eukaryotic transcriptional regulation but also will increase our knowledge on this family of serine/threonine protein kinases, several of them essential for all cells with a nucleus.

1.3 The essential role of PIKKs in the DNA damage response (DDR)

The maintenance of genome integrity is a fundamental cell process in all domains of life. DNA damage can be caused by a variety of sources and different alterations such as DNA base mismatch during DNA replication, cross-linking of chemicals to the DNA, UV damage, single and double strand DNA breaks and ionising radiation (IR) (Blackford and Jackson, 2017). Therefore, DNA repair is an essential process in all living organisms to prevent genome instability. A large number of proteins participate in the DDR pathway, some of which are involved in DNA damage recognition, whereas others can directly alter the DNA such as DNA endonucleases. In eukaryotes, we find two main DNA repair pathways in response to DNA double-strand break (DSB) lesions: homologous recombination (HR) and non-homologous end joining (NHEJ) pathways (Fig. 1.19). Several factors determine the pathway choice for DNA repair, such as the DNA damage complexity, the chromatin state and the cell cycle (Vignard, Mirey and Salles, 2013). In the case of HR, a sister chromatid is needed, i.e. during S or G2 phase of the cell cycle. The DDR also includes proteins that are able to pause the cell cycle if necessary, for example during replication, so that the damage can be properly fixed allowing to generate a faithful DNA copy. In addition to cell pause, apoptosis in metazoans is a potential response to prevent mutations that might lead to cancer.

PIKK participation in the DDR is of enormous importance (Hiom, 2005), since both the HR and NHEJ pathways are largely driven by different PIKK proteins. In the case of the NHEJ pathway in humans, the KU proteins (KU70/80) detect and bind to dsDNA ends at breaks (Fig. 1.19A), and this interaction is recognised by DNA-PKcs, which is able to phosphorylate Artemis for

DNA resection. The early steps in the NHEJ pathway are represented in Fig. 1.19. Although DNA-PKcs is absent is *S. cerevisiae*, the NHEJ pathway still operates, and in addition to the yeast Ku proteins, the MRX complex also participates in this case. In the HR pathway, DSBs are recognised by the MRN complex (MRX in yeast), and this interaction will activate the PIKK protein Tel1^{ATM} (Fig. 1.19). Upon resection, RPA-coated ssDNA further recruits ATR^{Mec1} to regulate other DDR proteins such as RAD51.

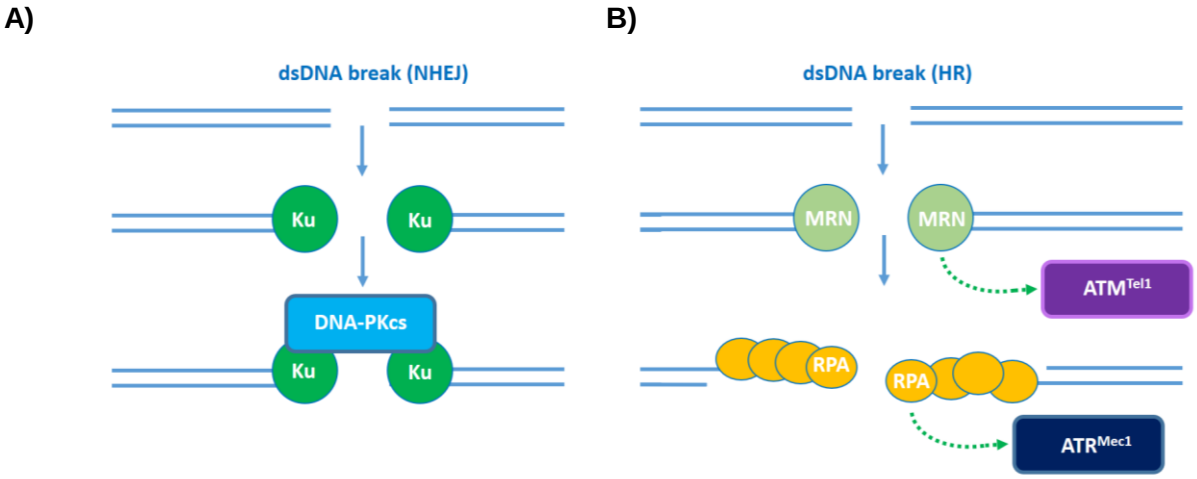


Figure 1.19 | The central role of PIKKs in the NHEJ and HR pathways. A) In the NHEJ pathway, a DSB is recognised by the KU proteins, binding to dsDNA ends. Then, DNA-PKcs interacts with the KU proteins to form a complex, and upon activation, DNA-PKcs can phosphorylate different substrates involved in the NHEJ pathway. **B)** The HR pathway requires the presence of a sister chromatid. The recognition of DSBs by the MRN complex activates ATM^{Tel1}. The exonuclease activity of the MRN complex generates ssDNA, which is coated by the trimeric RPA complex. ATR^{Mec1} is then recruited to RPA-coated ssDNA.

In addition to its histone acetylation activity, the human Tip60 complex has been shown to participate in the DDR by directly acetylating ATM (Sun, Jiang and Price, 2010). In yeast, the NuA4 complex has been shown to participate in DSB repair, recruiting the SWI/SNF chromatin remodeling complex to DNA lesion regions (Bennett and Peterson, 2015). Cells lacking Tip60 complex HAT activity present defects in DSB repair (Ikura et al., 2000). As previously mentioned (1.2.3.4; Fig. 1.16), in evolution the chromatin remodeler SWR1 fused to the NuA4 complex to form the Tip60 complex found in humans. Additionally, the SAGA complex has also been shown to play a role in DNA repair. In this regard, it has been reported that DSB repair is impaired in SAGA deubiquitinase deficiency situations (Ramachandran, Haddad, Li, Michael X Le, *et al.*, 2016). In this context, the Tra1 protein - TRRAP in humans -, could potentially be responsible for the recruitment of acetyltransferase complexes to DSBs. In the case of Tra1 as an essential NuA4 subunit, Tra1 participation is upstream of Mec1^{ATR}, at least as scaffold protein for the complex, given that NuA4 recruitment by the MRX complex to DSBs is previous to RPA binding to ssDNA and Mec1^{ATR} recruitment (Cheng *et al.*, 2018). However,

the participation of the putative pseudokinase Tra1^{TRRAP} in the DDR still needs to be fully dissected. Additional and essential biological processes are also regulated by PIKK proteins participating in the DDR, as is the case in V(D)J recombination, where NHEJ proteins such as DNA-PKcs participate to generate antibodies (Burma and Chen, 2004), and in meiotic recombination where ATR participates in the generation of gametocytes in sexual reproduction (Pacheco et al., 2018).

1.4 Multiple roles for the Essential Mec1-Ddc2 complex

1.4.1 The Mec1-Ddc2 complex in the DNA Damage Response (DDR) and the Replication Stress Response (RSR)

In vertebrates, at least three different PIKK proteins have been traditionally described as part of the DDR: ATM^{Tel1}, DNA-PKcs and the essential ATR^{Mec1} (Blackford and Jackson, 2017). Defects in ATR are known to be responsible for the Seckel syndrome 1 in humans (Zhou et al., 2006), and after more than 20 years since ataxia-telangiectasia mutated and Rad3-related (ATR) was cloned for the first time, still important questions related to the regulation of these protein kinases remain to be answered, due in a high percentage to the lack of structural information.

The Mitosis entry checkpoint (Mec1) protein, homolog to the human ATR, is an essential protein in all eukaryotic cells (Cimprich and Cortez, 2008) and it is composed of 2,368 residues in *S. cerevisia*e, associating in vivo with its 747 residue binding partner DNA damage checkpoint 2 (Ddc2), homolog to ATR-interacting protein (ATRIP) in humans. In yeast, Ddc2^{ATRIP} is also known as Lethal, checkpoint-defective, DNA damage-sensitive protein 1 (Lcd1). Mec1^{ATR} was originally identified in a genetic screen for checkpoint defects in S and G2 phases (Weinert, Kiser and Hartwell, 1994). Mec1^{ATR} mutants have hypersensitivity to UV, IR and hydroxyurea (HU), and this phenotype can be complemented by human ATR cDNA (Bentley et al., 1996; Cimprich, Shin and Keith, 1996). Contrary to ATM^{Tel1} and DNA-PKcs, ATR^{Mec1} is essential (1.1.2), and it is activated by a much wider variety of genotoxic agents. ATR^{Mec1} is recruited to ssDNA coated by the replication protein A (RPA) (Deshpande et al., 2017), which is a trimeric complex with affinity to ssDNA (Fig. 1.19). RPA-bound ssDNA is found at regions where DNA is damaged and it can also be found at stalled replication forks (Fig. 1.20).

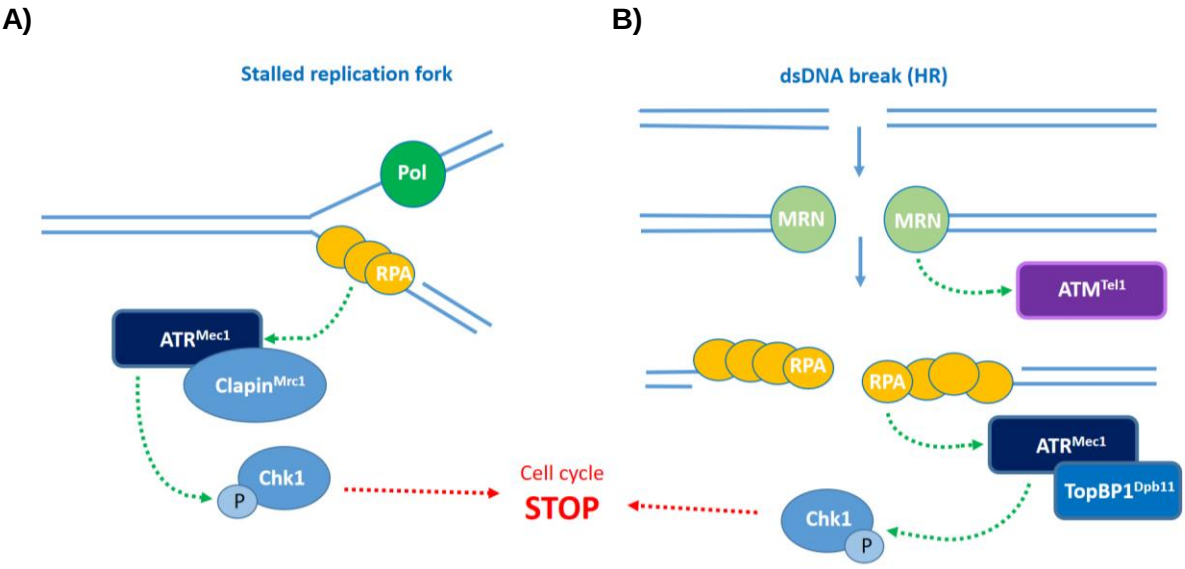


Figure 1.20 | ATR participates in the RSR* and the DDR. A) ATR^{Mec1} is recruited to stalled replication forks by the direct interaction with RPA-coated ssDNA. Claspin (Mrc1 in yeast) activates ATR, which in turn phosphorylates several substrates, including Chk1. **B)** In the HR pathway, ATR is activated by the adaptor protein TopBP1. The phosphorylation of Chk1 results in the detention of the cell cycle, so that the cell has time for either DNA repair or to solve the replication stress situation. *As previously mentioned, in the present text it has been considered the RSR as a separate process to the DDR for convenience purpose.

RPA is not sufficient for ATR activation and needs an adaptor protein to be fully active. The best-known ATR^{Mec1} adaptor protein in humans is TopBP1 (Mordes et al., 2008; Mordes, Nam and Cortez, 2008), known as Dpb11 in yeast (TopBP1^{Dpb11}) (Fig. 1.20). For this reason, TopBP1^{Dpb11} depletion has been shown to be lethal for the cell since it is also an essential protein, as ATR^{Mec1}. One of the main substrates of ATR^{Mec1} is Chk1, which is able to influence the cell-cycle (Rundle et al., 2017) (Fig. 1.20).

The phosphorylation of Chk1 would presumably provide time to the cell for DNA repair or during the RSR. In addition, the Mec1-Ddc2 complex has been also shown to stabilise the replication fork, so that DNA replication can efficiently restart after a replication fork blockage event. In extreme cases, genomic instability events such as the previously described can result in senescence or apoptosis.

In humans, ATR^{Mec1} also participates in DNA interstrand crosslink (ICL) repair. It regulates the Fanconi anemia (FA) pathway by phosphorylating directly or via its target effector Chk1 several components of the FA machinery (Shigechi et al., 2012). In addition, it is directly involved in ICL repair, participating in the HR pathway.

Before the present work was initiated, and previous to the cryo-EM structure of the Mec1-Ddc2 complex presented in this study (chapter 6), the only available structural information for the complex was a negative staining reconstruction (Sawicka et al., 2016). The structure shows

that the complex is assembled as a dimer, and resembles ATM architecture, as is its paralog PIKK protein. Distinct to ATM, which has been described sometimes in the literature to split into two monomers upon auto-phosphorylation, Mec1 has never been described before as a monomer, in concordance with our experimental data. The absence of structural information has strongly limited the knowledge we have so far on ATR^{Mec1}. Therefore, there is a strong need for structural analysis in order to obtain a deeper understanding of the biological processes in which the essential and conserved ATR^{Mec1} PIKK protein participates.

1.4.2 The role of the Mec1-Ddc2 complex in meiosis and telomere regulation

In addition to its participation in HR, it has been shown that ATR is required together with ATM for telomerase recruitment and assembly through the phosphorylation of TRF1 protein (Tong *et al.*, 2015). Furthermore, ATR also participates in meiosis (Pacheco et al., 2018; Widger et al., 2018), although these biological processes are less studied in comparison to its role in the DDR and the RSR.

1.5 Clinical application – Targeting PIKKs in cancer

Since many PIKK proteins have an essential role in genome regulation, targeting them has been shown to cause genome instability and cell lethality (Andrs et al., 2015). Given its size and their biological role, many of the PIKKs within the family can be defined as colossal guardians of the genome (Blackford and Jackson, 2017).

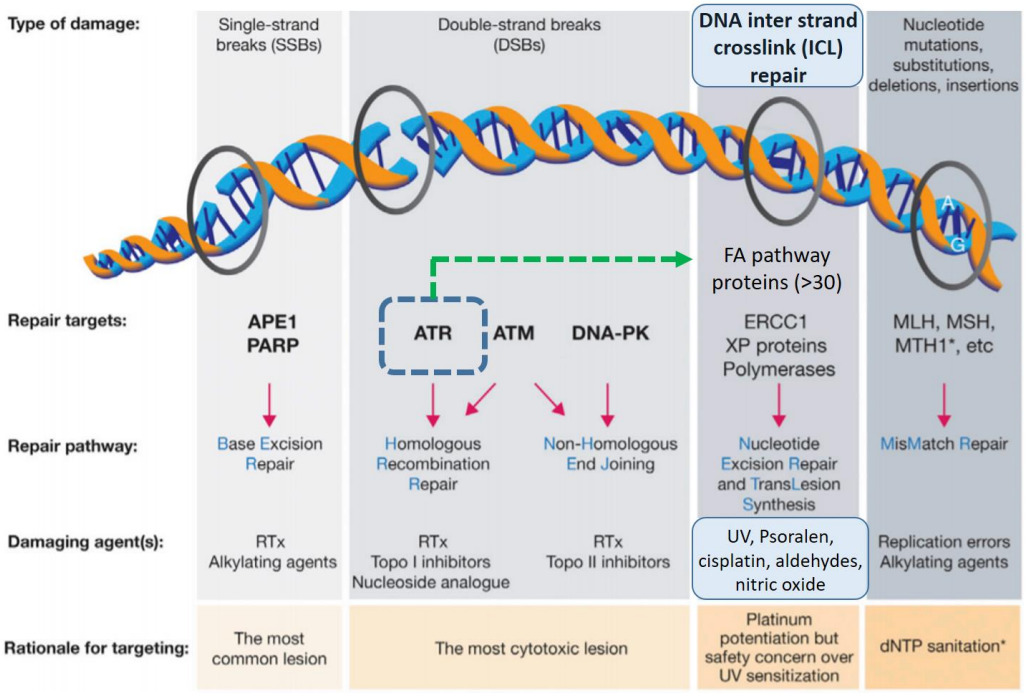


Figure 1.21 | Targeting PIKKs in cancer therapy. Different DNA lesions are shown together with its corresponding repair targets and repair pathways, in addition to the damaging agent and the rationale

for targeting. ATR also participates in the regulation of ICL repair (represented by a green arrow), which is activated upon the detection of DNA lesions caused by exogenous damaging agents (psoralen and cisplatin) and endogenous agents (aldehydes and nitric oxide). (Adapted from O'Connor, 2015**).**

In this regard, PIKKs participating in the DDR can be targeted for cancer treatment in dependency situations where one or more DNA repair pathways are not functional (Fig. 1.21), by targeting the remaining functional pathway(s) to compromise cell viability and induce apoptosis (O'Connor, 2015).

ATR^{Mec1} is the only essential and catalytically active PIKK protein directly participating in the DDR - in contraposition to the essential $TRRAP^{Tra1}$ which is predicted to be a pseudo-kinase - , participating in HR and also activating the FA pathway (1.4.1). In addition, ATR has a central role in the stabilisation of the replication fork upon replication stress situations, both directly by interacting with replication fork components such as Claspin (Mrc1 in yeast) (Fig. 1.20) and indirectly by regulating the FA pathway. Furthermore, its role in telomerase recruitment shows its fundamental contribution to the maintenance of genome stability (Tong *et al.*, 2015) (1.4.2). Therefore, ATR^{Mec1} is one of the most attractive targets for cancer therapy within the PIKK protein family (Fig. 1.21). Only ATP-analogues/ATP-competitive drugs are been tested at the moment in clinical trials as inhibitors of human ATR, which is the case for VX-970 and AZD6738 compounds (Lecona and Fernandez-Capetillo, 2018), in phase II of clinical trials for some cancer types. The difficulty to purify PIKK proteins has traditionally represented a handicap to develop new drugs and also to obtain structural information. New developments, including the present revolution we are living in cryo-EM, will potentially bring the opportunity to develop a new generation of allosteric PIKK inhibitors with more specificity and affinity to its target kinase. The research described in this work represents a proof of concept validating the drug development possibilities in the PIKK field, both in terms of biochemistry and structural biology.

In addition to the nuclear PIKKs, the cytoplasmic protein mTOR has been previously described as a potential target in cancer treatment given its role as essential metabolic and cell growth regulator (Sabatini, 2006). Furthermore, its paralog protein SMG1 has been shown to be a component of the non-sense mediated mRNA decay (NMD) pathway, which is involved in the removal of mRNA that bears premature stop codons. Targeting SMG1 enzymatic activity would potentially produce a cytotoxic effect in the cell by increasing the number of deleterious protein products, i.e. partially or incompletely expressed proteins lacking different regions might present an unfuctional or misfunctional behaviour and therefore interfere with multiple cell processes. In this regard, cells with the higher metabolic rates such as cancerous cells, with a high rate of protein synthesis, would be more significantly affected comparing to most

of non-cancerous cells. Since SMG1 is one of the less studied PIKKs, further research will also bring new therapeutic possibilities.

In addition, targeting essential acetyltransferase complexes such as the Tip60 complex (Zhang, Wu and Luan, 2017) - where TRRAP^{Tra1} is an essential and integral subunit - also has the potential to cause a very strong impact on genome instability and cell growth, converting these complexes in promising and very little exploited targets in cancer research. Hence, disrupting the interaction between TRRAP and oncogenic factors such as c-Myc has the potential to control oncogenic processes.

1.6 Aim of this work

In the PIKK protein family we find essential and highly conserved proteins from yeast to humans, and although progress has been made in the understanding of the biological processes where they are involved, no medium or high-resolution structural information was available for many of the PIKK members before the beginning of my PhD.

Regarding to the mechanisms of transcription initiation, the Tra1^{TRRAP} protein plays a fundamental role as direct target of transcription activators. Previous to the present work, no structural information was available for the single Tra1 subunit and only low resolution maps for the SAGA complex, where Tra1 is an essential subunit. Another central question to transcription initiation is the understanding of how transcription activators like the yeast Gal4 target their coactivator protein, and in this regard, Tra1 is a very well-known and general transcription coactivator described in the literature. For that reason, obtaining structural information of Tra1 was essential to improve our understanding of both the SAGA and NuA4 complexes where Tra1 is a shared subunit, and also to gain understanding of how activators directly interact with their targets to activate transcription. To accomplish that, I overexpressed and purified *S. cerevisiae* Tra1 polypeptide, and generated a high-resolution structure of the isolated polypeptide (chapter 4) by single-particle cryo-EM. In addition, I analysed the interaction of transcription activators with coactivators, incubating Gal4 TAD together with full-length Tra1 for biochemical and cryo-EM analysis (chapter 5).

In relation to the also essential Mec1^{ATR} polypeptide, which is recruited to RPA-coated ssDNA during the RSR and the DDR, crucial in the HR pathway, only low resolution structures were available before the present study begun. Important questions regarding to the complex architecture and assembly had to be addressed, and for that reason I recombinantly expressed and purified the protein complex from *S. cerevisiae*, and solved a high-resolution cryo-EM structure of the isolated full-length Mec1^{ATR} polypeptide in complex to its binding partner Ddc2^{ATRIP} (chapter 6).

In addition, there was need for structural knowledge on the PIKK protein family. The structures presented in this work greatly improve the knowledge we have so far in respect of the essential and conserved PIKK protein family, allowing for the first time to provide a first structural analysis of the PIKK protein family by the direct comparison to other recently released PIKK structures (Fig. 1.4). In addition, the presented atomic models allowed the analysis of key catalytic elements present in the Kinase domain, and to compare these structural elements to other PIKK members and ePKs (chapters 4 and 6), shedding light into the regulation and mechanism of action in the PIKK protein family.

2 Materials and methods

2.0 Materials

2.0.1 DNA cloning

Table 1 | PCR, Gibson Isothermal Assembly (GIA) and DNA gel electrophoresis

Name	Description or source
PCR	
Phusion High-Fidelity DNA polymerase	Thermo Fisher Scientific
5x Phusion HF Buffer	Thermo Fisher Scientific
dNTPs (100mM stock; 25 mM each dNTP)	Thermo Fisher Scientific
DMSO	Thermo Fisher Scientific
50 mM $MgCl_2$	Thermo Fisher Scientific
Q5 Hot-Start High-Fidelity DNA polymerase	New England BioLabs (NEB)
5x Q5 Reaction Buffer	New England BioLabs (NEB)
Taq DNA polymerase (recombinant)	In-house expression and purification
2.5x Taq DNA polymerase (recombinant) PCR mastermix	Phusion 2.5x HF buffer (NEB), 0.5 mM dNTPs (each), 2 µl DNA polymerase/50 µl of mastermix.
Gibson Isothermal Assembly (GIA)	
5x ISO Buffer (1 mL)	
PEG-8000	0.25 g (Sigma-Aldrich)
1 M Tris-HCl pH 7.5	0.5 mL (Sigma-Aldrich)
2 M $MgCl_2$	25 µl (Thermo Fisher Scientific)
100 mM dNTPs mix	40 µl (Thermo Fisher Scientific)
1 M DTT	50 µl (Sigma-Aldrich)
100 mM NAD	50 µl (Thermo Fisher Scientific)
2x Assembly Master Mix	
5x ISO Buffer (40 % v/v)	Handmade mix
10 U/ml T5 exonuclease	New England BioLabs (NEB)
2 U/mL Phusion polymerase	Thermo Fisher Scientific
40 U/mL Taq DNA ligase	New England BioLabs (NEB)

DNA gel electrophoresis	
10x TAE (1 L)	48.4 g of Tris base, 11.4 mL of glacial acetic acid (17.4 M), 3.7 g of EDTA, and deionized water.
Agarose (molecular biology grade)	Sigma-Aldrich
GeneRuler 1 kb DNA ladder	Thermo Fisher Scientific
GeneRuler 100 bp Plus DNA Ladder	Thermo Fisher Scientific
SYBR safe DNA gel stain, 10,000x	Invitrogen (Thermo Fisher Scientific)
6x DNA loading buffer	30% (v/v) glycerol, 0.25% (w/v), bromophenol blue, 0.25% (w/v) xylene cyanol FF

Table 2 | Bacterial plasmids

Name	Insert	Source
Plasmids used as template for cloning		
pET28a vector	Empty	Millipore
Plasmid no 17	3HA Tag- FL Tra1	A. Cheung
p3NH TrxA GFP		Waksman lab
Expression plasmids		
pET-28a(+)-Tra1-1	Nter 6xHis-3C-Tra1 F8 HEAT 1-7	This study
pET-28a(+)-Tra1-1-TrxA	Nter 6xHis-TrxA-3C-Tra1 F8 HEAT 1-7	This study
pET-28a(+)-Tra1-2	Nter 6xHis-3C-Tra1 F9 HEAT 8-9	This study
pET-28a(+)-Tra1-2-TrxA	Nter 6xHis-3C-TrxA-Tra1 F9 HEAT 8-9	This study
pET-28a(+)-Tra1-3	Nter 6xHis-3C-Tra1 F10 HEAT 8&9-DR1-HEAT 10&11	This study
pET-28a(+)-Tra1-3-TrxA	Nter 6xHis-TrxA-3C-Tra1 F10 HEAT 8&9-DR1-HEAT 10&11	This study
pET-28a(+)-Tra1-4	Nter 6xHis-3C-Tra1 F11 HEAT 10-16	This study
pET-28a(+)-Tra1-4-TrxA	Nter 6xHis-TrxA-3C-Tra1 F11 HEAT 10-16	This study
pET-28a(+)-Tra1-5	Nter 6xHis-3C-Tra1 F12 HEAT 17-22	This study
pET-28a(+)-Tra1-5-TrxA	Nter 6xHis-TrxA-3C-Tra1 F12 HEAT 17-22	This study
pET-28a(+)-Tra1-6	Nter 6xHis-3C-Tra1 F13 HEAT 23-28	This study
pET-28a(+)-Tra1-6-TrxA	Nter 6xHis-TrxA-3C-Tra1 F13 HEAT 23-28	This study

pET-28a(+)-Tra1-7	Nter 6xHis-3C-Tra1 F14 HEAT 28-37	This study
pET-28a(+)-Tra1-7-TrxA	Nter 6xHis-TrxA-3C-Tra1 F14 HEAT 28-37	This study
pET-28a(+)-Tra1-8	Nter 6xHis-3C-Tra1 F15 HEAT 38-41	This study
pET-28a(+)-Tra1-8-TrxA	Nter 6xHis-TrxA-3C-Tra1 F15 HEAT 38-41	This study
pET-28a(+)-Tra1-9	Nter 6xHis-3C-Tra1 F16 HEAT (38-40)-DR2-41&42	This study
pET-28a(+)-Tra1-9-TrxA	Nter 6xHis-TrxA-3C-Tra1 F16 HEAT (38-40)-DR2-41&42	This study
pET-28a(+)-Tra1-10	Nter 6xHis-3C-Tra1 F17 HEAT 43-47	This study
pET-28a(+)-Tra1-10-TrxA	Nter 6xHis-TrxA-3C-Tra1 F17 HEAT 43-47	This study
pET-28a(+)-Tra1-11	Nter 6xHis-3C-Tra1 F18 HEAT 48-53	This study
pET-28a(+)-Tra1-11-TrxA	Nter 6xHis-TrxA-3C-Tra1 F18 HEAT 48-53	This study
pET-28a(+)-Tra1-12	Nter 6xHis-3C-Tra1 F19 TRP1-14	This study
pET-28a(+)-Tra1-12-TrxA	Nter 6xHis-TrxA-3C-Tra1 F19 TRP1-14	This study
pET-28a(+)-Tra1-13	Nter 6xHis-3C-Tra1 F20 TRP1-8	This study
pET-28a(+)-Tra1-13-TrxA	Nter 6xHis-TrxA-3C-Tra1 F20 TRP1-8	This study
pET-28a(+)-Tra1-14	Nter 6xHis-3C-Tra1 F21 TRP9-14	This study
pET-28a(+)-Tra1-14-TrxA	Nter 6xHis-TrxA-3C-Tra1 F21 TRP9-14	This study
pET-28a(+)-Tra1-15	Nter 6xHis-3C-Tra1 F22 FRB-FATC	This study
pET-28a(+)-Tra1-15-TrxA	Nter 6xHis-TrxA-3C-Tra1 F22 FRB-FATC	This study
pET-28a(+)-Tra1-16	Nter 6xHis-3C-Tra1 F23 PI3KN-FATC	This study
pET-28a(+)-Tra1-16-TrxA	Nter 6xHis-TrxA-3C-Tra1 F23 PI3KN-FATC	This study
pET-28a(+)-Tra1-17	Nter 6xHis-3C-Tra1 F24 PI3KC-FATC	This study
pET-28a(+)-Tra1-17-TrxA	Nter 6xHis-TrxA-3C-Tra1 F24 PI3KC-FATC	This study
pET-28a(+)-Tra1-18	Nter 6xHis-3C-Tra1 F25 1905-2358	This study
pET-28a(+)-Tra1-19	Nter 6xHis-3C-Tra1 F26 2162-2588	This study
pET-28a(+)-Tra1-20	Nter 6xHis-3C-Tra1 F27 2478-2902	This study
pET-28a(+)-Tra1-21	Nter 6xHis-3C-Tra1 F28 2693-3140	This study
pET-28a(+)-Tra1-22	Nter 6xHis-3C-Tra1 F29 3092-3524	This study
pGEX-6p3	GST tag	Vaughan lab
pET-28a(+)-GST-VP16	GST-VP16 acidic activation domain (AAD), residues 413-490	This study

pET-28a(+)-GST-Gal4	GST-Gal4 acidic activation domain (AAD), residues 768-881	This study
pET-28a(+)-GST-Gcn4	GST-Gcn4 acidic activation domain (AAD), residues 9-172	This study
pET-28a(+)-GST-Hap4	GST-Hap4 acidic activation domain, residues 330-554 (AAD)	This study

Table 3 | Yeast plasmids

Name	Backbone	Insert	Source
Plasmids used as template for cloning			
pENT DQ			K. Nagai lab
pRS424			K. Nagai lab
pRS426			K. Nagai lab
Prs424 DQ	pRS424		This study
Prs426 DQ	pRS426		This study
pRS424 DQ GFP	pRS424	N-ter 3xFLAG-GFP	This study
pBS1479 TAP	pBS1479	TAP tag	A. Cheung
Expression plasmids			
pRS424-x3FLAG(N-terminal) Tra1	pRS424	x3FLAG-full-length Tra1	This study
pRS424-x3FLAG(internal 540) Tra1	pRS424	Full-length Tra1 (x3 FLAG insertion at 540)	This study
pRS424-x3FLAG(C-terminal) Tra1	pRS424	Full-length Tra1-3xFLAG	This study
pRS424-x3FLAG(N-terminal) Mec1	pRS424	x3FLAG-full-length Mec1	This study
pRS426-ProtA-TEV-(N-terminal) Ddc2	pRS426	ProtA-TEV-full-length Ddc2	This study

Table 4 | Cloning and sequencing primers

Name	Sequence	Source
Cloning primers		
AC132	cctggaggtgctgttccagggcccc	A. Cheung
AC133	ttcgggctttgttagcagccggatctca	A. Cheung

LMDS5	agcggtctggaagttctgtttcagggcccgatgtcactcactgagcag	This study
LMDS27	ttcctttcgggctttgttagcagccggatctcaTTGTAATTCAGATCTTATATTATGAA	This study
LMDS28	cctggaggtgctgttccagggccccGAATTACAACTAAGTGAAATTGA	This study
LMDS29	ttcgggctttgttagcagccggatctcaCTCACTCTCTTTGTCATTATCC	This study
LMDS30	ttcgggctttgttagcagccggatctcaGGAGGTTATTTCACTTGTTAAA	This study
LMDS31	cctggaggtgctgttccagggccccGACTTAAAAGTCTTTAACCCAC	This study
LMDS32	ttcgggctttgttagcagccggatctcaATTTCTACCTCCTAATTTGCC	This study
LMDS33	cctggaggtgctgttccagggccccAAATTAGGAGGTAGAAATCGTC	This study
LMDS34	ttcgggctttgttagcagccggatctcaATTTGCATTACTTAATTCGC	This study
LMDS35	cctggaggtgctgttccagggccccTTCGCGGATGTAAAAG	This study
LMDS36	ttcgggctttgttagcagccggatctcaAAATGGAGAATCCAGTTGA	This study
LMDS37	cctggaggtgctgttccagggccccTTATCTGACCACCAAAAGC	This study
LMDS38	ttcgggctttgttagcagccggatctcaAGGGAAATCAAACTTGGATAT	This study
LMDS39	cctggaggtgctgttccagggccccGATTTCCCTATTAAAGTTGT	This study
LMDS40	ttcgggctttgttagcagccggatctcaATTTACATTCGTAATTTCTAAG	This study
LMDS41	ttcgggctttgttagcagccggatctcaTGTCTTATTTTTGAAGAATACA	This study
LMDS42	cctggaggtgctgttccagggccccAATAAGACAAAAGAATGGAT	This study
LMDS43	ttcgggctttgttagcagccggatctcaAATTTCTGTGTTATTAAAATGT	This study
LMDS44	cctggaggtgctgttccagggccccAATAACACAGAAATTACCG	This study
LMDS45	ttcgggctttgttagcagccggatctcaTTCTTGTAGATTGACGTATAGT	This study
LMDS46	cctggaggtgctgttccagggccccAATACTAGCATCGACAATACA	This study
LMDS47	ttcgggctttgttagcagccggatctcaAGTAGTAGATTTGAAACGGTC	This study
LMDS48	ttcgggctttgttagcagccggatctcaTTCTATATTCGGTAGGGTG	This study
LMDS49	cctggaggtgctgttccagggccccACCCTACCGAATATAGAAATT	This study
LMDS50	cctggaggtgctgttccagggccccGGGGATAAACCAGATACAA	This study
LMDS51	ttcgggctttgttagcagccggatctcaGAACCATGGCATGAAG	This study
LMDS52	cctggaggtgctgttccagggccccAGAGCATCAAAAAAAGAAAA	This study
LMDS53	cctggaggtgctgttccagggccccAGTGTTTCTTTTACTACCCTTCA	This study
LMDS54	cctggaggtgctgttccagggccccGCGTACCTGGTTACATCA	This study
LMDS55	ttcgggctttgttagcagccggatctcaGCTTCTGGACATATTGACA	This study
LMDS56	cctggaggtgctgttccagggccccATGAATGCTTTGGATGT	This study

LMDS57	ttcgggctttgttagcagccggatctcaAGAAAGCAATGTAATGATAGA	This study
LMDS58	cctggaggtgctgttccagggccccGGTTCTTTTAATAGAGAGAAGGA	This study
LMDS59	ttcgggctttgttagcagccggatctcaTACTAGGTCATTCCACATATTG	This study
LMDS60	cctggaggtgctgttccagggccccTCATATGAACAAATTGGC	This study
LMDS61	ttcgggctttgttagcagccggatctcaTAGTTGAGGAATGAAAGTAATC	This study
LMDS62	cctggaggtgctgttccagggccccTATAAGAACTCGAAGATTAGGG	This study
LMDS63	ttcgggctttgttagcagccggatctcaTGATGGTACAAACATTGTTT	This study
LMDS64	AGTCACATCAAGATCGTTTATGG	This study
LMDS65	GCACGGAATATGGGACTACTTCG	This study
LMDS66	ACTCCACTTCAAGTAAGAGTTTG	This study
LMDS67	cagcagccatcatcatcatcactccgctagcccc	This study
LMDS68	GGGGCCCTGGAACAGCACCTCCAGcttagaaccactgccag	This study
AC267	GGTTTATTTATGTGTGTTTATTCG	
LMDS100	GACTACAAAGACCATGACGGTGATTATAAAGATCATGACA TCGACTACAAGGATGACGATGACAAGTAGGCTTAAGTCGA CTTTGTTCC	This study
LMDS101	TAGTTTCGAATAAACACACATAAATAAACCATGTCACTCAC TGAGCAGATC	This study
LMDS102	TTTATAATCACCGTCATGGTCTTTGTAGTCGAACCATGGCA TGAAGTT	This study
LMDS147	CATGACATCGACTACAAGGATGACGATGACAAGATGTCAC TCACTGAGCAGATC	This study
LMDS148	AAGTACAGTGGGAACAAAGTCGACTTAAGCCTAGAACCAT GGCATGAAG	This study
LMDS149	CTTGTCATCGTCATCCTTGTAGTCGATGTCATGATCTTTAT AATCACCGTCATGGTCTTTGTAGTCGATATCTTCTTTTTTA GGCTGTG	This study
LMDS150	GACATCGACTACAAGGATGACGATGACAAGAATGATAGTC CCGATGTTGA	This study
LMDS174	GGTTTATTTATGTGTGTTTATTCGAAACTAAGTTCT	This study
LMDS175	GCTTAAGTCGACTTTGTTCCCACTG	This study
LMDS176	TAAGCTTAAGTCGACTTTGTTCCCACT	This study
LMDS177	TAGTTTCGAATAAACACACATAAATAAACCATGTCCCCTAT ACTAGGTTATT	This study

LMDS178	AAGTACAGTGGGAACAAAGTCGACTTAAGCctacccaccgtactcg	This study
LMDS180	GGTTCACTGGAGGTGCTGTTCCAGGGCCCCTTTGCTTTAAATCCAATG	This study
LMDS181	TACAGTGGGAACAAAGTCGACTTAAGCTTATTTAGCATCTTCTAGAACAGG	This study
LMDS182	GGTTCACTGGAGGTGCTGTTCCAGGGCCCCAACGAAAATAATGATCTCTG	This study
LMDS183	AAGTACAGTGGGAACAAAGTCGACTTAAGCTCAAAATACTTGTACCTTTAAAA	This study
LMDS184	GGTTCACTGGAGGTGCTGTTCCAGGGCCCCGCCAATTTTAATCAAAGTG	This study
LMDS185	AAGTACAGTGGGAACAAAGTCGACTTAAGCTTACTCTTTTTTTGGGTTTG	This study
LMDS196	ggtatatctccttcttaaagttaaa	This study
LMDS197	gatccggctgctaac	This study
LMDS198	ttttgtttaactttaagaaggagatataccATGTCCCCTATACTAGGTTATT	This study
LMDS199	ttcctttcgggctttgttagcagccggatcctacccaccgtactcg	This study
LMDS200	ttcctttcgggctttgttagcagccggatcTTATTTAGCATCTTCTAGAACAG	This study
LMDS201	ttcctttcgggctttgttagcagccggatcTCAAAATACTTGTACCTTTAAAA	This study
LMDS202	ttcctttcgggctttgttagcagccggatcTTACTCTTTTTTTGGGTTT	This study
LMDS243	CATGGTTTATTTATGTGTGTTT	This study
LMDS244	GCTTAAGTCGACTTTGTTCC	This study
LMDS245	TTTCGAATAAACACACATAAATAAACCATGAAAACCGCGGCTCTT	This study
LMDS246	TGGAATATCATAATCACCCTGAAAATACAAATTCTCGAGCTCGGTTGACTTCCCCGC	This study
LMDS247	TTGTATTTTCAGGGTGATTATGATATTCCAACTACTGCTAGCATGAGACGAGAAACGGTG	This study
LMDS248	AAGTACAGTGGGAACAAAGTCGACTTAAGCTTACAGTCCCATTGAGATATATAA	This study
LMDS249	TTCGTCAAGATATTTGACGTGTGATTCCATCTTGTCATCGTCATCCTTG	This study
LMDS250	ATGTATATTGGTTGGCTTCCATTTTGGTAAGCTTAAGTCGACTTTGTTCC	This study

LMDS251	ATGGAATCACACGTCAAATAT	This study
LMDS252	TTACCAAAATGGAAGCCAAC	This study

General sequencing and colony PCR primers

T7-fwd	TAATACGACTCACTATAGGG	Eurofins A. Cheung
pET-RP*	CTAGTTATTGCTCAGCGG	Eurofins
T7-rev**	GCTAGTTATTGCTCAGCGG	A. Cheung
M13-rev	CAGGAAACAGCTATGACC	Eurofins A. Cheung

*pET-RP is a slightly shortened T7 terminator and it was used as the default sequencing primer to anneal over this region.

**T7-rev was used as the default primer for colony PCR analysis.

Tra1-specific sequencing primers

LMDS114	GATGATGTCAATATGTGGAA	This study
LMDS166	CCAAAGCATTCATGC	This study
LMDS167	GCCTCCTTATGTGATAGAG	This study
LMDS168	ACTTGTTGAATAATCTGTACAAT	This study
LMDS169	AGAGTTTTCAAACCTTGG	This study
AC15	ACCTTATAAACCAGACG	A. Cheung
AC16	CCCTGGCTTATAGTACTG	A. Cheung
AC17	TGATATATCAATGCCCAGC	A. Cheung
AC18	TTGACCGCCGAGTATTTTG	A. Cheung
AC19	GATTCTAGATGCTATTCCC	A. Cheung
AC20	CCACCAATATCCAAAAGAC	A. Cheung
AC21	GAGAAAGAACTCGACAAC	A. Cheung
AC22	CATAAAGCTCGAAGATACC	A. Cheung
AC23	AACTGGTTTACTTCGAG	A. Cheung
AC24	GTCAAAGCTATTCTACG	A. Cheung
AC25	CATCTTGGAATCCATTC	A. Cheung
AC26	GACCGTCTACCAAACAC	A. Cheung
AC27	ACCTCTCTATCACATAAGG	A. Cheung
AC28	TAGGCGTTTAATGATTCGG	A. Cheung

Mec1-specific sequencing primers		
LMDS253	GGTGGATTCAAAGGTG	This study
LMDS254	AATTTCCACAGCCCT	This study
LMDS255	TTTGTCATTTATGCGATATT	This study
LMDS256	GAGATATATCATCCCTTATGC	This study
LMDS257	ACTCTCTACTGTTATAGATAG	This study
LMDS258	GGCTAAGTTCAGAGTCA	This study
LMDS259	CCTCACGATAAGAACC	This study
LMDS260	GACTTCGTGCCAAAT	This study
LMDS261	TCCCACTTATATTTGGTAC	This study
LMDS262	TGGGATATTGGAAATGG	This study
Ddc2-specific sequencing primers		
LMDS263	GAAACGCCTTCATCC	This study
LMDS264	TTACAGAAACTGGAAGATG	This study
LMDS265	CAGATTTTCCAATATGAACTG	This study
LMDS266	ATGATAGAAAGGTACGTG	This study
LMDS267	AAGCATGCTTGCTTCACCT	This study
LMDS268	ATGAGACGAGAAACGGTGG	This study

2.0.2 Bacteria and yeast strains.

Table 5 | *E. coli* strains used for DNA cloning and protein expression

Name	Description	Source
XL1-Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB laclq ZΔM15 Tn10 (Tetr)].	A. Cheung lab
BL21 (DE3) pLysS	F - ompT hsdSB(rB - mB -) gal dcm l(DE3) tonA pLysS (CmR)	C. Vaughan lab
Rosetta2 (DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)	C. Vaughan lab
BLR (DE3)	F^- *ompT gal dcm hsdS*$_B$($r_K^-m_B^-$) Δ*(srl-recA)306*::Tn*10*(Tet^R) (DE3)	G. Waksman lab
C43 (DE3)	F^- ompT gal dcm hsdS$_B$(r_B^- m_B^-)(DE3)	G. Waksman lab
Lemo21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(CamR)	G. Waksman lab

λ DE3 = λ sBamHIo ΔEcoRI-B
int::(lacI::PlacUV5::T7 gene1) i21 Δnin5

pLemo = pACYC184-PrhaBAD-lysY

Table 6 | Yeast strains used for protein expression

Name	Description	Source
BC010	Mat**a**, pep4::HIS3/ prb1::LEU2, prc1::HISG, can1, ade2, trp1, ura3, his3, leu2-3,112	A. Cheung lab
BCY123	MATa pep4::HIS3 prb1::LEU2 bar1::HIS6 lys2::GAL1/10-GAL4 can1 ade2 trp1 ura3 his3 leu2-3,112	K. Nagai lab

2.0.3 Culture media

Table 7 | *E. coli* media

Name	Description
Luria-Bertani (LB) medium	1% (w/v) bactopeptone, 0.5% yeast extract (w/v), 0.5% NaCl (w/v)
ZYM5052 (auto-induction medium)	Yeast extract 0.5%, KH_2PO_4 50 mM, Na_2HPO_4 50 mM, $(NH_4)_2SO_4$ 25 mM, $MgSO_4$ 2 mM, Glucose 0.05 %, α-Lactose 0.2 %, Glycerol 0.5 %

Table 8 | *E.coli* medium additives

Name	Description
Ampicillin	100 mg/L (Sigma-Aldrich; ForMedium)
Kanamycin	50 mg/L (Sigma-Aldrich; ForMedium)
Chloramphenicol	25 mg/L (Sigma-Aldrich; ForMedium)
Tetracycline	10 mg/L (Sigma-Aldrich; ForMedium)
IPTG	0.1-1 mM (Sigma-Aldrich; ForMedium)
Rhamnose	100-2000 µM (Sigma-Aldrich; ForMedium)

Table 9 | *S. cerevisiae* media

Name	Description
YPD/YEPD	1% yeast extract, 2% bacto peptone, 2% gluclose
CSM-TRP (200 mL)	0.15 g CSM-TRP powder, 1.38 g Yeast Nitrogen Base without amino acids (YNB w/o aa), 180 H_2O, 20 mL 20% Glucose
CSM-TRP-URA (200 mL)	0.14 g CSM-TRP-URA powder, 1.38 g YNB w/o aa, 180 H_2O, 20 mL 20% Glucose
CSM-TRP agar plates (200 mL)	0.15 g CSM-TRP powder, 1.38 g YNB w/o aa, 4 g Agar, 180 H_2O, 20 mL 20% Glucose
CSM-TRP-URA agar plates (200 mL)	0.14 g CSM-TRP-URA powder, 1.38 g YNB w/o aa, 4 g Agar, 180 H_2O, 20 mL 20% Glucose
Casamino acids (-TRP selection) (1 L)	6.9 g YNB w/o aa, 11 g casamino acids, 55 mg adenine sulfate, 55 mg L-tyrosine, 20 mg uracil
Casamino acids (-TRP-URA selection) (1 L)	6.9 g YNB w/o aa, 11 g casamino acids, 55 mg adenine sulfate, 55 mg L-tyrosine

(Complete Supplement Mixture (CSM), YNB w/o aa and casamino acids from ForMedium).

Table 10 | *S. cerevisiae* medium additives

Name	Description
Ampicillin	100 mg/L (Sigma-Aldrich; ForMedium)
Adenine	50 mg/L (Sigma-Aldrich; ForMedium)
Raffinose (10x stock)	Filtered sterile 20% stock solution (ForMedium)
Galactose (10x stock)	Autoclaved 20% stock solution (ForMedium)

2.0.4 Transformation buffers

Table 11 | Buffers to prepare *E. coli* competent cells

Name	Description
TFB-I	30 mM K Acetate, 50 mM Mn_2Cl, 100 mM RbCl, 10 mM $CaCl_2$, 15 % (v/v) glycerol
TFB-II	10 mM MOPS, 75 mM $CaCl_2$,10 mM RbCl, 15% (v/v) glycerol

Table 12 | Buffers and components to prepare *S. cerevisiae* competent cells

Name	Description
Salmon sperm DNA	0.2% (w/v) solution (Sigma-Aldrich)
	Jansen group protocol
TELit pH 8.0	155 mM LiOAc ,10 mM Tris/HCl pH 8.0 and 1 mM EDTA pH 8.0
LitSorb solution	18.2 % (w/v) D-Sorbitol in TELit pH 8.0
LitPEG solution	40 % (w/v) PEG 3350 in TELit pH 8.0 filter sterilised
	Gietz lab protocol
PEG 3350 50% (w/v)	Sigma-Aldrich
Lithium Acetate 1M	Sigma-Aldrich
2x YPAD media (1 L)	20 g Yeast extract, 40g Peptone, 40g Glucose, 200 mg adenine hemisulphate,
	2000 mL distilled water
Transformation mixture (356 µl)	240 µl PEG 3350, 36 µl LiAc, 50 µl ssDNA, 30 µl water

2.0.5 Protein purification

Table 13 | General buffer components and reagents

Name	Description or source
Protease inhibitor (PI) homemade cocktail (PI[1])	1000x stocks: 210 mM AEBSF (water), 0.3 mM Aprotinin (water), 10.5 mM Leupeptin (water), 2.8 mM E-64 (water), 7.3 mM Pepstatin (ethanol). 100x stocks: 650 mM Benzamidine (water, fresh), 1150 mM PMSF (ethanol)
cOmplete tablets (PI[2])	Sigma-Aldrich
cOmplete ULTRA tablets (PI[3])	Sigma-Aldrich
Benzonase Nuclease	Sigma-Aldrich
TEV protease (turbo TEV)	MoBiTec Gmbh
3C protease	Recombinantly produced in-house

PBS (10x)	1.37 M NaCl, 27mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4
B-PER bacterial protein extraction reagent	Thermo Fisher Scientific
Y-PER yeast protein extraction reagent	Thermo Fisher Scientific
BS3 (bis(sulfosuccinimidyl)suberate)	Thermo Fisher Scientific

Table 14 | Buffers for the purification of Tra1 from *S. cerevisiae*

Name	Description
Lysis buffer [24 L]	125 mM HEPES 8.0, 250 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.1% NP40, 0.5 mM DTT, PI[1,2]
Lysis buffer [12 L]	50 mM Tris pH 7.5, 300 mM NaCl, 1.5mM $MgCl_2$, PI[1,2]
Washing buffer[24 L]	50 mM HEPES 8.0, 250 mM NaCl, 1.5mM $MgCl_2$, 10% glycerol, 0.3 mM DTT
Washing buffer[12 L]	50 mM Tris pH 7.5, 300 mM NaCl, 1.5mM $MgCl_2$, 0.1% NP40
Elution buffer[1] [(24 L)]	50 mM HEPES 8.0, 250mM NaCl, 1.5mM $MgCl_2$, 10% glycerol, 0.3 mM DTT, 500 µg/mL 3xFLAG peptide
Elution buffer[2] [(24 L)]	50 mM HEPES 8.0, 250mM NaCl, 1.5mM $MgCl_2$, 10% glycerol, 0.3 mM DTT, 200 µg/mL 3xFLAG peptide
Elution buffer [12 L]	50 mM Tris pH 7.5, 300 mM NaCl, 1.5mM $MgCl_2$, 200 µg/mL 3xFLAG peptide
Salt-free buffer to adjust NaCl to 100mM	50 mM HEPES 8.0, 1.5mM $MgCl_2$, 10% glycerol, 0.3 mM DTT
MonoS/Q equilibration buffer	50 mM HEPES 8.0, 100 mM NaCl, 1.5mM $MgCl_2$, 0.3 mM DTT
Superose 6 Increase buffer[24 L]	50 mM HEPES 8.0, 150 mM NaCl, 1.5mM $MgCl_2$, 0.3 mM DTT
Superose 6 Increase buffer [12 L]	50 mM Tris pH 7.5, 300 mM NaCl, 1.5mM $MgCl_2$

Table 15 | Buffers for the purification of acidic trans-activator domains from *E. coli* expression

Name	Description
Purification of Tra1 super-secondary structure fragments	
Lysis buffer	B-PER/Y-PER solution (3:1), 100 mM NaCl, PI[2], 0.25 mg/mL of lysozyme 0.25 mg/mL, 25U/mL of Benzonase, 2 mM $MgCl_2$, 10 mM imidazole
Buffer A	50 mM Tris pH 7.5, 100 mM NaCl, 0.01% Tween-20
Washing buffer	50 mM Tris pH 7.5, 100 mM NaCl, 0.01% Tween-20, 10 mM imidazole, PI[2]
Elution Buffer	50 mM Tris pH 7.5, 500 mM NaCl, 0.01% Tween-20, 10 mM imidazole, PI[2]
Purification of acidic trans-activating domains (AAD) fused to GST	
Lysis buffer	PBS, PI[1]
Wash buffer	PBS
Elution buffer	PBS, 30 mM GSH, glycerol 10 %
Superdex 200 buffer	PBS

Table 16 | Buffers for the purification of Mec1-Ddc2 from *S. cerevisiae*

Name	Description
Lysis buffer	125 mM HEPES 8.0, 250 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.1% NP40, 0.5 mM DTT, PI[1,3]
Washing buffer 1 (WB[1])	50 mM HEPES 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.3 mM DTT, PI[3]
Washing buffer 2 (WB[2])	50 mM HEPES 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.3 mM DTT
Elution buffer (EB[1])	50 mM HEPES 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.3 mM DTT, TEV protease (140 U/mL of EB), 500 µg/mL 3xFLAG peptide.
Elution buffer 2 (EB[2])	50 mM HEPES 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.3 mM DTT, 200 µg/mL x3 FLAG peptide.
Elution buffer 3 (EB[3])	50 mM HEPES 7.5, 600 mM NaCl, 1.5mM $MgCl_2$, 0.3 mM DTT
Gel filtration buffer	50 mM HEPES 7.5, 150 mM NaCl, 1.5mM $MgCl_2$, 0.3 mM DTT

2.0.6 Protein analysis

Table 17 | Acrylamide gel electrophoresis, protein staining and Western blot analysis buffers

Name	Description
Acrylamide gel electrophoresis, protein staining and others	
Stacking gel buffer (4%) (15 mL)	30% acrylamide/bis solution (1.98 mL) (Bio-Rad), 0.5 M Tris-HCl pH 6.8 (3.78 mL), 10% SDS (150 µl), dH_2O (9 mL), TEMED (15 µl), 10% APS (75 µl)
Separation/running gel buffer (12%) (15 mL)	30% acrylamide/bis solution (6 mL) (Bio-Rad), 1.5 M Tris-HCl pH 8.8 (3.75 mL), 10% SDS (150 µl), dH_2O (5.03 mL), TEMED (7.5 µl), 10% APS (75 µl)
MES-SDS running buffer (10x)	500 mM MES (2-[N-morpholino]ethanesulfonic acid), 500 mM Tris base, 10 mM EDTA, 1% (w/v) SDS
20x Bolt MES SDS running buffer	Invitrogen (Thermo Fisher Scientific)
Bolt 4-12% Bis-Tris Plus gels (10x and 15x wells)	Invitrogen (Thermo Fisher Scientific)
6x SDS loading buffer (Laemmli)	300mM Tris pH 6.8, 10.2 % SDS, 34.8% Glycerol, 600 mM DTT, bromophenol blue
PageRuler prestained protein ladder (10-170 kDa)	Thermo Fisher Scientific
BenchMark unstained protein ladder (10-220 kDa)	Novex (Life Technologies)
Hi-Mark prestained protein standard (31-460 kDa)	Thermo Fisher Scientific
Coomassie staining solution (1 L)	1 g of Coomassie Brilliant Blue (Bio-Rad), Methanol 50% (v/v), glacial acetic acid 10% (v/v), H_2O (10%)
Destaining solution	Methanol 50% (v/v), glacial acetic acid 10% (v/v), H_2O (10%)
InstantBlue Coomassie protein stain	Expedeon
Silver staining Kit	Invitrogen (Thermo Fisher Scientific)
Bradford reagent	Bio-Rad
Western blot analysis	
TCA precipitation buffer A (1 mL)	7.5% beta-ME (75 µl), 1.85 M NaOH (370 µl of 5 M stock) and H_2O to 1 mL

TCA precipitation buffer B (10 mL)	2.5 mL 20% SDS, 4 mL 0.5M tris pH 6.8, 20 µl 0.5M EDTA, 150 µl beta-ME, 4.8g Urea, trace amount of bromophenolblue and H_2O to 10 mL
20x NuPAGE buffer (125 mL)	Bicine 10.2g, Bis-Tris (free base) 13.1g, EDTA 0.75g and H2O to 125 mL
Transfer buffer	1x NuPAGE buffer 25 mL + 20% Methanol (100 mL) + H2O (375 mL)
Blocking solution	2% (w/v) skinny milk,1x PBS, Tween-20 0.05% (v/v)
Washing buffer	1x PBS, Tween-20 0.05% (v/v)
Antibody incubation solution	2% (w/v) skinny milk,1x PBS, Tween-20 0.05% (v/v)
Anti-rabbit HRP conjugate antibody	1:2000 (Thermo Fisher Scientific)
Anti-mouse HRP conjugate antibody	1:2000 (Thermo Fisher Scientific)
Anti-Flag M2 antibody (mouse monoclonal)	1:1000 (Sigma-Aldrich)
Anti-GAPDH antibody (mouse monoclonal)	1:1000 (Thermo Fisher Scientific)
Anti-GST HRP conjugate antibody	1:1000 (GE Healthcare)
Developing solution (2 mL)	1,880 µl of 1x PBS solution, 100 µl of 20 mM H_2O_2 and 20 µl of 10-Acetyl-3,7-dihydroxyphenoxazine (ADHP) solution (10 mM stock in DMSO) (Cambridge Bioscience)

2.0.7 Electron microscopy

Table 18 | Electron microscopy

Name	Description
Negative stain solution	2% (w/v) uranyl acetate dissolved in deionised H_2O

2.1 DNA cloning

DNA cloning was carried out by PCR amplification (2.1.1) of DNA fragments followed by the assembly of the amplified fragments in the Gibson Isothermal Assembly (GIA) reaction (2.1.4). The GIA reaction is a highly efficient DNA cloning method that allows the assembly of several DNA fragments simultaneously (Gibson, 2011), and in our case replaced the enzymatic digestion based system of DNA cloning in a similar way as the commercial in-Fusion (Takara) system.

DNA cloning, in addition to the PCR and GIA reaction, was assisted by other several techniques such as DNA agarose gel electrophoresis (2.1.2), DNA gel extraction (2.1.3) and purification (2.1.3), bacterial transformation (2.1.5 and 2.1.6), screening of transformants by colony PCR (2.1.7), DNA plasmid extraction (2.1.8) and finally DNA sequencing (2.1.9). The previously mentioned techniques will be sequentially described in the next sections.

2.1.1 PCR reaction

Polymerase Chain Reaction (PCR) primers were designed to be at least 15 bp long, to have a balanced content in C+G when possible, and anneal at temperatures from 50 to 70 °C (calculated using the online NEB T_m calculator service). Phusion High-Fidelity DNA polymerase (2.0.1, Table 1) was initially used for DNA cloning and subsequently it was substituted by Q5 Hot Start High-Fidelity DNA polymerase (2.0.1, Table 1) given its higher performance in terms of speed and mainly because of its ability to amplify challenging fragments. Note that optimal Q5 T_m temperatures tend to be higher comparing to other DNA polymerases. PCR reactions were performed on a Labtech G-Storm GS482 thermal cycler or an Agilent Technologies SureCycler 8800. Typically, 50 µl reactions were run on a thermocycler in the following conditions: 1x polymerase buffer, DNA template (200 ng in the case of genomic amplification and less than 1 ng for plasmid amplification), 0.5 µM of each primer, 200 µM dNTPs and 1 unit of polymerase (2.0.1, Table1). The PCR run with Q5 DNA polymerase consisted in an initial denaturation step at 98 °C for 30 secs, and then DNA melting at 98 °C for 10 secs, annealing of DNA primers to the DNA template at 50-72 °C and extension or DNA synthesis at 72 °C (20-30 secs/kb). A final extension step was carried out at 72 °C for 2 min, and then the reaction was hold at 4 °C. The PCR was typically set up for 30-35 cycles, following manufacturer recommendations.

2.1.2 DNA agarose gel electrophoresis

To visualise DNA by agarose gel electrophoresis, 1x **T**ris, **A**cetic acid, **E**DTA (TAE) buffer (2.0.1, Table 1) was used to resuspend agarose at 1% concentration (w/v) heating the solution

within a standard microwave and manually mixing the solution. Before pouring the agarose solution into the electrophoresis tank (Peqlab horizontal mini gel system), 1x SYBR Safe DNA gel stain (2.0.1, Table 1) was added to the mixture. 1x TAE buffer was used as running buffer and the DNA samples were loaded using 6x loading buffer (2.0.1, Table 1). The electrophoresis was typically run at 90 V in the small tank (50 mL Peqlab tank) and 100 V in the large tank (150 mL Pelab tank). GeneRuler 1kb DNA ladder (2.0.1, Table 1) was used as standard DNA marker. For small size DNA fragments, GeneRuler 100 bp Plus DNA Ladder was used instead. DNA was visualised on a UV trans-illuminator unit (UVP BioDoc-It Imaging Sytem). When needed, the DNA bands were extracted and purified using a gel extraction kit (2.1.3).

2.1.3 DNA purification of PCR products by gel extraction

The 50 µl PCR products were run in a 1x TAE 1% agarose gel (2.1.2) and analysed on the trans-illuminator unit. DNA bands were extracted using the QIAquick Gel Extraction Kit (QIAGEN). The final elution step from the micro-column was performed in 30 µl of milliQ H_2O and the DNA concentration (ng/µl) was measured analysing 2 µl of the elution on a Nanodrop Spectophotometer ND-1000 unit.

2.1.4 Gibson Isothermal Assembly reaction (GIA) for DNA cloning

2.1.4.1 Method description

This enzymatic method of DNA cloning consists in the amplification of DNA fragments that harbour an overlapping or complementary DNA region, so that 5' exonuclease digestion of the DNA ends generates ssDNA 3' overhangs. Because the reaction is performed at 50 ^{o}C, the exonuclease enzyme is active only for a very short period of time, so that the limited exonuclease activity generates short 3' overhangs to allow the annealing of complementary ssDNA regions and joining DNA fragments together (Fig. 2.1). DNA synthesis by a DNA polymerase generates a complete dsDNA molecule in the ssDNA regions after annealing. Finally, the ligase enzyme seals the remaining DNA nick between DNA chains and completes the assembly of the DNA fragments.

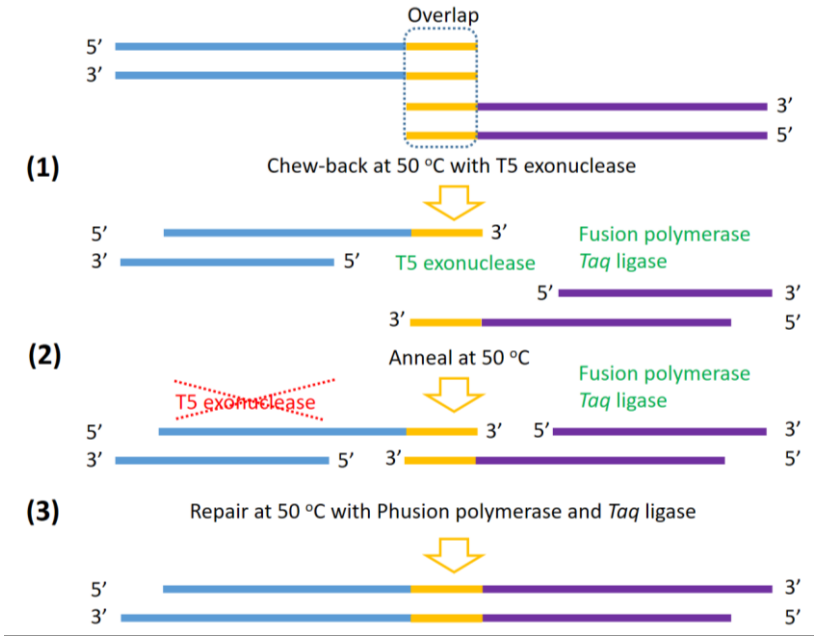


Figure 2.1 | The Gibson Isothermal Assembly reaction. 5' exonuclease activity (1) generates 3' overhangs. Complementary ssDNA regions anneal together (2). The DNA polymerase together with the DNA ligase repair the DNA to generate a dsDNA recombinant molecule (3). (Adapted from Gibson *et al.*, 2009).

2.1.4.2 Protocol

To perform the GIA reaction, PCR primers were designed with a 25-30 bp overlapping region for the assembly step in the reaction, in addition to the short annealing sequence (typically 15-20 bp) to amplify the desired DNA fragment by PCR (2.1.1). The amplified and purified DNA fragments were used for DNA cloning by GIA at the following insert/vector ratios: the insert concentration was 5x the vector concentration if the insert was less than 300 bp, 3x the vector concentration if the insert was 300-1000 bp, and finally, the insert concentration was the same as the vector concentration if the insert was bigger than 1000 bp. For a 5 µl reaction, 15.3 ng/kb of vector DNA were used, and the amount of insert DNA to be used was calculated according to its size as previously described.

To run the GIA reactions, the DNA fragments were mixed together - insert(s) and vector -, and H_2O was added to reach 2.5 µl final volume when needed. After the addition of 2.5 µl 2x GIA master mix (2.0.1 Table 1) to the mixture of DNA fragments, the reaction tube was then incubate for 60 min at 50 °C on the thermocycler unit (2.1.1). The assembled DNA fragments were transformed into *E. coli* XL1-Blue chemically competent cells (50 µl aliquots) (2.1.5) and plated according to the corresponding selection marker of the vector (2.1.6).

2.1.5 *E.coli* chemically competent cells protocol

To prepare *E. coli* chemically competent cells for DNA cloning, XL1-Blue strain was selected. For bacterial protein expression, competent cells of *E. coli* strains BL21, BLR, C43, Lemo21 and Rosetta2 were prepared (2.0.2, Table 5).

To prepare competent cells, 5 mL of LB (2.0.3, Table 7) were inoculated with the desired *E. coli* strain using an already existing aliquot of chemically competent cells, and grown at 37 °C and 200 rpm overnight (O/N). The 5 mL pre-culture was used the next day to inoculate 500 mL of LB medium with the corresponding antibiotic at 1:200 dilution (2.5 mL), which was grown at 37 °C and 180 rpm until a maximum OD_{600}=0.5 to ensure logarithmic growth. 200 mL of cell culture were chilled on ice for 20 min, and cells were spun down in 50 mL Falcon tubes at 4,000 rpm for 10 min at 4 ºC. The pellet was gently rinsed with prechilled TFB-I (2.0.4, Table 11) to remove residual LB medium, and 20 mL of TFB-I were utilised to resuspend the cell pellet. After 10 min on wet ice, the resuspension was spun down at 4 ºC and 4,000 rpm for 5 min, and 8 mL of prechilled TFB-II (2.0.4, Table 11) were used to resuspend the cell pellet corresponding to 200 mL of the LB culture. After incubation on ice for 30 min, cells were aliquoted in sterile prechilled Eppendorf tubes, 50 µl each. Approximately ~ 180 tubes were generated from 200 mL of LB culture at OD_{600} 0.5. Aliquots were stored at -80 ºC. To test the competent cells, 80-160 ng of plasmid DNA were transformed, generating a Petri dish with many colonies in close proximity to each other (typically more than 100).

2.1.6 Bacterial transformation in *E. coli* chemically competent cells

To amplify plasmid DNA – either from a pure plasmid stock or GIA cloning reaction – or to express protein in *E. coli* for protein purification, a 50 µl aliquot of chemically competent cells (2.1.5) from -80 ºC was thawed on ice for 10 min. Then, pure plasmid DNA (20-50 ng) or a GIA reaction product was gently mixed within the Eppendorf tube, and left on ice for 20 min more. After a heat-shock consisting in 30-60 secs at 42 ºC on a thermo-block unit, the sample was incubated on ice for 2 min. 200 µl of LB medium – antibiotic free- were added to the Eppendorf tube and incubated for 45-60 min at 37 ºC and 200 rpm. After centrifuging the culture for 2 min at 5,000 rcf in an Eppendorf 5415 D bench micro-centrifuge, the cell pellet was resuspended in 100 mL of H_2O. The cell resuspension was split into 20 µl and 80 µl, and plated separately – especially for DNA cloning – to prevent the presence of colonies in close proximity. Plating was performed on a 2% LB agar Petri dish containing the corresponding antibiotic.

2.1.7 *Screening of positive bacterial transformants by colony PCR*

After the transformation of XL1-Blue cells (2.1.6) with the product of GIA reaction (2.1.4), the resulting colonies were screened by colony PCR. Primers were designed to amplify the full-length of the insert DNA fragment, annealing to the vector at the flanking regions of the insert (typically T7 fwd/rev primers for bacterial expression plasmids, and T7 and M13RP for yeast expression plasmids, (2.0.1, Table 4)). For the amplification of DNA fragments, an in-house recombinantly expressed DNA polymerase was initially used (2.0.1, Table 1), but it was replaced by the Q5 Hot Start High Fidelity Polymerase (2.0.1, Table 1) given its very high-performance and low cost when 10 µl reactions were run for screening. The master mix for the in-house polymerase is in 2.0.1, Table 1. For colony PCR, typically two different isolated colonies were scratched with a sterile tip (0.2-2 µl tips) and resuspended into 20 µl of sterile H_2O. Q5 Hot Start High Fidelity Polymerase was used for 35 cycles in 10 µl reactions final volume, adding 0.4 µl of the resuspended colony solution to the PCR reaction tube. The initial PCR step consisted in 5 min at 98 ^oC for cell lysis. 6 µl of PCR product were loaded for analysis into a DNA agarose gel (2.1.2). Positive colonies were grown and selected for plasmid DNA extraction (2.1.8) and posterior DNA sequencing (2.1.9). Given the high rate of positive transformants obtained with the GIA reaction and the low cost of DNA sequencing, the colony PCR step was frequently skipped and replaced by a DNA plasmid miniprep of two different colonies followed by DNA sequencing, providing almost 100% of positive transformants. For difficult assemblies with multiple DNA fragments and/or very different lengths, colony PCR was still useful.

2.1.8 Extraction of plasmid DNA from *E. coli*

Plasmid DNA from *E. coli* was extracted using the QIAprep Spin Miniprep Kit (Qiagen). Typically, 5 mL of sterile LB were poured into 12 mL plastic tubes with the corresponding antibiotic. The medium was then inoculated with the selected colony scratching the agar plate with a sterile plastic pipette tip, a sterile plastic loop or taking 5-10 µl of the colony resuspension previously used for colony PCR (2.1.7). After overnight growth at 37 ^oC and 180-200 rpm, the culture was centrifuged at maximum speed in an Eppendorf 5804 R bench centrifuge, and the cell pellet was resuspended following the kit protocol. DNA elution was usually carried out in 50 µl of dH_2O, and the resulting DNA concentration extensively varied depending on the vector architecture. After measuring the DNA concentration on the nanodrop machine (2.1.3), the extracted plasmid DNA was stored at -20 ^oC for future use either as a template for PCR (2.1.1) or to transform either an *E. coli* (2.1.6) or a *S. cerevisiae* (2.3.1) strain, typically for protein expression.

2.1.9 Sanger DNA sequencing

Sanger DNA sequencing was carried out by the “GATC” company (later Eurofins) using either custom primers (2.0.1, Table 4) or commonly used primers available at the company. The sequencing was designed so that each sequencing reaction (estimated to be approximately 750 bp long on average) would overlap to the adjacent DNA region to be subsequently sequenced. Flanking primers at the insertion points in the corresponding vector confirmed the directionality and the right insertion of the inserted sequence(s). “ApE” software (Davis, M.W., 2012. ApE: a plasmid editor) was used for DNA sequence analysis purpose of the “.fas” files - containing the sequence - and “.ab” files - containing the chromatogram - generated from the sequencing reaction. Both, the generated sequence after DNA sequencing and the quality of the chromatogram were carefully analysed for each reaction. The analysis consisted in the sequence alignment in ApE software of the generated sequence(s) in the sequencing reaction(s) to the designed construct sequence and also the peak intensities and level of overlap of the adjacent peaks in the chromatogram were analysed for full coverage of the sequencing at high confidence level.

2.2 Expression and purification of recombinant protein fragments in *E. coli*

2.2.0 The *E. coli* expression system

The bacterial expression system used to express and purify proteins in *E. coli* was based on the pET28a vectors, modified to express either a 6xHis or GST N-terminal tag. The expression is driven by a T7 promoter in DE3 lysogenic strains, expressing T7 RNA polymerase from a bacteriophage polymerase gene integrated into the bacterial genome. It is characterised for presenting high transcription levels, and expression can be tuned modifying the IPTG concentration in the culture medium (Fig. 2.2).

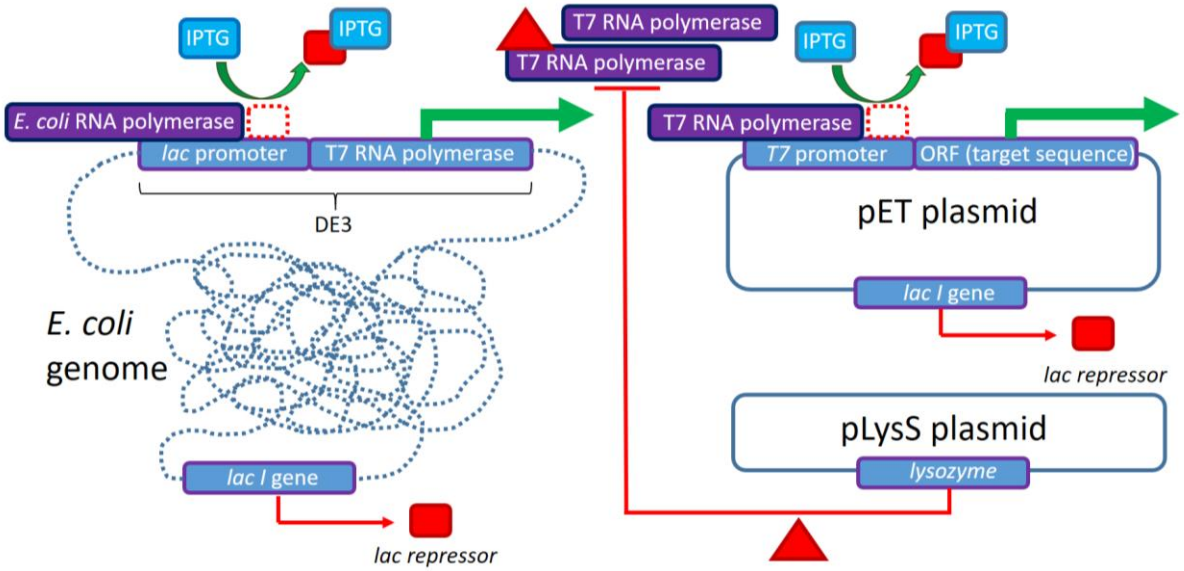


Figure 2.2 | Control of protein expression in *E. coli* by the T7 RNA polymerase/T7 lysozyme system. The sequence coding for the T7 RNA polymerase is integrated within the bacterial genome (DE3 lysogen), and the T7 lysozyme is coded within the transformed vector in the case of the pLysS plasmid, together with the open reading frame (ORF) that we want to express in a separate plasmid, pET28A in this case. Upon IPTG induction, the overexpression of T7 RNA polymerase cannot be neutralised by the lysozyme, so that expression of the ORF of interest occurs from the T7 promoter in the plasmid. Please note that the expression levels are dependent on the IPTG concentration.

Expression strains containing the pLysS plasmid express the T7 lysozyme (2.0.2, Table 5), a natural inhibitor of the T7 RNA polymerase (Fig. 2.2). The constitutive expression of T7 lysozyme removes basal polymerase activity in the absence of protein induction signal (IPTG). The addition of IPTG increases T7 RNA polymerase levels so that there is not enough T7 lysozyme to maintain repression. This system allows for a tighter transcriptional control and it is recommended for the expression of toxic proteins that at low basal expression levels might affect *E. coli* growth before protein induction.

For a more efficient expression of eukaryotic proteins in *E. coli*, the pRARE2 plasmid was sometimes used (2.2.1 and 2.2.2), which contains seven rare tRNAs (AGA, AGG, AUA, CUA, GGA, CCC, and CGG), in addition to code for a T7 lysozyme like the pLysS plasmid. The pLysS and pRARE2 plasmids, when transformed into BL21 DE3 cells, help to optimise protein expression and create *E. coli* expression strains BL21 (DE3) pLysS and Rosetta2 (DE3), respectively (2.0.2, Table 5).

2.2.1 Expression and purification of Tra1 super-secondary structure elements

An expression library consisting of 22 Tra1 fragments was created, amplifying the desired Tra1 DNA sequence from plasmid DNA containing full-length Tra1 sequence (2.0.1, Table 2, plasmid no17) and cloning the amplified fragments into the *E. coli* pET28a expression vector

(2.0.1, Table 2). The 6xHis-3C-Tra1 fragments (2.0.1, Table 2 and 4.1, Fig. 4.1) were initially expressed transforming the desired chemically competent *E. coli* cell strain (2.1.5 and 2.1.6). To express the different Tra1 fragments, although the Rosetta2 (DE3) strain was initially used for testing, and C43 strain was aslo tested, the *E. coli* BLR strain generated the best expression results (2.0.2, Table 5).

For test expression, 5 mL of LB medium (2.0.3, Table 7) supplemented with kanamycin (2.0.3, Table 8) were inoculated from the transformation plate, and grown overnight at 37 ºC and 200 rpm. The next day, 5 mL of fresh LB Kan medium were inoculated from the preculture at 1:1000 dilution. Protein induction was performed with 0.4 mM of IPTG (2.0.3, Table 8) at OD_{600} ~ 0.5, and the culture was grown at 18 ºC and 180 rpm overnight (O/N). Cells were centrifuged for 10 min at 4,000 rfc in an Eppendorf 5804 R bench centrifuge. The cell pellet was resuspended in 500 µl of Lysis buffer (2.0.5, Table 15), and incubated at room temperature (RT) for 20 min. Then, the sample was centrifuged at 13,000 rpm for 5 min in an Eppendorf 5415 D bench microcentrifuge and the 10 µl of supernatant were kept for SDS-PAGE analysis (2.4.2). 500 µl of "Buffer A" (2.0.5, Table 15) were added to the pellet, which was dissolved incubating the resuspended mixture at 70 ºC and maximum rpm for 10 min on an Eppendorf TermoMixer F1.5/F unit; 12 µl were used for SDS-PAGE analysis. The supernatant was incubated with 30 µl of pre-equilibrated nickel magnetic beads (MagneHis, Promega) washing them with "Buffer A" (2.0.5, Table 15). The mixuture of beads and supernatant was inverted 10 times and incubated for 2 min at RT. Then, the flowthrough was kept for analysis and the beads were washed three times with 500 µl of "Washing buffer" (2.0.5, Table 15) each time. For elution, 100 µl of "Elution buffer" (2.0.5, Table 15) were added to the beads and the mixture was inverted 10 times, and then incubated for 3 min at RT. 20 µl of elution were loaded into a gel for SDS-PAGE analysis. To analyse the magnetic beads, 20 µl of "Buffer A" were added in addition to 6x loading buffer (2.0.6, Table 17), and the sample was boiled for 5 min at 95 ºC. As negative control, in parallel a 5 mL LB culture was also grown but without IPTG induction. Pellet, supernatant, flow-through, magnetic beads and elution fractions of cultures with and without IPTG induction were analysed simultaneously by SDS-PAGE electrophoresis followed by Coomassie staining (2.4.3).

Test expressions with BLR *E. coli* cells provided high expression levels, but the protein was always found to be insoluble. To increase the solubility of the fragments, different NaCl concentrations were tested in the purification buffers (50, 500 and 1000 mM NaCl). In addition, a solubility test consisting in the presence of 500 mM D-sorbitol in the LB culture medium was performed. However, all tested protein fragments were still found in the pellet fraction.

To increase the probability of obtaining soluble protein, a more controlled system was tested expressing the Tra1 fragments in Lemo21 (DE3) *E. coli* cells. The expression conditions consisted in 6 hours and overnight induction at 30 ^{o}C with 0.4 mM of IPTG at different rhamnose concentrations (100, 250, 500, 750, 1,000 and 2,000 µM). Although protein expression was detected, the protein was also found in the pellet fraction.

Finally, a thioredoxin (TrxA) solubility tag was added to the already existing constructs to create a new library of 22 TrxA-fused Tra1 fragments (2.0.1, Table 2). Elution from nickel magnetic beads was performed this time by 3C protease digestion (10 U/reaction) for 30 min and 1 hr. The 3C protease utilised for the digestion was recombinantly expressed and purified in-house (2.0.5, Table 13)). Although protein expression was detected, Tra1 fragments were found to be insoluble in all cases but one (TrxA-3C-H8-H9 Tra1 fragment; 4.1.1, Fig. 4.2), and after 3C cleavage, the fragment was found to precipitate.

2.2.2 Expression and purification of transcription activation domains (TADs) in *E. coli*

2.2.2.1 Protein expression

Rosetta2 cells (for autoinduction media expressions) or BL21 pLysS cells (for LB media expressions) (2.0.2, Table 5) were transformed with the corresponding pET28a construct carrying a GST tag fused to a transcriptional activation domain (TAD) (2.0.1, Table 2), i.e. Gal4 768-881, Gcn4 9-172, Hap4 330-554 or VP16 413-490. After transformation (2.1.6), one single colony was selected to inoculate a 5 mL preculture of LB containing kanamycin and chloramphenicol (2.0.3, Table 8).

For test expressions in auto-induction medium, 1 L of fresh medium (2.0.3, Table 7) was inoculated from an O/N pre-culture at 1:500 or 1:1000 dilution and grown at 37 ^{o}C and 200 rpm until OD_{600} ~ 1. Then, it was moved to 25 ^{o}C and 180 rpm for protein expression over 25 hrs before cell harvesting.

For protein expression in LB, 1 L of fresh medium for test expressions (grown in a glass flask) were inoculated from an O/N pre-culture at 1:500 or 1:1000 dilution and grown at 37 ^{o}C and 180 rpm until OD_{600} 0.6-0.9. Induction with 1 mM of IPTG (2.0.3, Table 8) was carried out at 37 ^{o}C for 3 hrs. For the pull-down assays and single-particle cryo-EM analysis, 2 L of LB medium were used in the case of GST-Gal4 TAD purification and 12 L for GST-VP16 TAD, always using glass flasks with 1 L of culture capacity.

Cell harvesting was performed in a Beckman Coulter J6-MI centrifuge (JS-4.2 rotor) at 4,000 rpm for 20 min. The supernatant was discarded and the cell pellet was stored at -80 ^{o}C for further processing.

2.2.2.2 Protein purification

To lyse bacterial cells, the pellet was thawed and resuspended in 8 mL of Lysis Buffer (2.0.5, Table 15) for test purifications, 40 mL for Gal4-TAD 2 L expression and 2x 50 ml for VP16-TAD 12 L expression, and supplemented with 0.5 µl of benzonase per 10 mL of lysis buffer (2.0.5, Table 13). A sonication step was performed in a Vibra-Cell sonicator unit (Sonics, In.) consisting in four cycles at 65% amplitude, running each cycle for two seconds “on” and two seconds “off”. The material was clarified in a Thermo Scientific Sorvall RC 6 plus centrifuge (F21-8x50y rotor) at 17,000 rpm for 45 min. The supernatant was incubated for 2 hrs at 4 ^{o}C with 0.5 mL of bed Glutathione Sepharose 4B affinity resin (GE Healthcare) for the test purifications, 2 mL in the case of GST-Gal4 TAD 2 L expression and 2x 2 ml for GST-VP16 TAD 12 L expression. The material was poured into a 25 mL Econo-Pac chromatography column (Bio-rad) and washed with 15 CVs of “Washing buffer” supplemented with PI^{2} (2.0.5, Table 15) followed by 10 CVs of PI free “Washing buffer”. The elution was performed in batches of one CV each with “Elution buffer” containing 30 mM of GSH (2.0.5, Table 15). In the case of the pull-down expressions, the purified material was applied to a MonoQ 5/50 GL column (GE Healthcare), previously equilibrated with PBS buffer (2.0.5, Table 13), and the material was eluted in a linear gradient of 137mM-1000 mM NaCl. The cleanest fractions corresponding to GST-TAD were collected (5.2, Fig. 5.2), concentrated using an Amicon Ultra-15 mL 10 kDa MWCO (Merk Millipore) spin concentrator at 3,500 rpm in an Eppendorf 5804 R centrifuge (S-4-72 swinging-bucket rotor) and injected into a Superdex 200 Increase column (GE Healthcare) (5.2). Aliquots containing pure VP16-TAD at ~ 0.55 mg/ml and Gal4-TAD at ~ 0.3 mg/ml were stored at -80 ^{o}C in 10% glycerol (5.2).

2.3 Protein expression and purification in *S. cerevisiae*

2.3.0 General description

To express and purify protein in *S. cerevisiae*, the recombinant plasmids coding for the proteins of interest (2.0.1, Table 3) were transformed into yeast cells (2.3.1). Transformant cells were selected in the appropriate drop out medium (2.3.3.1 and 2.3.4.1) and grown in the presence of glucose free medium containing raffinose as alternative carbon source to avoid expression repression by glucose upon induction. The induction of protein expression was carried out by the addition of 2% galactose to the medium (2.0.3, Table 10), with variable induction times depending on the protein(s) to be expressed (2.3.3.1 and 2.3.4.1). The cell

pellet (2.3.2.4) was lysed in a Freezer/Mill cryogenic grinder machine (2.3.2.5) and subsequent steps comprised affinity gel purification and ÄKTA chromatography, combined with protein concentration steps when needed (2.3.3 and 2.3.4).

The mentioned steps involved in protein expression and purification will be independently and sequencially described below.

2.3.1 Yeast transformation protocols

Yeast transformation of *S. cerevisiae* was carried out following two different protocols. The Jansen group protocol was initially used for yeast transformation, and it was replaced later on by the Gietz group protocol for simplicity and efficiency reasons.

2.3.1.1 Jansen group protocol

The preparation of chemically competent yeast cells and transformation was carried out according to a protocol from the Jansen group (adapted from Knop *et al.*, Yeast 15, 963-972 (1999)). To transform *S. cerevisiae* with recombinant plasmid DNA, 20 mL of YEPD medium (2.0.3, Table 9) supplemented with ampicillin (2.0.3, Table 10) were inoculated with BCY123 strain (2.0.2, Table 6) and incubated overnight at 30 °C and 150 rpm. The next day, 50 mL of YPED were inoculated with the overnight preculture to reach an OD_{600} ~ 0.2, and the medium was incubated at 30 °C and 150 rpm until OD_{600} = 0.5 – 0.7. Then, the culture was transferred to a 50 mL Falcon tube, and it was centrifuged in an Eppendorf 5804 R bench centrifuge for 5 min at 4,000 rpm and room temperature (RT). The cell pellet was resuspended in 25 mL of sterile H_2O, centrifuged again, resuspended in 5 mL of LitSorb (2.0.4, Table 12), centrifuged again, and finally resuspended in 360 µl of LitSorb. Salmon sperm DNA (2.0.4, Table 12) was heated for 10 min at 95 ^oC on a thermoblock unit and chilled on ice. 40 µl of salmon sperm DNA were added to the yeast cell resuspension and 50 µl aliquots were prepared, either to use fresh or stored at -80 ^oC. To transform the competent yeast cells, 2 µl of plasmid DNA were added to 10 µl of cells. After the addition of 72 µl of LitPEG, the mixture was incubated for 30 min at RT. Then, 8 µl of DMSO were added and incubated for 15 min more at 42 ^oC in an Eppendorf TermoMixer F1.5/F. Then, the sample was centrifuged for 3 min at 2,000 rpm and RT in an Eppendorf 5415 D bench microcentrifuge, and the cell pellet was resuspended in 100 µl of sterile H_2O and plated into the corresponding selection medium. Plates were grown for 2-3 days at 30 ^oC.

2.3.1.2 Gietz group protocol

To transform *S. cerevisiae* BCY123 strain, the -80 ^oC stock was streaked out on an agar plate containing YPAD medium and ampicillin (2.0.4, Table 12), and grown at 30 ^oC for 2-3 days.

Then, 10 mL of 2x YPAD medium supplemented with ampicillin were used to inoculate a 50 mL Falcon tube, and grown overnight at 200 rpm and 30 ^{o}C. The next day, 1 mL of preculture (OD_{600} at ~ 13.8) was used to inoculate a 50 mL culture, at a starting OD_{600} of ~ 0.3, and incubated at 200 rpm and 30 ^{o}C for 3-5 hrs until OD_{600} ~ 1.2. Then, the yeast cells were spun down in a 50 mL Falcon tube at 3,000 x *g* in an Eppendorf 5415 D bench microcentrifuge for 5 min, and the pellet was washed with 50 mL of sterile water and centrifuged again, and finally resuspended in 1 mL of H_2O. The 1 mL mixture was split into 100 µl samples and spun down for 1 min at 3,000 x *g*. Subsequently, 356 µl of transformation mixture were added (2.0.4, Table 12), in addition to 2-4 µl of plasmid DNA. After vortex mixing, the mixture was incubated at 42 ^{o}C for 40 min. Then, the cells were spun down at 3,000 x *g* for 30 secs and resuspended in 1 mL of sterile H_2O. 10 µl and 200 µl of the cell resuspension were plated using the corresponding selective medium (2% agar, 2% glucose and CSM-TRP or -TRP-URA; 2.0.3, Table 9), and grown at 30 ^{o}C for 2-3 days to detect colonies.

2.3.2 Protein expression in *S. cerevisiae*

2.3.2.1 The *Multi* Sac yeast expression system (K. Nagai lab)

Protein expression in *S. cerevisiae* was carried out by the overexpression of the desired protein(s) from plasmid(s) containing a galactose inducible promoter. The yeast expression system used in our lab, named as *Multi* Sac system, was kindly provided by Kiyoshi Nagai and consisted in plasmids containing a GAL-GAPDH hybrid promoter and a 3' UTR sequence derived from a vector developed at Somatogen, Inc. by Dr. Gary Stetler and colleagues.

We adapted the system to our GIA based cloning method so that the cassette in the pENTDQ plasmid which contained the hybrid GAL-GAPDH galactose inducible promoter and the 3' UTR sequences were cloned by GIA into the 2 µ-based pRS vectors to generate the final expression vectors (Fig. 2.3). The yeast expression strain utilised by the *Multi* Sac expression system is BCY123 (2.0.2, Table 6), so that the pRS424 plasmid was able to complement tryptophan (TRP) auxotrophy and pRS426 was able to complement uracil (URA) auxotrophy. To select for cells containing the plasmid(s) coding for the protein(s) of interest, CSM (Complete Synthetic Medium) was used: CSM-TRP for pRS424 selection and double drop-out CSM-TRP-URA medium in the case of double plasmid transformation (2.0.3, Table 9). The CSM medium was subsequently replaced by casaminoacid expression medium (Table 9; 6.3). All media were always supplemented with adenine to complement the strain auxotrophy and ampicillin to prevent contamination (2.0.3, Table 10).

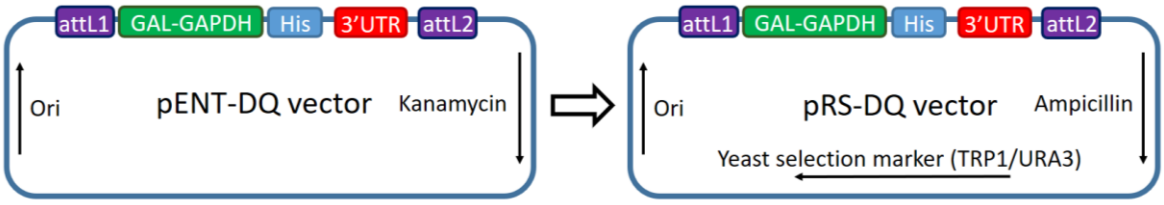


Figure 2.3 | The pRS-DQ yeast expression vector. A cassette containing the galactose inducible GAL-GAPDH hybrid promoter and 3' UTR was transferred from pENT-DQ vector to pRS424/pRS426-Dest vectors to generate the pRS-DQ expression vectors; the His tag showed in the figure was removed and replaced by different tags. The ORF of interest was inserted in between the tag and the 3' UTR regions. (Adapted from Kiyoshi Nagai group's *Multi* Sac system manual).

The modified plasmids were design in this study to contain either a 3xFlag tag (N-Met-Asp-Tyr-Lys-Asp-His-Asp-Gly-Asp-Tyr-LysAsp-His-Asp-Ile-Asp-Tyr-Lys-Asp-Asp-Asp-AspLys-C) or ProteinA tag, ready to clone an open reading frame (ORF) by a two fragments and one-step cloning reaction.

2.3.2.2 Yeast test expression

After transformation of *S. cerevisiae* BCY123 with the desired pRS-DQ plasmid(s) containing the ORF(s) of interest, a single colony was selected and grown in a 50 mL Falcon tube containing the corresponding drop-out medium and supplemented with 2% glucose. After 24-48 hours of incubation time at 30 ºC and 180-200 rpm, 0.75 mL of culture were taken and mixed with sterile 50% glycerol solution, and stored at -80 ºC as stock.

To test for protein expression, small scale cultures (typically 50 mL of culture) were grown from -80 ºC yeast stocks streaking the yeast cells on 2 % agar CSM drop-out medium and leaving them to growth for 2-3 days at 30 ºC. To prevent repression of protein expression, the liquid pre-culture contained 2% of raffinose (2.0.3, Table 10) as carbon source instead of glucose, and cell growth was carried out at 180 rpm and 30 ºC in either CSM drop out medium or casamino acid medium (2.0.3, Table 9). The inoculation of the culture from the pre-culture was performed at an initial OD_{600}=0.2. Induction was carried out with 2% of galactose at OD_{600}=0.8-1.2. To check for protein expression, recombinant protein expression was analysed by TCA precipitation of three OD_{600} units (2.4.4) at different expression times, followed by Western blot analysis (2.4.5). For the lysis of the yeast material, Y-PER reagent (2.0.5, Table 13) was used following the indications of the manufacturer.

2.3.2.3 Yeast medium-large scale expression

After test expression, purifications consisting of 12 L were carried out, and increased to 24 L if needed. Yeast culture volumes were incubated at 180 rpm and 30 ºC growing 1 L of culture

in individual glass flasks within incubators with 12-13 flask clamps (Multitron standard incubators, Infors HT). The inoculation from one to three 1 L flasks of preculture and the induction with 2% galactose was carried out as previously described for the test expression protocol (2.3.2.2). Therefore, 2% of raffinsose was used as alternative carbon source instead of glucose, and the expression medium consisted in either CSM drop out medium (for Tra1) or casamino acids medium (for Mec-Ddc2 12 L expressions) (2.0.3, Table 9).

2.3.2.4 Cell harvesting

Cell harvesting was performed in a Beckman Coulter J6-MI centrifuge (JS-4.2 rotor) for 30 min at 4,000 rpm and 4ºC using 1 L centrifuge tubes, followed by a second centrifugation step in 50 mL Falcon tubes on an Eppendorf 5804 R bench centrifuge (4-S-72 swinging bucket rotor) at 4,000 rpm. Alternatively, 500 mL centrifuge tubes - instead of the 1 L tubes - were used in a Thermo Scientific Sorvall RC 6 plus (SLA-3000 rotor) for 8 min at 8,000 rpm to reduce the resuspension effect and increase cell recovery. The resulting yeast cell pellet was frozen and stored at -80 ºC for further processing.

2.3.2.5 Yeast cell lysis

Either fresh cell pellet or frozen material was used for cell lysis. Cell pellets at -80 ºC were thawed placing the 50 mL Falcon tubes containing the yeast cell material in a room temperature (RT) water bath with magnetic stirring.

To prepare the cell pellet for lysis, the corresponding lysis buffer was added to the cell pellet at 1:1 ratio (v/v) and mixed (2.0.5, Tables 14-16). The resulting mixture was dropped into a liquid nitrogen container using a Pasteur pipette to generate beads of frozen material that could be stored at -80 ºC for further processing or immediately for lysis.

The frozen beads were lysed in a SamplePrep 6870 Freezer/Mill cryogenic grinder unit (SPEX, Inc.) running four grinding cycles (two min “on” and two min “off” each cycle), with an impact rate of 15 Hz. The resulting product was a frozen powder containing a mixture of cell debris and lysis buffer. This powder was stored at -80 ºC for further processing.

2.3.3 Purification of full-length Tra1 from *S. cerevisiae*

2.3.3.1 Protein expression

The pRS424-DQ-3xFLAG-Tra1 plasmid (2.0.1, Table 3) was transformed into *S. cerevisiae* BCY123 strain (2.0.2, Table 6). For 12 and 24 L protein expressions, CSM-TRP medium (2.0.3, Table 9) was inoculated in 1 L culture glass flasks at an initial OD_{600}=0.2, as previously described (2.3.2.3). Protein expression was induced adding 2% of galactose to the medium at

OD_{600}=0.8. Induction was carried out overnight for the 12 L expressions and later optimised to 6 hrs for the 24 L expressions, at 30 ºC and 180 rpm.

2.3.3.2 Affinity tag purification

Tra1 cell pellet was resuspended in "Lysis buffer" (2.3.2.5; 2.0.5, Table 14) and cell lysis was performed in a Freezer/Mill cryogenic grinder unit as previously described (2.3.2.5). In addition to the increased ratio of lysis buffer/cell pellet (from 1/3 (v/v) to 1/1 (v/v)), additional modifications were added to the optimised purification as described in 4.1.6. Firstly, after thawing the freezer-milled material (stored at -80 ºC) in a water bath, benzonase (2.0.5, Table 13) was added to the melted lysate material solution and the mixture was incubated for 15 min at 4 ºC. Then, a sonication step in a Vibra-Cell sonicator unit (Sonics, In.) followed consisting in three cycles at 65% amplitude and one cycle at 85% amplitude, running each cycle for two seconds "on" and four seconds "off". The sonicated material was ultra-centrifuged in a Beckman Coulter Optima L-90K Ultracentrifuge unit (Type 45 Ti rotor) at 34,000 rpm for 2 hrs and the supernatant was recovered for further processing. The pH of the solution was properly adjusted before the resin binding step, and 1 mL of equilibrated anti-FLAG-M2 affinity agarose gel resin (Sigma) was incubated in batch with the lysate in the 12 L expressions and 2 mL of affinity gel resin in the 24 L expressions. The incubation was carried out at 4 ºC and mixing for 2 hrs. Then, the resin was recovered, centrifuging the material at 500 x *g* for 5 min at 4 ºC in an Eppendorf 5804 R bench centrifuge and removing the supernatant with a pipette, and finally recovering the resin at the bottom with a 25 mL plastic pipette. The resin was poured into a 25 mL Econo-Pac chromatography column (Bio-rad). For the 12 L purification, the washing step was performed applying five column volumes (CV) of "Washing buffer" (WB) (2.0.5, Table 14) to the affinity resin column followed by 15 CVs of detergent-free WB. For the 24 L purification, the washing step was replaced by 20 CVs of detergent-free WB (2.0.5, Table 14), and for elution, 2.5 CVs of "Elution buffer 1" (EB^1) (2.0.5, Table 14) were incubated with the washed resin for 15-30 min. Then, five additional CVs of "Elution buffer 2" (EB^2) (2.0.5, Table 14) were applied to remove most of attached material to the resin. In the case of the 12 L purification, the first elution fraction consisted in 1 CV of EB, followed by 4 CVs of the same EB. The eluted material was injected into a gel filtration column (2.3.3.4) in the case of the 12 L purification. For the 24 L purification, the material was applied first to a MonoS 50/50 GL column (GE Healthcare) (2.3.3.3) without any delay or intermediate freezing steps, followed by additional purification steps.

2.3.3.3 IEX chromatography

After affinity purification (2.3.3.2), ion exchange chromatography (IEX) was the next purification step for the Tra1 affinity purified material for the 24 L expressions. Typical protein yield after affinity purification in the case of the optimised (24 L) purification consisted in 4 ml of elution fraction 1 (E1) at ~ 0.4 mg/ml and 2 ml of E2 at ~ 0.2 mg/ml, and lower than 0.05 mg/ml for the other fractions, as measured by Bradford assay (2.4.1). The sample was sequentially applied to MonoS 5/50 GL and MonoQ 5/50 GL columns (GE Healthcare) using an ÄKTA Pure protein purification system (GE Healthcare). Before sample injection to the MonoS column, the affinity purified protein was balanced to 100 mM NaCl concentration using salt-free buffer (2.0.5, Table 14) and the corresponding column was equilibrated in "Gel filtration buffer" (2.0.5, Table 14). After sample application, the column was washed with "MonoS/Q equilibration buffer" (2.0.5, Table 14) and the elution was carried out applying a 100-1,000 mM NaCl linear gradient to the column. The elution fractions were analysed by SDS-PAGE (2.4.2) and the fractions containing pure Tra1 were collected from the rack and pooled together (4.1.6). After balancing the NaCl concentration in the pooled fractions to 100 mM NaCl, the material was applied this time to a pre-equilibrated MonoQ 5/50 GL column and eluted using a 100 mM-1,000 mM NaCl linear gradient. As previously described, fractions were analysed by SDS-PAGE, and the fractions containing the Tra1 peak were pooled together for gel filtration chromatography (2.3.3.4).

2.3.3.4 Size-exclusion chromatography

The last purification step consisted in a gel filtration (GF) chromatography using a Superose 6 Increase 30/100 GL column (GE Healthcare). Because of the limited injection volume, a spin concentrator Amicon Ultra-15 mL 10 kDa MWCO (Merk Millipore) was used to concentrate the Tra1 sample from the MonoQ 5/50 GL (GE Healthcare) step to 0.5 mL. Two peaks were present in the chromatogram, the first peak corresponding to aggregated protein and the second peak to monodisperse Tra1 protein. The resulting fractions were analysed by SDS-PAGE and the fractions containing monodisoperse Tra1 (approx. at 12.5-14.75 mL of elution volume) were pooled together and concentrated to 150 µl using a micro-centrifugal filter Amicon Ultra-0.5 mL with 10 kDa MWCO (Merck Millipore). The freshly concentrated material was used for cryo-EM sample preparation on a FEI Vitrobot Mark IV system (Thermo Fisher Scientific) to vitrify monodisperse Tra1 in liquid ethane (3.4.1.1).

2.3.4 Purification of full-length Mec1-Ddc2 from *S. cerevisiae*

2.3.4.1 Protein expression

pRS424-DQ-3xFLAG-Mec1 and pRS426-DQ-ProtA-Ddc2 plasmids (2.0.1, Table 3) were cotransformed into *S. cerevisiae* BCY123 strain (2.0.2, Table 6). The preculture was prepared as previously described (2.3.2.3) and 12 L of casamino acids medium (2.0.3, Table 9) were inoculated into 1 L culture glass flasks at an initial OD_{600} ~ 0.1. Protein expression was induced adding 2% of galactose to the medium at OD_{600}=1.2 and grown for 24 hours at 30 ºC and 180 rpm before harvesting. As additive, 0.05 % of glucose was added to the expression medium.

2.3.4.2 Affinity tag purification

Mec1-Ddc2 cell pellet was resuspended into "Lysis buffer" (2.0.5, Table 16) and the subsequent steps previous to the binding to the resin were performed as described before (2.3.3.2), incubating the lysate with benzonase for 30 min instead. For the initial 12 L purification, 1.5 mL of equilibrated anti-FLAG-M2 affinity gel resin (Sigma) were incubated in batch with the lysate, and 2.2 mL of resin for the next purification. The incubation was performed at 4 ºC and mild mixing for 2 hrs. Then, the resin was recovered centrifuging the material at 500 x *g* for 5 min at 4 ºC in an Eppendorf 5804 R bench centrifuge, removing the supernatant with a 1,000 µl pipette and finally recovering the resin at the bottom with a 25 mL plastic pipette. The resin was poured into a 25 mL Econo-Pac cheomatography column (Biorad) and the washing step was performed applying sequentially to the column 5 CVs of WB^1 and 20 CVs of WB^2 (2.0.5, Table 16). To elute the protein complex, 1.5 mL of "Elution buffer 1" (EB^1) (6.3.1) or 5.6 mL of EB^1 (6.4) were incubated in batch with the washed resin overnight at 4 ºC and gentle mixing. To elute the protein from the resin, EB^1 contained 500 µg/mL of 3xFLAG peptide and TEV protease at 140 U/mL (2.0.5, Table 16). The eluted material was recovered by pouring the mixture of resin and elution buffer into a 25 mL Econo-Pac column (Biorad). After filtering the mixture, 5 CVs of "Elution buffer 2" (EB^2) (2.0.5, Table 16) were applied to the column to compete the elution of the material from the resin. Protein yield after affinity purification in the case of the optimised purification consisted in 6 ml of E1 at ~ 0.3 mg/ml, 5 ml of E2 at ~ 0.15 mg/ml and 5 ml of E3 at ~ 0.05 mg/ml. The eluted fractions were stored at -80 ºC for further processing.

2.3.4.3 Size-exclusion chromatography

Given the high-level of purity after the affinity purification step (6.3 and 6.4), only one additional gel filtration (GF) chromatography step was required. Initially, a Superose 6 Increase 10/300 GL column (GE Healthcare) was used for test purifications and a XK 16/70 column (GE

Healthcare) for structural studies. When the sample was BS3 cross-linked, 0.75-6 mM of BS3 solution (2.0.5, Table 13; 6.3.2) was incubated for 15 min at 30 ^{o}C and the sample was quenched using 25 mM Tris pH 7.0. Then, the sample was centrifuged for 30 min at 4 ^{o}C in an Eppendorf 5415 D bench microcentrifuge at full speed before injection into the gel filtration column. Gel filtration columns were pre-equilibrated with “Gel filtration buffer” (2.0.5, Table 16), and two peaks were consistently present in the gel filtration chromatograms; the first peak corresponding to aggregated Ddc2 protein and the second peak to monodisperse Mec1-Ddc2 protein complex (6.3.1, Fig. 6.3). The resulting fractions were analysed by SDS-PAGE and fractions containing monodisoperse complex (approx. at 11-14 mL of elution volume in the Superose 6 Increase 30/100 GL column and ~ 68-78 mL in the XK 16/70 column) were pooled together. A final concentration step consisted in the incubation of the pooled fractions with 100 µl of Q sepharose resin pre-equilibrated with gel filtration buffer (2.0.5, Table 16). After 2 hrs of incubation in batch at 4 ^{o}C and mixing, the recovery was performed applying the batch solution to a Pierce micro-Spin column (Thermo Fisher) and eluting in 100 µl of “Elution buffer 3” (EB^{3}) containing 600 mM NaCl (2.0.5, Table 16). The concentrated material was vitrified in liquid ethane for single-particle cryo-EM analysis using a FEI Vitrobot Mark IV system (Thermo Fisher Scientific) (3.4.3.1).

2.4 Protein analysis

2.4.0 General description

Protein analysis was performed by SDS-PAGE protein gel electrophoresis (2.4.2) followed by Coomassie staining (2.4.3), Western blot analysis (2.4.5) or Silver staining (2.4.6), and also by the measurement of protein concentration using chemical and physical methods (2.4.1). The 260 nm and 280 nm reading by the spectrophotometer of the ÄKTA machine during sample elution also provided very valuable biochemical information. Purified protein was imaged by electron microscopy negative staining (3.1) as a quality control method for preliminary analysis before cryo-sample preparation or vitrification (3.2 and 3.4) previous to structural analysis by single-particle cryo-electron microscopy (cryo-EM) (3.4).

2.4.1 Measurement of protein concentration

The main two methods to measure protein concentration consisted in Bradford chemical analysis (Bio-Rad, (2.0.6, Table 17)) and the nanodrop machine (2.1.3).

For Bradford analysis, a 1:5 water dilution (v/v) of Bradford reagent was mixed with normally 2-8 µl of purified protein solution, and the mixture was read on a spectrophotometer at 595 nm wavelength using a blank solution as reference, consisting of diluted Bradford reagent with no

protein. The calibration curve was performed using a BSA protein solution diluted at concentrations that provided values in the linear range of the Beer-Lambert equation.

For nanodrop measurements, 1.5-2 µl of purified protein solution were applied for protein measurement, and the resulting plot was analysed for both 280 nm and 260 nm signals, including their ratio. The reading value on the nanodrop was used in combination with the extinction coefficient obtained in ProtParam (on-line program from the Expasy server) to calculate a more accurate protein concentration value. In the case of Tra1, this value is similar to the nanodrop output value, where Abs 0.1% (=1 g/l) 0.973, assuming all pairs of Cys residues form cystines, and 0.968 assuming that all Cys residues are reduced.
In the case of the presence of detergent within the buffer, Bradford analysis was the method of choice given that the high level of background signal coming from the buffer prevents an accurate measurement by the nanodrop.

2.4.2 SDS-PAGE protein analysis

Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) was used as standard method to analyse protein purity and quality, and also as a non-quantitate method for visual assessment of protein amount.

Before loading the samples into the SDS-PAGE system, the protein sample was mixed with 6x Laemnli loading buffer (2.0.6, Table 17) and heated at 95 ^{o}C for 5 min on a thermoblock unit.

Initially, acrylamide gels were prepared in-house, polymerising acrylamide in stacking and running buffers (2.0.6 Table 17). The “running solution” or “resolving gel solution” was polymerised after the addition of TEMED and 10% APS (2.0.6, Table 17), so that the mixture was immediately poured into a FastCast acrylamide gel casting system (Bio-Rad). After the polymerisation of the resolving gel solution, the stacking solution was poured on top and polymerised in a similar way. The gel was sealed using sample wells (10x or 15x) and left for complete polymerisation before utilisation. 1x MES running buffer (2.0.6, Table 17) was used to run the hand casted polyacrylamide SDS-PAGE gels on a mini-PROTEAN electrophoresis system (Bio-Rad), typically at 120V for 1-1.5 hrs. The in-house hand casted acrylamide gels were soon substituted by precast gradient gels (Bolt 4-12% Bis-Tris Plus gels, 2.0.6, Table 17). Vertical protein electrophoresis was performed in this case on a Bolt Mini Gel Tank 500VDC electrophoresis system (Novex LifeTechnologies, Thermo Fisher Scientific) containing 1x Bolt MES SDS running buffer (2.0.6, Table 17), and the run was typically performed at RT for 30-60 min at 180 V.

2.4.3 Coomassie staining

Initially, in-house Coomassie staining solution (2.0.6, Table 17) was used for protein staining of SDS-PAGE gels. The staining solution was applied to the gel inside a small container and heated for a few seconds (10-20 s) in a standard microwave unit. Then, the container was covered with aluminium foil and incubated in agitation for 15-30 min. To destain the gel and see protein bands, destaining solution (2.0.6, Table 17) was applied to the Coomassie stained gel.

The homemade staining and destaining solutions were replaced later on by commercial InstantBlue stain (2.0.6, Table 17). After 5 min of incubation, many bands were already visible and overnight staining at RT and shaking typically yielded excellent results.

2.4.4 TCA precipitation from yeast cell cultures

To analyse protein expression from yeast cells, 3 OD_{600} units[2] of yeast culture were centrifuged in a 1.5 mL Eppendorf tube at 6,000 x *g* for 5 minutes in an Eppendorf 5415 D bench microcentrifuge. The cell pellet - either fresh or frozen - was used as starting material for cell lysis followed by TCA precipitation. To lyse yeast cells, the pellet was resuspended in 1 mL of water until homogeneity and then, 170 µl of "Buffer A" (2.0.6, Table 17) were added. The mixture was mixed by vortexing, followed by 15 min of incubation on ice. Then, 80 µl of 100% TCA were added, and the mixture was vortexed again and left for 15 minutes on ice. The Epperdorf tube was centrifuged for 10 minutes at full speed in a microcentrifuge at 4 ºC and the supernatant was discarded. The pellet was washed with 500 µl of cold acetone and the tube was centrifuged again at full speed at 4 ºC for 5 minutes. The supernantant was discarded and the pellet dried out on an Eppendorf TermoMixer F1.5/F at 37 ºC. Then, the pellet was resuspended in 50 µl of "Buffer B" (2.0.6, Table 17) and incubated on the thermoblock at 65 ºC for 10 minutes at 600 rpm. After resuspension, 20 µl were used for SDS-PAGE analysis followed by Coomassie staining (2.4.3) or Western blot analysis (2.4.5), so that 1.2 OD_{600} units were loaded into the gel.

2.4.5 Western blot immunodetection of protein epitopes

Western blot analysis of pure protein or yeast cell lysates after TCA precipitation was performed by running the samples in SDS-PAGE precast gels (Bolt 4-12% Bis-Tris Plus (Invitrogen); 2.0.6, Table 17) at 180 V for 40 min. After removing the plastic cassette, the gel was equilibrated for 10 min in "Transfer Buffer" (TB) (2.0.6, Table 17) with agitation. Transfer

[2] Three OD_{600} units are defined as: [measured OD_{600}] x [culture volume (mL)] = 3. Therefore, 1 mL of culture at OD_{600}=1 is equal to one OD_{600} unit.

sponges (Invitrogen) were soaked into the transfer buffer together with four sheets of Whatman blotting paper. The Immobilon-FL PVDF transfer membrane (Millipore, Inc.) was pre-wetted in pure methanol for 12 s for activation and then transferred to 1x TB for 5 min. Two sponges were placed on the anode tray, followed by two Whatman blotting papers and the transfer membrane on top, and followed by the gel. Air in between the membrane and the gel was carefully removed by gently applying pressure. Finally, two more wet Whatman blotting papers and sponges were placed on the cathode side. The cassette, containing the transfer sandwich inside, was assembled into the XCell II blot module (Invitrogen) and locked by the built-in wedge. The transfer sandwich was covered with TB (2 cm of buffer above the cassette). The outer chamber was filled with 600 mL of deionised water for heat dissipation. The protein electrophoresis or protein "transfer " system was run at 180 mA (starting electric current), for 1 hour or for 2hrs and 30 minutes in the case of large size proteins such as Tra1 and Mec1. After transfer, the protein gel was stained with InstantBlue (2.0.6, Table 17) to check for protein bands.

The membrane was blocked overnight in "Blocking solution" (2% of PBS-T milk**;** 2.0.6, Table 17) with shaking at 4 ^{o}C. The next day, the membrane was washed with 40-50 mL of "Washing solution" (PBS-T; 2.0.6, Table 17) for 10 minutes before the application of the primary antibody, consisting in 6 mL of PBS-T milk with the corresponding antibody dilution (2.0.6, Table 17). After incubation at RT for 1 hour with gentle shaking, the membrane was washed again with 40-50 mL of "Washing buffer", three times for 10 minutes each. The secondary antibody (2.0.6, Table 17) was incubated at RT for 1 hour with gentle shaking, and then washed again three times in "Washing buffer". An additional wash with 1x PBS (2.0.6, Table 17) for 10 min was carried out before developing the membrane.

To develop the membrane, fresh "Developing solution" (2.0.6, Table 17) was prepared and applied evenly to the membrane. The membrane was incubated for 5 min in the dark, then placed over absorbing tissue, and left in the dark 5 min more. The membrane was imaged on a Fujitsu FLA 3000 scanner (excitation wavelength at 532 nm and filter at 580 nm).

2.4.6 Silver staining

Silver staining was performed to detect protein bands from SDS-PAGE analysis at nanogram (ng) amounts. SilverQuest staining kit (Invitrogen) was used for silver staining following the manufacturer manual.

2.4.7 Mass spectrometry

Purified Tra1 was analysed by SDS-PAGE and the corresponding protein band was extracted from gel with a scalpel and sent to mass spectrometry (MS) analysis (MS unit at Cambridge University, Department of Chemistry). MALDI-TOF analysis of tryptic digest on a “Waters Micromass MALDI micro MX” device generated a mass fingerprint. The Mascot search engine (Matrix Science) for peptide and protein identification processed the generated values. The NCBInr database was used (National Center for Biotechnology Information, non-redundant protein accession number), obtaining 587 highest scoring protein hits evenly distributed from N to C terminus along *S.c.* Tra1 polypeptide, and confirming that the band extracted from SDS-PAGE analysis is pure full-length Tra1.

2.4.8 Electron microscopy (EM)

Negative staining (NS) EM was used for protein analysis (please see section 3.1) to assess homogeneity and monodispersion of the sample, and to obtain preliminary structural information. For single-particle cryo-EM analysis, the sample was plunge-frozen into liquid ethane using a Vitrobot unit as described in 3.2 and 3.4 (3.4.1.1, 3.4.2.1 and 3.4.3.1). Data processing (3.4.1.2, 3.4.2.2 and 3.4.3.2), atomic model building (3.4.4) and maps and models interpretation (3.4.5) are also described in the next sections.

3 Electron Microscopy: Theory and Practice

3.0 Introduction: Electron microscopy and single-particle analysis

During the last decades, X-ray crystallography has been a mainstream technique for structural analysis at atomic resolution. Although it has brought enormous knowledge during the past years, structural analysis is very often impeded by the difficulty in crystallising the protein sample. Another limitation resides in the need to purify protein in milligram quantities for crystallisation assays. In this respect, single-particle cryo-EM offers nowadays the possibility to achieve structural knowledge at near-atomic resolution of large protein complexes from microgram amounts of protein samples, many of which are unamenable to crystallography.

3.1 Negative staining

3.1.1 Method description

After protein purification, the purified protein was imaged by negative staining electron microscopy to analyse protein quality and mono-dispersion, and also to obtain preliminary, low resolution information regarding to the overall protein shape. Negative staining is a fast method to quickly analyse protein, and needs a very low amount of protein (less than 0.1 mg/mL of protein can be imaged). The method is based on the differential deposition of heavy metal ions on a continuous carbon support, where protein particles are attached to a previously glow-discharged surface. The presence of a particle will create a low electrodense area in comparison to the regions where there is no protein, which contain a thicker layer of heavy metal in comparison to the layer deposited over the protein.

3.1.2 Protocol

Uranyl acetate solution (2.0.7, Table 18) was used for negative staining. To negative stain particles, Carbon support film type B 400 Mesh (Agar Scientific) grids with continuous carbon were glow discharged in a easiGlow (PELCO) glow discharge cleaning unit at 30 mA for 1 min. Then, 3 to 5 µl of protein solution (typically at 0.1-0.3 mg/mL) were applied to the grid and incubated for 1 min (or longer if needed because of low protein concentration). The applied drop to the grid was side blotted with Whatman filter paper (more than one application can be performed if necessary). Then, two 10 µl drops of buffer (the same as the protein solution) were placed into a ParaFilm strip to sequentially wash the grid on each drop. After blotting the grid again, two 10 µl drops of 2% uranyl acetate solution (2.0.7, Table 18) were deposited on a Parafilm strip, the first drop for a washing step of 5 seconds and the second drop for staining over 1 min. After a final blotting step, the grid was dried out for at least 5 min at RT before imaging on the electron microscope. Typically, several grids at different protein

concentrations were prepared. Images were collected on a 100 kV FEI Tecnai T10 or 120 kV FEI Tecnai T12 microscope equipped with a 4k x 4k CCD camera (Gatan, Inc.), or a 200 kV FEI Tecnai F20 mounted with a DE20 direct electron detector.

3.2 Vitrification in single-particle cryo-EM analysis

Single-particle cryo-EM is an electron transmission technique based in the ability of electrons to cross a very thin layer of vitreous ice embedded protein (Fig. 3.1A). The protein sample (typically 2-3 µl) is applied to a grid of 3.05 mm in diameter.

The classical method of vitrification for cryo-EM has largely relied on a blotting step to remove the excess of protein solution from the grid utilising filter paper (which removes more than 90% of the applied protein) (Fig. 3.1A). Although initially the blotting step was manual, it has been replaced by a robotic arm to provide more control over the blotting and plunge-freezing steps, including humidity conditions. The structures presented in this study were solved operating a FEI Vitrobot Mark IV system (Thermo Fisher Scientific) unit for vitrification (settings in section 3.4).

After blotting, the grid is plunged into liquid ethane, rapidly vitrifying the sample, and preventing water molecules in the buffer to organise and form crystals, hence preserving the protein structure and allowing to image the sample under near-native conditions (Fig. 3.1A). If the layer of vitreous ice is too thin, the protein or complex stability can be compromised, and in some cases it can force preferred orientation on the grid. If the layer is too thick, high-resolution information can be compromised, and in extreme cases thick ice can prevent the obtainment of any structural information due to the impediment for the electrons to go through the sample. Therefore, the optimisation of the ice thickness is a critical parameter in any single-particle cryo-EM project, and in many cases this optimisation step represents a substantial part of the time spent during a structural biology project.

The grid type is also a crucial element for the success of a cryo-EM project. Support free versus continuous support grids can radically change particle distribution and also orientation. Either a thin layer of carbon (~ 3 nm thick) or graphene (Russo and Passmore, 2014) can be used, as well as different glow-discharge methods that invert the grid charge from negative in most cases to positive. Also, the grid material has been shown to play an important role in the image quality by reducing specimen movement when gold material is used (Russo and Passmore, 2014), allowing to increase the resolution of the reconstruction.

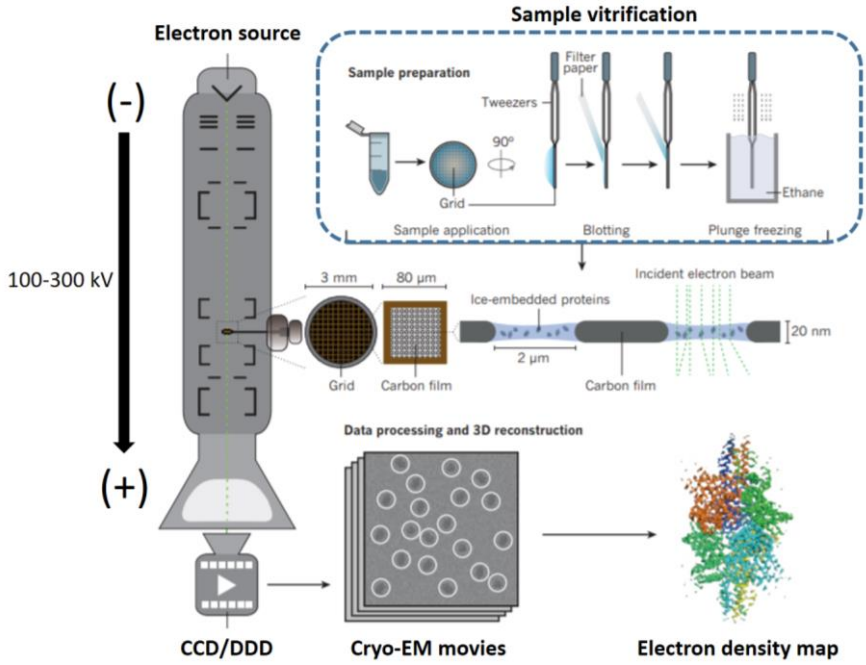
New methods are under development to automate vitrification and increase the level of control over the process. These new blot-free methods can apply (VitroJet system, Cryosol) or spray (Chamaleon system, Ttp Labtech) the sample on the grids with several advantages comparing

to paper blotting. These advantages include a reduced amount of sample, increased control over ice thickness, several application conditions can be quickly screened in one single grid, it can potentially reduce the preferred orientation problem and finally, time resolved experiments can be performed by controlling the mixing time of different proteins and buffers (Kontziampasis *et al.*, 2019).

3.3 Image formation and image processing in cryo-EM

After sample vitrification, the grid is inserted within the electron microscope column in high-vacuum conditions, so that the vitreous ice where the particles are embedded is preserved, and electrons can travel unimpeded from the electron source to the detector (Fig. 3.1A). Electrons are accelerated within the column vacuum from a filament or a field emission gun (FEG) in high-end microscopes by a potential difference generated in the column. After crossing the sample, electrons hit the camera – CCD camera or direct detector device (DDD) - that is located at the bottom of the microscope (Fig. 3.1A).

A)



B)

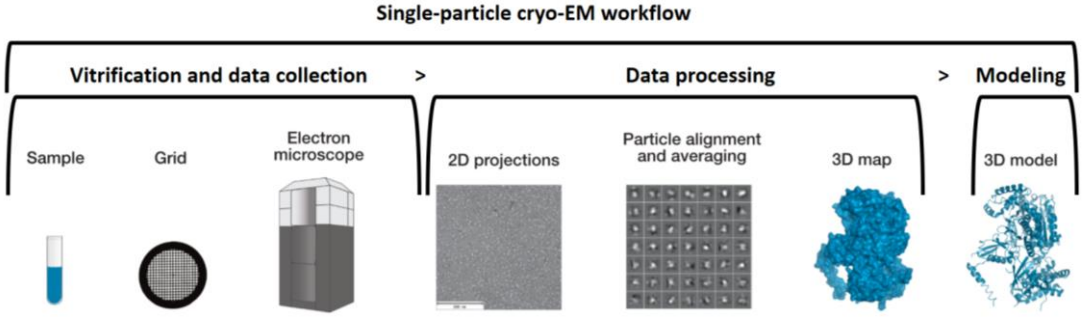


Figure 3.1 | Schematic representation of the workflow in single-particle cryo-EM. A) The sample preparation procedure for single-particle cryo-EM analysis is shown, consisting in the blotting of the protein sample before plunge-freezing of the grid in liquid ethane to fast freeze the particles embedded within a thin layer of buffer. A schematic representation of the microscope shows the electron source on top, the camera at the bottom (CCD or DDD), the magnetic lenses and the sample within the vacuum column. The different frames represent the fact that movies, instead of single-frames, are collected by the camera in a DDD. A high-resolution reconstruction of the glutamate dehydrogenase is also shown, as a product of the image processing process. **B)** Schematic representation of a single-particle cryo-EM project, starting from the purified sample and including grid vitrification and optimisation, image collection, image processing (2D and 3D), and finally atomic model building in the case of near-atomic resolution maps or map regions. (Adapted from Fernandez-Leiro and Scheres, 2016 and Doerr, 2015).

As will be discussed in section “no 3”, there is very little difference in electrodensity between vitreous ice and protein particles, therefore providing very little amplitude contrast to the image, and in addition, the weak scattering generated from EM samples also provides little phase contrast. Therefore, typical EM images are poor in contrast and present high level of noise. Because proteins are degraded after the exposure to the accelerated electrons (known as electron damage), protein contrast cannot be increased by imaging the sample with higher electron dose. Therefore, poor contrast raw images in single-particle analysis are noisy and represent 2D projections of the protein particles embedded in the vitreous ice. To generate a 3D reconstruction of the particle of interest, many 2D particle projections are combined (Fig. 3.1B; section “no 5”).

In addition to the noisy nature of the cryo-EM images, several other factors play a crucial role in single-particle cryo-EM analysis, and they will be discussed one by one in order as the electron path is described from its source to the detection device, including the cryo-EM processing of the collected data.

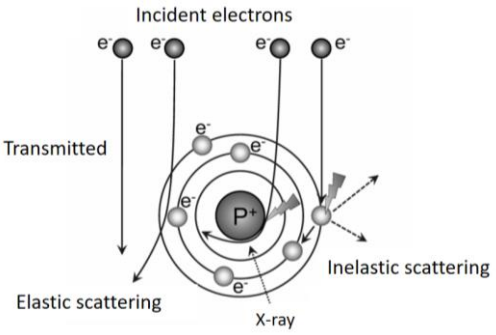
1) **The electron source**. Within the microscope vacuum column we find an electron source at the top (Fig. 3.1A). The cheapest electron source is a heated filament, and in this case, the coherence of the electron beam is very poor comparing to other electron sources, limiting the ability of the microscope to potentially collect high-resolution images. For this reason, field emission electron guns (FEG) are used in high-performance microscopes, which can generate a much more coherent electron beam. A last generation of cold (cathode) field-emission guns (CFEG) has been developed (Jeol company), with a higher level of electron beam coherence comparing to the previous design.

   If we consider the different components potentially affecting to the resolution limit in single particle analysis, and define them from a mathematical point of view, an envelope function can be used to describe the effect of those elements over the overall

resolution. Therefore, electron beam coherence can act as a first envelope function limiting resolution.

2) **Electron scattering**. Electrons are accelerated within a high-vacuum column, typically at 200kV or 300kV potential difference in high-performance microscopes. The electron wavelength depends on the acceleration voltage (equation in Fig. 3.2A), and at the previously described voltage values, the electron wavelength is small enough to properly image the objects of interest – smaller than the atomic size in this case-. When electrons hit the sample, two different scattering events can be found: inelastic scattering, which is when electrons lose energy and suffer a wavelength shift, damaging the sample without contributing to the image, and the other possible event is elastic scattering, where the electron energy is not affected, contributing to create the image (Fig. 3.2A). Different features of the sample will scatter differently i.e. high-resolution features in the sample as side chains will scatter electrons poorly compared to low and medium resolution features as protein domains (Fig. 3.2B). This physical effect related to the differential and decreased scattering ability of high-resolution features in the particles has an effect in the resolution of the images, given that the high-resolution signal is weaker in comparison to low and medium resolution information.

A)



$$\lambda = \sqrt{\frac{1.5}{V}}\,\text{nm}$$

B)

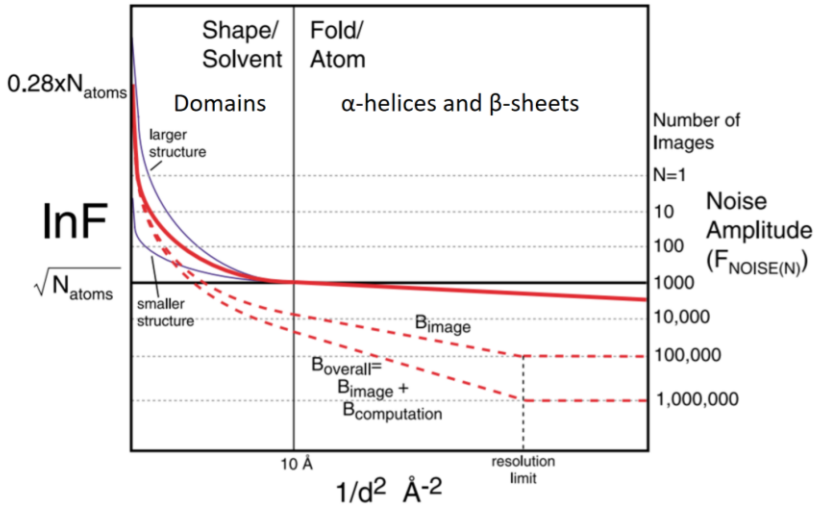


Figure 3.2 | The interaction of electrons with the sample. A) The possible outcomes upon the interaction of the electrons with the sample are highlighted, as explained in the text; transmitted electrons include elastically scattered and unscattered electrons. In addition, the equation that correlates the electron wavelength (λ) as a function of the voltage or electric potential difference (V) is also shown. (Adapted from Orlova and Saibil, 2011). **B)** Guinier plot showing the signal intensity measured as the spherically averaged structure factor amplitude (F) of a protein against resolution measured as $1/d^2$ where distance is in ångström units. The low resolution region of the plot includes protein domains and the high-resolution region of the blot (higher than 10 Å) includes secondary structure elements (such as α-helices and β-strands), and extending to atoms in the highest resolution region of the plot. (Adapted from Rosenthal and Henderson, 2003).

3) **Phase contrast, defocus, optical aberrations and the contrast transfer function.** Because biological molecules consist of light atoms, i.e. H, O, N and C, and they are embedded in water molecules - also composed of light atoms -, they do not produce very much amplitude contrast. For that reason, in EM the contrast of the image is mainly originated by phase contrast (Fig. 3.3). In the scattering event, electrons experience a phase shift of 90^o. In addition, the different path of electrons after elastic scattering comparing to the unscattered beam generates an additional phase shift that contributes to phase contrast, and the value of the additional path length due to scattering will depend on the particular scattering angle (Fig. 3.3A, B). Each Fourier component from the sample will scatter differently, and the contribution of each spatial frequency (*k*) to the image contrast is represented in the contrast transfer function (CTF) (Fig. 3.3C). For a given electron wavelength (λ), the CTF is influenced by the defocus value (Δ£) and the spherical aberration (C_s) of the lens (equation in Fig. 3.3C). Not all scattered electron beams will contribute to generate contrast at a given defocus value (Fig. 3.3C, D), generating zero values in the CTF at particular spatial frequencies. In this regard, the CTF function describes the amount of contrast that is transferred to the image as a function of the spatial frequency. A value of one in the CTF for a particular spatial frequency means that contrast is fully transferred and the particular Fourier component is fully present in the image, or expressed in a different way, that a particular Fourier component from the sample that is present at a given spatial frequency has generated the maximum possible contrast in the image (100% contrast). On the other hand, zero values in the CTF represent invisible Fourier components or spatial frequencies and are not present in the image. For that reason, data collection is performed at different defocus values to produce a shift in the CTF sinusoidal function and cover zero values. Thus, each image will contribute to the reconstruction with different Fourier components of the sample or spatial frequency information, and the combination of the different images during image processing has the potential to generate a complete reconstruction with no missing information. Therefore, phase contrast results from the interference of (phase shifted) elastically scatted electrons with unscattered electrons at the image plane (Fig. 3.3B, D) (Orlova

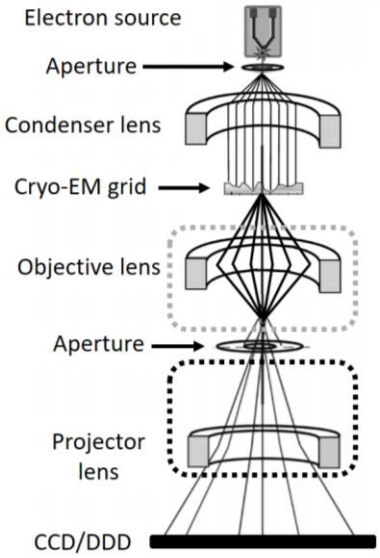
and Saibil, 2011; G. Jensen's EM course at Caltech). The amplitude value for the scattered electrons is much smaller comparing to the unscattered ones because the thin transparent samples in EM scatter electrons weakly, and therefore are described as weak phase objects. In the case of electron beams experiencing a total or accumulated phase shift of (+/-) 90 degrees, the interference with the unscattered beam at the image plane will result is virtually no difference in terms of amplitude (Fig. 3.3C, D). Likewise, if an electron beam is scattered at a very low angle, the Fourier component that it represents is almost invisible in the image because we can consider a similar scenario as for the 90^{o} shift. All different phase shift scenarios in between the previous ones (with a total phase shift value distinct to +/- 90 degrees) will contribute to the image to some extent (Fig. 3.3C, D).

Scattered electrons at different angles are focused by an extensive set of magnetic lenses (Fig. 3.3A, B, and E). For that reason, electron beam alignment and astigmatism correction, among others, are fundamental to avoid image aberrations coming from the lenses and provide a coherent electron beam, which can limit otherwise the resolution by imposing another envelope function to the system (Fig. 3.3F). In this regard, coma-free alignment is also very important, which consists in configuring the electron beam so that the electron path is perpendicular to the sample. Small deviations from the theoretical perpendicular axis can limit resolution significantly, especially at near-atomic resolutions. For that reason, newly developed image processing software allows to estimate beam-tilt values in the images and correct them (Zivanov *et al.*, 2018). If the microscope alignment is not ideal, the improvement after the application of this computational correction can be very significant.

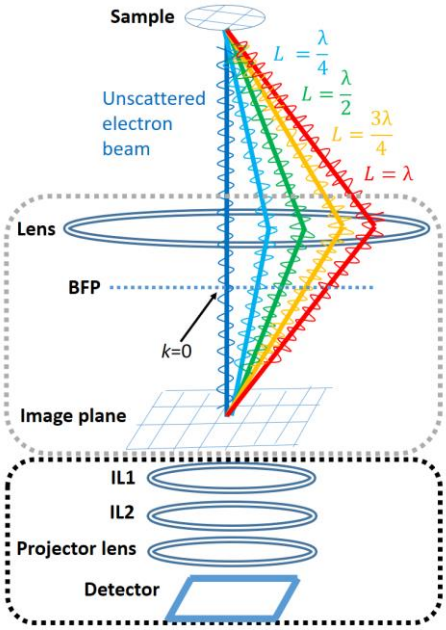
After electrons have physically interacted with the sample, they are scattered and need to be focused to form an image. Because electrons that are scattered at a different angle will be focused on a different point when they experience the same magnetic field from the lenses, a spherical aberration (C_s) effect is generated (Fig. 3.3E). This effect is considered during image processing and increases the contrast of the sample in combination with defocus by additional phase shifts, being the defocus intentionally applied to the sample (Orlova and Saibil, 2011). This value is C_s=2.3 mm on a 300kV FEI Tecnai G^2 Polara microscope and C_s=2.7 mm on a 300 kV FEI G^3i Titan Krios (Thermo Fisher Scientific). The defocus value can be estimated from the fitting of the Thon rings present in the power spectrum of the Fourier transformed image (Fig. 3.3G), a process called CTF estimation, and this information will be used to restore the image during image processing. In addition, spatial frequency information can be obtained, so that CTF estimation also allows to estimate resolution when the cross-correlation

between the theoretical (estimated) CTF and the power spectrum of the image is performed at a specific cut-off value (Rohou and Grigorieff, 2015).

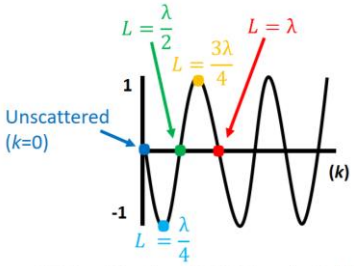
A)



B)

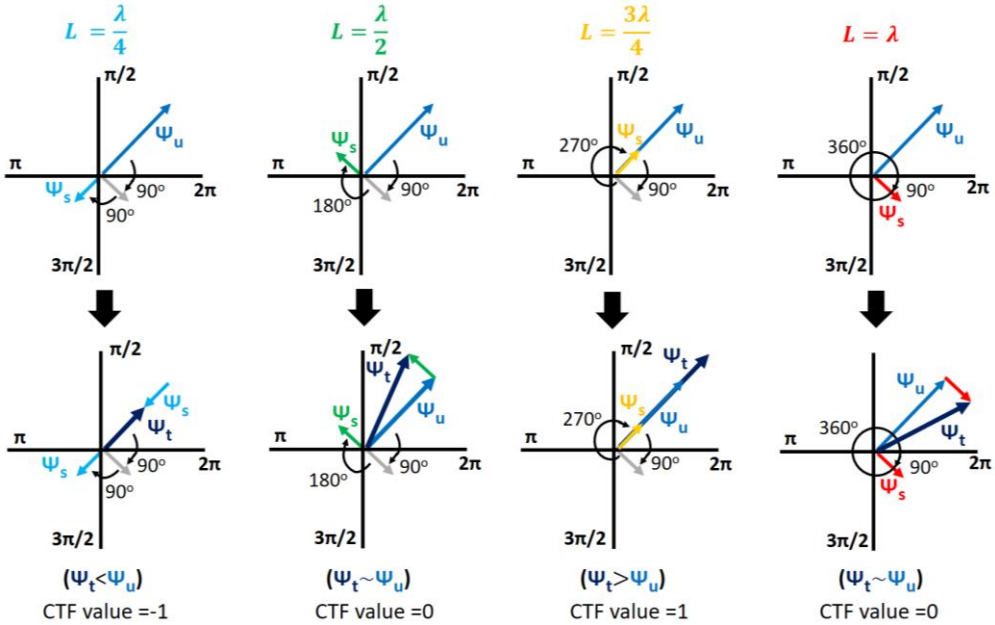


C)



$$CTF = Sin\ [\ -\pi\Delta f\lambda k^2 + (\pi C_s\lambda^3 k^4/2)]$$

D)



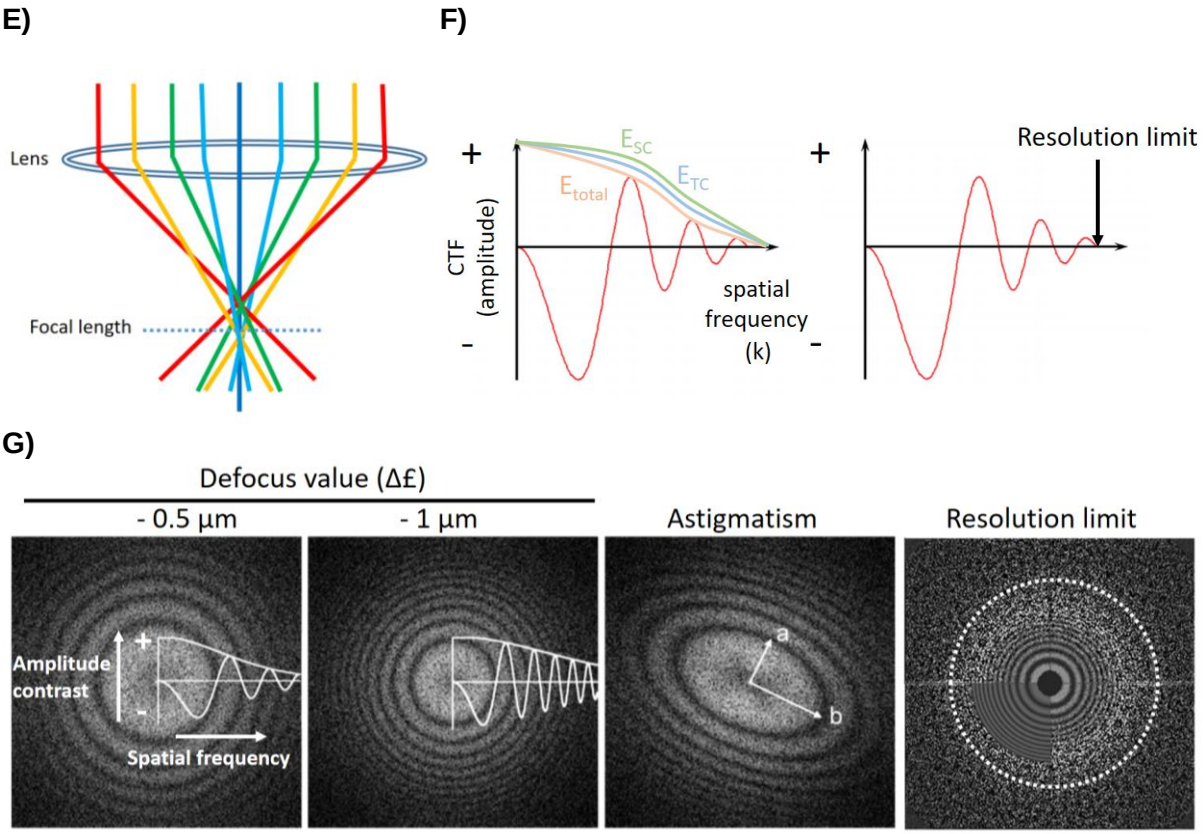


Figure 3.3 | Image formation and phase contrast in EM. A) Column anatomy of the microscope. The path of electrons from the source to the camera is focused by an extensive set of magnetic lenses, including condenser and projector lens. Areas enclosed within dotted lines are represented in more detail in **B**. **B) Different scenarios of elastic electron scattering** are represented, each one affecting differently to phase contrast formation in the image, and representing unique Fourier components of the sample. Scattered electron beams go through a longer path comparing to unscattered beams, creating an additional phase shift that is related to the additional distance (L) that electron beams travel in the column before reaching the image plane. L is represented in the figure relative to the electron wavelength (λ), which is the same in all cases (scattered and unscattered electron beams). Four different scattered electron beams are depicted, presenting an additional path of L = λ/4, L = λ/2, L = 3λ/4 and L = λ. Please note that high-frequency components in the sample such as atoms will scatter electrons at higher angles comparing to other components such as α-helices or protein domains. The physical process of scattering reveals the FT of the sample, which is represented at the back focal plane (BFP) of the microscope. In this regard, the CTF, as a function of spatial frequency, can be analysed at the BFP, where the unscattred beam in the centre represents a spatial frequency value of zero. The image is filtered or convoluted by the CTF, and a real space image is formed at the image plane. IL=intermediate lens. **C)** The **phase Contrast Transfer Function (CTF)** is represented as a function of spatial frequencies (*k*), and expressed in 1/Å units. Whereas a *k* value of zero can be considered for the unscattered beam, high-frequency components are located distantly in space, and present higher *k* values (please see Power Spectrum in **G**). In the equation we can see that the CTF function allows to calculate the defocus value (Δ£) of an image given a particular electron wavelength (λ) and a spherical aberration value (C_s). The represented function in the image corresponds to a theoretical (unaltered) CTF function, different to a typical experimental CTF as shown in **F**. **D) The corresponding Argand diagram for each elastically scattered electron beam** in **B** is plotted, representing electron beam phases and amplitudes. In addition to the 90º phase shift suffered by electrons after the scattering event (grey arrows), the additional path length (L) experienced by the scattered electrons creates an additional phase shift. The phase shifted scattered beam (ψ_s) will interfere at the image plane with the unscattered electron beam (ψ_u), generating a sum wave (ψ_{total}). In the case of $L = \lambda/2$ and $L = \lambda$, the resulting wave ψ_t is practically identical to ψ_u, and therefore the Fourier components of the sample that scattered the beam at those angles will not contribute to the image (zero contrast and therefore a CTF value of zero). **E) Spherical aberration (C_s).** Ray diagram of the spherical aberration that occurs when electrons with parallel paths are focused with the magnetic

lenses. **F)** A typical CTF plot indicating the resolution limit of the plotted function from where the amplitude value for the function is zero (no signal). Different to **C**, where an ideal CTF function is plotted, the analysis of the power spectrum on experimental images shows CTF functions with reduced amplitudes in the high-frequency range, and due to factors that can be defined as envelope functions of the CTF. Two envelope functions (E) are depicted to show their effect on the CTF, affecting high-frequency information and limiting resolution. Spatial coherence envelope (E_{SC}) and time coherence envelope (E_{TC}) functions combine together in this example to generate an overall envelope function (E_{total}) with a larger effect on the final CTF function. **G) Different Power Spectrum patterns.** The power spectrum of images collected at different defocus values are shown, including the CTF estimation for the images. In addition, the power spectrum of an image with astigmatism is also shown, as well as the information limit in the power spectrum which relates to the resolution information of the image, as defined by the spatial frequency *k*. (**A, E** and **G** are adapted from Orlova & Saibil, 2011, **B**-**D** and **F** from G. Jensen's EM course at Caltech).

Finally, electron beam coherence can be improved by adding an energy filter to the microscope, with the aim of imaging exclusively electrons presenting the same energy after going through the sample (Gubbens *et al.*, 2010).

4) **Imaging**. For image processing, photographic film was initially used to record images, and CCD cameras were incorporated later on, which allowed for automation and speed (Fig. 3.4). CCD cameras record electrons indirectly, so that the interaction of electrons with the scillintator component in the CCD camera generates photons, which are detected by the device. This system generates noise, mainly coming from the so-called back scattering effect (Fig. 3.4), i.e. one electron can potentially generate signal that will be recorded as two independent events.

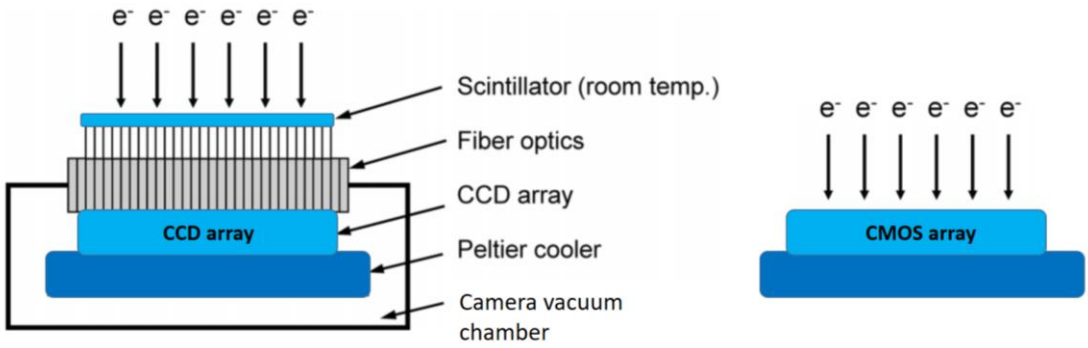
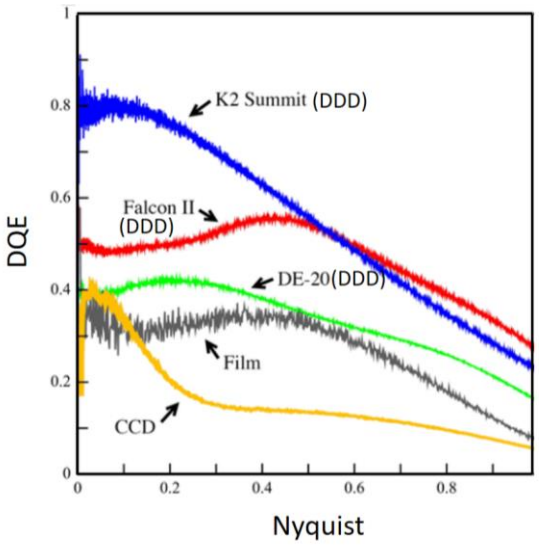


Figure 3.4 | CCD (charge-coupled device) versus CMOS (complementary metal-oxide semiconductor). On the left, a CCD camera is shown, including the fibre optics to guide photons generated in the scillintator. On the right, the CMOS camera is able to directly detect electrons as they hit the camera, reducing the level of noise in the images. (Adapted from Orlova and Saibil, 2011).

Even though CCD cameras allow automation in image processing, there is a limitation in resolution coming from this effect. For that reason, the film system still presents a higher detective quantum efficiency (DQE), providing higher resolution images (Fig. 3.5). The DQE describes the ability of the detector to record the incoming signal (SNR_{in}), providing information of efficiency at different resolutions or spatial

frequencies (SNR_{out}/SNR_{in}) (Clough, Moldovan and Kirkland, 2014) (equation in Fig. 3.5).



$$DQE(\omega) = \frac{(SNR_{out})^2}{(SNR_{in})^2}$$

Figure 3.5 | DQE curve for direct electron detector devices (DDD), CCD camera and film. The lowest detective quantum efficiency or DQE values correspond to the CCD camera, as shown in the chart. The equation defining the DQE parameter is also shown (Adapted from McMullan, Faruqi and Henderson, 2016).

With the development of direct electron detectors devices (DED or DDD), a revolution occurred in single-particle cryo-EM, also driven by the development of new processing software. DDD present a much higher DQE comparing to the previous systems (Fig. 3.5), generating higher resolution images.

In addition, the fast read-out of these devices allows to collect movies, consisting of several frames per second recorded over several seconds. Although movies were initially recorded to correct for mechanical drift over long exposures (Li *et al.*, 2013), a so-called "beam-induced particle-motion" effect was observed, that was generated after the interaction of the electron beam with the specimen. Therefore, recording movies instead of single frame images allows to partially correct for this effect (Campbell *et al.*, 2012; Li *et al.*, 2013; Scheres, 2014). Although computational methods have been proven effective in correcting particle translations over the different fames, the particle rotation effect can not be corrected "*in-silico*". In this regard, the development of new grid technology such as gold specimen support prevents motion during electron irradiation of the sample as previously mentioned (3.2) (Russo and Passmore, 2014) and results not only in an increase of resolution but also allows to solve challenging, small size protein particles.

Another advantage of recording movies instead of single-frames is related to the fact that the differential radiation damage effect in proteins also imposes another limitation

to resolution. Because high-resolution features are degraded more rapidly by the electron beam in comparison to medium and low resolution features, an electron dose weighting approach can be used over the different frames to generate a final single frame containing stronger high-resolution signal (Scheres, 2014; Zheng *et al.*, 2017).
In addition, the camera is able to record low and medium resolution information better than high-resolution information (reflected in the DQE value of the camera). Therefore, it also imposes an envelope function that limits resolution, and depends on the specific detection device. On the other hand, actual DDD devices only detect the 50% of electrons reaching the camera, limiting the achievable resolution (Eisenstein, 2016). Because the sample suffers from radiation damage after electron exposure, we are limited in the number of electrons that we can use to image our sample, as previously described. Hence, an increase in only a 20% of recorded electrons would represent a second revolution in cryo-EM and structural biology. It would have a very big impact not only in the single-particle cryo-EM field but also in tomography, given that one of the most challenging steps is to properly align the noisy and poor contrast micrographs obtained during the tilt series, and to accurately estimate the defocus value (CTF estimation).

5) **Image processing**. This step also represents another envelope function to the resolution. Our ability to better process the noisy projection images that we obtain from the microscope largely depends on the development of new processing algorithms. While in X-ray crystallography we need to solve a phase problem in order to solve the structure, in transmission cryo-electron microscopy (cryo-TEM) we face an orientation problem, so that the relative orientation of each particle projection needs to be assigned in order to compute a three-dimensional model of the particle. The experimental noisy images corresponding to particle projections obtained from the electron microscope need to be compared to each other, and the combination of typically tens of thousands of particle projections can generate nowadays near-atomic resolution 3D reconstructions with the available DDD devices in the market.
If we consider that the noise present in the image is random, by averaging many equivalent projections it is possible to increase the signal corresponding to the particles and reduce the signal coming from random noise. The first step after image processing is movie alignment, to correct for motion that takes place during electron beam exposure and data collection (**2**). A particle picking step follows on the aligned single frames or micrographs – either manually, in a semi-automated procedure or fully automated. Machine learning based particle-picking software has been shown to be very effective for automated particle picking including Warp (Tegunov and Cramer,

2019) and crYOLO (Wagner *et al.*, 2019). After particle extraction of the picked coordinates, an initial 2D classification step is typically performed, which consists in comparing and averaging equivalent projection images of single particles. Several rounds of 2D classification can clean “bad” particles, e.g. not corresponding to our protein sample but representing contaminants or noise. The generated 2D classes can be used for a new particle picking step, in an iterative process, and improve the quality of the particles by a reference based particle picking. The 2D classified particles are then used to generate a 3D reconstruction assessing the position of each projection relative to each other in the Fourier space, using the so-called back projection theorem (Orlova and Saibil, 2011). After the generation of an initial model, 3D classification can be used to sort out heterogeneity, which is essential to generate high-resolution maps from collected data. In the case of cryoSPARC, ab-initio classification generates at the same time a classified set of particles and a *“de novo”* reconstruction of the particle.

Several image processing software packages are available, but most of the high-resolution structures deposited in the EM data bank (EMDB) at near-atomic resolution have been solved by RELION so far. Based on a maximum-likelihood algorithm, it uses a Bayesian approach to automatically set the values of a dynamic 3D Wiener filter at every iteration (Scheres, 2016). CryoSPARC is a more recently released software, therefore the number of deposited structures in the EMDB is growing every day. The main advantages are its high-performance in terms of processing quality and also its processing speed. CryoSPARC processing is run by deep learning algorithms, where stochastic gradient descent (SGD) combined with Bayesian marginalization and branch-and-bound maximum likelihood optimization algorithms (Punjani *et al.*, 2017) are combined together, extensively increasing processing speed.

In addition, the ability of the software to correct magnification anisotropy and high-order aberrations (Zivanov *et al.*, 2019) from collected images allows to quantify and correct optical aberrations including trefoil (three-fold astigmatism) and coma, greatly improving resolution, and especially when reconstructions are in the high-resolution range (Bromberg *et al.*, 2019).

Therefore, our ability to reduce the computational envelop function and also to computationally reduce the effect of the other envelop functions or resolution limitations coming from the microscope by improving the processing software has been shown to play a critical role in cryo-EM, in addition to the increase in the processing speed.

In extreme cases, the processing of the same dataset can lead to an improvement of more than 2 Å resolution within the near-atomic resolution range when the detected

optical aberrations are corrected (Bromberg *et al.*, 2019). This difference represents a massive improvement in the building of atomic models and the biological interpretation. In addition, the implementation of on-the-fly processing systems is becoming mainstream, such as Warp for pre-processing and Live cryoSPARC for full processing of data, including 3D reconstructions. Therefore, full automation of datasets is becoming a reality, although for challenging and difficult cases still the manual intervention of an expert is needed for image processing.

6) **An ideal system with no envelop functions or resolution limitations**. Given the intrinsic flexibility of many protein domains and complex subunits, and because single-particle cryo-EM relies on averaging out multiple particle projections, even in the situation where our system is near to ideal with no envelops, we might still be limited in resolution. In this regard, the signal subtraction approach has been shown to be very effective, increasing resolution in cases where a region of the particle moves independently respect to the main density (Nakane *et al.*, 2018); however, this region still requires a minimum of mass in order to be properly aligned and processed independently. Nevertheless, still valuable information can be obtained from cryo-EM maps even in extreme cases where some regions of the protein are totally missing in the reconstruction due to flexibility, whereas in X-ray crystallography the formation of the crystal is a "*sine qua non*" or necessary condition acting as a bottleneck to obtain any structural information.

3.4 Single-particle cryo-EM analysis

3.4.1 Cryo-EM analysis of isolated full-length Tra1 from *S. cerevisiae*

3.4.1.1 Sample vitrification and data collection[3]

Before cryo-EM analysis, preliminary examination by negative staining was performed as previously described (3.1.2). Immediately after gel filtration, the concentrated Tra1 peak (2.3.3.4) at ~ 0.1 mg/mL was applied to negatively glow discharged Lacey grids coated with an ultrathin layer of carbon (~ 3 nm thick) (Agar Scientific, UK) and vitrified in liquid ethane using a Vitrobot Mark IV system (Thermo Fisher Scientific) set to 4 °C and 95% humidity. Due to protein solubility issues the protein was used at low concentration, so that the sample was applied twice to the grid in order to optimise the density of particles. Each application of 3 µl was set to 20 s waiting time, followed by a short 0.5 s blotting after the first application and 5 s blotting after the second.

A first test dataset of 40 movies generated a ~9 Å map. Cryo-EM data collection was performed in-house at Birkbeck College (University of London) on a 300 kV FEI Tecnai G^2 F30 Polara microscope (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector and GIF Quantum energy filter (slit width 20 eV) (Gatan, Inc.). The images were collected in counting mode at a nominal magnification of 125,000x, corresponding to 1.39 Å per pixel. The electron dose rate on the specimen was set to ~ 3 electrons per $Å^2$ per s with a total dose of ~ 46 e/$Å^2$ fractionated over 50 frames. The dataset was collected at a nominal defocus range of −1.5 µm to −4 µm. Data collection parameters are shown in Table 4.1 (4.2).

A second dataset generated a high-resolution cryo-EM map at 3.7 Å. For this second dataset, data collection was performed at eBIC-DLS (Electron Bio-Imaging Centre; Diamond Light Source, Didcot) on a 300 kV FEI Titan Krios microscope (Thermo Fisher Sceintific) equipped with a GIF Quantum energy filter (slit width 20 eV). The images were collected on a K2 Summit direct electron detector (Gatan) operated in counting mode at a nominal magnification of 130,000x, corresponding to 1.06 Å per pixel. The electron dose rate on the specimen was set to 5.5 electrons per $Å^2$ per s with a total dose of ~44 e/$Å^2$ fractionated over 32 frames. The dataset was collected at a nominal defocus range of −1.5 µm to −3.5 µm using EPU software (FEI). Because Lacey grids are irregular, data collection was performed in semi-automated mode, consisting in the manual selection at low magnification of suitable areas for data

[3] The vitrification of the samples presented in this study was carried out by Natasha Lukoyanova, as well as the in-house data collection of cryo-EM data on a direct detector device (Birkbeck College, University of London). Alistair Siebert performed the cryo-EM data collection at Diamond Light Source (DLS) (Didcot, Oxford) of the large Tra1 dataset.

collection presenting thin carbon. A total of ~ 1,600 areas were manually selected, and additional areas were added later on during the session. Data collection parameters are shown in Table 4.2 (4.2). A total of 2,449 movies were collected.

3.4.1.2 Data processing

The in-house dataset collected on the Polara microscope and consisting of 40 movies was pre-processed in MotionCor (v1) (Li *et al.*, 2013) using a 5x5 patch to generate the aligned single frame micrographs. The coordinates generated during particle picking in Gautomatch (Urnavicius *et al.*, 2015) using a built-in Gaussian reference were imported into RELION 1.4 (Scheres, 2012) for reference-free 2D classification. Contrast transfer function (CTF) was estimated using CTFFIND3 (Mindell and Grigorieff, 2003). The best 2D classes were used iteratively as reference for particle picking instead of the firstly used Gaussian reference, until no further improvement was observed. Secondary structure elements were already visible in the 2D classes (4.2, Fig. 4.11C, D) and to avoid the "Einstein from noise" effect (Henderson, 2013), the selected 2D classes were low-pass filtered to 20 Å prior to be used as references for particle picking in Gautomatch. Particles contributing to the best 2D classes (7,680 particles in total) were selected for 3D auto-refinement in RELION. An existing 13 Å EM map of Tra1 human paralog DNA-PKcs (EMD-1102) was low-pass filtered to 40 Å resolution and used as reference volume for the 3D auto-refinement. The resulting consensus model refined to 9.16 Å resolution. All reported resolutions in the present work use gold standard Fourier Shell Correlation (FSC) curves at 0.143 criterion (Scheres and Chen, 2012), calculated from two models that are refined independently, each one using half of the data. The map presented visible secondary structure elements, especially in the alpha-solenoid region, and allowed to identify the Head region containing the kinase domain (4.11C).

The 2,449 movies collected at DLS on a Titan Krios were initially aligned in MotionCor (v1) as previously described. Because the irregular nature of the Lacey grids (4.2, Fig. 4.11B and 4.12A), the aligned single frame micrographs were manually screened and images containing mostly thick carbon support were discarded, resulting in a final manually curated dataset consisting of 1,734 micrographs. Particle picking was performed in Gautomatch, and particle coordinates were thoroughly screened by manually removing particle coordinates present in thick carbon areas. Particle coordinates were imported into RELION 2.0 (Kimanius *et al.*, 2016) for reference-free 2D classification, and the best 2D classes were low-pass filtered and used as reference for a new round of particle picking in an iterative process, as previously described. CTF was estimated using CTFFIND4 (Rohou and Grigorieff, 2015). 418,339 particles contributing to the best 2D classes were selected for 3D auto-refinement in RELION using the previously obtained ~ 9 Å Tra1 map, low-pass filtered to 40 Å resolution before to

be used as reference volume. The consensus map was utilised as reference volume for 3D classification using four different classes, providing one single high-resolution class reconstructed from 182,285 particles that refined to 3.9 Å resolution. Sub-classification of the best 3D class did not provide any further improvement in terms of resolution and/or map quality as assed by visual inspection in Chimera. Micrographs corrected with a newly released dose-weighting feature improved the map resolution to 3.7 Å (4.2, Fig. 4.12B) when the particles contributing to the final reconstruction were substituted by particles processed in MotionCor2 (Zheng *et al.*, 2017). The map was automatically sharpened during post-processing in RELION applying a negative B-factor, and the local resolution map (4.2, Fig. 4.13D) was calculated using a built-in program available in the RELION 2.0 package (β-version) as a testing feature and implemented later on in following versions.

3.4.2 Cryo-EM analysis of isolated Tra1 incubated with Gal4 TAD

3.4.2.1 Sample vitrification and data collection

The Tra1 sample (~ 0.1 mg/mL) was purified as previously described (2.3.3) and mixed with Gal4 TAD (2.2.2) at 1:10 molar excess of activator. The sample was vitrified as in 3.4.1.1.

A cryo-EM dataset of 94 movies was collected in-house at Birkbeck College on a 300 kV FEI Tecnai G^2 F30 Polara microscope (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector and GIF Quantum energy filter (slit width 20 eV) (Gatan, Inc.). The images were collected in counting mode at a nominal magnification of 125,000x, corresponding to 1.39 Å per pixel. The electron dose rate on the specimen was set to 2.8 electrons per $Å^2$ per s with a total dose of 50.4 e/$Å^2$ fractionated over 50 frames. The dataset was collected at a nominal defocus range of −1.8 µm to −3.5 µm. Data collection parameters are shown in Table 5.1 (5.4).

3.4.2.2 Data processing

The 94 movies were processed as described before for isolated Tra1 (3.4.1.2). 16,733 particles contributing to the best 2D classes generated a consensus map at ~ 7 Å resolution (Fig. 5.5), that was used as reference volume for 3D classification using four different classes. Two 3D classes showed defined Tra1 features and refined independently to 8.34 and ~ 9 Å resolution, and the two classes together generated a map from 13,966 particles that refined to 7.0 Å resolution.

3.4.3 Cryo-EM analysis of the Mec1^{ATR}-Ddc2^{ATRIP} complex

3.4.3.1 Sample vitrification and data collection

Before cryo-EM analysis, preliminary inspection by negative staining was performed as previously described (3.1.1). For the first in-house test dataset (6.3.3), the Mec-Ddc2 peak (<0.1 mg/mL) from the Superose 6 Increase 10/300 GL column was applied to negatively glow discharged Lacey grids coated with an ultrathin layer of carbon (~ 3 nm thick) (Agar Scientific, UK). The grids were vitrified in liquid ethane using a Vitrobot Mark IV system (Thermo Fisher Scientific) set to 4 °C and 95% humidity. Because of the protein concentration, the sample was applied five times to the grid (3 µl each time) in order to optimise particle density. Each application was set to 30 s of waiting time, followed by a short 0.5 s blotting step after each application, and a last blotting step of 7 s.

For the high-resolution Mec1-Ddc2 cryo-EM maps (6.5 and 6.6), ~ 0.9 mg of affinity purified and BS3 crosslinked complex (2.3.4.2 and 6.4) were injected into a KX 16/70 gel filtration column. The peak corresponding to the complex was concentrated to ~ 140 µl (2.3.4.3 and 6.5) and 3 µl of sample were applied to UltraAuFoil R 1.2/1.3 Gold 300 mesh grids (Quantifoil) coated with a thin layer of graphene oxide (GO) using a publicly available protocol (LMB, Cambridge; http://dx.doi.org/10.6084/m9.figshare.3178669). The sample was applied once to the grid and after 30 s of waiting time, and 0.5 s of blotting, the high-salt concentration of the sample was washed on the grid applying a second 3 µl drop of "gel filtration buffer" (2.0.5, Table 16) to the grid before a second 7.5 s blotting step. Then, the sample was plunge-frozen into liquid ethane.

A first test dataset of 185 movies generated an initial map at ~ 7 Å. Cryo-EM data collection was performed in-house at Birkbeck College on a 300 kV FEI Tecnai G^2 F30 Polara microscope (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector and GIF Quantum energy filter (slit width 20 eV) (Gatan, Inc.). The images were collected on Lacey grids in counting mode at a nominal magnification of 125,000x, corresponding to 1.39 Å per pixel. The electron dose rate on the specimen was set to 2.95 electrons per Å^2 per s with a total dose of 44.25 e/Å^2 fractionated over 50 frames. The dataset was collected at a nominal defocus range of −1.5 µm to −4 µm. Data collection parameters are shown in Table 6.1 (6.3.3).

A second, larger dataset of 1,500 movies resulted in a final cryo-EM map at ~ 4.1 Å resolution. Data collection was performed in-house on the 300 kV F30 Polara microscope using a K2 camera and energy filter as previously. The images obtained from UltraAuFoil GO grids were collected in counting mode at a nominal magnification of 125,000x, corresponding to 1.39 Å per pixel. The electron dose rate on the specimen was set to 3.3 electrons per Å^2 per s with a

total dose of 49.5 e/Å^2 fractionated over 50 frames. The dataset was collected at a nominal defocus range of −1.5 µm to −4 µm. Data collection parameters are shown in Table 6.2 (6.5.1).

The highest resolution structure consisted in a final cryo-EM map at 3.3 Å resolution. Cryo-EM data collection was performed in-house at Birkbeck College, this time on a 300 kV FEI Titan Krios microscope (Thermo Fisher Sceintific) equipped with a GIF Quantum energy filter (slit width 20 eV). The images from UltraAuFoil GO grids were collected on a K2 Summit direct electron detector (Gatan) operated in counting mode at a nominal magnification of 165,000x, and corresponding to 0.85 Å per pixel. The electron dose rate on the specimen was set to 3.4 electrons per Å^2 per s with a total dose of ~60 e/Å^2 fractionated over 50 frames. The dataset was collected at a nominal defocus range of −1 µm to −4 µm using EPU software (FEI). Data collection parameters are shown in Table 6.3 (6.6).

3.4.3.2 Data processing

Single frame micrographs were generated in MotionCor2 from the alignment of 185 movies that were collected in-house on the Polara microscope using Lacey grids. The coordinates generated during particle picking in Gautomatch (Urnavicius *et al.*, 2015) and using a built-in Gaussian reference, were imported into RELION 2.1 (Kimanius *et al.*, 2016) for reference-free 2D classification. CTF was estimated using CTFFIND4. 2D classifications ignoring the CTF until the first peak generated good results (6.3.3, Fig. 6.5B). The best 2D classes were used iteratively for new rounds of particle picking and 2D classification as previously described (3.4.1.2). A total of 8,511 particles contributing to the best 2D classes were imported into cryoSPARC (v1) (Punjani *et al.*, 2017) to perform ab-initio reconstruction and generate a *"de novo"* 3D reconstruction. The 3D map was low-pass filtered to 40 Å resolution and used as reference volume for 3D auto-refinement in RELION. The resulting consensus model at ~7 Å resolution (FSC=0.143 threshold) presented clear secondary structure elements, revealing the unique alpha-solenoid architecture present in the complex (6.3.3, Fig. 6.5 and 6.6). A 3D classification with three different classes provided the best quality 3D class containing 4,353 particles, that refined to 7.2 Å resolution (FSC=0.143).

A dataset collected on UltraAuFoil GO grids was obtained from the 300 kV Polara microscope. Movies were aligned in MotionCor2 as previously described. Because some areas presented bent GO or were support free areas, the aligned single frame micrographs were manually screened and images containing nicely distributed GO were selected, resulting in a final dataset of 891 micrographs. Particle picking was performed in Gautomatch as previously described, using initially a built-in Gaussian reference. CTF of the aligned images was estimated with CTFFIND4. Particle coordinates were imported into RELION 2.1, and particles were extracted from the micrographs. The extracted particles were imported into cryoSPARC

(v1) for 2D classification. The best 2D classes were used iteratively for new rounds of particle picking and 2D classification as previously described (3.4.1.2). A total of 140,990 particles contributing to the best 2D classes were selected for ab-initio 3D classification in cryoSPARC. The best result was obtained using four different classes, and 73,163 particles contributing to the best ab-initio map were imported into RELION for a final 3D auto-refinement. The resulting 4.1 Å resolution reconstruction (FSC=0.143 criterion) presented visible side chains in many regions of the map, revealing the complex architecture at near-atomic level (6.5.2, Fig. 6.11). Map sharpening was performed automatically in RELION during post-processing applying a negative B-factor. The map was locally filtered using a built-in program in the RELION package that applied a spherical mask to the different regions of the map to filter each region to its estimated local resolution, that was also calculated by the previously mentioned built-in feature in RELION (6.5.1, Fig. 6.10).

Movies collected on the 300kV Titan Krios from UltraAuFoil GO grids were aligned in cryoSPARC (v2) using a built-in movie alignment program, as well as the subsequent processing steps. After CTF estimation and particle picking utilising a Gaussian reference, 2D classification was performed in cryoSPARC. Secondary structure elements were clearly visible in the high-resolution 2D classes (6.5.1, Fig. 6.9). The best 2D classes were selected, filtered to 25 Å using a built-in feature in cryoSPARC for particle picking and used as 2D references for new rounds of particle picking in an iterative process until no better 2D classes were obtained. 191,211 particles contributing to the best 2D classes were selected for ab-initio classification, and the best result was obtained using four different classes. The best ab-initio class (containing 107,156 particles) was refined to 3.5 Å applying C1 symmetry and 3.3 Å applying C2 symmetry to the reconstruction (FSC=0.143). The sub-classification of this 3D class using two different classes resulted in a better quality map, generated from 92,242 particles. Although the resolution estimation was similar, visual inspection in Chimera showed better resolved regions. Signal subtraction and focus refinement on the FATKIN domain and also on the Mec1 N-terminal region together with Ddc2 did not improve the map on those areas, suggesting that the relative mobility between these protein regions is restricted, as shown in the local resolution map (6.6.1, Fig. 6.15). Map sharpening was performed automatically in cryoSPARC after the 3D refinement, applying a negative B-factor. The local filtering built-in feature was very useful for the manual model building process performed in COOT (3.4.4.2), facilitating the interpretation of the map in the lower resolution regions. Local resolution estimations of the maps were also calculated in cryoSPARC.

3.4.4 Atomic model building

3.4.4.1 Tra1 atomic model building[4]

The final high-resolution map at 3.7 Å presented clear density for β-strands, α-helices and loops, which allowed to build *"de novo"* 3,474 of 3,744 residues (4.3.1, Fig. 4.15). COOT software (Emsley *et al.*, 2010) was used for manual building, which was assisted by the secondary structure prediction servers PSIPRED (Buchan and Jones, 2019) and JPRED (Cole, Barber and Barton, 2008), and from existing Tra1 secondary structure analysis works (Knutson and Hahn, 2011). The high content in α-helices facilitated the assignment of sequence register, and given that the Kinase domain is highly conserved within the family, existing crystal structures of mTOR kinase domain (Yang *et al.*, 2013) initially guided the building of this region. Sharpening in Phenix (DiMaio *et al.*, 2013) utilising different B-factors and the low-pass filtering of the map facilitated the modeling process. The manually built model was finally refined in *phenix.real_space_refine* using secondary structure restraints. For the validation of the atomic model, the extended and generally used cross-validation method in cryo-EM was employed to monitor overfitting during refinement, as previously described (Amunts *et al.*, 2014; Fernández *et al.*, 2014). The method consists in the calculation of three different model versus map FSC curves. The first FSC calculation is performed between the refined model and the final map (FSC_{total} or FSC_{final}). Next, the position of each atom within the atomic model is randomly displaced by up to 0.5 Å, to remove model bias from the prior refinement that used the - all particles - final map. Then, the model is refined within one of the half-maps, and used for FSC calculation with the mentioned refinement half-map (FSC_{work}). Finally, the resulting refined model is directly used without any additional refinement step for FSC calculation with the other, refinement-free half-map (FSC_{test} or FSC_{free}). The half-maps used in the validation process for model versus map FSC calculation are the same half-maps utilised during the gold standard FSC calculation for resolution estimation of the - all particles - final map. Similarity between the three different FSC curves, i.e. $FSC_{total/final}$, FSC_{work} and $FSC_{test/free}$, indicates that no significant overfitting is present. For the validation of Tra1 atomic model, the resolution limit was set to 3.7 Å in all refinements, and the comparison of the three FSC curves indicates that no significant overfitting was present during the refinement (Fig. 4.13F; Díaz-Santín *et al.*, 2017). In addition, the validation of the atomic model was completed using MolProbity structure validation software (Chen *et al.*, 2015), and the validation report is publicly available (PDB 5ojs). Validation curve and statistics are also provided in Figure 4.13 and Table 4.2.

[4] The atomic model building of Tra1 polypeptide was carried out by Alan C.M. Cheung.

3.4.4.2 Mec1-Ddc2 atomic model building

An initial atomic model of the *S. cerevisiae* Mec1-Ddc2 complex was built using a previously published model as starting point for model building (PDB 5x6o; Fig. 6.12). The first high-resolution map of Mec1-Ddc2 at 4.1 Å was used to build the complete atomic model of Mec1, and the building process was assisted by the secondary structure prediction severer PSIPRED (Fig. 6.13) as previously described (3.4.4.1), in addition to an extensive amount of available literature (6.5.2). Discrepancy between the atomic models generated in this study and the publicly available (PDB 5x6o) atomic models was found in both Mec1 and Ddc2 polypeptides. In Mec1, a long 235-reside region at the N-terminal was built *"de novo"* in COOT as before (3.4.4.1). Due to flexibility, the disordered linker connecting Ddc2 coiled-coil region to Ddc2 Solenoid domain made the coiled-coil domain invisible by cryo-EM (Fig. 6.11C), so that model building of the Ddc2 polypeptide started from residue Asn189 (6.6.3, Fig. 6.19). Therefore, the first visible residues at the N-terminal region of Ddc2 structured linker were built *"de novo",* revealing the interaction of this N-terminal region with a hydrophobic pocket present in Mec1 (Fig. 6.18 and 6.19).

Using the 3.3 Å cryo-EM map, a more complete atomic model was built with many more side chains and also loops that were missing in previous models of Mec1 and Ddc2. The Kinase region largely benefited from this improvement in resolution (6.6), where the density for the activation loop was fully present and many more side chains could be built in the active site. The locally filtered map in cryoSPARC (v2) facilitated the modelling task in COOT, filtering out the noisy signal in the mobile regions located at the periphery of the complex (Fig. 6.15A). The manually built atomic model was finally refined in *phenix.real_space_refine* using secondary structure restraints as previously described (3.4.4.1), and validation statistics are provided (Table 6.3). In addition, the validation of the generated atomic models was also performed as before, calculating model versus map FSC curves (Fig. 6.15D). The analysis of the three resulting FSC curves indicates that no significant overfitting was present during the refinement.

3.4.5 Maps and models interpretation and analysis

UCSF Chimera (Goddard *et al.*, 2018) was used to analysed and interpret cryo-EM maps and atomic models, together with COOT, and Chimera was used to create figures. The comparison of atomic models was carried out using the MatchMaker tool in Chimera. Sequence alignment analysis was performed in Clustal Omega (Sievers and Higgins, 2018) and rendered in Jalview (Waterhouse *et al.*, 2009) and ESPrit3 (Gouet, Robert and Courcelle, 2003). Sequence alignment was plotted into the corresponding atomic structures using the AL2CO program (Pei and Grishin, 2001) implemented in Chimera. The Consurf server (Ashkenazy *et al.*, 2016) was also used to render the evolutionary analysis of conserved residues in the atomic structure.

4 Single-particle cryo-EM analysis of full-length *S. cerevisiae* Tra1

4.1 Tra1 expression and purification

4.1.1 Bacterial expression of Tra1 super-secondary structure elements

Because of its very large size of 3,744 residues, structural and biochemical information on Tra1 was extremely limited. Expression in bacteria as full-length protein was likely to be intractable. For that reason, the initial approach to get structural knowledge from Tra1 was to design bacterial expression constructs to produce Tra1 fragments in bacteria. Cloning of Tra1 regions was guided by secondary structure prediction to prevent truncation within a HEAT repeat (Knutson and Hahn, 2011). Complete coverage of the Tra1 sequence (YHR099W) was represented in 17 overlapping ORFs, each one cloned and expressed individually in *E. coli* (2.0.1, Table 2) (Fig. 4.1).

The 17 Tra1 DNA fragments were amplified by PCR (2.1.1) using LMDS5 and LMDS27-LMDS53 primers (2.0.1, Table 4) from plasmid DNA containing native *S. cerevisiae* Tra1 DNA sequence (2.0.1, Table 2, FL Tra1 (plasmid no 17) and cloned into pET-28a(+) vector containing a 6xHis tag for protein purification (Millipore, 2.0.1, Table 2, “cloning plasmids”). The vector backbone was linearised with primers AC132 and 133. PCR products were run in a DNA gel electrophoresis (2.1.2) and the bands corresponding to the amplified fragments were gel purified (2.1.3). A two fragments GIA reaction was performed (2.1.4) to clone each amplified Tra1 DNA fragment into the pET-28a(+) vector containing a 6xHis tag. The reaction product for each construct was transformed into *E. coli* chemically competent cells (2.1.5 and 2.1.6), and transformant colonies were screening by colony PCR (2.1.7) using the T7 fwd and T7 rev sequencing primers (2.0.1, Table 4). Colonies that generated PCR products of the expected size were selected for miniprep (2.1.8) to perform plasmid DNA extraction. The purified material was sent for Sanger sequencing (GATC company, now Eurofins Genomics) (2.1.9), and mostly performed using primers T7-fwd and pET-RP (2.0.1, Table 4). The expressed fragments were designed to be 15 to 72 kDa size for efficient expression in *E. coli*, and were designed according to Tra1 domain organisation as previously mentioned, including HEAT repeats, TPR repeats, the Kinase domain with and without the FRB motif, and the C-terminal lobe of the kinase domain (Fig. 4.1). From previous expression experiences on other proteins, *E. coli* strain Rosetta2 (2.0.2, Table 5) was initially selected for testing (2.2.1) but the best protein expression results were obtained with *E. coli* strain BLR (Fig. 4.2). However, the expressed Tra1 fragments were generally insoluble in virtually all cases, so that the strain Lemo21 (2.0.2, Table 5) was also tested in the search of soluble expression, with negative results (2.2.1). For that reason, a new library of Tra1 fragments was cloned with a thioredoxin (Trx) solubility tag. To do so, the vectors containing the Tra1 DNA fragments (2.0.1, Table 2)

were individually amplified from the 3C cleavage site with the reverse primer LMDS70 and from the stop codon and 3' UTR region with the forward primer LMDS69 (2.0.1, Table 4, "cloning primers"). The thioredoxin solubility tag was amplified with primers LMDS67/68 from plasmid p3NH TrxA GFP (kindly provided by the Waksman lab, 2.0.1, Table 2). A two fragments GIA reaction (2.1.4) was performed to assemble the amplified DNA sequences, and subsequent cloning steps including DNA sequencing were carried out as previously described (2.1). After the test expression (2.2.1) of the 17 newly cloned expression plasmids (2.0.1, Table 2) containing the thioredoxin solubility tag, only one construct was found to express soluble protein, present in the supernatant (Fig. 4.2, TrxA-H8-H9, lanes 4 and 6). This construct corresponded to a small solenoid region (HEAT repeats 8-9), but after thioredoxin tag cleavage by 3C protease the mentioned Tra1 fragment was shown insoluble (Fig. 4.2, TrxA-H8-H9, lane 8). Alpha solenoids typically contain a hydrophobic core running along their axis, hence the expression of short solenoid segments is likely to expose the hydrophobic core to the solvent, and therefore causing precipitation. Also, it is very likely that the folding of the large alpha solenoid found in PIKKs needs to be assisted by a specific chaperone system absent in *E. coli*, as for example the RT2P-TTT complex, with a previously described role in PIKK stabilisation (Kakihara and Houry, 2012; von Morgen, Hořejší and Macurek, 2015).

Given the limitations of *E. coli* as expression system in terms of maximal protein size to be efficiently expressed and the available chaperone system for the proper folding of eukaryotic proteins, *S. cerevisiae* was chosen as the new expression system, which is the original source organism of the selected Tra1 sequence for expression. The full-length version of Tra1 is unlikely to be amenable to crystallography, as can only be purified in small quantities from recombinant expression within yeast. However, because of its very large size – it is the third biggest proteins in *S. cerevisiae* genome and the biggest PIKK -, it represents an excellent target for study by single-particle cryo-electron microscopy (cryo-EM).

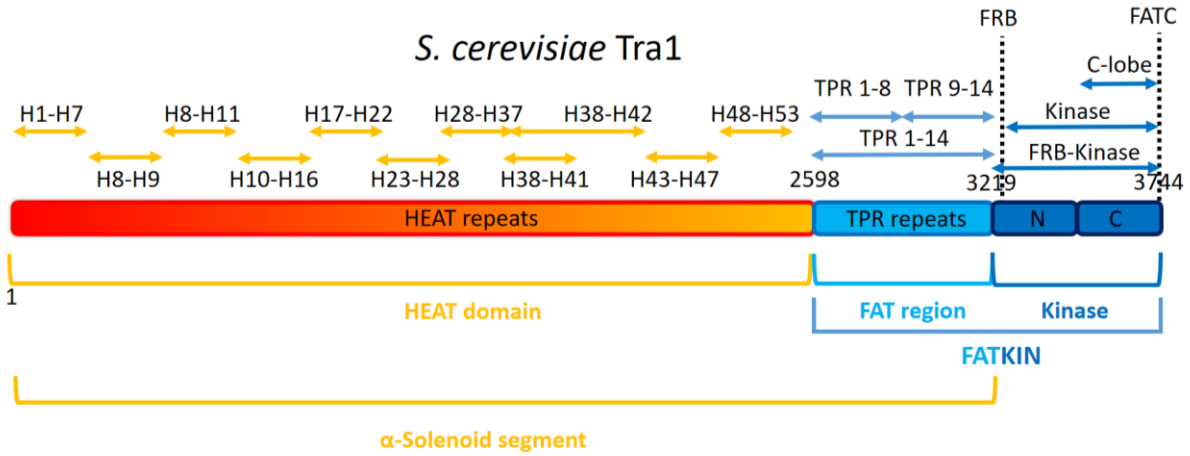


Figure 4.1 | Bacterial expression of *S.c.* Tra1 super-secondary structure elements. The arrows represent DNA fragments covering all super-secondary structure elements present in *S.c.* Tra1. Each

fragment was independently cloned into pET-28a(+) for test expression (2.2.1). A previously published secondary structure analysis of Tra1 (Knutson and Hahn, 2011) was used to design the seventeen DNA constructs covering the full-length of the Tra1 polypeptide (2.0.1, Table 2). The elements present in each construct are also described (H: Heat repeat). The position for the FRB and FATC motifs is also depicted. In addition to the fragments shown in the figure, an additional set of five constructs were cloned (2.0.1, Table 2, pET-28a(+)-Tra1-18 to pET-28a(+)-Tra1-22), coding for a different combination of Tra1 structural elements.

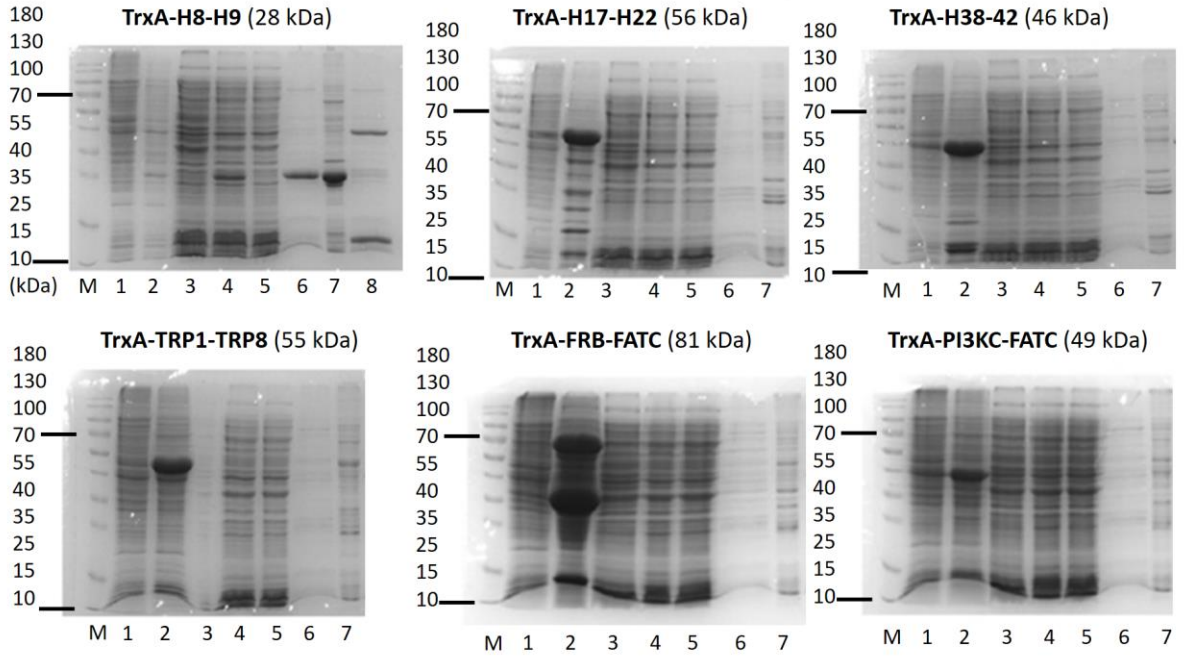


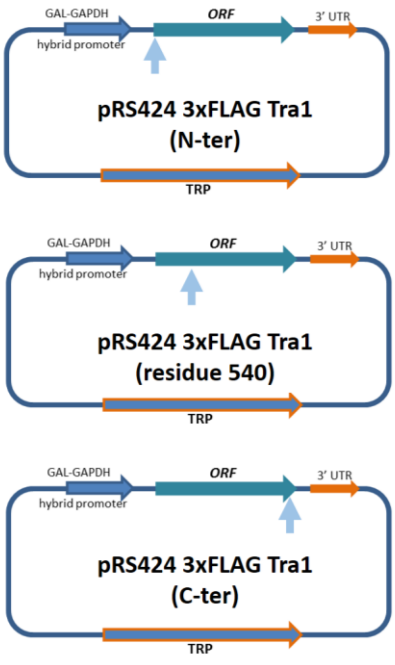
Figure 4.2 | Expression of *S.c.* Tra1 super-secondary structure elements in *E. coli*. 12% SDS-PAGE analysis (2.4.2) and Coomassie Brilliant Blue staining (2.4.3) of typical test expressions of the Tra1 fragments shown in Fig. 4.1, fused to thioredoxin (2.0.1, Table 2). For test expression (2.2.1), *E. coli* BLR cells were transformed. Magnetic nickel beads were used for 6xHis tag purification (MagneHis, Promega). Typical expressions consisted in insoluble protein products, as the protein can only be found in the pellet fraction (lane 2). Only one single construct expressed soluble protein in the tested conditions (TrxA-H8-H9), and after 3C cleavage the product was shown to be insoluble (the H8-H9 fragment is missing in the supernantant, lane 8). TrxA: thioredoxin. M: molecular marker (protein ladder). 1: pellet (w/o IPTG), 2: pellet (IPTG induction), 3: supernatant (w/o IPTG), 4: supernatant (IPTG induction), 5: flowthrough (IPTG induction), 6: elution, 7: beads after elution, 8: supernatant after 3C cleavage. (Please note that the size of the TrxA solubility tag is ~ 12 kDa and ~ 58 kDa in the case of the GST-3C protease (TrxA-H8-H9, lane 8)).

4.1.2 Tra1 yeast constructs and test expressions

Yeast expression and purification was achieved cloning native Tra1 DNA sequence coding for the full-length Tra1 protein into the pRS424-DQ expression vector with a 3xFLAG tag, and under the control of a galactose-inducible GAL-GAPDH hybrid promoter. The expression system (2.3.2.1), including the plasmids (2.0.1, Table 3) and the yeast expression strain BCY123 (2.0.2, Table 6) was a gift from Kiyoshi Nagai (MRC LMB Cambridge). To create the expression plasmid pRS424-DQ (2.0.1, Table 3), the GAL-GAPDH hybrid promoter and 5'/3' UTR regions present in the pENT-DQ plasmid were amplified and cloned into the pRS424 vector (2.3.2.1) by GIA as previously described. A GFP polypeptide was cloned into the pRS424-DQ expression plasmid with an N-terminal 3xFLAG tag for test expression (2.3.2.2),

resulting in high-expression levels of GFP. Three different versions of the pRS424-DQ Tra1 expression plasmid were designed placing 3xFLAG tag at the N-terminal, C-terminal or within an internal position (at residue 540) predicted to be disordered by the secondary structure prediction server PSIPRED (Buchan and Jones, 2019) (2.0.1, Table 3) (Fig. 4.3). The N-ter tag vector was made amplifying FL Tra1 from “FL-Tra1” plasmid (2.0.1, Table 4, “cloning plasmids”) using primers LMDS101 and LMDS102, and the vector pRS424-DQ was linearised by the reverse primer LMDS100 (containing a 3xFLAG tag) and the forward primer AC267 (2.0.1, Table 4, “cloning primers”) to amplify from the codon stop. To assemble vector and insert, GIA was performed as previously described (2.1.4), including posterior cloning steps (2.1). Colony PCR was used to detect positive transformant colonies using primers T7-fwd and M13-rev (2.0.1, Table 4). After plasmid DNA miniprep, Sanger sequencing was performed using primers T7 (fwd), M13-rev, LMDS114, LMDS166-169 and AC15-28 (2.0.1, Table 4). Although initially the yeast strain CB010 was used for plasmid transformation, it was replaced by BCY123 (2.0.2, Table 6) to use the complete K. Nagai’s expression system. To transform the plasmids into yeast, the Jansen group protocol was initially used (2.3.1.1) and it was replaced by the Gietz lab protocol (2.3.1.2) for simplicity reasons. After transformation into the yeast strain BCY123, plasmid selection was achieved via a tryptophan marker. To test for protein expression, a transformant colony was selected to inoculate 50 ml of yeast expression medium, following the protocol in 2.3.2.2. Yeast cell lysis was followed by TCA precipitation (2.4.4) and Western blot analysis (2.4.5) of the yeast cell material, confirming the proper expression of each individual construct (Fig. 4.3).

A) **B)**



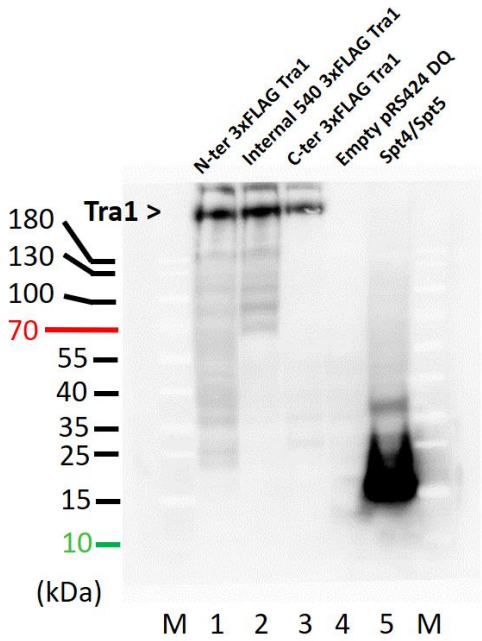


Figure 4.3 | Tra1 expression constructs and test expression analysed by anti-FLAG antibody. A) Tra1 expression vector architecture for the N-terminal (**B**, lane 1), internal (**B**, lane 2) and C-terminal (**B**, lane 3) 3xFLAG tagged Tra1 constructs. **B)** Western blot analysis showing clear protein expression for each construct in **A** (lanes 1-3). 50 ml of yeast culture (CSM-TRP medium) were used for test expression (2.3.2.2) of the three different constructs (2.0.1, Table 3, "expression vectors"). Induction was carried out with 2% of galactose overnight and 3x OD_{600} units of cell culture were used for yeast cell lysis followed by TCA precipitation (2.4.4). For analysis, 1.2x OD_{600} units were run in a SDS-PAGE (2.4.2) and Tra1 protein was detected by Western blot analysis (2.4.5). As positive control for the Western blot, a previously tested construct expressing the 3xFlag-Spt4/Spt5 complex was used (kindly provided by Alan C.M. Cheung, lane 5), and an empty pRS426-DQ vector as negative control (lane 4) (2.0.1, Table 3, "cloning plasmids"). Anti-Flag antibody M2 was used at 1:1000 dilution for 30 min at RT and anti-mouse HRP conjugated antibody (2.0.6, Table 17) at 1:2000 dilution for 1 hr previous to membrane development by the addition of ADHP developing agent (2.0.6, Table 17) (2.4.5). For imaging, a Fujitsu FLA3000 fluorescence scanner was used at sensitivity level F10. M: molecular marker (protein ladder).

4.1.3 Large scale expression and purification

After confirmation of Tra1 expression (Fig. 4.3), a purification was carried out using 12 L of yeast culture expressing N-terminally tagged 3xFLAG Tra1 protein, as previously described (4.1.2). After protein expression (2.3.3.1) by overnight induction using 2% of galactose in the medium, cells were harvested (2.3.2.4) and yeast cell lysis was performed in a Freezer Mill unit (2.3.2.5). Then, the lysate product was frozen at -80 ^{o}C for further processing. For affinity purification, the material was thawed and centrifuged, and the supernatant was incubated with anti-FLAG M2 affinity gel (Sigma) (2.3.3.2). The eluted material was analysed by SDS-PAGE gel electrophoresis (2.4.2) followed by Coomassie staining (2.4.3), showing a clearly visible band corresponding to Tra1 (Fig. 4.4A, lanes 1, 4-7). An additional purification step was carried out by gel filtration chromatography on a Superose 6 Increase 10/300 GL column (GE Healthcare) (2.3.3.4) (Fig. 4.4B) and the material was analysed this time by both Coomassie (Fig. 4.4C) and Silver staining (2.4.6; Fig. 4.4D). The bands corresponding to Tra1 were identified by mass spectrometry (2.4.7), confirming the presence of intact, full-length *S.c.* Tra1 polypeptide. Given that monodisperse Tra1 protein is expected to elute at 13-14 ml from the Superpose 6 column, the gel filtration chromatography profile shows the presence of high-molecular weight species, suggesting the presence of Tra1 aggregates and was confirmed by negative staining (NS) analysis (3.1) of the three main peaks identified in the gel filtration run (Fig. 4.4B and 4.5).

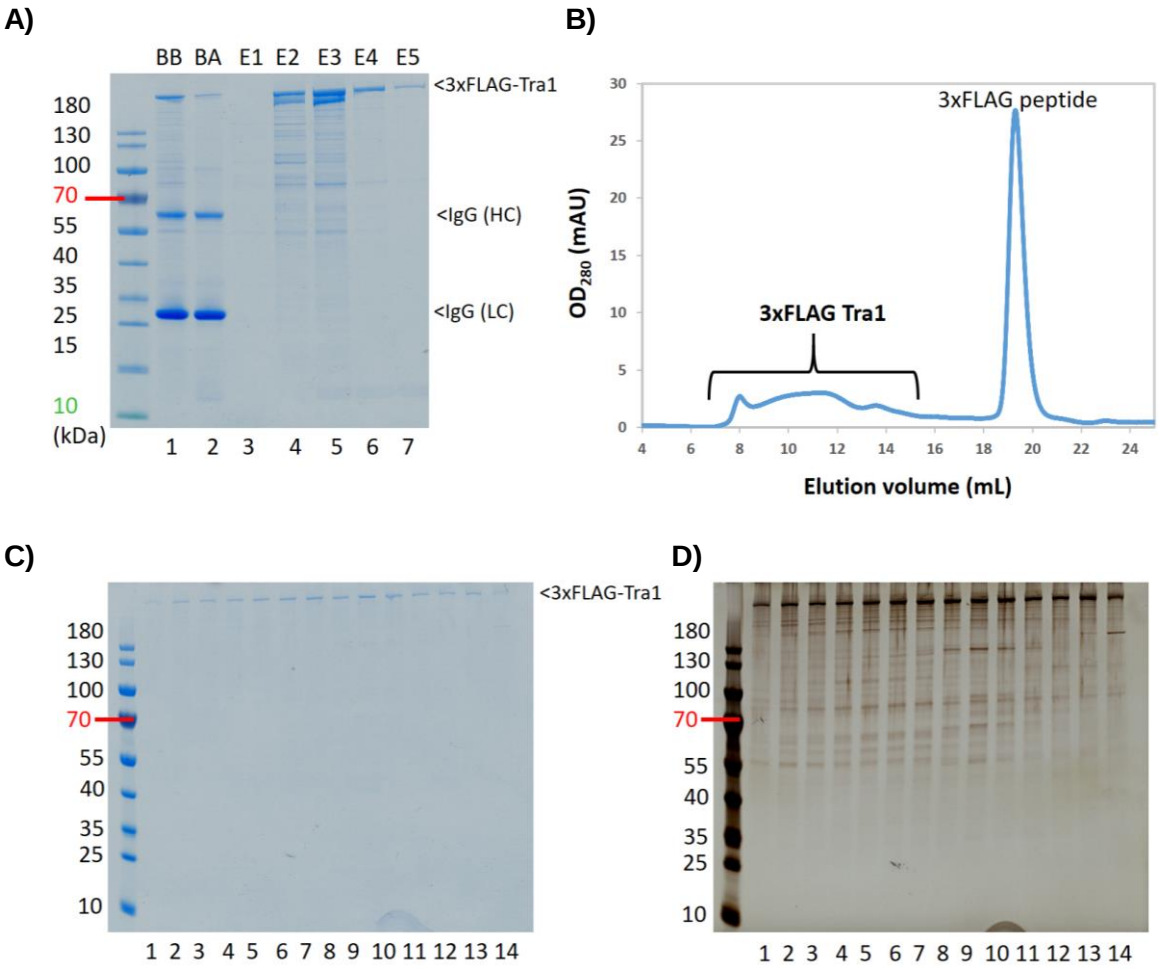


Figure 4.4 | 3xFLAG Tra1 purification. A) 4-12 % SDS-PAGE analysis of affinity purified 3xFLAG-Tra1 followed by Coomassie staining with InstantBlue (2.4.3). BB: beads before elution (lane 1); BA: beads after elution (lane 2); E1-E5: elution fractions 1-5 (lanes 3-7); IgG (HC): immunoglobulin G (heavy chain); IgG (LC): immunoglobulin G (light chain). **B)** Chromatography profile corresponding to the Superose 6 Increase 10/300 GL column step. **C)** 4-12% SDS-PAGE analysis of **B.** Lanes 1-14: 0.5 ml fractions eluting between 7.5 and 14.5 ml, covering Tra1 main peaks. **D)** Silver staining analysis (2.4.6) of **C**.

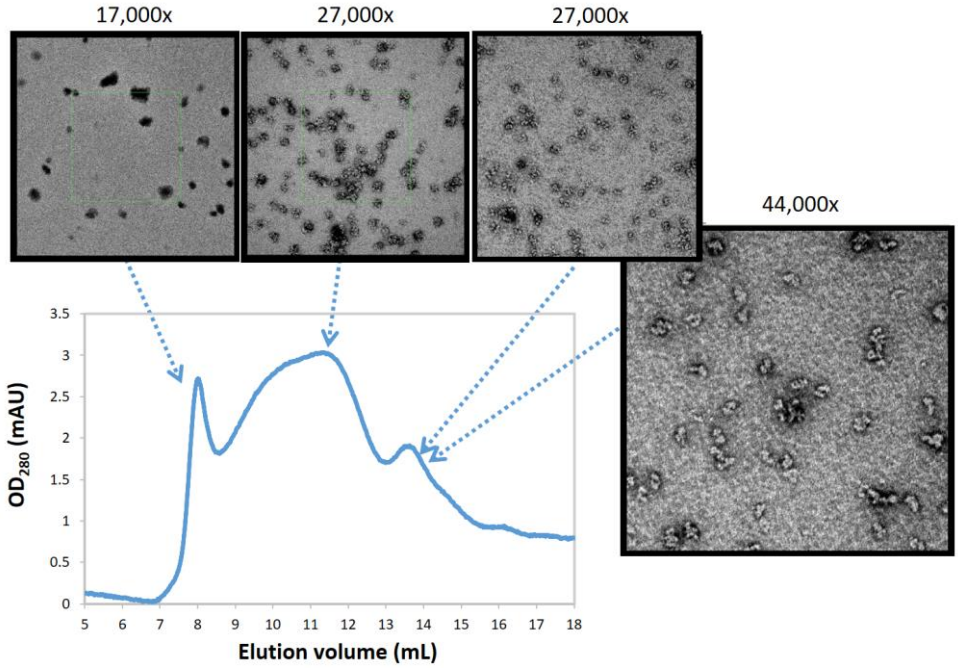


Figure 4.5 | Negative staining analysis of Tra1 main peaks on the Superose 6 column. The main peaks of the chromatography profile shown in Fig. 4.4B were analysed by negative staining (3.1) and the grids were imaged on a 100 kV FEI T10 microscope equipped with a CCD camera. The images show the presence of high molecular weight species corresponding to Tra1 aggregates (first and second peak) as well as monodisperse Tra1 polypeptide (third peak) that elutes at the expected retention volume for a ~433 kDa protein in the tested Superose 6 Increase 10/300 GL column.

4.1.4 Optimisation assays

To optimise the binding of 3xFLAG-Tra1 to anti-FLAG M2 resin, a binding assay was performed using different lysate/resin ratios and also incubation times (Fig. 4.6). Analysis of Tra1 in the supernantant by Western blot showed that increasing resin ratios did not significantly decrease Tra1 protein from the supernatant (Fig. 4.6A, lanes 2-4). Although Tra1 is bound to the resin, because very large amounts of anti-FLAG M2 resin were used for the assay without large differences, Tra1 binding to the resin was clearly shown to be inefficient.

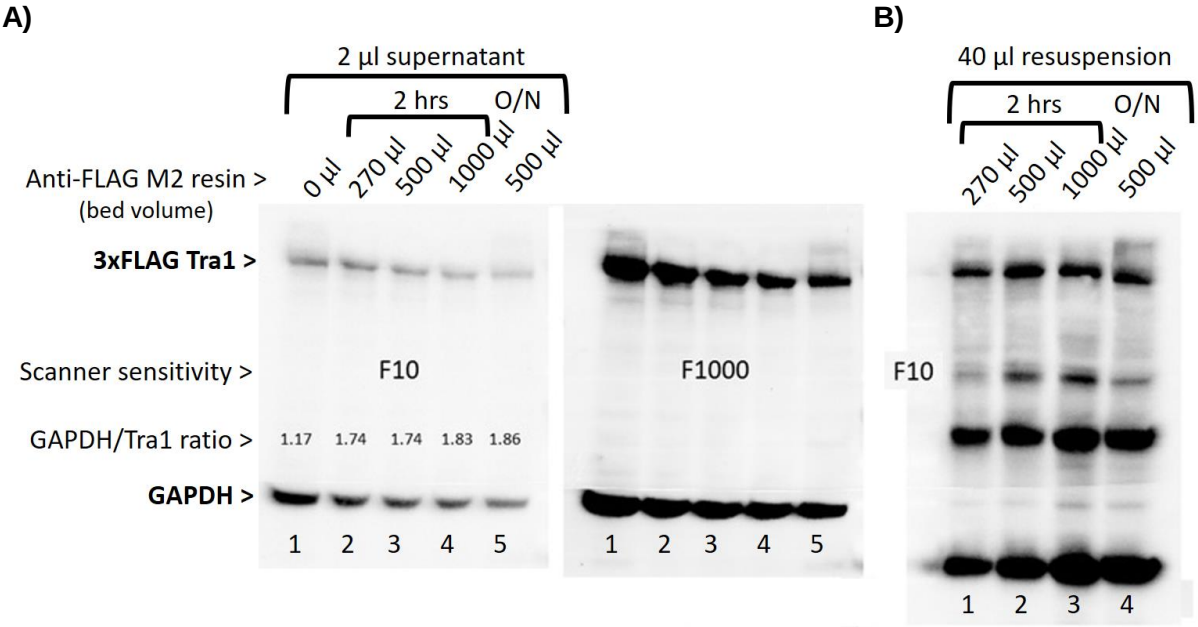


Figure 4.6 | Binding assay of Tra1 to anti-FLAG M2 affinity gel. A) Different anti-FLAG M2 resin volumes (270, 500 and 1000 µl, lanes 2-5) were incubated for 2 hrs (lanes 1-4) and overnight (O/N) (lane 5) with 3 ml of yeast lysate containing recombinantly expressed 3xFLAG-Tra1 (2.3.2.2). The presence of 3xFLAG-Tra1 in the supernatant was analysed by Western blot with anti-FLAG M2 antibody (1:1000) and anti-GAPDH antibody (1:2000) was used as loading control (2.0.6, Table 17); as secondary antibody, anti-mouse HRP conjugated was used (1:2000) (2.0.6, Table 17). The ratio of GAPDH/Tra1 signal was measured using ImageJ software (Schneider, Rasband and Eliceiri, 2012). **B)** The batch mixture of beads and lysate for the 2 hrs incubation (lanes 1-3) and O/N incubation (lane 4) was analysed by Western blot to test for potential degradation during the longest incubation time.

In addition, reducing or removing the presence of Tra1 aggregates would potentially improve the purification, because the presence of Tra1 aggregates largely decreases the amount of effective or useful Tra1 that could be used for electron microscopy analysis i.e. monodisperse protein (Fig. 4.5). With the aim of reducing the presence of aggregates, an additional protein expression test was performed to optimise Tra1 expression. The aim was to harvest yeast

cells at the optimum induction time to potentially reduce Tra1 aggregates species and also to provide protein that has been synthesised during cell exponential growth. The expression test (2.3.2.2) consisted in a time course experiment performed growing 1 L flask instead of 50 mL as in the previous test expression, to simulate large-scale growth conditions. Overall expression was analysed by Western blot at different time points after induction of Tra1 expression with 2% of galactose (Fig. 4.7).

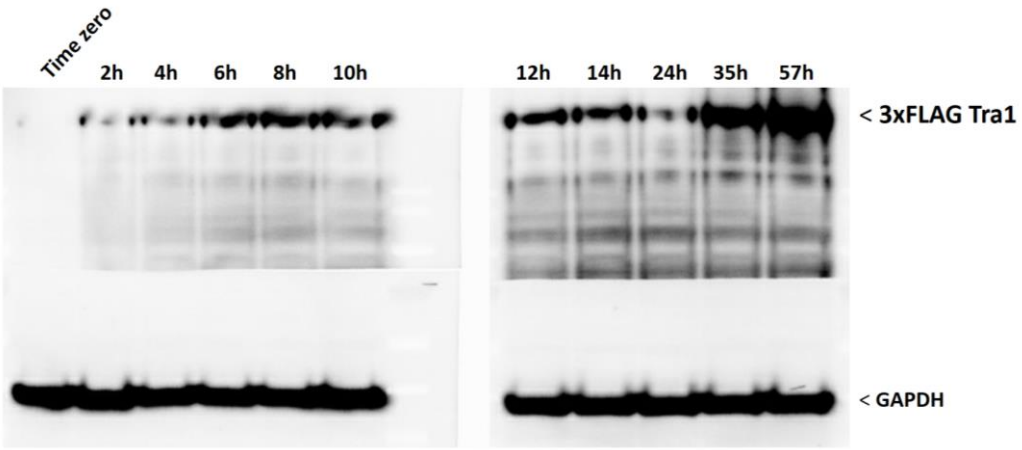


Figure 4.7 | Time course of Tra1 expression. 1 L of cell culture was grown at 30 ºC, and 3x OD_{600} units were taken at the indicated time points and TCA precipitated for Western blot analysis, as previously described (4.1.2).

Additional assays were carried out to test the effect of different salt concentrations and pH on the aggregates. Affinity purified sample was run on a gel filtration column (Superose 6 Increase 10/300 GL) using different salt concentrations in the “Gel filtration buffer” (2.0.5, Table 14) but containing instead 50 mM and 800 mM NaCl (Fig. 4.8) or a different pH condition (pH 8.45) comparing to the previously tested (300 mM NaCl and Tris pH 7.5). Although different chromatography profiles were obtained, no significant effect was observed in the presence of Tra1 aggregates when the affinity purified protein was exposed to these particular buffer conditions during the size exclusion chromatography (SEC) step.

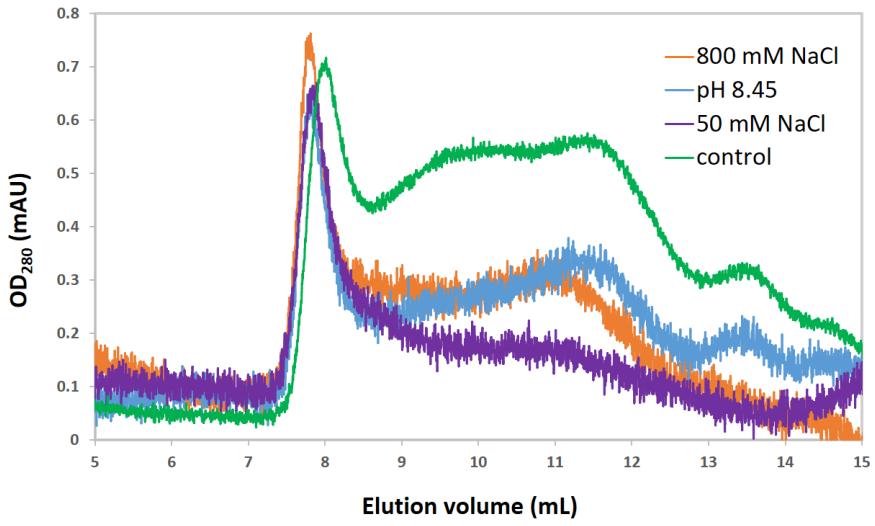


Figure 4.8 | Effect of salt concentration and pH in the SEC profile. Three different salt concentrations (50 mM in purple, 300 mM in green and 800 mM in orange) were tested as buffer conditions for gel filtration analysis, to check for a potential disruption of aggregates; in addition, a more basic pH (blue) was also tested. No significant changes were observed in the disruption of aggregates in the tested conditions. (The control run (green) was downscaled for comparison purpose to the other three curves).

4.1.5 A near optimum-purification

The time course of protein expression for the N-terminally tagged construct (Fig. 4.7) showed that 6 hrs of induction time was enough to express the protein in abundant quantities. The expression volume was increased to 24 L following the new induction parameters. The purification protocol also required optimisation across many parameters. An optimised purification that included benzonase treatment, sonication and increased pH buffering of the lysate yielded much more material after affinity purification, together with the increase in the volume of yeast culture from 12 L to 24 L (2.3.3). The optimised purification also included extra purification steps (MonoS followed by MonoQ 5/50 GL before gel filtration in this case), leading to the production of higher quality sample for electron microscopy and a reproducible purification protocol (Fig. 4.9).

Buffering was replaced initially from 50 mM Tris pH 7.5 to 50 mM HEPES 8.0, and later on it was modified by both increasing the concentration of buffer from 50 mM HEPES 8.0 to 125 mM HEPES 8.0 in the lysis buffer and also increasing the ratio of lysis buffer to cell pellet from 1:3 to 1:1 (v/v).

To reduce viscosity, the lysate material was incubated for 30 min at 4^{o}C with mixing in the presence of benzonase (2.0.5, Table 13). Its addition has an instant effect on the viscosity of the lysate, and extra sonication steps were also carried out performing 4 cycles of sonication for a total of 8 min accumulated time prior to cell debris centrifugation in a ultracentrifuge at 34 k rpm (2.3.3.2).

Because these steps lead to an increase in purified protein after affinity purification, white precipitates were observed for Tra1 when a 1:1 (v/v) ratio of elution buffer and anti-FLAG M2 agarose gel (bed resin) was used to elute the protein with 3xFLAG peptide. The increase in affinity-purified material is substantial when compared to the previous purification (Fig. 4.9A, lanes 4-8), allowing to add additional purification steps.

Because of the increased yield of purified material and to further improve purity, additional steps of ion exchange chromatography (IEX) on MonoS and MonoQ 5/50 GL columns (2.3.3.3) were carried out prior to the Superose 6 Increase 30/100 GL column step, as will be discussed in the next section (4.1.6).

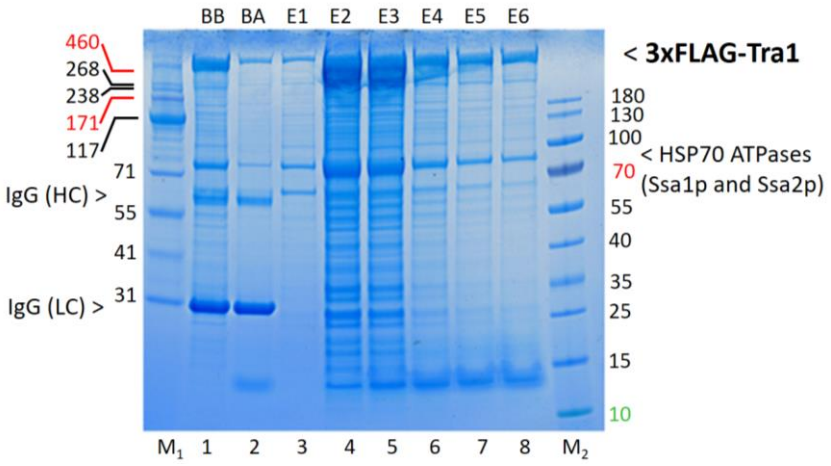


Figure 4.9 | Optimised purification of the *S.c.* Tra1 polypeptide. Affinity purification of 3xFLAG-Tra1 by anti-FLAG M2 agarose affinity gel. The modification of several steps in the purification resulted in an increased yield comparing to previous purifications (Fig. 4.4A), allowing to add more chromatography steps and to obtain monodisperse Tra1 protein for cryo-EM analysis (4.1.6). M: molecular marker. BB: beads before elution (lane 1), BA: beads after elution (lane 2), E1-E6: elution fractions 1-5 (lanes 3-8); IgG (HC): immunoglobulin G (heavy chain); IgG (LC): immunoglobulin G (light chain).

4.1.6 An optimised Tra1 purification for single-particle cryo-EM analysis

For cryo-EM grid preparation, 24 litres of culture were grown and the purification was carried out following the optimised protocol (2.3.3 and 4.1.5) (Fig. 4.10). In addition to the previously described modifications, the elution volume for FLAG affinity chromatography was increased to maintain a lower protein concentration and reduce the precipitation observed during the previous purification (4.1.5), and presumably because of the low Tra1 solubility - at least under the tested buffer conditions. Therefore, this time a mild elution in batch was performed instead, using 2.5 CVs of “Elution buffer 1” (2.0.5, Table 14; 2.3.3.2) for the first elution fraction and resulting in more diluted protein fractions as was pursued. The affinity purified material was balanced to a conductivity of ~ 74 µS/cm, matching 100 mM NaCl in our affinity purification buffer conditions. Unexpectedly, Tra1 appeared in both flowthrough and elution fractions from the MonoS 5/50 GL step, and were subsequently pooled and diluted back to 74 µS/cm (2.3.3.3). This salt balanced material was applied to a MonoQ 5/50 GL column (GE Healthcare) (Fig. 4.10A), and the material was eluted using a 100-1000 mM NaCl linear gradient. The peak corresponding to Tra1 at ~ 211-217 ml of elution volume (Fig. 4.10A, B, gel lanes 1-13) was recovered and concentrated to 0.5 mL using a spin concentrator (2.3.3.4), and injected into a Superpose 6 Increase 10/300 GL column (Fig. 4.10C, D) (2.3.3.4).

As mentioned before (4.1.3), Tra1 aggregates preceded monodisperse Tra1 protein in the Superpose 6 increase 10/300 GL chromatography, but this time the monodisperse protein was found to be segregated from the aggregation peak (Fig. 4.10C, D, peak at ~ 14 mL, lanes 6-8), allowing for the recovery of pure, monodisperse and homogeneous protein for electron microscopy analysis. The peak was then recovered and the material was concentrated using

a micro-spin column (2.3.3.4) to 150 µl final volume. The protein concentration as measured by the nanodrop was ~ 0.1 mg/mL (2.4.1).

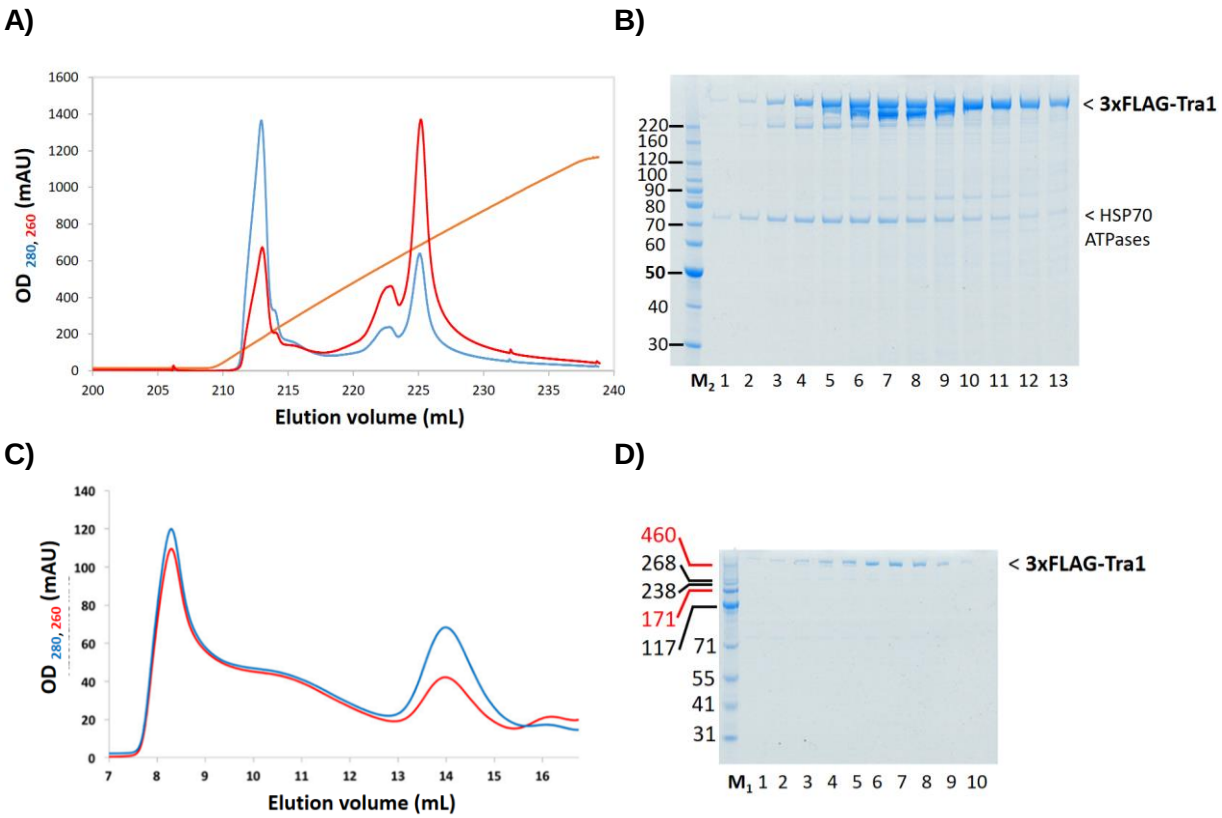


Figure 4.10 | Tra1 purification for cryo-EM analysis. A) MonoQ 5/50 GL column chromatography step. **B)** 4-12 % SDS-PAGE and Coomassie staining of **A**. Lanes 1-13: 0.5 ml fractions eluting between 210.5 and 217 ml, corresponding to the Tra1 peak. The lower protein band in the gel corresponds to yeast chaperone (as analysed by mass spectrometry; Fig. 4.9). **C)** Superose 6 Increase 10/300 GL chromatography step. The second peak in the chromatogram corresponds to monodisperse Tra1 protein. **D)** 4-12 % SDS-PAGE and Coomassie staining of **C.** Lanes 1-10: 0.25 ml fractions eluting between 12.5 and 14.75 ml. M_1: 31-460 kDa protein ladder. M_2: 30-220 kDa protein ladder (2.0.6, Table 17).

4.2 Electron microscopy and image processing

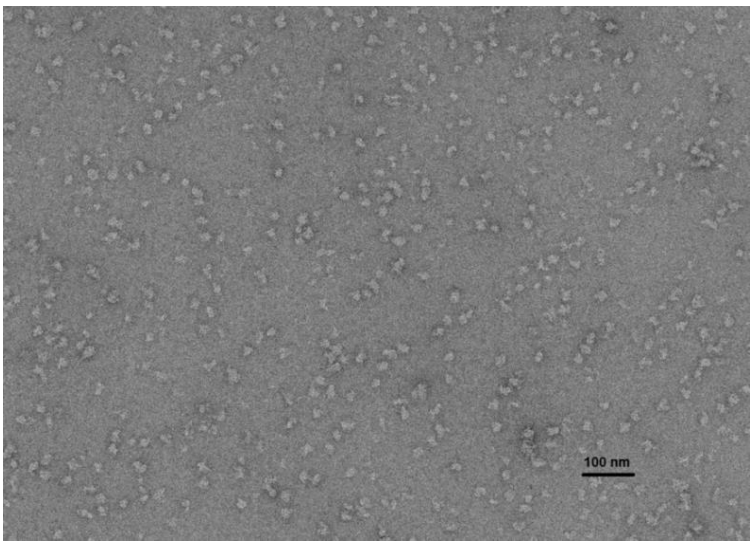
Tra1 monodisperse peak was recovered (Fig. 4.10 C, D) and concentrated to 150 µl as previously described (2.3.3.4). The concentrated material was used for negative stain electron microscopy (3.1.2) by adsorbing the sample to the grid for 5 min of incubation time (Fig. 4.11A). The purified yeast material was used by Natasha Lukoyanova to vitrify monodisperse Tra1 and to collect in-house high-resolution cryo-EM data (3.4.1.1). For cryo-EM grid preparation, a continuous support grid with a 3 nm continuous layer of carbon (Lacey grids, Agar Scientific) was used to absorb the protein to the surface, and several applications were needed to get an optimum particle density (3.4.1.1) (Fig. 4.11B).

As a proof of concept, I generated a first map from only 40 movies collected in-house on a 300 kV FEI Polara microscope mounted with a K2 detector (Gatan) operated in counting mode (Fig. 4.11B and C; Table 4.1).

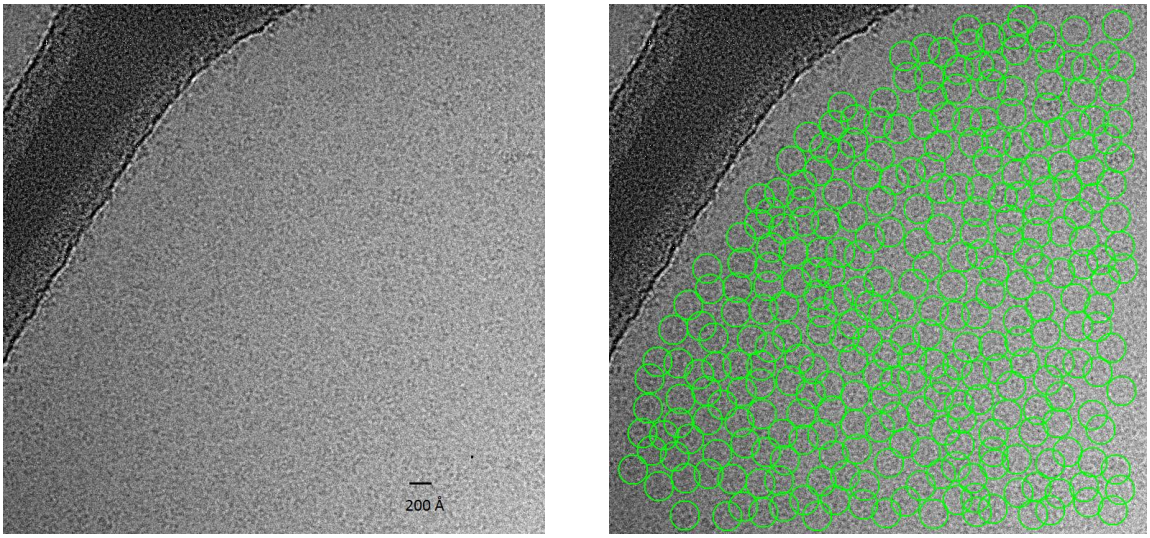
Table 4.1 | Cryo-EM data collection and processing of *S. cerevisiae* Tra1 (9 Å map)	
Nominal magnification	125,000x
Voltage (kV)	300
Electron exposure (e^{-}Å^{-2})	~ 46
Defocus range (μm)	-1.5 to -4
Calibrated pixel size (Å)	1.39
Symmetry imposed	C_1
Final particle images (no.)	7,680
Map resolution (Å) at FSC=0.143	~ 9

After movie alignment with MotionCor (version 1) and particle picking in Gautomatch, reference-free 2D classification of the picked and extracted particles was carried out in RELION 1.4 (CPU powered version) (Fig. 4.11C). The resulting 2D classes showed clear secondary structure elements (Fig. 4.11C, D). As the distinctive Tra1 structure is largely composed of α-solenoid super-secondary structure elements, this yielded 2D classes with repetitive structural features, some of them resembling a “spring” like structure (Fig. 4.11D). The consensus 3D map presented a reported resolution of ~ 9 Å, generated from approximately ~ 7,7 k particles (Fig. 4.11C; Table 4.1). (For a more detailed description of EM processing please see section 3.4.1.2).

A)



B)



C) **D)**

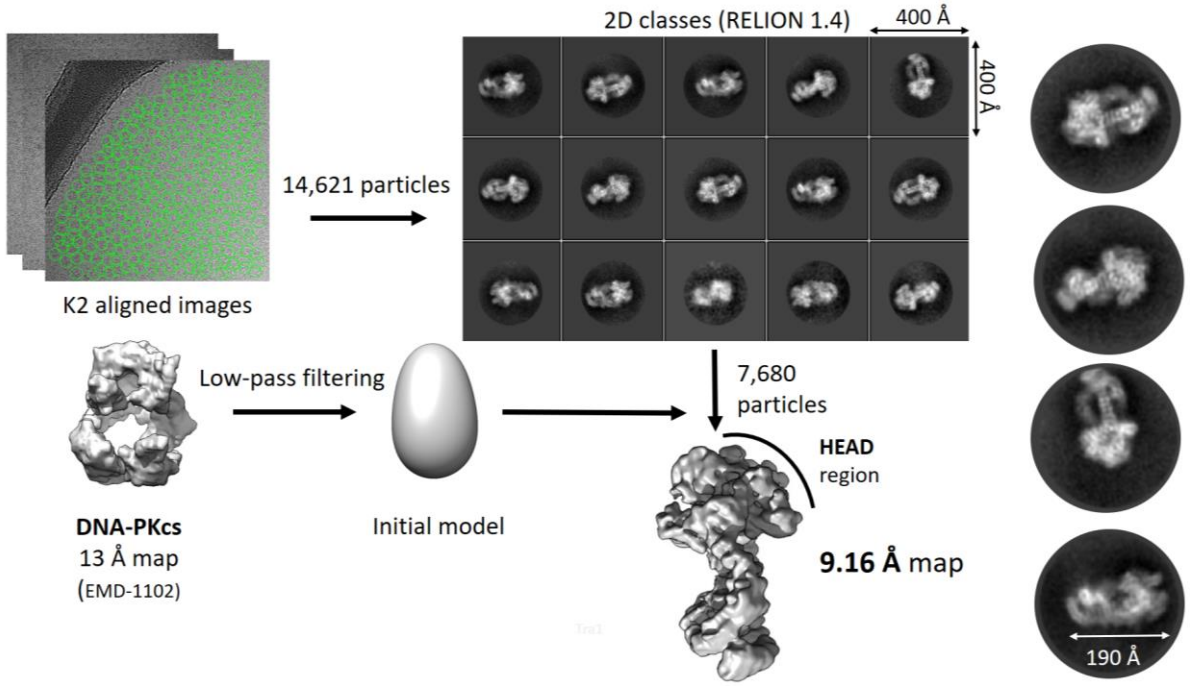
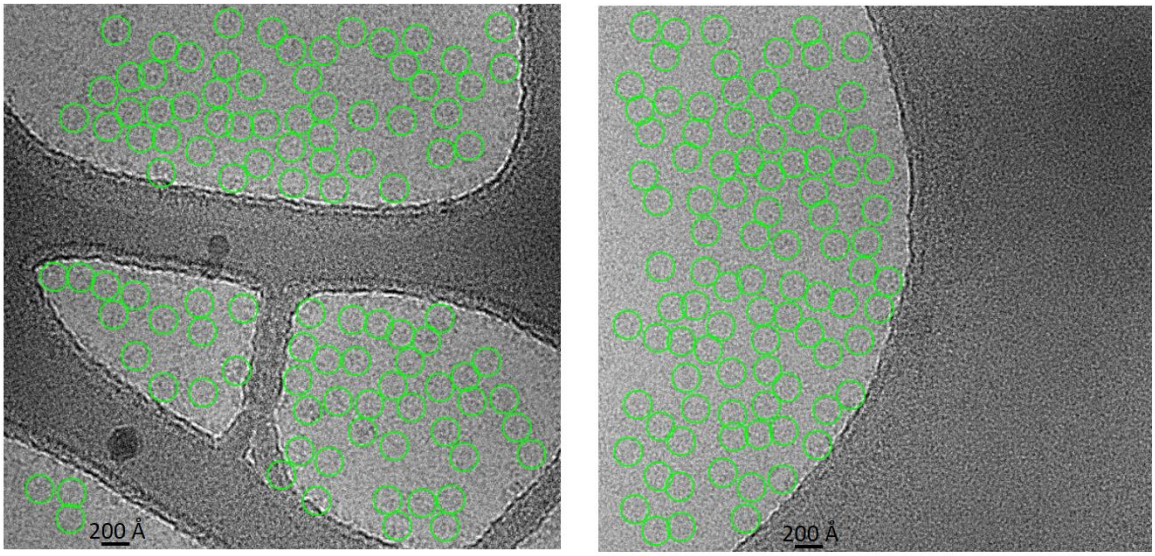
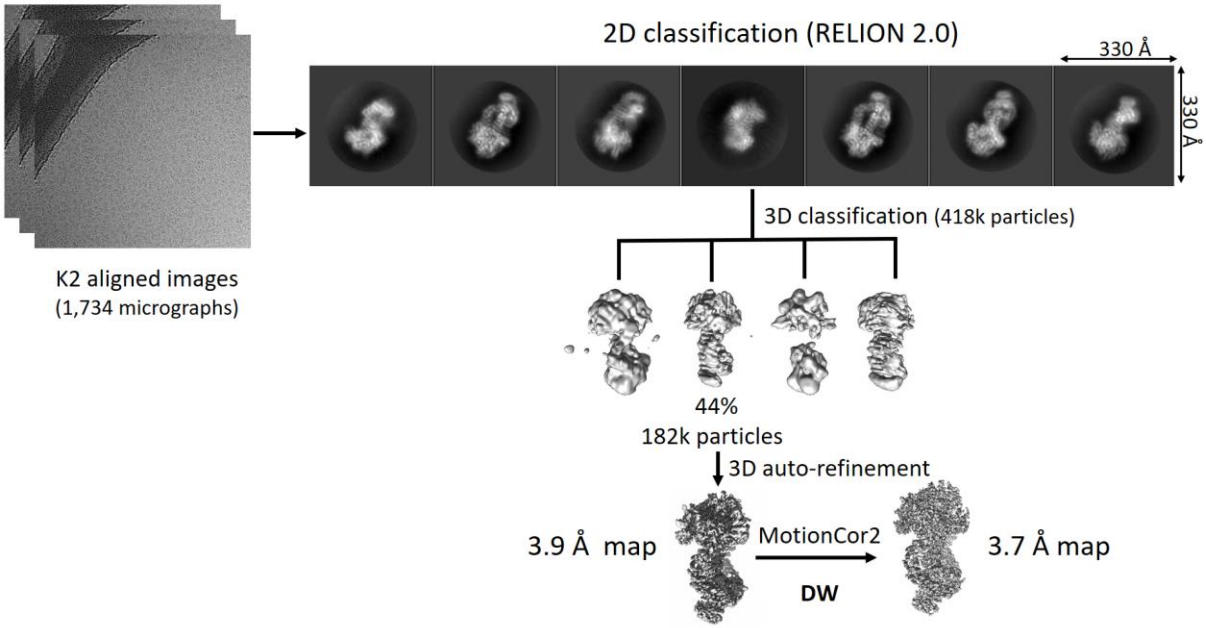


Figure 4.11 | In-house dataset and first Tra1 cryo-EM map. A) Preliminary negative staining analysis (3.1) of purified Tra1, showing monodiserse protein. Images were collected on a 200kV FEI F20 microscope equipped with a DDD camera (3.1). **B)** Representative cryo-EM micrographs of Tra1 on Lacey grids coated with a thin layer of carbon. Images were collected at Birkbeck College (University of London) on a direct detector device (DDD). Each circle is 200 Å in diameter and represents particle picking coordinates. **C)** A ~ 9 Å cryo-EM map of Tra1 was obtained from the 40 movies collected in-house. Particles were picked in Gautomatch and processed in RELION 1.4. For the 3D-autorefinement, a previously published low resolution DNA-PKcs structure (EMD-1102) was low-pass filtered and used as reference volume. **D)** Amplified image of the obtained 2D class averages, showing secondary structure features. (As mentioned in the main text, vitrification and high-resolution cryo-EM data collection was performed by Natasha Lukoyanova, whereas I performed the protein purification, negative stain electron microscopy analysis and data processing).

A)


200 Å
200 Å

B)


2D classification (RELION 2.0)
330 Å
330 Å
3D classification (418k particles)
K2 aligned images
(1,734 micrographs)
44%
182k particles
3D auto-refinement
3.9 Å map
MotionCor2
DW
3.7 Å map

C)

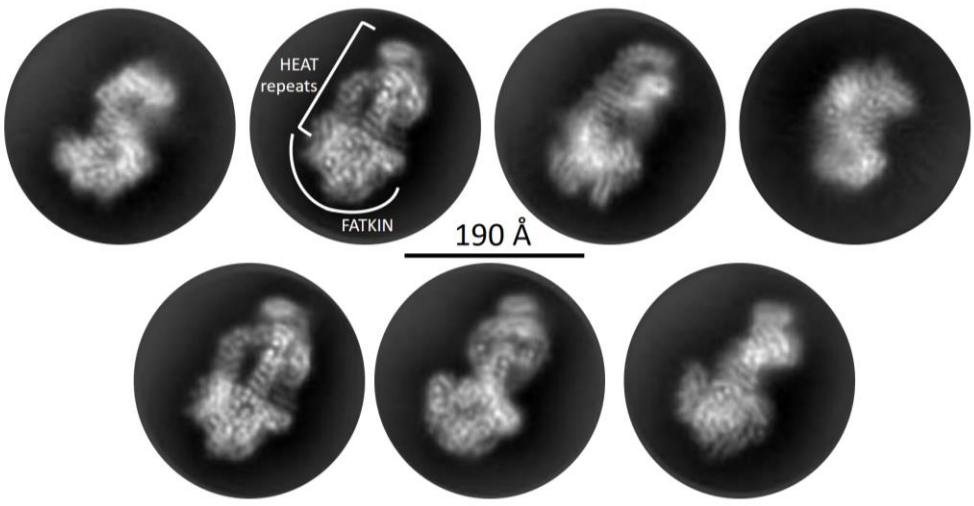
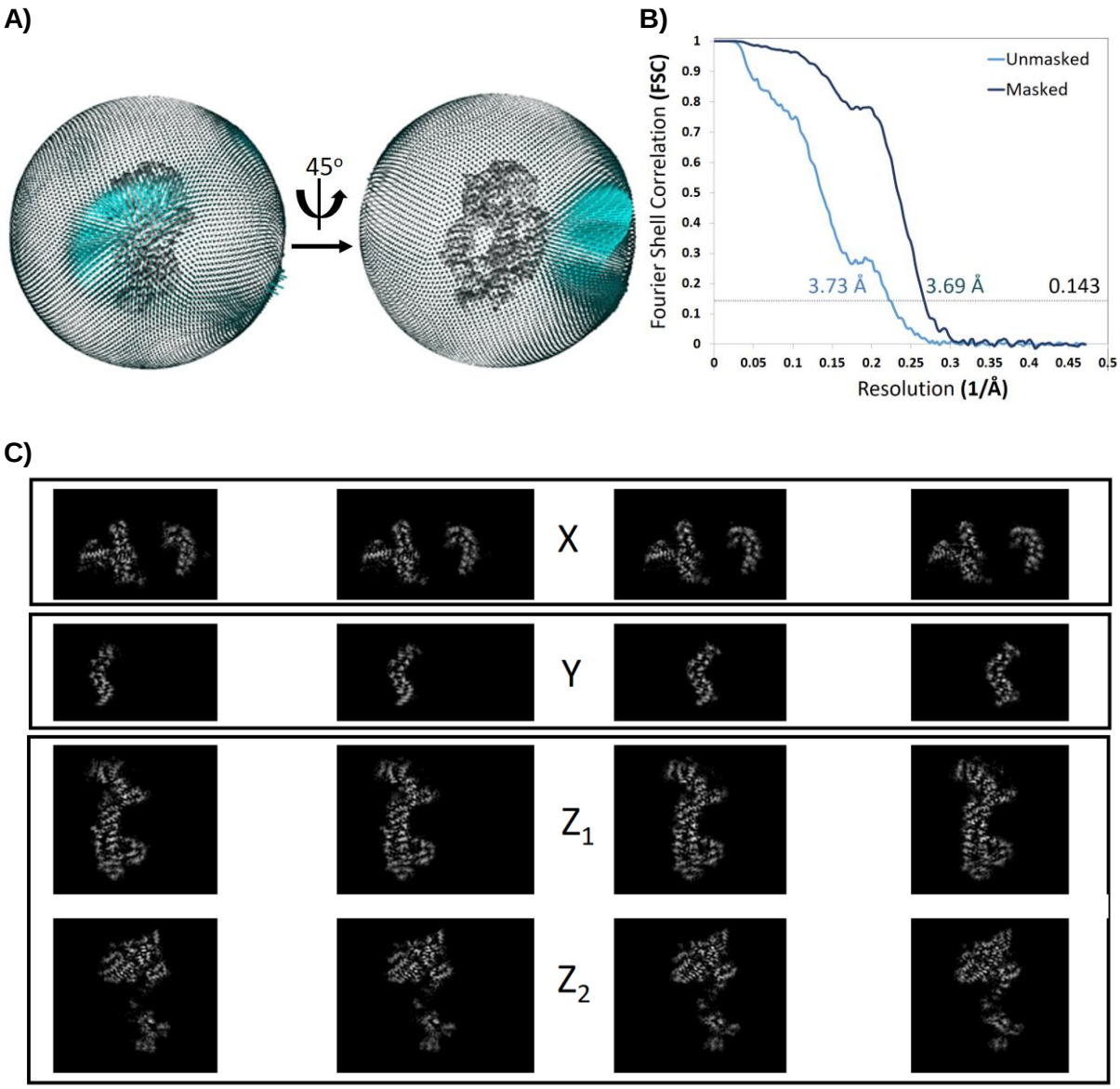

HEAT
repeats
FATKIN
190 Å

Figure 4.12 | Diamond Light Source (DLS) dataset and high-resolution cryo-EM map of Tra1. A) Two representative cryo-EM micrographs (Lacey grids). Particle picking was performed in Gatotomatch (200 Å diameter circular mask) and supervised in RELION 1.4. **B)** A summary of the processing is shown. After manual supervision, a set of 1,734 micrographs were selected for processing. After particle picking in Gautomatch and 2D classification in RELION, particles contributing to the best classes were selected and 3D classified. One 3D class refined to high-resolution, and dose weigthed particles generated a final map at 3.7 Å resolution. **C)** Zoomed 2D classes showing diverse views and high-resolution features. The vitrification of the sample was performed by Natasha Lukoyanova and high-resolution cryo-EM data collection was performed by Alistair Siebert on a 300 kV FEI Titan Krios at DLS, Didcot (Oxford).



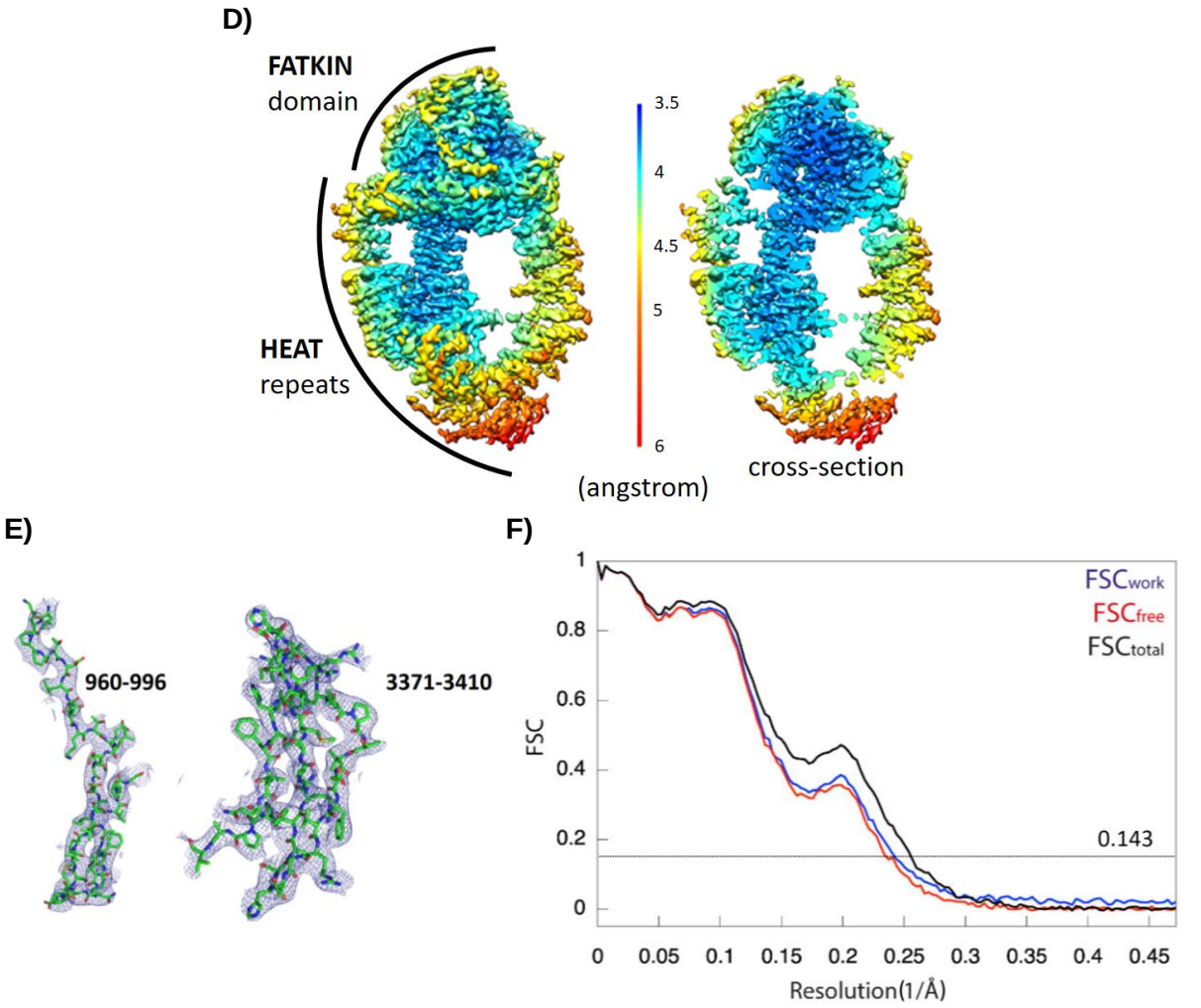


Figure 4.13 | Tra1 cryo-EM map analysis. A) Angular distribution of particles contributing to the final map. **B)** Fourier Shell Correlation (FSC) analysis showing 3.7 Å resolution at 0.143 threshold. **C)** Cross-sections of the 3.7 Å map. SPIDER software (Baxter, Leith and Frank, 2007) was used to generate cross-sections of the Tra1 cryo-EM map, showing its high content in α-helix. Contiguous sections in the X, Y and Z planes are shown. In the case of the Z-axis, the last two rows represent distant cross-sections within the same axis. **D)** Local resolution map of Tra1 cryo-EM reconstruction. **E)** Details of the *"de novo"* generated atomic model fitting into the electron density map. **F)** Model vs map FSC validation curves (**D**-**F** are adapted from Díaz-Santín *et al.*, 2017).

This preliminary ~ 9 Å cryo-EM map obtained from the small test dataset allowed the identification of secondary structure elements, especially in the alpha-solenoid region. It also allowed the identification of the Head region containing the kinase domain surrounded by the FAT domain (Fig. 4.11C) and provided confidence that high resolution was achievable.

The next step was to increase the size of the dataset in order to obtain a near-atomic resolution map, enabling for the generation of a full atomic model of *S. cerevisiae* Tra1. To accomplish that, a dataset was collected at the Electron Bio-Imaging Centre (eBIC) in Diamond Light Source (DLS). Data collection was performed on a 300 kV FEI Titan Krios microscope mounted with a K2 camera operated in counting mode, resulting in a dataset of 2,449 movies (Table 4.2).

Table 4.2 | *S. cerevisiae* Tra1 (3.7 Å map): cryo-EM data collection, processing, refinement and validation	
Data collection and processing	
Nominal magnification	130,000x
Voltage (kV)	300
Electron exposure (e^{-}Å^{-2})	~ 44
Defocus range (μm)	-1.5 to -3.5
Calibrated pixel size (Å)	1.06
Symmetry imposed	C_1
Final particle images (no.)	~ 182 k
Map resolution (Å) at FSC=0.143	3.7
Refinement	
Model resolution (Å) (FSC=0.143)	3.7
Map CC (whole unit cell)	0.817
Average B-factor (Å^2)	79.2
RMS deviations – Bonds (Å)	0.010
RMS deviations – Angles (deg)	1.41
Validation	
Molprobity Score	2.29
Clashscore	8.27
Ramachandran plot (%)	
Favoured	84.6
Allowed	14.8
Outliers	0.64
C-beta deviations	0
Rotamer Outliers (%)	1.72

Because the irregular nature of the Lacey grids (Fig. 4.11B, 4.12A), a semi-automated procedure was applied, based on the manual selection at low magnification of suitable areas for data collection. This selection was performed in the FEI EPU software through the manual selection of thin carbon regions. A total of 1,600 areas were manually selected, and more areas were also scheduled for data collection during the three-day session to reach a final

number of 2,449 movies. After movie alignment by MotionCor v1 (3.4.1.2), the 2,449 aligned images were visually inspected and particle picking was performed over a selected subset of "good" 1,734 micrographs (Fig. 4.12B). Particle picking was performed with Gautomatch, given the very good results provided on the in-house dataset. Although optimisation in particle picking was intended, it was not possible to computationally exclude picking on thick carbon areas without compromising the picking in the thin carbon region of the grid. Therefore, to avoid missing particles, the strategy was to optimise the picking on the thin carbon region and manually remove the picked coordinates on the thick carbon over the 1,734 micrographs (Fig. 4.12A). The supervised picking improved the results of the 2D classification in RELION 2.0 (beta-version), and after several rounds of 2D classification, the best 2D classes were used as templates for a last round of particle picking in a feed-back process (3.4.1.2). Therefore, another round of manual supervision of the picked particles was carried out to ensure the quality of the particle picking before a new round of 2D classification. The resulting increased number of particles contributing to good 2D classes was subjected to 3D-autorefinement in order to obtain a consensus 3D map previous to 3D classification. The reference volume used for refinement was the in-house 9 Å resolution Tra1 cryo-EM map, which was low-pass filtered to 40 Å (generating a featureless blob). Reference-based 3D classification in RELION resulted in one single "good", high-resolution 3D class. Subsequent subclassification of the particles contributing to this map did not provide any different 3D classes or conformations, and a last 3D auto-refinement step lead to a 3.9 Å resolution map. Dose-weighting calculation was implemented in the recently released MotionCor2 (Zheng *et al.*, 2017) so that the micrographs were movie aligned by the new version of the software and the particles contributing to the 3.9 Å resolution map were extracted from the MotionCor2 processed dataset and 3D-autorefined. Exhaustive processing resulted in a 3.7 Å resolution final map of the Tra1 protein (Fig. 4.12B and 4.13A-D; 3.4.1.2). The angular distribution of particles on the grid for the high-resolution 3D reconstruction of Tra1 shows that there is a range of closely related stable orientations in space (Fig. 4.13A) - given Tra1's elongated shape. However, the presence of different orientations, as shown by the diversity of different 2D classes obtained (Fig. 4.12B, C), generated a near-atomic resolution map by fulfilling the Fourier space during the 3D reconstruction. The resulting map was used by Alan C.M. Cheung to build a complete *"de novo"* atomic model of the Tra1 polypeptide (Fig. 4.13E-F).

The comparison of our high-resolution data against previously available structures for the SAGA and NuA4 complexes shows that Tra1 density can be identified in previously published EM images (Fig. 4.14A), although it could not be recognised as such at the time. EM data corresponding to structures that were published after the release of our Tra1 3.7 Å resolution map (Díaz-Santín *et al.*, 2017) is also rendered in the figure for comparison purpose (Fig. 4.14B), to show that the density corresponding to Tra1 can be clearly identified. The fitting of

the newly generated *"de novo"* atomic model of Tra1 (section 4.3) into an existing EM map of the SAGA complex allowed to identify for the first time the density corresponding to Tra1 within the complex (Fig. 4.14C), previously described as "Lobe A" in the low resolution map.

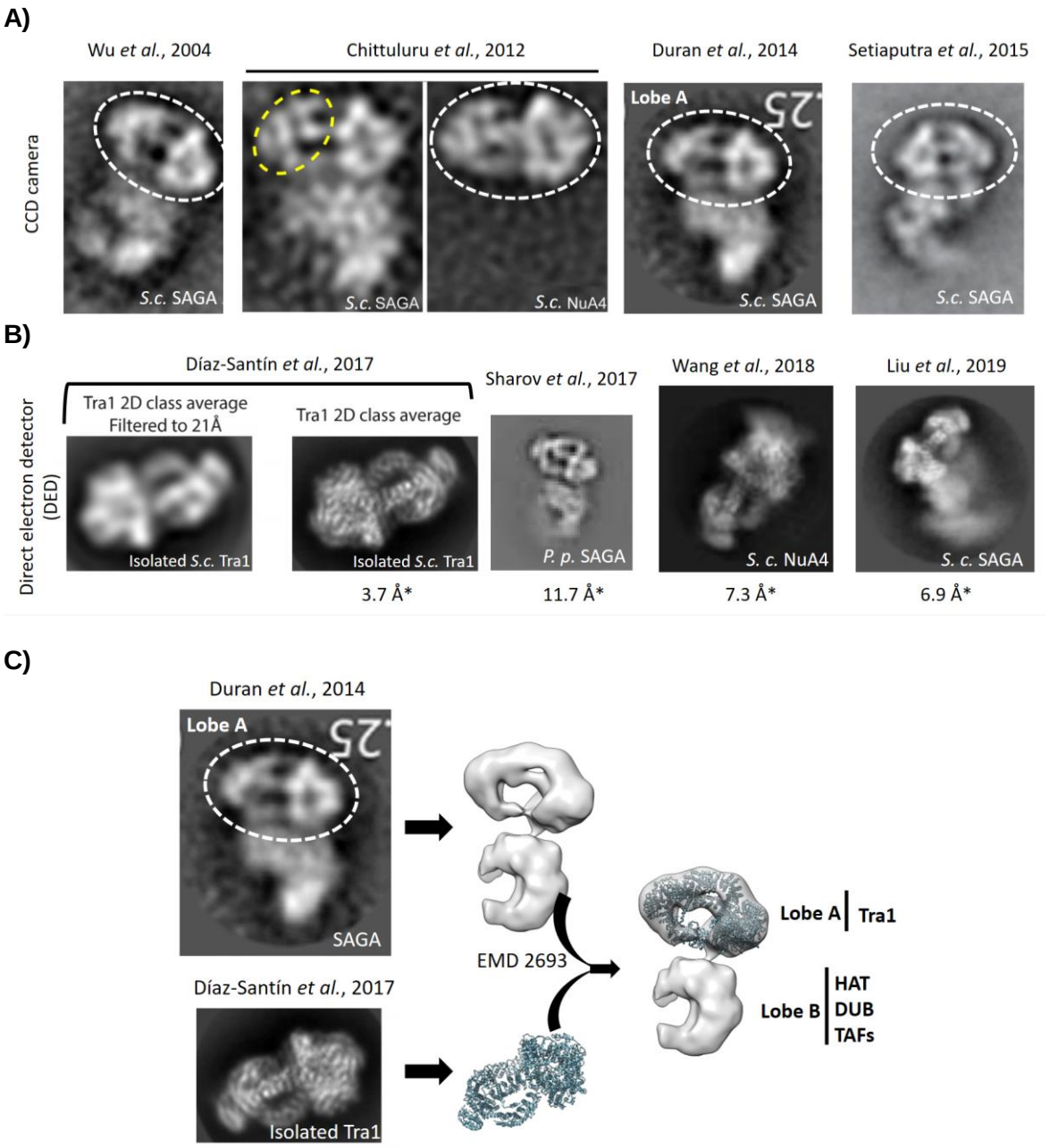


Figure 4.14 | Timeline of released Tra1, SAGA and NuA4 EM structures. Published EM data obtained by CCD camera (**A**) or direct electron-detector device (DED or DDD) technology (**B**). **A)** Previously available 2D EM images corresponding to the SAGA and NuA4 complexes as shown in Fig. 1.18. **B)** 2D EM images of isolated Tra1 (this study), the SAGA complex and the NuA4 complex (available structures). A high-resolution 2D class of Tra1 low-pass filtered to 21 Å (adapted from Díaz-Santín *et al.*, 2017) shows its similarity to previously published EM data (**A**). The density for Tra1 can be unambiguously identified in all images (**A**, **B**). *S.c.: Saccharomyces cerevisiae. P.p.: Pichia pastoris.* *Overall reported resolution for the 3D EM reconstructions. **C)** The fitting of Tra1 atomic model (this study, 4.3) into an existing map of the SAGA complex allows to identify the so-called "Lobe A" in the complex as the Tra1 subunit. HAT: histone acetyltransferase module. DUB: deubiquitinase module. TAFs: TBP-associated factors.

4.3 Tra1 atomic model and biological insights

4.3.1 Overview of full-length Tra1 atomic model and its unique α-solenoid region

The overexpression of full-length Tra1 within its original source organism and the purification of the monodisperse protein form, followed by single-particle cryo-EM analysis, generated a high-quality electron density map at a nominal resolution of 3.7 Å, allowing for the generation of an accurate atomic model that covers the 93% of the 3,744 residues composing the gigantic Tra1 polypeptide in *S. cerevisiae* (Fig. 4.15; 3.4.4.1).

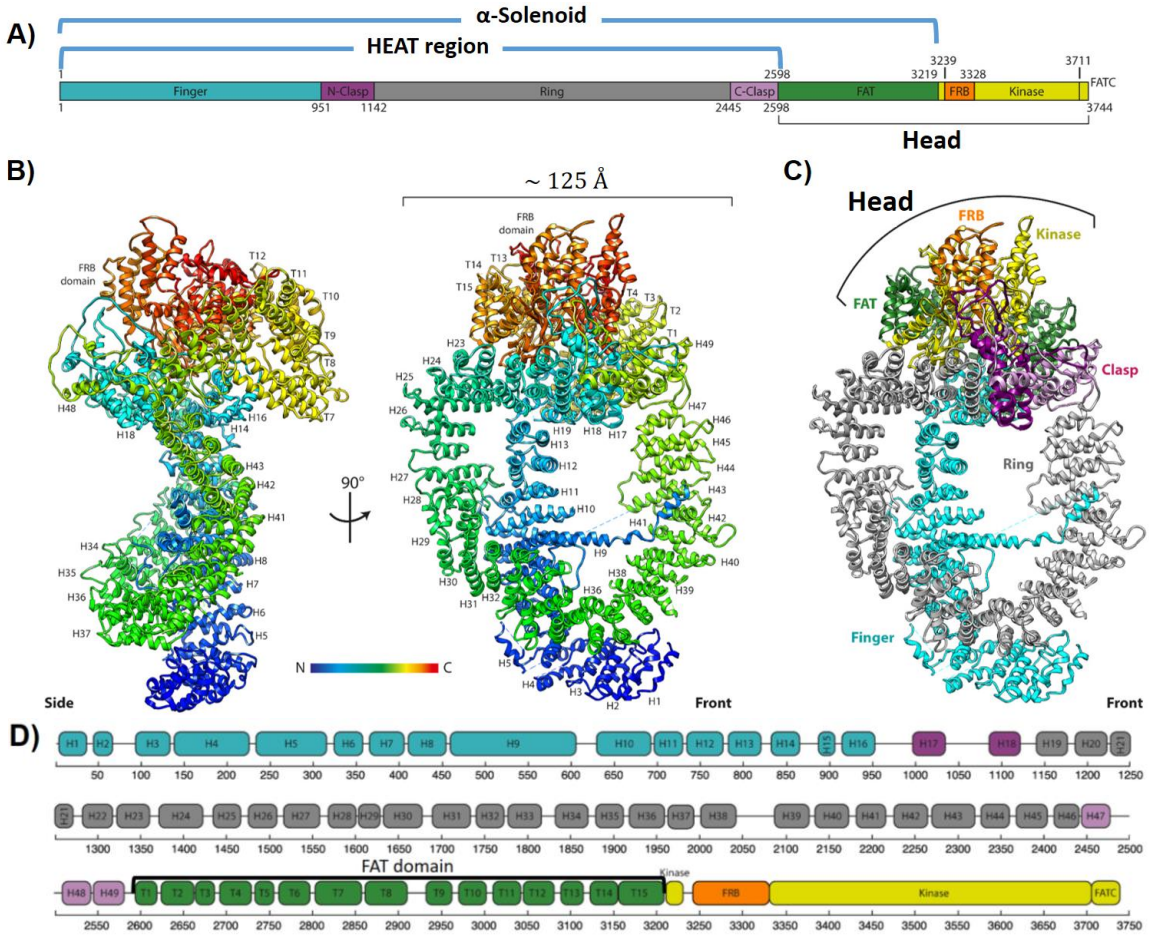


Figure 4.15 | Tra1 atomic model. A) Domain organisation from N to C-terminal comprising the Finger, Clasp, Ring, FAT and Kinase regions. **B)** Tra1 atomic model is coloured from N to C-terminal (blue to red) as can be read in the legend. **C)** Tra1 atomic model coloured by regions as in **A**. **D)** Detailed organisation of the primary sequence into HEAT and TPR repeats, as coloured in **A**. (Adapted from Díaz-Santín *et al.*, 2017).

This is the first atomic structure for this member of the family and the first full-length PIKK to be solved at a resolution better than 4 Å. The unresolved 270 residues (7% of Tra1) are not concentrated within any particular region, meaning that there is not any particular missing domain or secondary structure element; instead, the missing segments are spread along 15

different regions that are likely to be loops or disordered protein, as inferred from secondary structure prediction.

The atomic model shows a very unique architecture for Tra1. As previously described in the introduction, we can identify from N to C-terminal the following regions in any PIKK: two solenoid tracks, the first one is a large segment named HEAT domain, and a second, short segment named FAT, preceding the Kinase domain at the C-terminal region of the protein. FRB (Fig. 4.15, 4.16 and 4.17), LBE (Fig. 4.17) and FATC (Fig. 4.15, 4.17 and 4.18) structural motifs are found in the Tra1 Kinase domain. A globular region containing the FAT (consisting of 15 TPR repeats in Tra1) and the Kinase domain – the two together named FATKIN or Head onwards - is very similar and conserved along the different PIKKs. However, the solenoid region containing the 49 HEAT repeats in Tra1 adopts a very unique structure when compared to other members, and it is composed of two different regions: an N-terminal Finger region extending and pointing towards the FATKIN domain and inserted into a second region, consisting in an extensive circular Cradle (Fig. 4.15 B, C; Fig. 4 16). The structure is stabilised and locked by a Clasp region (Fig. 4.16). Therefore, Tra1 is topologically composed of three different regions: the Finger, Cradle or Ring, and the Head or FATKIN region.

4.3.2 Organisation of the HEAT domain: the Finger and the Ring regions

The N-terminal end of the HEAT domain is the Finger region, composed of HEAT repeats H1-H16, and comparable to the “Spiral” in ATR and mTOR or the “Arm” region of DNA-PKcs, also called “Bridge”. The N-terminal part of the finger (repeats H1-H6) is the most flexible region of the protein as shown by the local resolution map (Fig. 4.13D), and for that reason presents the lower local resolution reported for the map (red colour). The intrinsic flexibility of this region might have a role in the interaction with other proteins (discussed later). Following this region, HEAT repeats H7-H16 connect to the FATKIN domain (Fig. 4.15). At the H9 repeat we can find a very long insertion containing 99 residues (positions 482-580); 2/3 of this insertion are visible in the cryo-EM map, being possible to identify the contact of this insertion to the opposite side of the Cradle with HEAT repeats H42-H43.

H17 and H18 (N-clasp) together with H47 to H49 (C-clasp) play an important role in maintaining the stability of the Cradle or Ring region, a very large solenoid segment comprising HEAT repeats H19 to H46 (Fig. 4.16). This extensive track of HEAT repeats forms a ring-like shape of ~ 125 Å in diameter, where the two ends of the solenoid are locked by the reciprocal interaction of the previously described N and C-clasps, the last one preceding the FAT domain that belongs to the Head region containing the Kinase domain. The Ring, Clasp and FAT together create a figure of eight conformation where the Clasp region represents the cross-over (Fig. 4.16). Therefore, the overall structure shows a very large solvent accessible area

along the polypeptide where protein-protein interactions can occur, as is typically expected from HEAT repeat containing proteins.

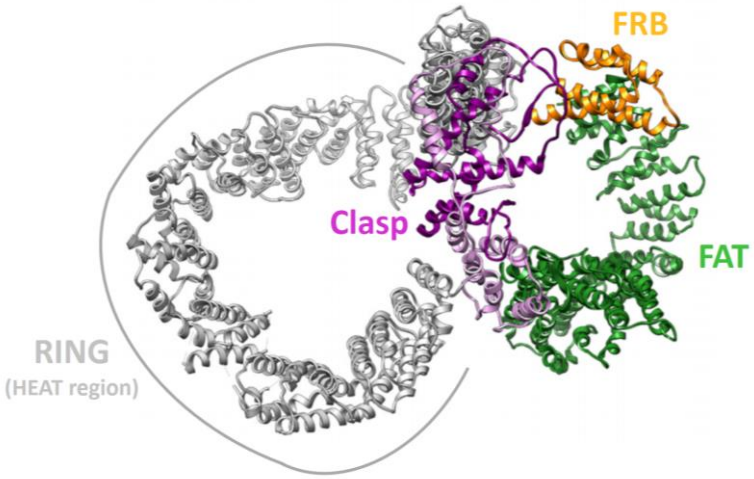
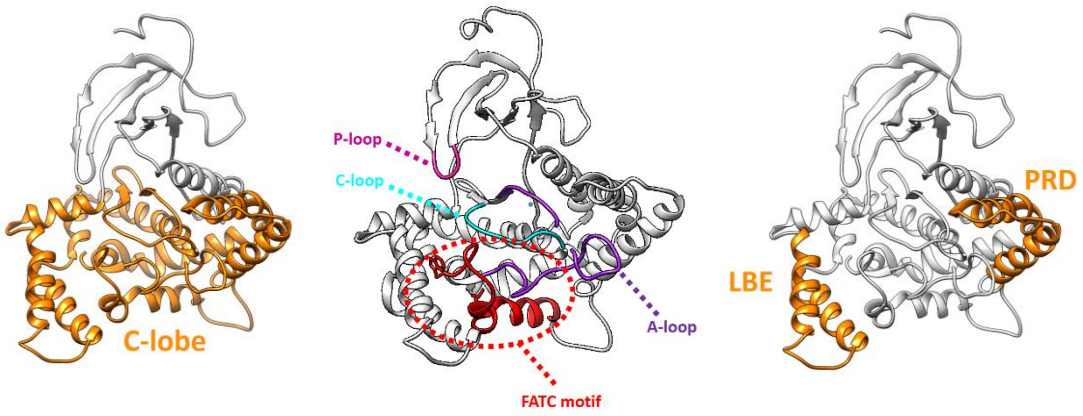


Figure 4.16 | The Cradle or Ring region and the FAT domain. The Finger and Kinase domains are removed in the figure for clarity. The circular Cradle (grey) is stabilised by the Clasp (purple), which precedes the FAT region (green). The FRB motif (orange) is found at the end of the FAT region. (Adapted from Díaz-Santín *et al.*, 2017).

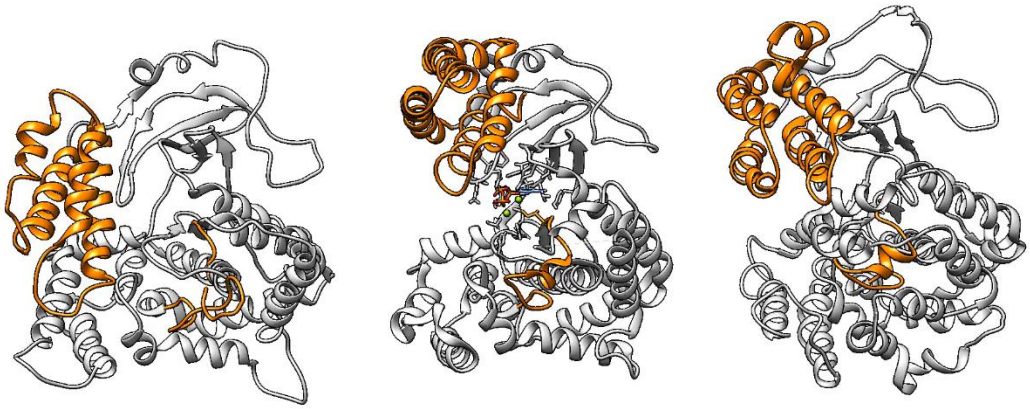
4.3.3 The Head or FATKIN region contains a putative inactive kinase domain

This region is essentially composed of the FAT and the Kinase domain. The FAT region is made of 15 TPR repeats consisting of pairs of anti-parallel alpha-helices organised around the kinase domain and connected to the clasp region, increasing the strength and stability of the structure. The kinase domain is stabilised by the FAT region and inserted within the space contained in between the FAT domain and the Clasp (Fig. 4.16). The kinase contains at least three main regulatory elements also found in other PIKK members: FRB, LBE and FATC motifs (Fig. 4.17 and Fig. 4.18). These elements are distributed along the N and C lobes of the Kinase domain (1.1 and Fig. 4.17A, B). A PRD (PIKK regulatory element) motif has also been described in ATM and ATR proteins (Fig. 4.17). The PRD motif is poorly conserved within the PIKK protein family in terms of sequence, especially in the loop connecting the two α-helices (Fig. 4.17C), and little literature can be found for other members apart from ATM and ATR. However, the PRD motif is structurally conserved in all PIKKs, as is possible to identify the two alpha-helices in the same position in all released structures (some of them shown in Fig. 4.17B). The PRD motif sometimes shows variations that consist in the presence of an insertion in between the two alpha-helices, as is the case for mTOR and DNA-PKcs (Fig. 4.17 B, C). The FRB motif consists of a bundle of four alpha-helices in close proximity to the kinase active site and its position within Tra1 varies comparing to other family member like DNA-PKcs and mTOR, where its relative orientation to the active cleft complicates any possible substrate access, unlike the other two proteins (Fig. 4.17B).

A)



B)



Tra1 **mTOR** **DNA-PKcs**

C)

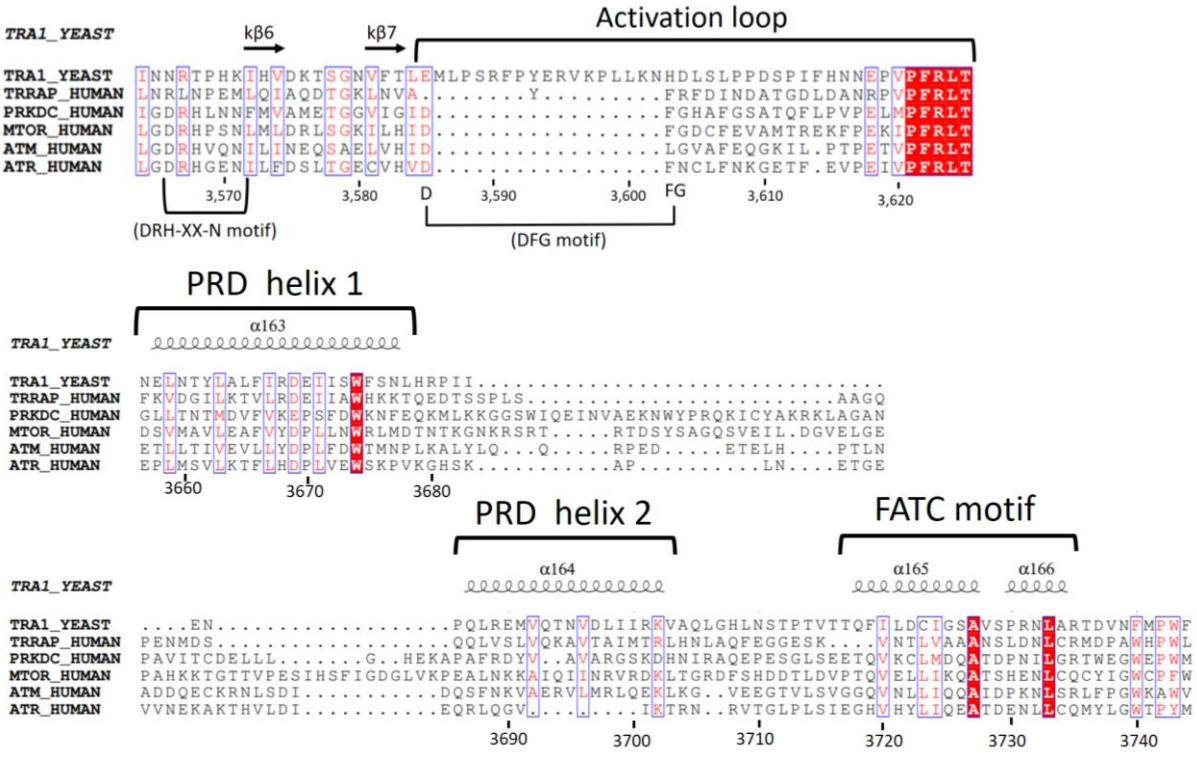


Figure 4.17 | The Kinase domain in Tra1 versus other PIKKs. A) On the left, the Kinase domain, as in other eukaryotic protein kinases (ePK), is composed of an N-terminal lobe (in grey) and a C-terminal lobe (in orange). **In the middle**, the Tra1 catalytic site showing the P-loop in pink, the activation loop (A-loop) in purple and the catalytic loop (C-loop) in light blue; in addition, the FATC motif (residues 3,717-3,744) at the very C-terminal region found in PIKKs is highlighted in red. **On the right**, Tra1 LBE (residues 3,482-3,521) and PRD (residues 3,657-3,703) motifs within the putative pseudokinase domain are coloured in orange. **B)** The FRB domain in Tra1 (this study) versus mTOR (PDB 6bcx) and DNA-PKcs (PDB 5luq). The FRB and activation loops are highlighted in orange. In Tra1, the FRB motif (residues 3,242-3,325) closes the active site, potentially preventing the access of a putative substrate when compared to the other two PIKKs. **C)** Sequence alignment in Clustal Omega of the catalytic loop, activation loop, PRD and FATC motifs of *S. cerevisiae* Tra1 and human TRRAP, DNA-PKcs, mTOR, ATR and ATM. The analysis shows that key catalytic elements such as the DRH-XX-N (HRD motif in canonical ePKs) in the catalytic loop and the DFG motif in the activation loop are not conserved in Tra1^{TRRAP}; however, a conserved PFRLT motif at the end of the activation loop is present in all members. In addition, the loop connecting the two α-helices in the PRD domain is poorly conserved among PIKK members. The alignment was rendered using the ESPrit 3 server (Gouet, Robert and Courcelle, 2003).

We can identify in Tra1 the three catalytic elements present in other protein kinases and their relative positions are conserved comparing to other members (Fig. 4.17A); these elements are the catalytic loop (C-loop), the activation loop (A-loop) and the phosphate binding loop (P-loop). However, critical residues necessary for ATP and Mg^{++} binding and catalysis are not present in Tra1; more specifically, the “DRH-XX-N” and “DFG” conserved motifs in other PIKKs are missing (Fig. 4.17C) and for that reason it has been always described as a putative pseudo-kinase. In addition, a 17-residue insertion is found within the A-loop when compared to other members (Fig. 4.17B, C), although this is poorly conserved between Tra1 homologs.

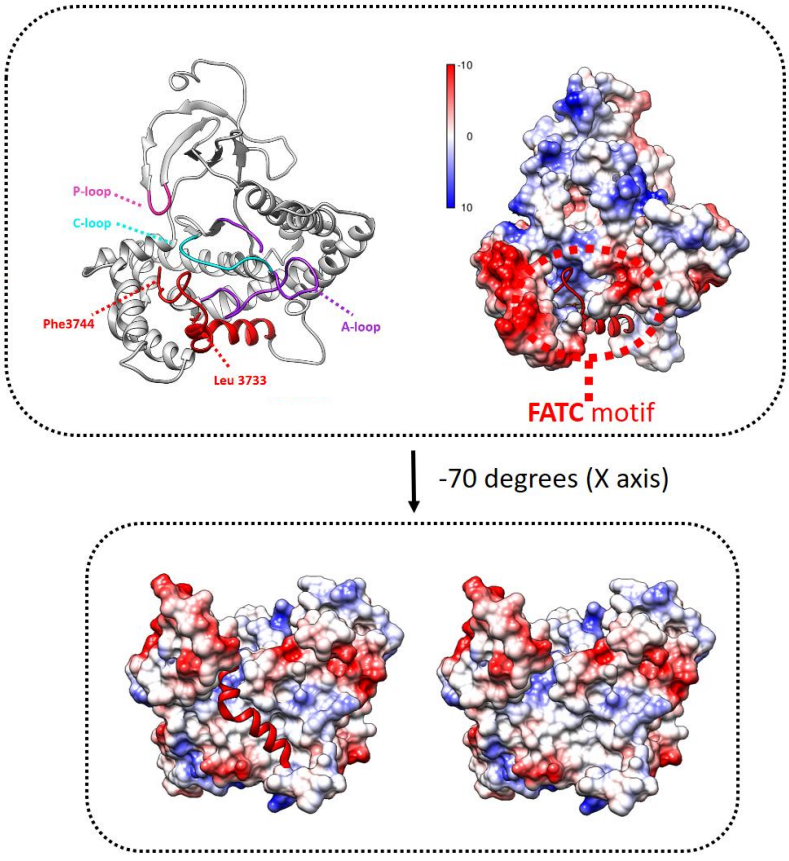


Figure 4.18 | The Tra1 FATC domain is inserted into a hydrophobic pocket. Tra1 catalytic elements are shown on the left as in Fig. 4.17A, and previously described residues shown to be relevant in the stability of the 28-residue FATC motif are highlighted in red. The electrostatic potential of Tra1 Kinase domain surface is rendered using Coulombic surface representation (± 10 kcal/mol · e). The FATC motif is rendered in ribbon to show the electrostatic potential on its docking region within the Kinase.

To complete the description of Tra1 Kinase domain, it is necessary to describe the FATC motif at the C-terminal region of the polypeptide, made of only two alpha-helices (Fig. 4.17A and 4.18). It is inserted inside a hydrophobic cavity between the LBE and the kinase C-terminal lobe, and single residue substitution mutagenesis consisting in L3733A or F3744 are sufficient to cause transcriptional changes (Hoke *et al.*, 2010) (Fig. 4.18). Although the FATC domain has been shown to be critical in regulating catalytically active PIKKs, Tra1 lacks key conserved catalytic residues; hence, this motif is also likely to have an essential function in maintaining the structural integrity of the kinase region. Although several roles have been proposed for the FATC motif, including participation in protein-protein interaction or even membrane insertion (Dames *et al.*, 2005; Jiang *et al.*, 2006; Dames, 2010; Hoke *et al.*, 2010; Sommer, Schaad and Dames, 2013; Ogi *et al.*, 2015), there is no evidence so far for a clearly defined function of the FATC motif.

In the human genome at least 48 proteins (nearly 10% of known kinases) have been identified as pseudo-kinases since they lack key catalytic residues (Boudeau *et al.*, 2006), mainly contained in previously described canonical catalytic motifs such as the VAIK, HRD and DFG motifs (Scheeff *et al.*, 2009). Among these putative pseudo-enzymes is Tra1 human homolog TRRAP, which in spite of presenting an inactive kinase domain, it has been shown to be essential, given that its deletion causes cell lethality (1.2.4). In humans, TRRAP is associated to cancer, in addition to other pseudo-kinases (Zhang *et al.*, 2012). Many of them participate in multi-protein complexes, as is the case for Tra1^{TRRAP}, participating in SAGA and NuA4 (1.2.4). In addition to their structural role - Tra1 is a scaffolding protein in NuA4 (4.3.7) -, it has been shown that some pseudo-kinases function as a hub for the integration of cellular signals (Zeqiraj and van Aalten, 2010). In addition, allosteric activation of proteins by pseudo-kinases has also been reported (Rajakulendran and Sicheri, 2010). Because Tra1 interacts with multiple transcription activators (sections 1.2.4 and 5), the very large Tra1 surface could function as a platform that integrates different signals to finely tune transcription initiation and respond to specific transcriptional requirements. Finally, it has been previously reported that some putative inactive kinases can display catalytic activity in vitro (Kannan and Taylor, 2008), and in this regard, experimental confirmation is still necessary for Tra1. Even if they lack kinase activity, the ability to bind ATP is still retained by some pseudo-kinases and has been shown to be critical for activation of cellular processes (Hammarén *et al.*, 2015). Therefore, both catalytic activity and ATP binding assays need to be specifically addressed for Tra1^{TRRAP}.

4.3.4 Tra1 is highly conserved in evolution from yeast to humans

Protein sequence alignment of Tra1 against its human homolog TRRAP shows 27% of identity, and secondary structure prediction analysis allows the alignment of the HEAT repeat region almost entirely between the two sequences, indicating that Tra1 presents an identical and very well conserved protein architecture in humans (Knutson and Hahn, 2011).

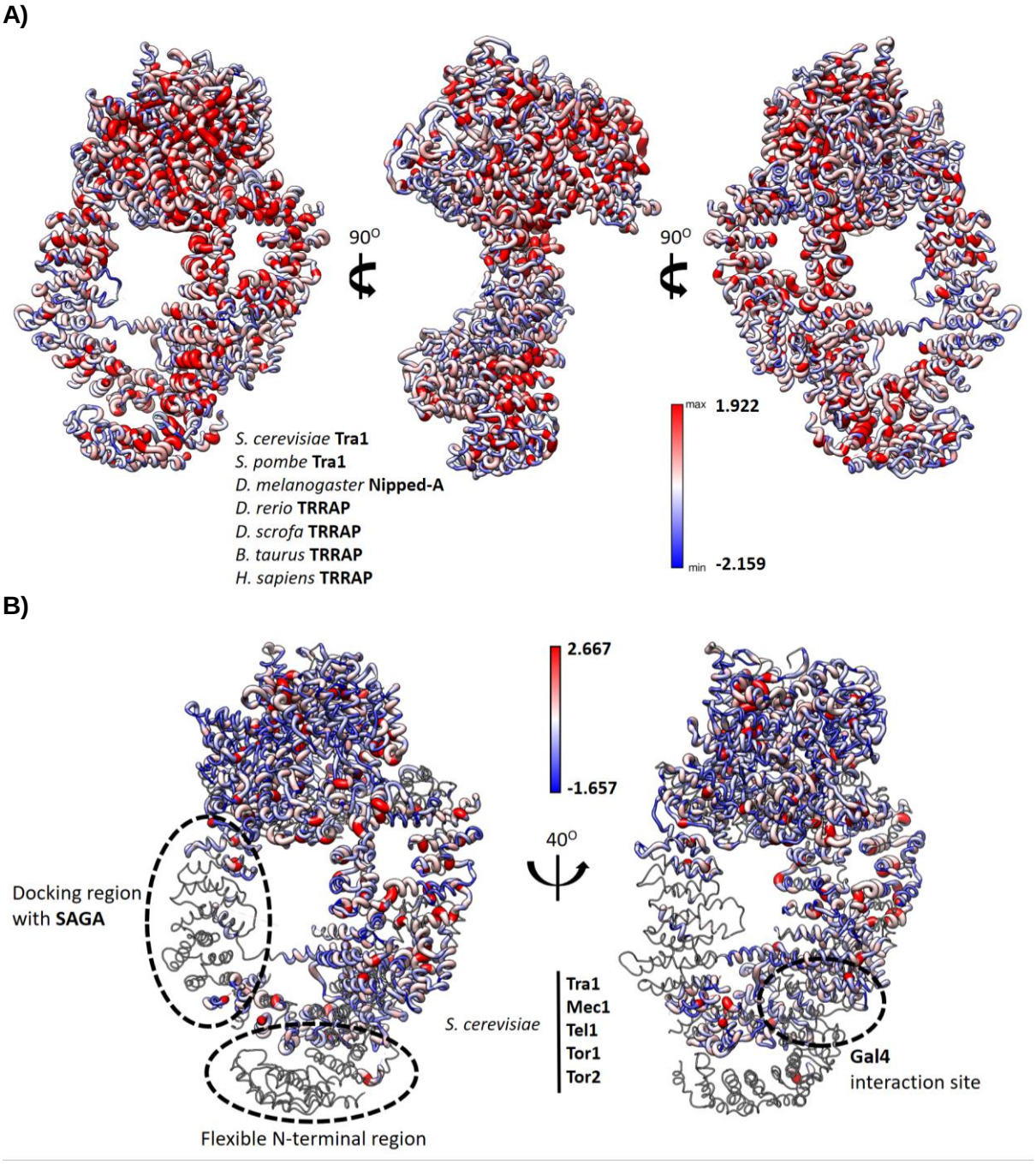


Figure 4.19 | Tra1 protein sequence alignment in Clustal Omega. A) Tra1 evolutionary analysis from yeast to humans. Clustal Omega (Sievers and Higgins, 2018) alignment of Tra1 in *S. cerevisiae* (budding yeast), *S. pombe* (fission yeast), *D. melanogaster* (fly), *D. rerio* (fish), *S. scrofa* (pig), *B. taurus* (cow) and *H. sapiens*. Tra1 residue conservation analysis was plotted in Chimera using the values calculated by the AL2CO algorithm. White colour corresponds to an intermediate value (-0.1185) in between maximum and minimum computed values; red colour indicates fully conserved positions. **B)** Tra1 sequence alignment against other PIKKs in *S. cerevisiae*, i.e. Mec1, Tel1, Tor1 and Tor2. Because

Tra1 is the largest PIKK in *S. cerevisiae*, some regions did not align to any other PIKK sequence. These regions are rendered as wires, and match regions in Tra1 that are involved in the interaction with SAGA and also the binding with transcription activators such as Gal4. White colour corresponds to an intermediate value of 0.505.

Sequence alignment in Clustal Omega (Sievers and Higgins, 2018) of Tra1 homologs from yeast to humans shows several regions distributed across the structure that are fully-conserved in evolution (Fig. 4.19A), and play a functional role as will be discussed in the next section (4.3.5). The alignment of Tra1 against other PIKKs in *S. cerevisiae* (Fig. 4.19 B) shows two regions of Tra1 that do not align to any other PIKK sequence. One of these regions is at the N-terminal, matching the HEAT segments where transcription activators directly interact with Tra1 (4.3.5). The other region is the HEAT region comprised between residues 2,159 to 2,379, involved mainly in maintaining the architecture of the Tra1 Ring (Fig. 4.19B). Interestingly, this region is involved in the integration of Tra1 within the SAGA complex as is suggested by the fitting of Tra1 atomic model into existing low resolution cryo-EM maps (4.3.6), with a role in the direct interaction with other SAGA subunits. Intriguingly, in the case of NuA4, the Piccolo sub-complex is expected to interact with the same Tra1 region implicated in the interaction with SAGA, ruling out the possibility of Tra1 being shared at the same time by the two complexes (4.3.8).

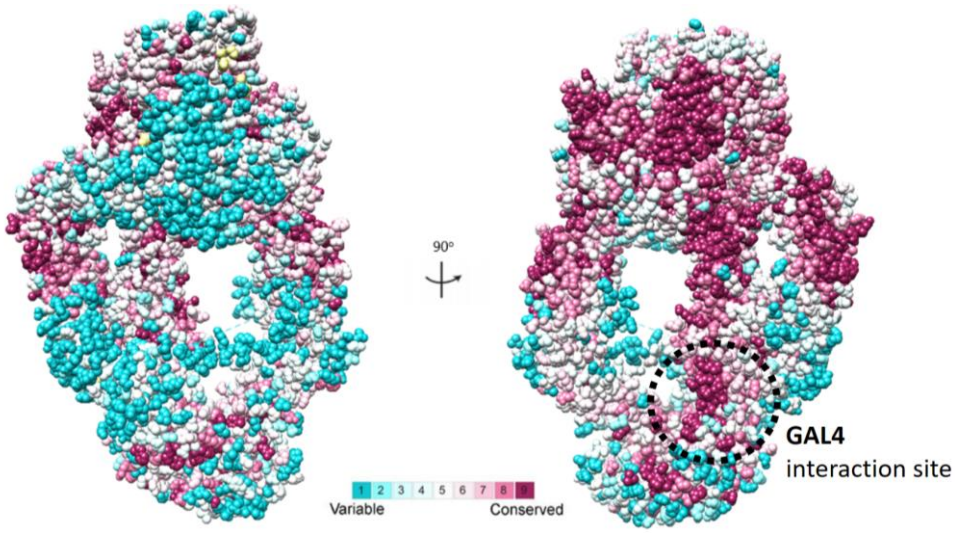


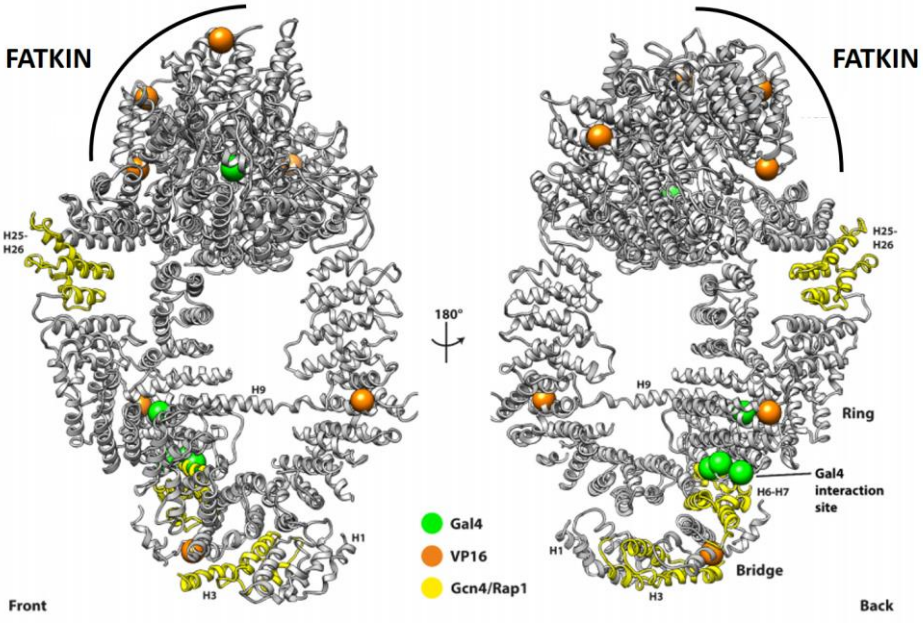
Figure 4.20 | Evolutionary conservation analysis of Tra1 by the ConSurf server. To analyse residue conservation in *S. cerevisiae* Tra1, ConSurf (Ashkenazy *et al.*, 2016) searched homolog sequences using HMMER (Finn *et al.*, 2015), and 150 unique sequences were chosen, presenting a maximal %ID of 95% and a minimal of 35%. The best evolutionary model was selected to be JTT, and the calculated conservation scores were projected into a surface representation of the Tra1 atomic model for display in Chimera. The dotted lined circle shows a previously described Gal4 interaction site on Tra1 surface (Lin *et al.*, 2012) (4.3.5). Interestingly, one side of Tra1 is highly conserved versus the other side, which is generally poorly conserved. The Gal4 interaction surface is found in the highly conserved side of Tra1. The colour code can be read in the legend at the bottom of the figure following a simple score from 1 for less conserved positions to 9 for the most conserved regions of the protein.

Evolutionary analysis of Tra1 residue conservation performed by the ConSurf server (Ashkenazy *et al.*, 2016) (Fig. 4.20) shows that one side of Tra1 is much more conserved than the other, and this side coincides with a Gal4 interaction site (next section).

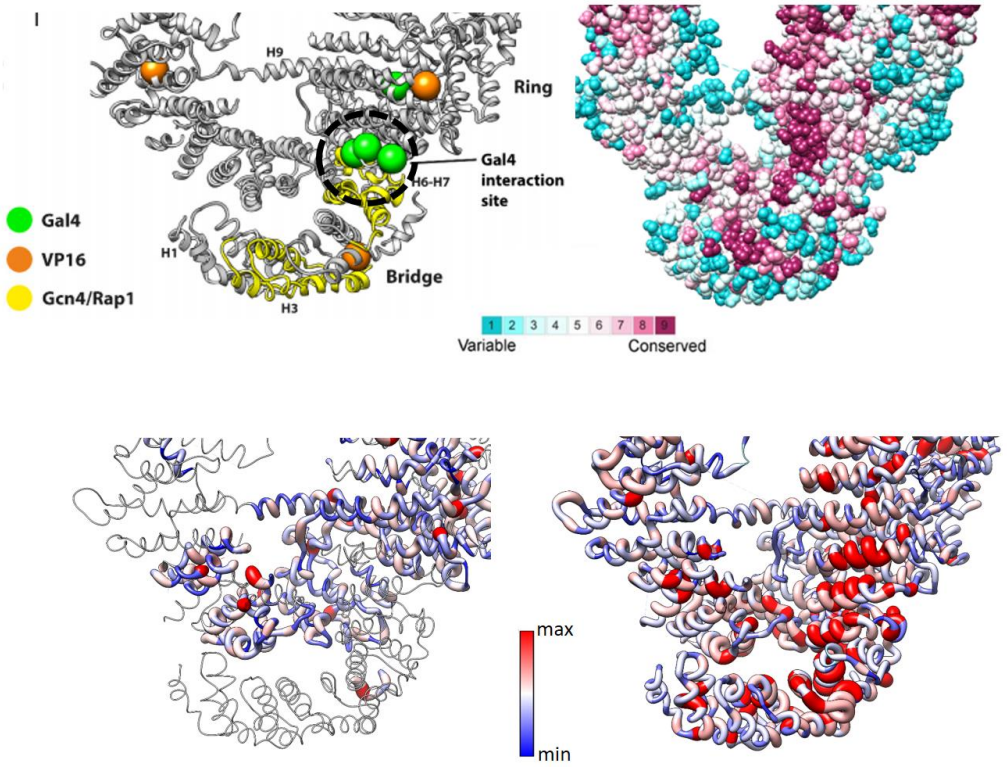
4.3.5 The Tra1 atomic model allows for the mapping of activator binding sites

Tra1 is able to directly interact with transcription activators (hereafter referred to as 'activators') (Brown *et al.*, 2001), which are modular polypeptides composed of a DNA binding domain (DBD) and a transactivation domain (TAD) that can interact with different transcription co-activators upon activator binding to the their cognate DNA sites (1.2.1). Previously published works have focused on mapping relevant Tra1 regions responsible for activator binding via mutagenesis (Knutson and Hahn, 2011; Lin *et al.*, 2012). The availability of a complete atomic model of Tra1 (this study) allowed the mapping of these mutations directly to the structure. These are distributed across Tra1 surface, but also cluster within the N-terminal and adjacent regions as a main target area on Tra1 for activator binding (Fig. 4.21A). More specifically, HEAT repeats H7 and H8 cluster five positions where amino acid substitution impairs Gal4 interaction.

A)



B)



C)

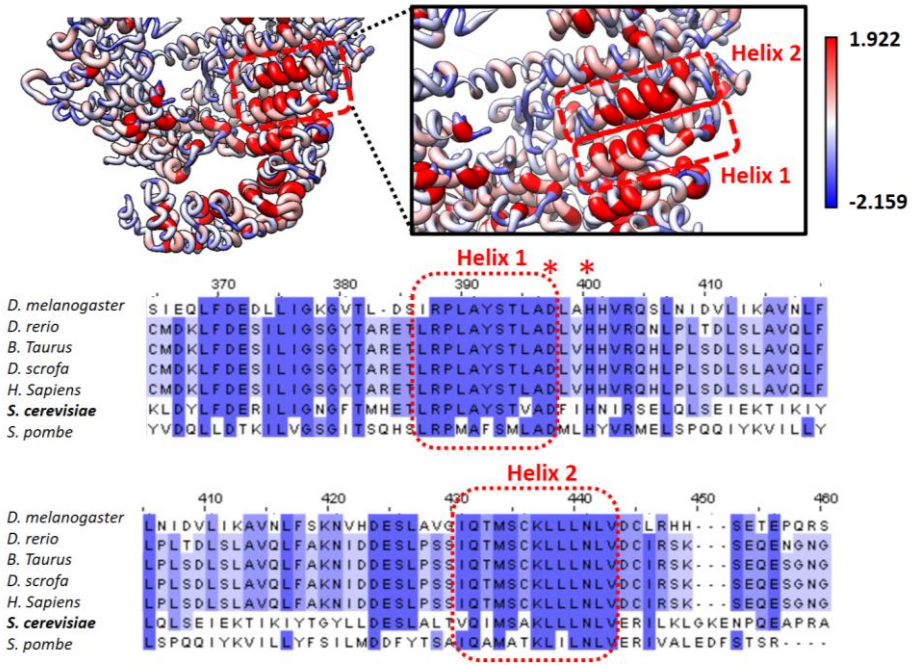


Figure 4.21 | Mutations in Tra1 affecting the interaction with transcription activators and analysis of residue conservation in evolution. A) Activator interaction sites on Tra1. Spheres represent amino acid substitution mutations and yellow ribbons represent deletions affecting SAGA/NuA4 targeting. Mutations affecting the interaction with Gal4, VP16 and Gcn4/Rap1 are represented following the colour code in the legend. A cluster of Gal4 mutations maps into HEAT repeats H7 and H8, suggesting a specific Gal4 interaction region. Mutations affecting the interaction with activators are distributed across the entire structure. (Adapted from Díaz-Santín *et al.*, 2017). **B)** Enlarged view of Tra1 mutations affecting Gal4 interaction (**top left**) as in **A**. Evolutionary analysis (**top right** and **bottom left**) and comparison to other PIKK members (**bottom right**) as shown in Fig. 4.19 and 4.20 are zoomed into the

Gal4 interaction site. **C)** Sequence alignment in Clustal Omega of Tra1 region where mutations affect the interaction with the transcription activator Gal4 (Tra1 Helix 1 in the figure, that belongs to Heat repeat 7). In addition, an adjacent helix (Helix 2, that belongs to Heat repeat 8) also presents solvent exposed and fully-conserved residue positions in evolution from yeast to humans, as is the case for Helix 1. "Star" positions in the alignment of Helix 1 indicate previously described point mutations in Tra1 affecting the interaction with the transcription activator Gal4 (Lin *et al.*, 2012). These positions correspond to Tra1 residues D397 and H400, that are found to be also fully-conserved in evolution. The sequence alignment was rendered in Jalview (Waterhouse *et al.*, 2009).

These Tra1 mutations affecting Gal4 interaction (Fig. 4.21A, B and C) are specific for Gal4 interaction with Tra1 and does not seem to affect the interaction with other activators such as Gcn4 and Rap1 (Knutson and Hahn, 2011). Sequence alignment analysis (Fig. 4.21B, C) shows that this Gal4 interaction site on Tra1 is evolutionary conserved, and sequence alignment of Tra1 against other PIKKs indicates that this region does not align to the sequence of any other member in the family. Some point mutations in Tra1 previously described to disrupt Gal4 interaction (Lin *et al.*, 2012) correspond to fully-conserved residue positions in evolution from yeast to humans (D397 and H400, Fig. 4.21C). Although Gal4 is a yeast activator, the Gal4 interaction region in Tra1 might represent a conserved activator-binding site in evolution for other activators that interact in a similar mode as Gal4 does with Tra1.

In the Tra1 human homolog TRRAP, the oncogenic protein c-Myc has been reported to interact with HEAT repeats H36 to H38 (Park *et al.*, 2001), and interestingly the very N-terminal region of Tra1 contacts the Ring repeats H31 to H38, corroborating the idea that this pole of the protein is generally targeted by activators. Similarly, we find mutations affecting Gcn4 and Rap1 interaction with Tra1 (Knutson and Hahn, 2011), positioned at H3 and H6 to H7 respectively, but do not disrupt coactivator recruitment by Gal4, suggesting that there are specific regions within Tra1 responsible for the binding of different transcriptional activators. Furthermore, the presence of several activator binding regions with a very low level of overlap can function as a tuneable mechanism of transcriptional activation and DNA damage response to assist depending on the specific nuclear requirements. Since Tra1 is part of two different chromatin modifying complexes involved in the regulation of different biological processes, the presence of different activator binding sites in Tra1 could potentially lead to differential recruitment of either SAGA or NuA4.

The analysis of Tra1 surface charge shows a positively charged channel made by the interaction of the Finger and the Ring regions (Fig. 4.22). It possibly represents a binding region for the acidic residues typically found in the transactivation domain of several activators that target Tra1, such as Gal4.

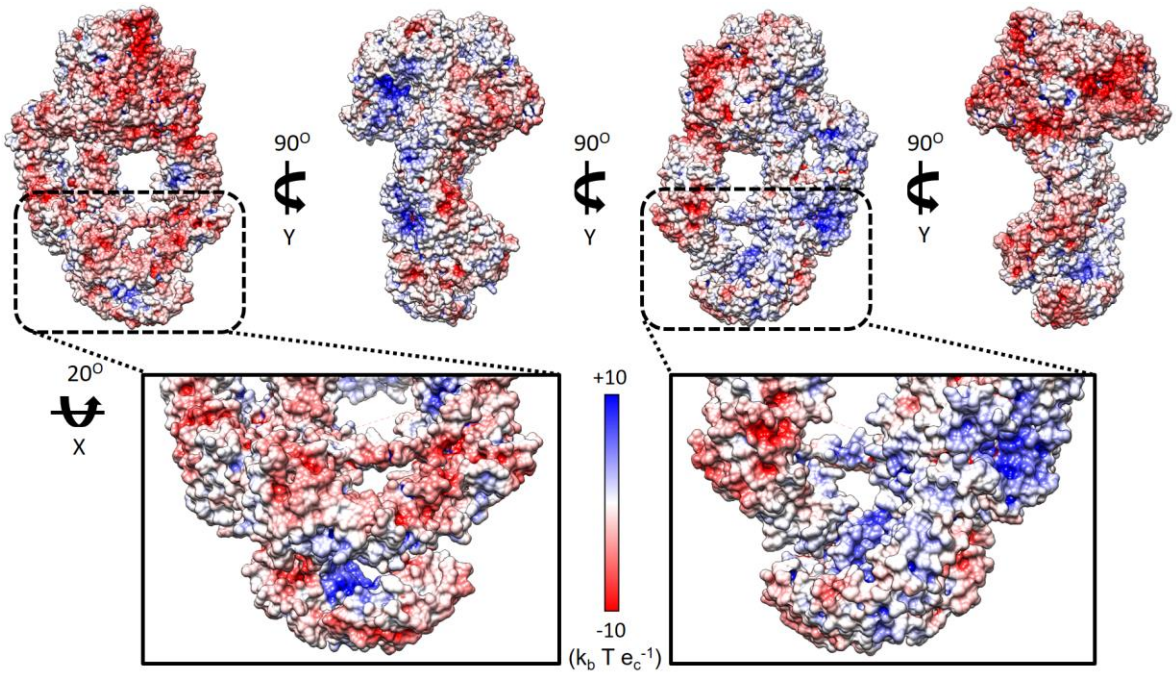


Figure 4.22 | Electrostatic potential analysis of Tra1 surface. A default pH value of 7.0 and the PARSE force field for protonation were selected to perform calculations with PDB2PQR (Dolinsky et al., 2007), implemented in Chimera. To compute the electrostatic potential of Tra1 surface, the calculations generated by PDB2PQR were used as input for APBS (Baker et al., 2001), a tool also implemented in Chimera, to solve the Poisson-Boltzmann equation for the input protonated molecule. As can be inferred from the analysis, the less conserved Tra1 side (as shown in Fig. 4.20) is dominated by negatively charged surfaces, whereas on the more conserved Tra1 side there are two well defined regions with opposite surface charge. The amplified image corresponds to Tra1 N-terminal region and shows the positively charged tunnel contributed by the Finger and Cradle elements. (k_b=Boltzmann constant; T=temperature in K; e_c=electron charge).

It has been shown that activators are able to interact not only in one but in different Tra1 regions (Lin *et al.*, 2012) spread across the structure. For example, in the case of the VP16 activator, its TAD is able to target the C-terminal region of Tra1 contained in the Head domain (Fig. 4.21A). Another possible function for this channel is the accommodation of nucleic acids, given that Tra1 is a subunit present in chromatin interacting complexes and responsible for their recruitment to chromatin. An available structure of Tra1 paralog DNA-PKcs, in complex with the KU70/80 proteins and DNA, supports this possibility (Yin *et al.*, 2017), showing that the accommodation of the DNA could be made by an equivalent channel region in Tra1 (discussed in more detail in section 4.3.9**).**

4.3.6 Tra1 position within the SAGA complex is peripheral

Given its conservation and essentialness amongst eukaryotes, the original hypothesis for Tra1 function was that it plays a role as scaffolding protein within the SAGA complex around which the other subunits would assemble, due its very large mass relative to the complex. However, the fitting of Tra1 atomic model into existing low and medium resolution EM maps of SAGA

(Durand *et al.*, 2014; Sharov *et al.*, 2017; Liu *et al.*, 2019) shows that Tra1 has rather a peripheral position within the complex with limited contact points (Fig. 4.23).

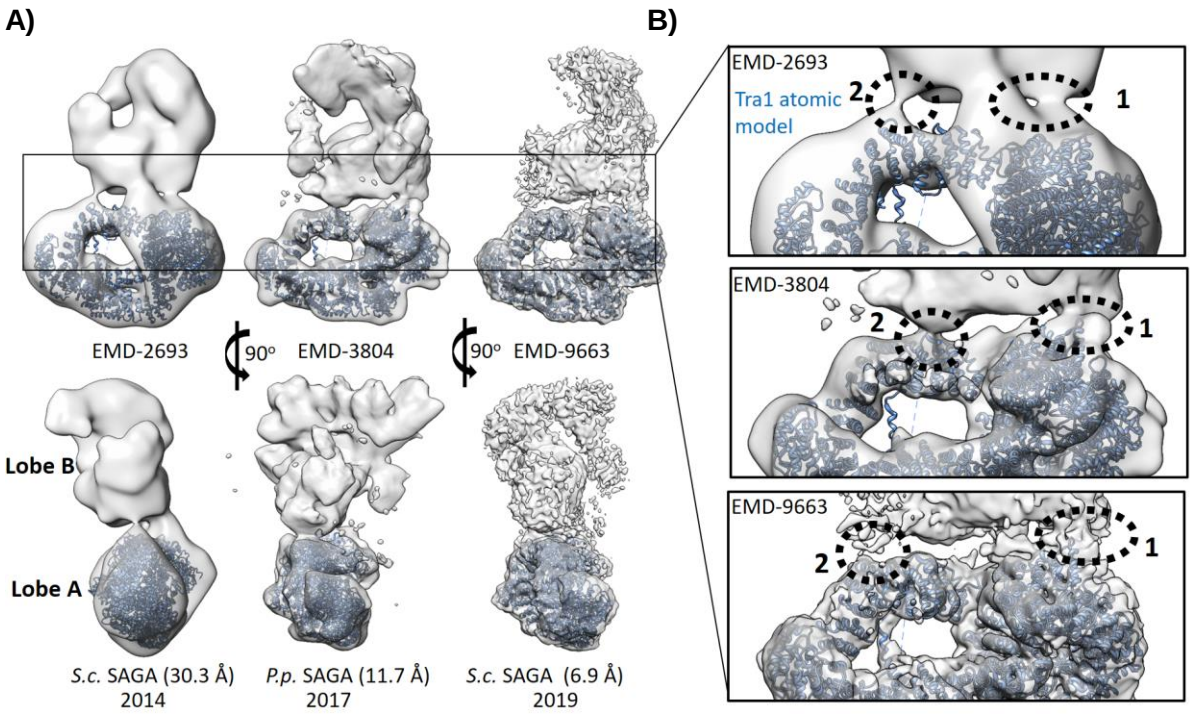


Figure 4.23 | Fitting of Tra1 atomic model within available cryo-EM structures of the SAGA complex. A) Tra1 atomic model (this study) was fitted into available EM maps of the SAGA complex. In the three cases, the Tra1 atomic model fits unambiguously into the SAGA region defined as "Lobe A". **On the left**, Tra1 atomic model fitting into a 30 Å negative stain EM reconstruction of SAGA (EMDB-2693). In the **middle**, Tra1 atomic model fitting into a cryo-EM reconstruction of *P. pastoris* SAGA (EMDB-3804). **On the right**, Tra1 atomic model fitting into a cryo-EM reconstruction of *S. cerevisiae* SAGA (EMDB-9663). **B)** Enlarged view of **A** showing the two main integration points of Tra1 within SAGA. "Position 1" presents the highest electron density values. "Position 2" corresponds to one of the Tra1 regions that did not align to any other PIKK sequence (Fig. 4.19B).

The SAGA complex was previously described to have two lobes, where each lobe was named "Lobe A" and "Lobe B", and Tra1 unambiguously fits into the "Lobe A" (Fig. 4.23 and 4.24). TPR repeats T1 to T7 within the FAT domain and HEAT repeats H41 to H44 within the Ring are regions responsible for the interaction of Tra1 with the SAGA complex (Fig. 4.23B), and previously published BS3 cross-linking experiments (Y. Han *et al.*, 2014; Setiaputra *et al.*, 2015) detected several residues located in close proximity to the interface region, validating the available structural data. In the mentioned BS3 cross-linking analysis, the SAGA subunits Taf12, Spt20, Ada1 and Sgf73 were identified as presenting intermolecular crosslinks to Tra1, and Taf12 in particular is the subunit with the highest number of detected crosslinks to Tra1, meaning that its position within the complex is within or close to the interface area, in concordance with the previously reported models for SAGA. In addition to the interface region, the BS3 cross-linking analysis also detected Tra1 residues on the opposite side to the

interface with SAGA, and cross-linking with subunits located in the “Lobe B”. Although it might seem contradictory, flexible or mobile elements in SAGA could potentially reach this region, as present data suggest that it is a very flexible complex (Setiaputra *et al.*, 2015). The limited contacts of Tra1 with the rest of the SAGA subunits can be analysed by fitting Tra1 atomic model into existing SAGA cryo-EM maps (Fig. 4.23B and 4.24).

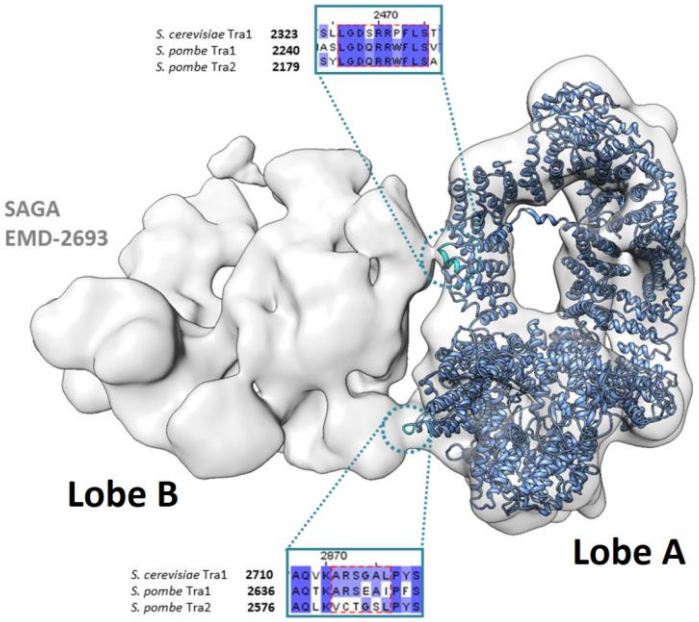


Figure 4.24 | Tra1 integration within the SAGA complex. Tra1 atomic model was fitted into EMD-3804 as in Fig. 4.23, and the two integration regions of *S. cerevisiae* Tra1 within SAGA are outlined. Sequence alignment analysis in Clustal Omega using *S. cerevisiae* Tra1 and *S. pombe* Tra1 and Tra2 is shown for each of the two integration points. A loop within Tra1 likely involved in the integration or interaction with the rest of SAGA subunits is composed of residues R2715, S2716, G2717, A2718, L2719 in *S. cerevisiae* (sequence alignment at the bottom). Interestingly, the equivalent segment in *S.p*. Tra2, a Tra1 homolog that specifically loads into the NuA4 complex, is less conserved in comparison.

4.3.7 Tra1 has a scaffolding role within the essential NuA4 complex

We find two different “Tra” genes in *S .pombe*: Tra1 and Tra2, the first one specifically associates with SAGA and the second one with NuA4 (Fig. 4.24). Interestingly, the deletion of *S. pombe* Tra1 from the genome does not cause lethality. This is probably due to the fact that Tra1 occupies a peripheral position within the SAGA complex (4.3.6), and its role as coactivator and direct target of transcriptional activators can be partially compensated by the SAGA subunits Ada1, Gcn5, Taf6, and Taf12 (Klein *et al.*, 2003; Fishburn, Mohibullah and Hahn, 2005; Reeves and Hahn, 2005; Zhang *et al.*, 2014) (1.2.3.3). However, the function of *S. cerevisiae* Tra1 in NuA4 (Tra2 in *S. pombe*) is essential as scaffolding protein of the complex (Setiaputra *et al.*, 2018; Wang *et al.*, 2018) (Fig. 4.25). Furthermore, the deletion of the Tra2 gene in *S. pombe* makes the cell unviable. Because the NuA4 subunit Esa1 is the only essential acetyltransferase in *S. cerevisiae*, Tra1 mutations affecting the proper function

and/or assembly of the NuA4 complex will cause lethality in the cell, and the same is true for its homolog protein Tra2 in *S. pombe*.

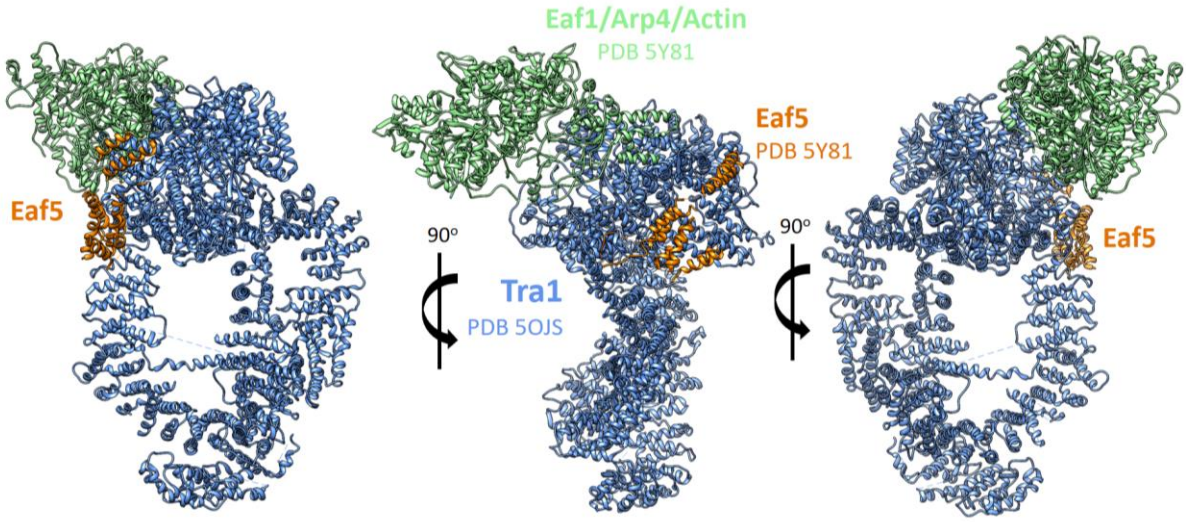


Figure 4.25 | Tra1 within the essential NuA4 complex and its role as scaffolding protein. Tra1 atomic model (blue; this study) was fitted into an existing NuA4 cryo-EM map at 4.7 Å resolution (EMD-6816) and the atomic models for the NuA4 subunits Eaf1, Arp4, Act1 and Eaf5 (PDB 5Y81) are also shown. Eaf5 subunit is in orange, and overlaps one of the areas where Tra1 interacts with the SAGA complex (Fig. 4.23 and 4.24), so that the Eaf5 subunit in NuA4 represents a steric impediment for the SAGA complex to interact with Tra1 when it is already integrated within NuA4 (4.3.8).

4.3.8 Different roles for Tra1 within the SAGA and NuA4 complexes: a comparative structural analysis

The interaction of Tra1 with the SAGA and NuA4 complexes is mediated by partially overlapping regions on Tra1 surface (Fig. 4.26), which might explain why Tra1 does not interact simultaneously with the two complexes, and avoiding the formation of a chimeric complex that would bring them together. In the case of the NuA4 complex, the Eaf5 subunit mediates the interaction with Tra1 in an overlapping region with the SAGA complex "(position 1", Fig. 4.23), and creates a steric impediment for the simultaneous interaction of Tra1 with both NuA4 and SAGA. For SAGA, because of the lower resolution map (11.7 Å), the assignment of the subunits responsible for the interaction with Tra1 is more difficult compared to NuA4, but a clear clash of Eaf5 NuA4 subunit and the region of SAGA interacting with Tra1 (Fig. 4.26A, B) can be observed. A loop in Tra1 potentially involved in the interaction with SAGA can also be identified, covering residues R2715, S2716, G2717, A2718 and L2719 (Fig. 4.24 and 4.26A), and flanked by the very well conserved K2713 and P2720 residues. Sequence alignment of this Tra1 loop shows conserved positions for *S. cerevisae* Tra1 and *S. pombe* Tra1 (only present in SAGA), whereas *S. pombe* Tra2 (only present in NuA4) is poorly conserved in this region in comparison to the previous two, suggesting a role for this loop in the interaction with the SAGA complex (Fig. 4.24).

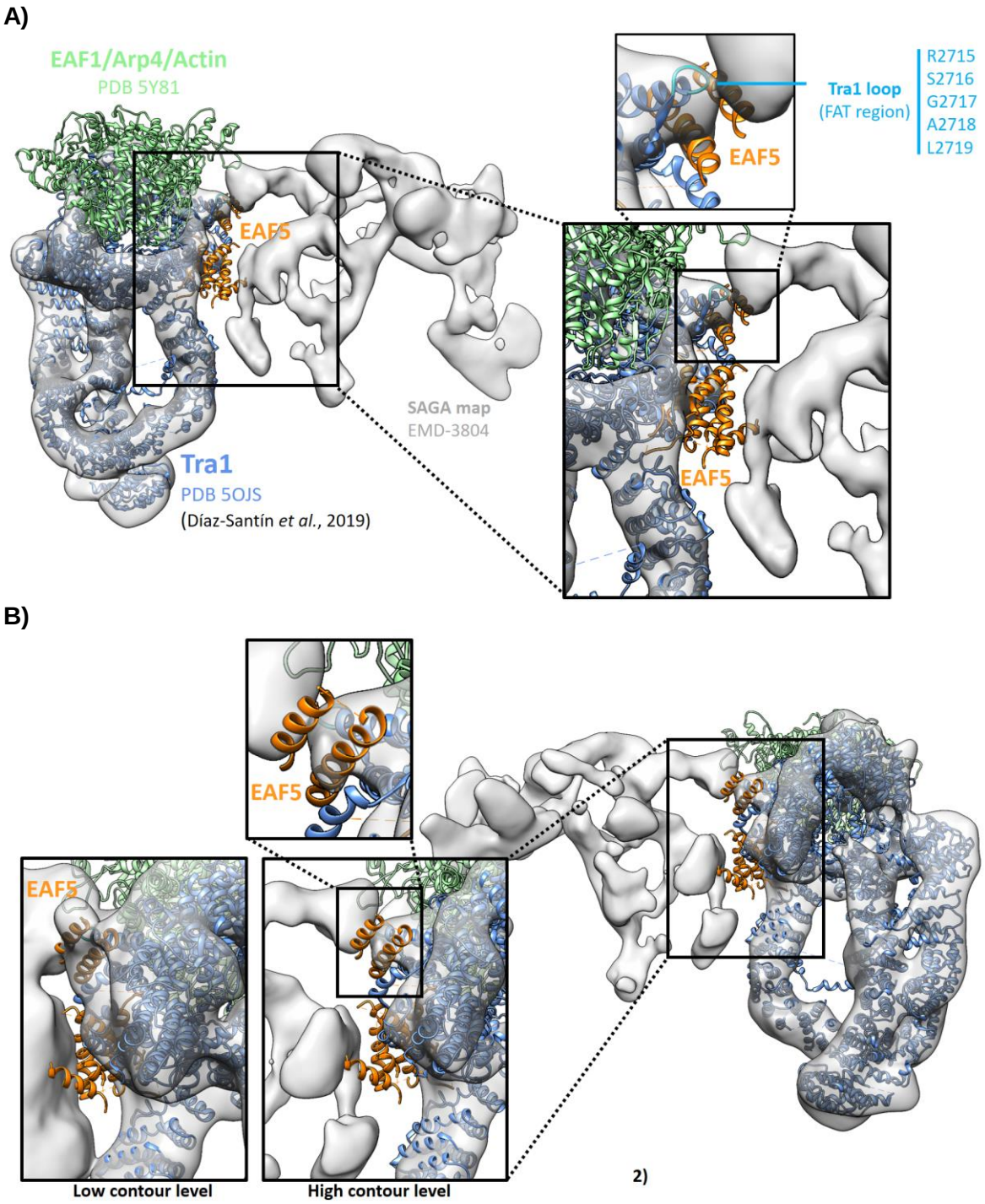


Figure 4.26 | The NuA4 Eaf5 subunit prevents simultaneous loading of Tra1 within SAGA and NuA4. The atomic model of NuA4 Eaf1, Eaf5, Arp4 and Act1 (PDB 5Y81; Wang *et al.*, 2018), together with the atomic model of Tra1 presented in this study (PDB 5OJS, Díaz-Santín *et al.*, 2017) and a cryo-EM map of the SAGA complex at 11.7 Å resolution (EMD-3804, Sharov *et al.*, 2017) are shown in the figure. **A)** The atomic model for a NuA4 core complex composed of Eaf1, Eaf5, Arp4 and Act1 and Tra1 (NuA4 EEAAT) is overlapped with a cryo-EM map of the SAGA complex, showing that the Eaf5 subunit largely prevents the interaction of Tra1 with the SAGA complex when Tra1 is part of NuA4. In addition, a loop found in Tra1 FAT region fits in the density of the map that corresponds to the main integration point of Tra1 with SAGA ("position 1" in Fig. 4.23 and sequence alignment at the bottom in Fig. 4.24). The residues participating in the loop are shown in cyan. **B)** Same as **A** but rotated 180º in the Y axis, showing the clashes of Eaf5 with the main integration point of Tra1 within SAGA (zoomed image). In addition, Eaf5 also overlaps over a more extended region on Tra1 that interacts with SAGA (rendered at different contour levels in the figure).

When overlapping SAGA and NuA4 maps with Tra1 atomic model, Eaf5 looks like a wedge in between SAGA and Tra1 that blocks any potential interaction (Fig. 4.26). The NuA4 cryo-EM map shown in the figure corresponds to a core complex described as NuA4 TEEAA sub-complex and composed of five different subunits (out of the 13 described subunits constituting the native complex). The Piccolo subcomplex, missing in the figure, is expected to bind to a SAGA interacting region on Tra1 (Fig. 4.19), which also prevents any possible interaction between Tra1 and SAGA when Piccolo is interacting with Tra1. There is no available atomic model for NuA4-TEEAA-Piccolo and the available 7.6 Å cryo-EM map contradicts a previously published NuA4 map which shows a different location for the Piccolo subcomplex within NuA4 (Setiaputra *et al.*, 2018). The quality of the available maps have not enough resolution on that particular region to unambiguously fit an available Piccolo atomic model (Xu *et al.*, 2016). Whereas SAGA DUB and HAT modules mediate the interaction of the complex with chromatin, in the case of NuA4 the nucleosome-interacting sub-complexes are Piccolo and TINTIN, and interestingly, the location of these two sub-complexes remains controversial given the contradictory data that has been recently released. However, for the rest of the complex (mainly the SWR1-like module) the two different maps show an identical position of subunits within the complex, mutually confirming the observed data for this large region of the complex.

4.3.9 Tra1 spatial organisation is very similar to DNA-PKcs

The recent abundance of PIKK cryo-EM structures allows the comparative analysis of PIKK members. Because all PIKKs share a highly conserved kinase domain stabilised by a short solenoid region composed of TPR repeats (the so-called FAT region), what differentiates one PIKK to another is the three dimensional organisation of the extensive HEAT repeat solenoid region. This HEAT region folds in a unique architecture depending on the particular PIKK and makes each PIKK structurally unique. In some cases, this region contributes to dimerisation – which is the case for Tel1^{ATM}, Mec1^{ATR} and (m)TOR members. This structural diversity supports the different biological processes of PIKKs where the solenoid region plays an important role as interface for protein-protein interactions (Yoshimura and Hirano, 2016). As previously described in the “Introduction” (1.1; Fig. 1.4), in this regard, we can clearly identify pairs of similar or structurally equivalent PIKKs, where the three pairs of paralog proteins are Tra1^{TRRAP} and DNA-PKcs, Mec1^{ATR} and Tel1^{ATM}, and (m)TOR and SMG1. The similarity between Tra1 and DNA-PKcs (Fig. 4.27) could indicate a putative function on DNA repair given that DNA-PKcs is involved in the DNA damage response triggered by double strand DNA breaks, participating in the NHEJ pathway.

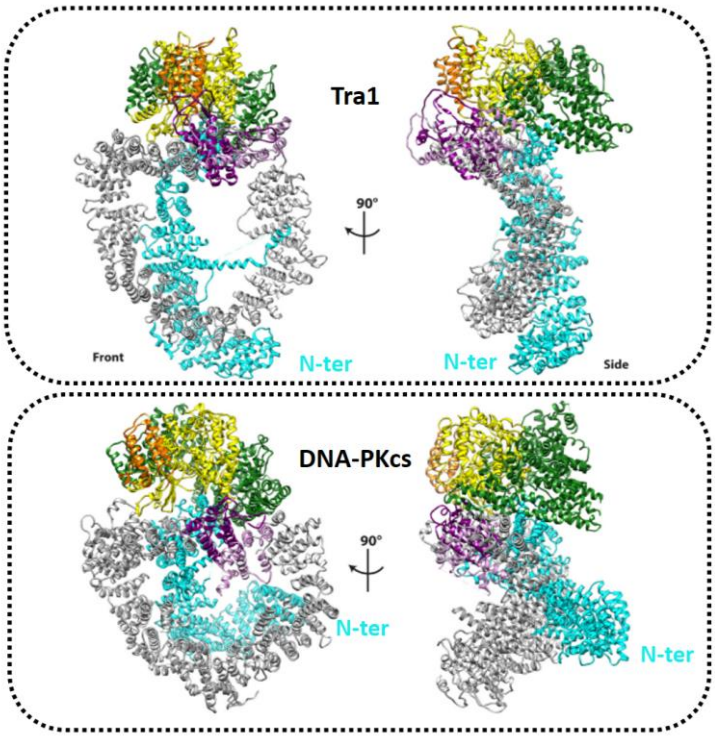


Figure 4.27 | Comparison of Tra1 against DNA-PKcs. DNA-PKcs PDB 5LUQ was used for comparison to Tra1 atomic model (this study). The atomic models are coloured by equivalent regions as in Fig. 4.15. Front and side views of the two proteins show the high level of topological similarity. The N-terminal end of each polypeptide is also indicated in the figure. (Adapted from Díaz-Santín *et al.*, 2017).

Because Tra1 is part of NuA4 and it has been described to participate in DSB repair, the similarity to DNA-PKcs reinforces the hypothesis of Tra1 participating in the DNA damage response. We can identify the same topological regions in the two PIKKs: Finger, Clasp, Ring and Head (Fig. 4.27). The main difference remains at the N-terminal region or Finger, where the very end of Tra1 contacts the Ring at the distal part of the Head region, unlike to DNA-PKcs, where the N-terminal end instead of contacting the equivalent region, turns and contacts a more proximal region to the Head on the Ring (Fig. 4.27 and 4.28). A published structure of DNA-PKcs in complex with the KU70/80 proteins and DNA shows that the Finger and the distal part of the Ring interact with DNA through the channel comprised between them (Yin *et al.*, 2017) (Fig. 4.28A, B). In Tra1, the equivalent channel region is too small to accommodate dsDNA, but the local resolution map (Fig. 4.13D) indicates that this region is the most flexible within the protein, suggesting that it could undergo a conformational change, potentially enabling for the accommodation of dsDNA (Fig. 4.28C) as is the case for DNA-PKcs.

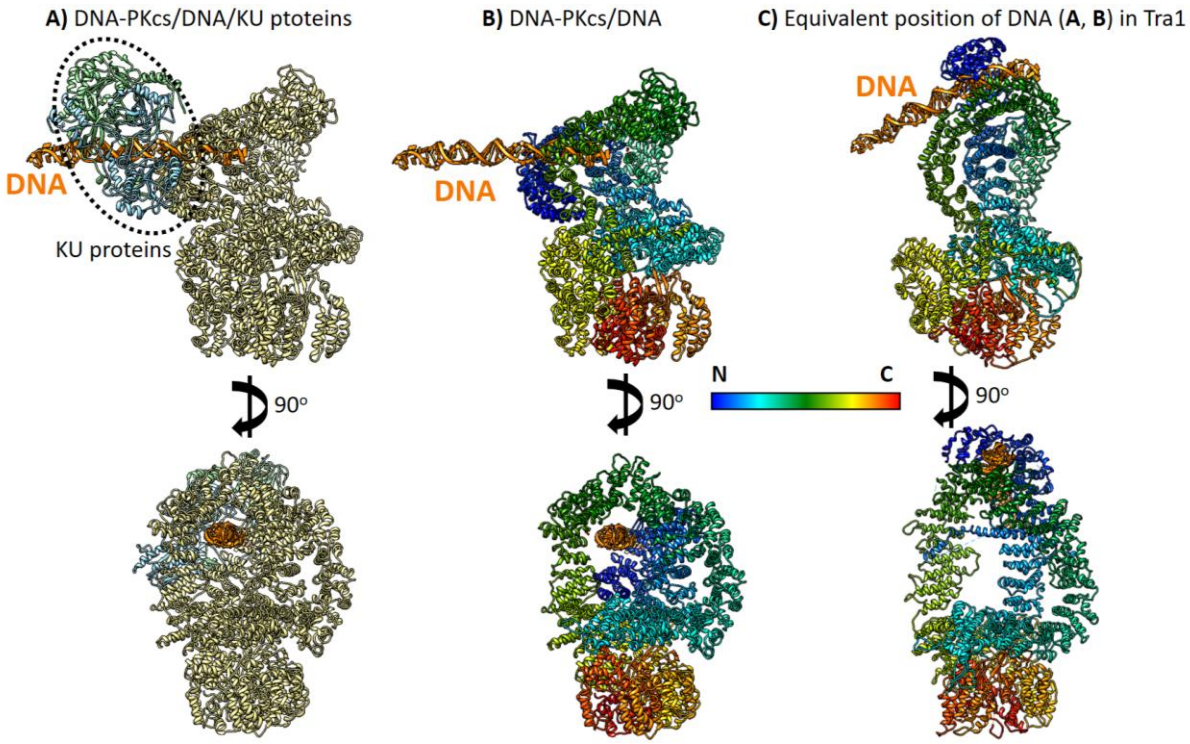


Figure 4.28 | The [DNA]-[KU70/80]-[DNA-PKcs] complex compared to the Tra1 polypeptide. A) A pseudo-atomic model of the [DNA]-[KU70/80]-[DNA-PKcs] complex from a 6.6 Å cryo-EM map is shown (PDB 5Y3R). **B)** Same as **A** but showing the DNA-PKcs polypeptide coloured from N to C-terminal. The atomic model for the KU proteins was removed for clarity, showing only the DNA (in orange). **C)** Tra1 atomic model is shown and coloured from N to C-terminal. For explanation purpose, the position of the DNA in the figure was manually placed in Chimera at the equivalent region in DNA-PKcs, creating a speculative model to illustrate the possibility of Tra1 N-terminal channel interacting with DNA, as is the case for DNA-PKcs. The positively charged tunnel at Tra1 N-terminal region where the DNA was manually docked is shown in Fig. 4.22.

4.3.10 Concluding remarks: The role of Tra1 in transcription initiation

The pre-initiation complex (PIC) in eukaryotes is an assembly of multiprotein complexes with a fundamental role in the activation of gene transcription across the eukaryotic genome (1.2.2), so that transcription by RNA Pol II is achieved with the participation of many other proteins involved in transcription initiation. Mediator (1.2.3.2), a transcriptional coactivator with a role in the stabilisation of Pol II interaction with TFIIH, was initially identified as an essential part of the PIC. Therefore, its function can be partially understood considering Mediator as a "transcriptional chaperone" from a functional point of view only, meaning that it contributes to the formation of TFIIH-cPIC (1.2.3). Other essential complexes also involved in the initiation of transcription of most genes in eukaryotes have been described and solved at the structural level over the recent years, suggesting a more complex model where the originally described PIC participates with other multi-protein complexes to activate transcription. In this model, the role of the conserved Tra1^{TRRAP} subunit is essential given its up-stream role in transcriptional activation by directly interacting with transcription activators. This interaction recruits SAGA to UASs at the promoter DNA region, chemically modifying chromatin through acetylation and

deubiquitination of DNA bound histones and allowing with this modifications further recruitment of chromatin remodeling complexes, which are able to recognise acetylated lysines and remodel chromatin architecture to allow RNA Pol II transcription (Fig. 4.29). Therefore, Tra1^{TRRAP} mutations affecting the interaction with transcription activators or other subunits in the SAGA complex have a deleterious effect on the general transcription levels in the cell (Baptista *et al.*, 2017).

In addition to SAGA (1.2.3.3), the TFIID complex (1.2.3.1) also interacts with transcription activators, and recognises the core promoter DNA to load TBP previous to the assembly of the PIC with the RNA Pol II at its core. Eventually, RNA Pol II in complex with Mediator and TFIIH will be recruited to the promoter DNA to initiate transcription. The previously described model for transcription activation is summarised in Fig. 4.29.

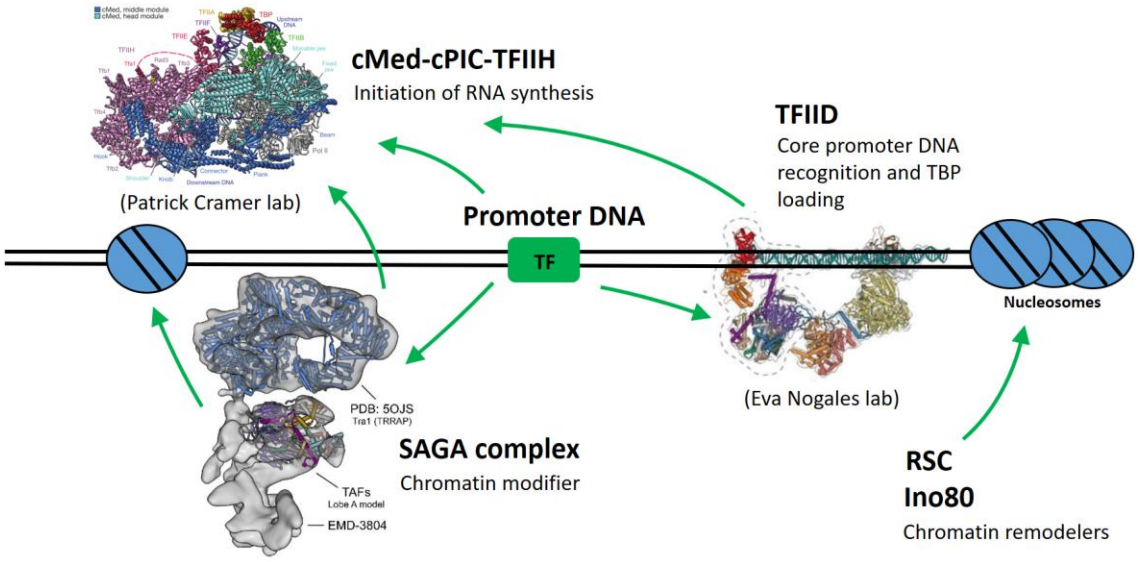


Figure 4.29 | A model for transcription activation in eukaryotes: the role of Tra1 in the assembly of the Pre-Initiation Complex (PIC). A model for the formation of the PIC is depicted schematically representing promoter DNA, chromatin in the form of DNA bound nucleosomes and the different complexes involved in eukaryotic transcription activation. Available structures of the SAGA complex (EMDB-3804), TFIID and cMed-cPIC-TFIIH are shown in the figure. The co-activator complexes SAGA, TFIID and Mediator physically and directly interact with transcription activators, which are bound to the UASs at the promoter DNA, contributing to the recruitment of these co-activator complexes to initiate transcription. Green arrows represent physical interactions resulting in transcription activation. TF: transcription activation factor. (TFIID and the SAGA complex, including Tra1 atomic model (this study; PDB 5OJS) and the TAF proteins were adapted from Patel *et al.*, 2018, and the cMed-cPIC-TFIIH structure was adapted from Schilbach *et al.*, 2017).

5 The interaction of isolated Tra1 with transcription activation domains (TAD)

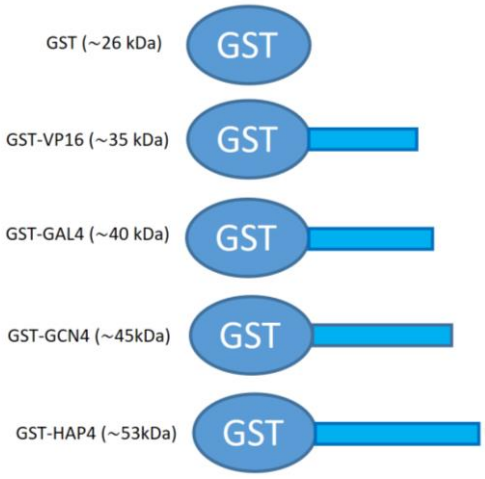
5.0 Introduction

Upon the stimulation of transcription in yeast, the recruitment of transcriptional coactivators correlates with the presence of chromatin remodelers and the general transcription machinery (Galarneau *et al.*, 2000; Peterson and Workman, 2000; Neely *et al.*, 2002). This recruitment is initiated by transcription activators (1.2), proteins composed of two different and well defined regions: a DNA binding region or domain (DBD) which binds to upstream activating sequences (UAS) and a transcription activation or transactivation domain (TAD) that directly interacts with transcriptional coactivators such as SAGA, NuA4, Mediator or TFIID (1.2, Fig. 1.5-1.6, 1.10-1.11, 1.14-1.15). Many yeast activators are rich in acidic residues and in those cases they are also called acidic activation domains (AADs). Because the mechanism of activator binding and specificity to a particular target is not fully understood yet, and because Tra1 is a direct target of several transcription activators (Brown *et al.*, 2001), the goal was to obtain structural insight analysing the interaction of Tra1 with TADs that are previously described in the literature.

5.1 Cloning design for the expression of transcription trans-activation domains (TADs)

Well described and studied yeast activators are Gal4 (YPL248C), Gcn4 (YEL009C), Hap4 (YKL109W) and herpesvirus VP16 (Barley *et al.*, 1995). The activation domain of these transcription activators co-precipitates with purified Tra1 (Brown *et al.*, 2001), a direct target of activators present in both the SAGA and NuA4 coactivator complexes. For that reason, a library of transcription activation domains fused to GST protein (Fig. 5.1A, B) was generated, expressed and purified from *E. coli* (5.2) for *"in vitro"* interaction assays (5.3) and structural analysis by single-particle cryo-EM (5.4).

A)



B)

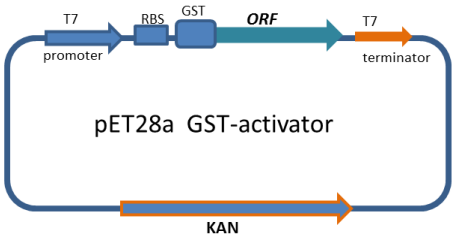


Figure 5.1 | Library of Transcription Activation Domains (TADs). A) Schematic representation of the GST- activation domain library. Different activation domains were cloned as a fusion protein to GST. The approximate size of each fusion protein is shown in the figure. **B)** The pET28a vector is controlled by a T7 promoter for expression in *E. coli* and was used for the expression of the different transcription activation domains in **A**. RBS: ribosome binding site. GST: glutathione S-transferase affinity tag. ORF: open reading frame. KAN: kanamycin selectable marker.

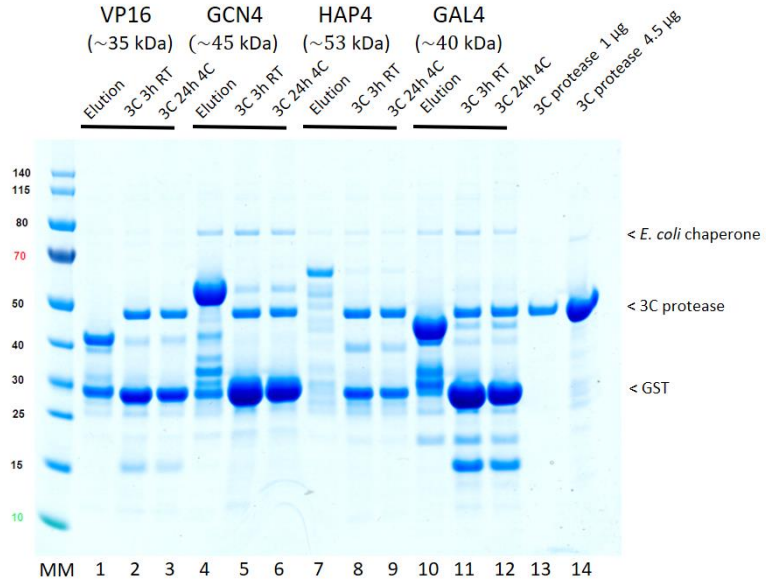
To create a library of TADs fused to GST protein (Fig. 5.1A), the TAD of VP16 (413-490), GCN4 (9-172), HAP4 (330-554) and GAL4 (768-881) were cloned into the bacterial expression vector pET28a(+) to create the "pET28a(+)-GST-3C-TAD" plasmids (Fig. 5.1B). A synthetic "GST-3C-VP16_413-490" DNA sequence (Eurofins, Inc.) was used as template for a PCR reaction (2.1.1) with primers LMDS177/178 (2.0.1, Table 4) to amplify the synthetic sequence for GIA (2.1.4). Initially, the fragment was fused to the yeast expression vector pRS426DQ, and the vector was amplified using LMDS175/176 primers. The amplified DNA fragments were analysed by DNA gel electrophoresis (2.1.2) and gel purified (2.1.3). A two fragments GIA reaction was performed from the purified fragments to generate the "pRS426DQ-GST-3C-VP16 413-490" plasmid (2.0.1, Table 3) as previously described (2.1.5-9). To clone GCN4 (9-172), HAP4 (330-554) and GAL4 (768-881), primers LMDS180/181, LMDS182/183 and LMDS184/185 were used, respectively, to amplify the corresponding TAD from *S. cerevisiae* genomic DNA. To clone the TADs into the pRS426DQ-GST-3C vector (2.0.1, Table 3), primers LMDS174/176 were used to amplify the vector backbone for GCN4 cloning and LMDS174/175 for HAP4 and GAL4. A two fragments GIA reaction generated the plasmids "pRS426DQ-GST-3C-GCN4_9-172", "pRS426DQ-GST-3C-HAP4_330-554" and "pRS426DQ-GST-3C-GAL4_768-881" (2.0.1, Table 3). Finally, the "pET28a(+)-GST-3C-TAD" bacterial expression vectors were generated using primers LMDS196/197 to amplify pET28a(+) (2.0.1, Table 2) and primers LMDS198/199, LMDS198/200, LMDS198/201, LMDS198/202 to amplify GST-3C-VP16_413-490, GST-3C-GCN4_9-172, GST-3C-HAP4_330-554 and GST-3C-GAL4_768-881, respectively, from the newly created "pRS426DQ-GST-3C-TAD" vectors. Two fragments GIA reactions generated the bacterial expression vectors "pET28a(+)-GST-3C-VP16_413-490", "pET28a(+)-GST-3C-GCN4_9-172", "pET28a(+)-GST-3C-HAP4_330-554" and "pET28a(+)-GST-3C-GAL4_768-881" (2.0.1, Table 2). Positive transformant colonies were analysed by colony PCR using T7-fwd and T7-rev primers. Following plasmid DNA miniprep, Sanger sequencing was performed using primers T7-fwd and pET-RP (2.0.1, Table 4).

5.2 Expression and purification of transcription trans-activation domains (TADs)

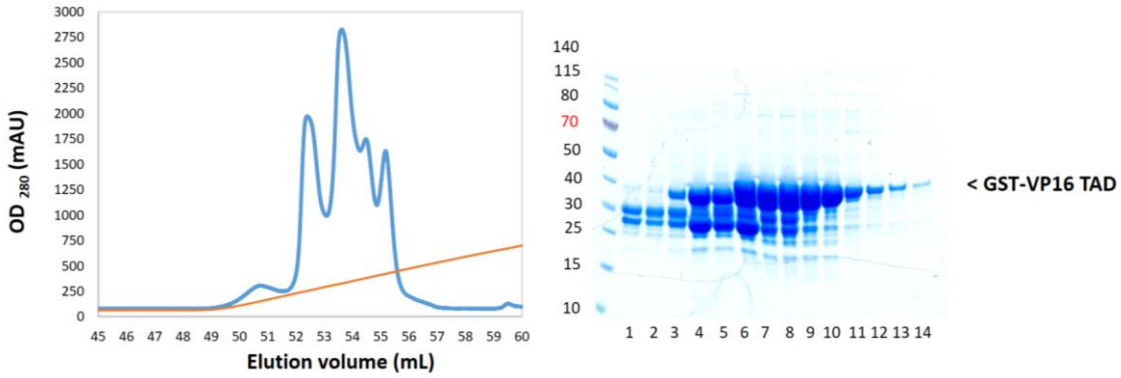
The expression and purification of VP16, Gcn4, Hap4 and Gal4 transcription activation domains was initially performed in Rosetta2 cells and auto-induction medium (2.2.2.1), and it was changed to BL21 pLysS cells and LB medium because it provided cleaner material in the

GST affinity purification step (Fig. 5.2A, lanes 1, 4, 7 and 10). After test expressions, VP16 TAD and Gal4 TAD were selected for further purification steps (Fig. 5.2B-2E). The affinity purified material was applied to a MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with PBS buffer (2.0.5, Table 13) and eluted in a linear salt gradient of 137-1000 mM NaCl (Fig. 5.2B and D). The cleanest fractions of VP16 and Gal4 (Fig. 5.2B, lanes 6-14; 5.2D, lanes 1-8) were collected and subjected to a last purification step through the injection into a Superdex 200 Increase column (GE Healthcare), previously pre-equillibrated with PBS buffer (Fig. 5.2C, E). Clean GST-TAD fractions were stored at -80 ºC in 10% glycerol for further use.

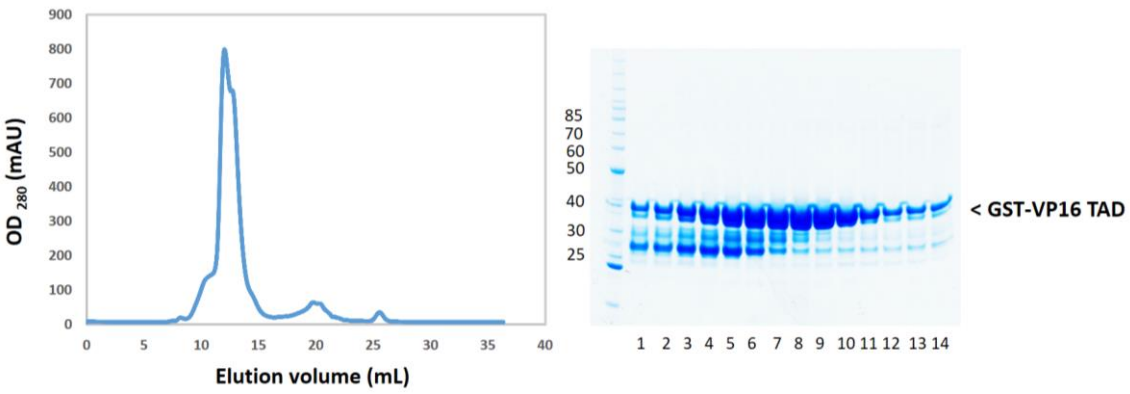
A)



B)



C)



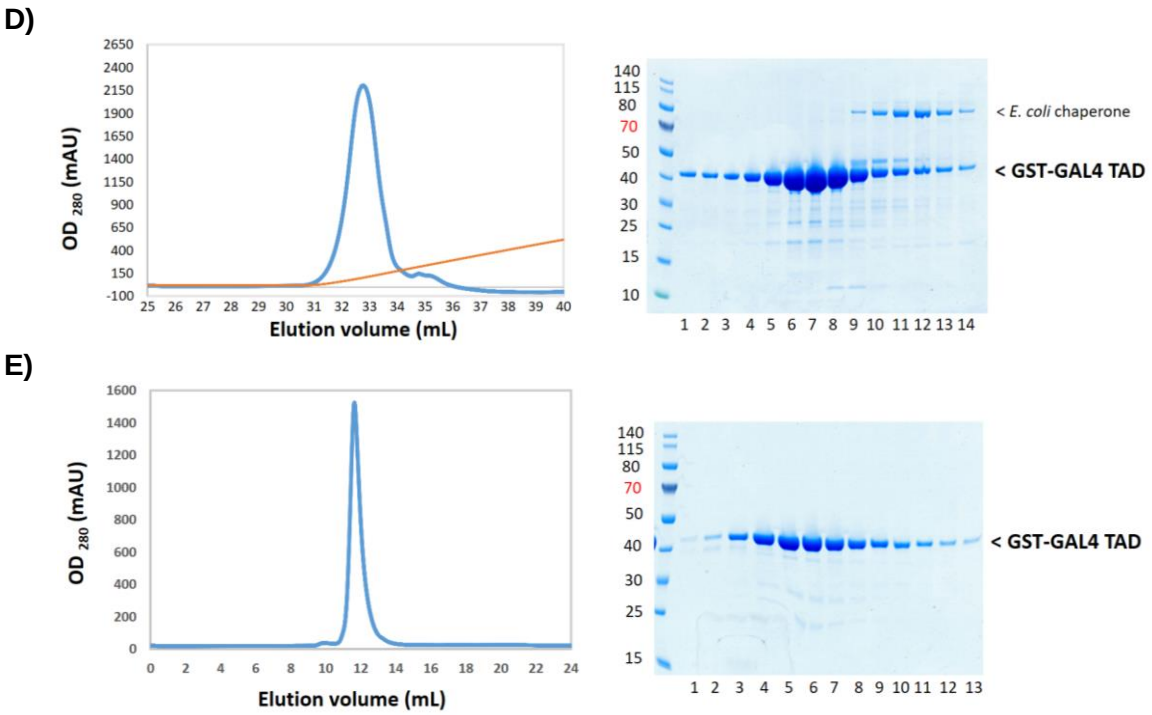


Figure 5.2 | Purification of TADs fused to GST protein. Protein purification of transcription activation domains fused to GST. **A)** Test purification of VP16 TAD (lanes 1-3), Gcn4 TAD (lanes 4-6), Hap4 TAD (lanes 7-9) and Gal4 TAD (lanes 10-12) analysed by 4-12% SDS-PAGE followed by Coomassie staining. GST tag cleavage with GST-3C protease was carried out for analytical purpose only, digesting the purified material in two different conditions (for 3 hrs at RT and for 24 hrs at 4 ^{o}C). **B)** Anion exchange chromatography of affinity purified GST-VP16 on a MonoQ 50/50 GL column. Lanes 1-14: 0.5 ml fractions eluting between 50 and 57 ml. **C)** Superdex 200 Increase column purification of **B** (lanes 6-14). Lanes 1-14: 0.25 ml fractions eluting between 11 and 14.5 ml**. D)** Anion exchange chromatography of affinity purified GST-Gal4 on a MonoQ 50/50 GL column. Lanes 1-14: 0.5 ml fractions eluting between 30 and 37 ml. **E)** Superdex 200 Increase column purification of **D** (lanes 1-8). Lanes 1-13: 0.25 ml fractions eluting between 10.5 and 13.75 ml.

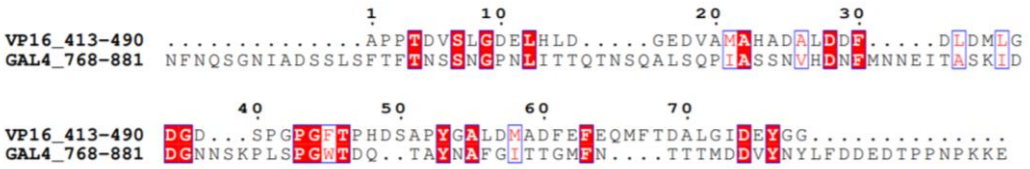
5.3 Testing the Tra1-activators interaction: pull-down assay

From the library of purified TADs (2.2.2), VP16 and Gal4 TADs were selected for initial analysis of the *"in vitro"* interaction with isolated *S.c* Tra1. The pull-down assay served as a test to confirm the interaction previous to the cryo-EM analysis of Tra1 incubated with TADs (5.4). As will be discussed (5.5), it has been described in some cases in the literature the formation of a α-helix in the TAD upon the interaction with the coactivator target. Sequence alignment in Clustal Omega (Sievers and Higgins, 2018) of the purified VP16 and Gal4 TADs (Fig. 5.3A) shows a moderate to low level of homology between the two TADs tested *"in vitro"*. In addition, secondary structure (SS) prediction using the PSIPRED server (Buchan and Jones, 2019) suggests that the content in SS elements is low (Fig. 5.3B), as previously described in the literature (Wright and Dyson, 1999).

To perform the pull-down assays, the interaction of Tra1 with VP16 and GAL4 trans-activation domains fused to GST was assayed immobilising the GST-TAD fusion proteins to glutathione agarose beads and incubating them with MonoQ purified Tra1 as previously described (2.3.3

and 4.1). For the assay, 10 µg of purified GST-TADs (2.2.2) and 10 µg of GST protein (as negative control) were incubated with 10 µl of Glutathione Sepharose 4B resin (GE Healthcare) for 1 hr at 4 ºC. In addition, 10 µl of free Glutathione beads were also used as negative control. After incubation, the beads were washed two times with 100 µl of wash buffer (50 mM HEPES, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 1x cOmplete PI (Roche) and 10% glycerol). Then, 6 µg of purified Tra1 were incubated with each batch of beads (GST-VP16 TAD bound beads, GST-Gal4 TAD bound beads, GST bound beads and free beads) for 1 hr at RT, and the beads were washed three times with 250 µl of wash buffer. Then, the samples were analysed by Western blot as previously described (2.4.5) using anti-FLAG and anti-GST antibodies, showing clear and specific Tra1 binding to both activators, whereas no binding was detected in the case of the negative controls (free agarose beads and GST protein immobilised on agarose beads) (Fig. 5.4).

A)



B)

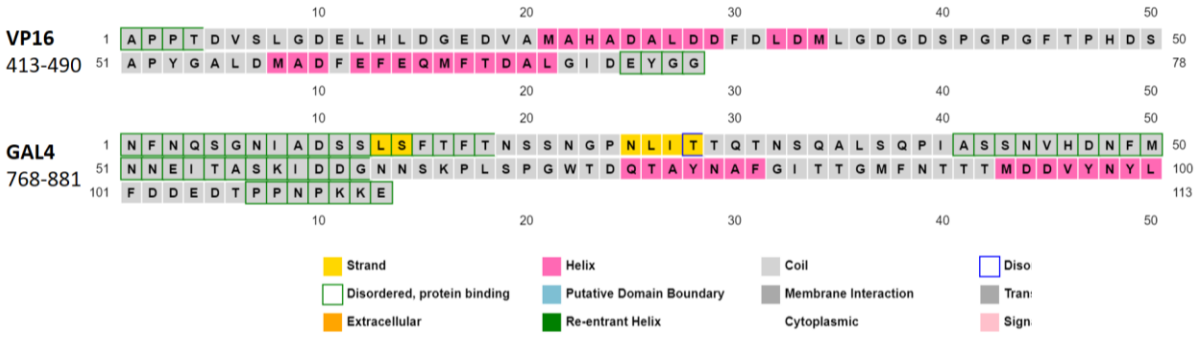


Figure 5.3 | Sequence alignment and secondary structure (SS) prediction of V16 and Gal4 TADs. **A)** Sequence alignment in Clustal Omega of VP16 and Gal4 TADs. The alignment was rendered using ESPrit 3 server (Gouet, Robert and Courcelle, 2003). **B)** Secondary structure prediction analysis of VP16 and Gal4 TAD by PSIPRED.

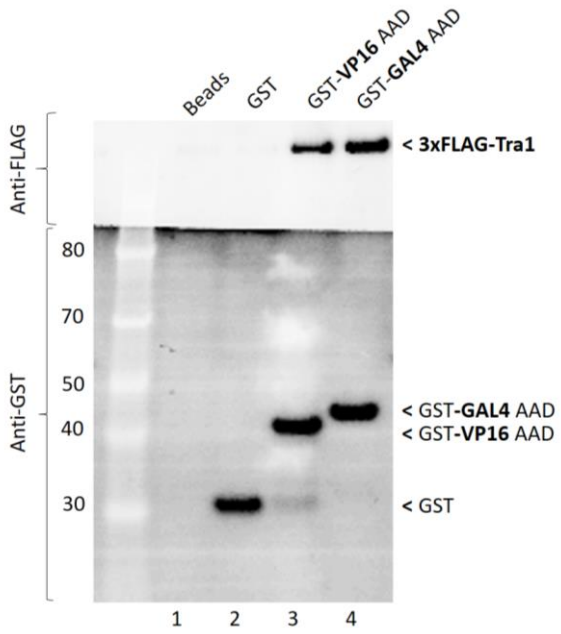


Figure 5.4 | Tra1-activators pull-down assay and Western blot analysis. The transcription activation domains of VP16 and Gal4 fused to GST protein were immobilised in Glutathione Sepharose 4B agarose beads and incubated with MonoQ purified 3xFlag-Tra1 protein. After washing three times the beads, the material was analysed by Western blot using anti-FLAG and anti-GST antibodies (2.0.6, Table 17). As negative controls, free agarose beads (first lane) and GST protein immobilised on agarose beads (second lane) were used. No Tra1 protein was detected in the negative control lanes (1 and 2), whereas a clear band corresponding to the Tra1 polypeptide was visible in the case of the immobilised TAD domains (lanes 3 and 4), showing specific interaction. AAD: acidic activation domain.

5.4 Cryo-EM reconstruction of Tra1 bound to Gal4 TAD

Tra1 was purified as previously described (4.2) and the resulting Tra1 material from the gel filtration step consisting in pure and monodisperse Tra1 protein (at ~ 0.1 mg/mL) was mixed at 1:10 molar excess of Gal4 TAD fused to GST (2.2.2). For cryo-EM grid preparation, the Tra1-Gal4 TAD mixture was applied to Lacey grids with a continuous carbon support film support (~ 3 nm thick, 3.4.2.1). Data collection was performed on a 300 kV FEI Tecnai G^2 Polara microscope equipped with a K2 camera (Gatan) operating in counting mode (Table 5.1). Sample vitrification and data collection on a direct-detector device were performed by Natasha Lukoyanova (3.4.2.1).

Table 5.1 | Cryo-EM data collection and processing of *S. cerevisiae* Tra1 bound to Gal4 TAD	
Nominal magnification	125,000x
Voltage (kV)	300
Electron exposure (e^{-}Å^{-2})	50.4
Defocus range (μm)	-1.8 to -3.5
Calibrated pixel size (Å)	1.39

Symmetry imposed	C_1
Final particle images (no.)	13,966
Map resolution (Å) at FSC=0.143	~ 7 Å

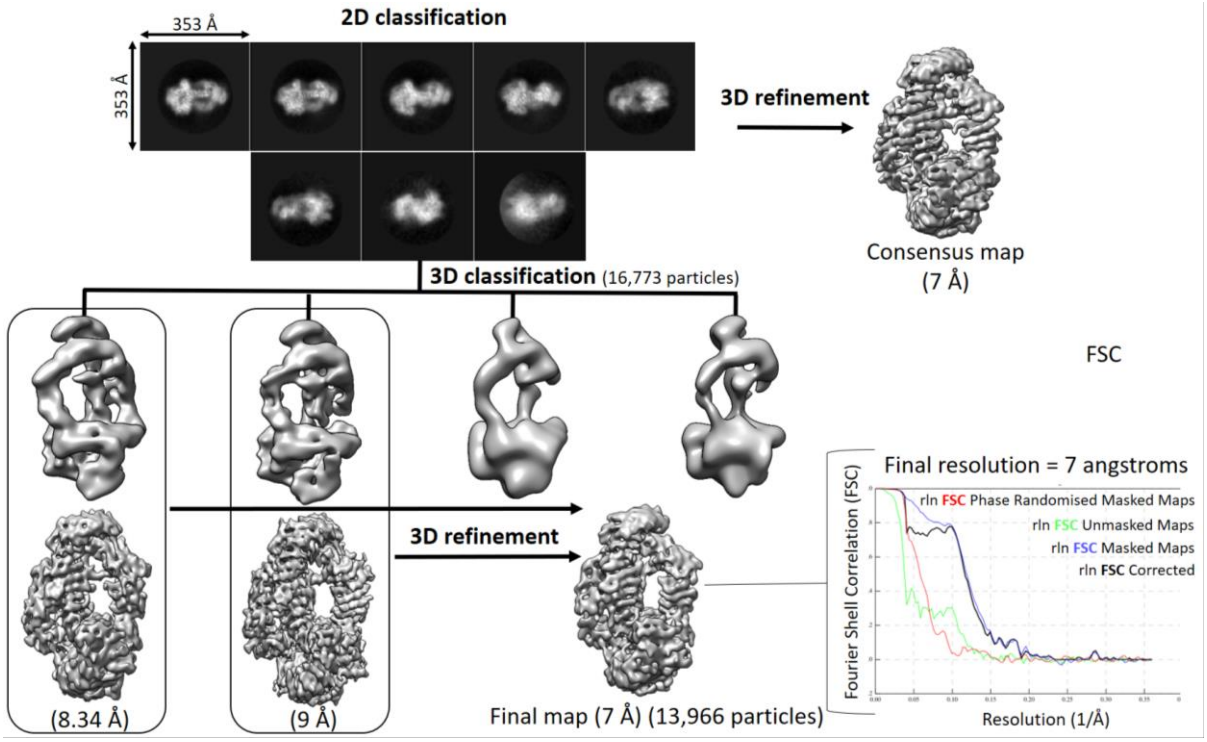
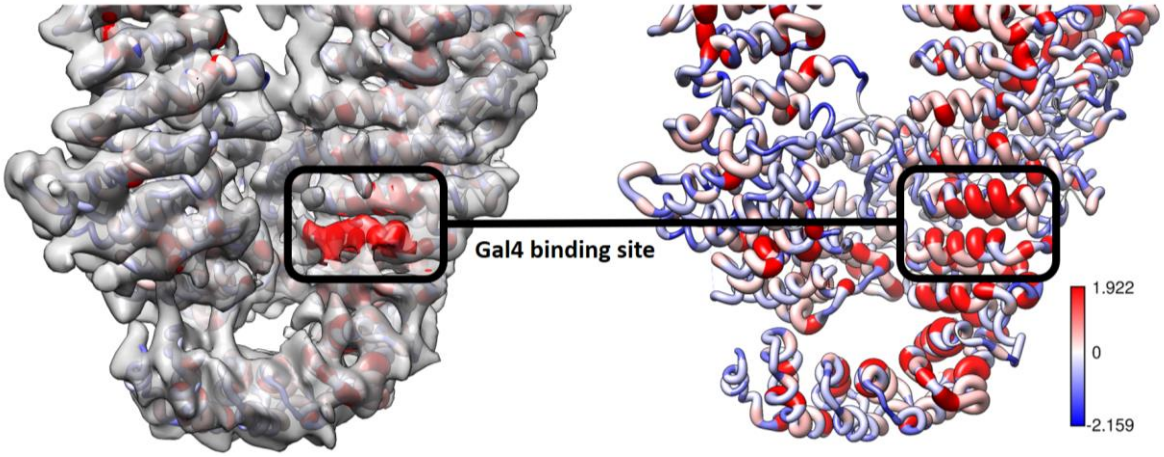


Figure 5.5 | Cryo-EM map of Tra1 bound to Gal4 TAD. The map of Tra1-Gal4 TAD was generated from an in-house dataset (Birkbeck College, University of London) collected on DDD camera. The dataset was processed in RELION 2.0 to a final resolution of 7.0 Å for the final map (at FSC=0.143 criterion). The number of particles for the 3D classification and the final 3D reconstruction are indicated.

A test dataset consisting of 94 cryo-EM movies collected for the Tra1-Gal4 TAD sample were aligned in MotionCor2, and particle picking and 2D classification in Gautomatch and RELION 2.0 were performed as previously described (3.4.2.2). After the selection of the best 2D classes, the consensus map refined to ~ 7 Å resolution (Fig. 5.5). 3D classification into four different classes generated two 3D classes presenting secondary structure features (8.34 and 9 Å resolution). The two classes together refined to 7.0 Å resolution (Fig. 5.5), providing a map with an extra density at Tra1 N-terminal region (Fig. 5.6), and the location of the density is found at the predicted Gal4 binding site on Tra1 surface (Fig. 5.6; Fig. 4.21). Interestingly, when the atomic model generated from the high-resolution apo-Tra1 reconstruction (4.3.1) was fitted into the cryo-EM map of Tra1-Gal4 TAD, no large conformational changes were observed in the protein.

The extra density identified in the Tra1-Gal4 TAD map might potentially correspond to a α-helix, although there is not enough resolution to fit Gal4 residues corresponding to the transactivation domain.

A)



B)

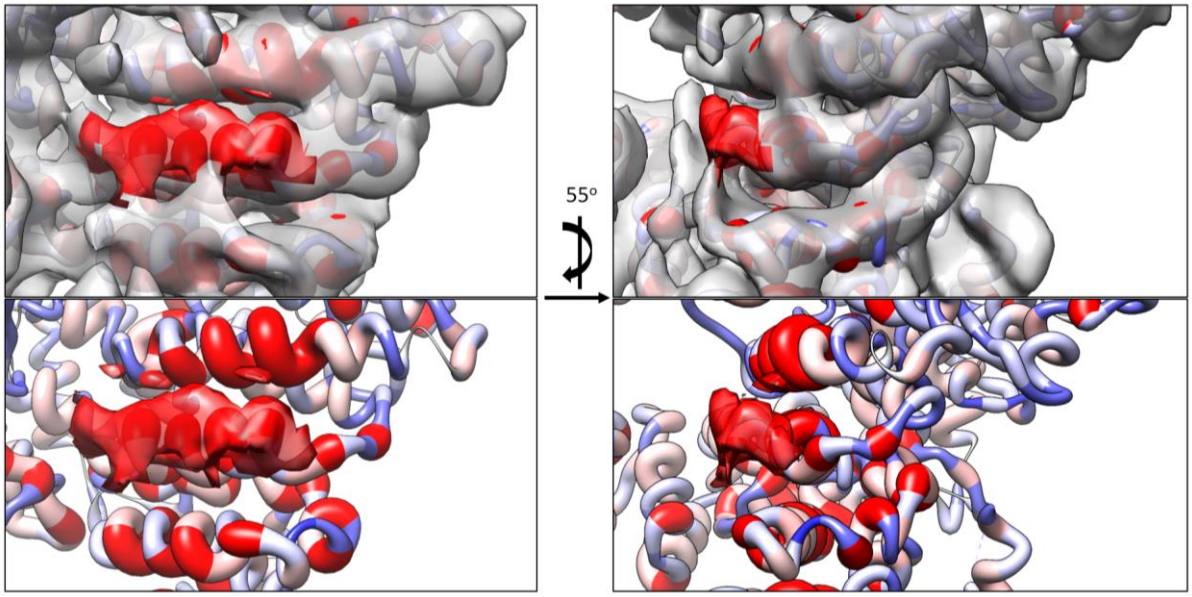


Figure 5.6 | Cryo-EM map of full-length Tra1 incubated with GST-Gal4 TAD. Tra1 atomic model (this study) was fitted into the cryo-EM reconstruction of Tra1 incubated with GST-Gal4 TAD at 7.0 Å resolution (FSC=0.143 criterion). **A)** An extra-density was detected in the cryo-EM map (density shown in red colour) when compared to isolated or apo-Tra1 (4.2 and 4.3). The location of this density corresponds to a region previously described as a Gal4 interaction site (4.3.5). Sequence alignment analysis of Tra1 in different organisms from yeast to humans is plotted as in Fig. 4.21, showing residue conservation in the region where the extra-density is found in the 3D map. Conservation values generated by AL2CO algorithm are shown in the figure legend; red colour means fully conserved residue positions (maximum conservation). **B)** Close up view of the density shown in **A**. Upper images show Tra1 atomic model fitted within the cryo-EM map, whereas lower images only show the atomic model together with the extra-density, for clarity purpose.

5.5 Discussion

Although TADs have been previously described as a nine amino acid domain present in yeast and human transcription activators (Piskacek *et al.*, 2007), the fact that the sequences of the TADs are not well conserved and the coactivators do not present related sequences make it difficult to understand how these activators interact selectively with their targets (Hahn and Young, 2011). Moreover, one activator is able to interact with different unrelated targets and even different binding regions within the same protein. So far, the best-known example of activator-coactivator interaction is between Gcn4 and the Gal11/Med15 subunit of the Mediator coactivator complex (Brzovic *et al.*, 2011; Tuttle *et al.*, 2018). NMR analysis shows that its binding is unstable, presenting micromolar affinity and a half-life shorter to a millisecond. The proposed model is that Gcn4 TAD samples several activator binding regions on Gal11 and the recruitment of Gal11 parent complex Mediator is carried out without a stable protein-protein interaction and dominated by weak and transient interactions. Moreover, NMR analysis reveals that although upon Gcn4 binding a track of approximately eight unstructured residues fold into a helix that interacts with its target protein Gal11, in fact multiple conformations co-exist at the same time, and at least two of these alternative conformations bind to an opposite site on Gal11. Therefore, this activator-coactivator interaction is a so-called "fuzzy complex", where there is no single defined conformation state (Tompa and Fuxreiter, 2008).

The mentioned NMR study used a simplified system - Gcn4 central activation domain and Gal11/Med15 activator binding domain (ABD) - and distinct to Tra1, which is a rigid and very large protein, Gal11/Med15 is a very small and dynamic protein. The cryo-EM reconstruction of Tra1 incubated with Gal4 TAD presented in this study represents to my knowledge the first attempt to analyse an activator bound to its full-length coactivator target using direct detector technology. Although an extra density has been observed matching Gal4 binding site on Tra1 surface, further analysis is still necessary to structurally confirm and validate that the observed extra density corresponds to Gal4 TAD. Several methods can be used to address the analysis. In this regard, cross-linking (CL or XL) is probably the most obvious method, both for cryo-EM analysis to potentially improve the density of the activator on the target in a new reconstruction with increased resolution and also to couple CL to mass spectrometry analysis (CLMS or XLMS) in order to obtain information of Gal4 binding sites on Tra1 surface. Another method for validation could be hydrogen deuterium exchange (HDX) coupled to mass spectrometry analysis (HDX-MS) (Vadas and Burke, 2015), where single residue exchange rate information can be obtained (Masson *et al.*, 2019).

6 The complete atomic model of Mec1^{ATR} in complex to Ddc2^{ATRIP} by single-particle cryo-EM

6.0 Introduction

Of all biological processes where PIKKs are present, DNA repair is one of the best-studied, with many years of accumulated literature for ATR^{Mec1}, ATM^{Tel1} and DNA-PKcs (Blackford and Jackson, 2017) (1.3), being ATR^{Mec1} the only essential member of these three DDR proteins (1.1.2, Fig. 1.3). Although ATR^{Mec1} has been originally described as a canonical DNA repair enzyme, there is also an extensive amount of literature describing its role during the S-phase (Friedel, Pike and Gasser, 2009), specifically in facilitating replication fork progression during replication stress situations as a component of the RSR (1.4.1). ATR in particular has been extensively studied at the cellular and biochemical levels, and there is clear evidence respecting to its involvement in checkpoint signalling (Cortez *et al.*, 2001) and also in activating the Fanconi anemia (FA) DNA repair pathway (Shigechi *et al.*, 2012) (1.4.1).

Mec1^{ATR} has been previously purified in complex with its binding partner Ddc2, (known as ATRIP in humans) and they collectively contain over 3,100 residues (Fig. 6.1A). Before the beginning of the project, only negative stain analysis of the complex was available at structural level (Sawicka *et al.*, 2016). Because atomic knowledge is needed to completely understand the function of this essential and highly-conserved protein in evolution, which functionally overlaps with other PIKK members such as Tra1, the knowledge obtained during the previous project was used to purify *S. cerevisiae* Mec1-Ddc2 complex for single-particle cryo-EM analysis. The overexpression and purification of *S.c.* Mec1-Ddc2 was carried out in its natural host, and the high-resolution structure of the complex is presented here, revealing a dimer of Mec1-Ddc2 heterodimers where the full-length of the Mec1 polypeptide was entirely visible by cryo-EM, representing to my knowledge the first full-length atomic structure of the Mec1 polypeptide.

Before this project began, no high-resolution structure for Mec1 was available. However, a cryo-EM map of the Mec1-Ddc2 complex has since been published (Wang *et al.*, 2017) before this work was complete. Nevertheless, the work shown here resolves the full-length Mec1^{ATR} polypeptide in complex to its binding partner Ddc2^{ATRIP}, in contrast to the published work. The complete atomic structure of Mec1 (6.5.2) provides the structural basis to understand for the first time how the dimer of Mec1-Ddc2 heterodimers is assembled (6.6.2), and shows how the unstructured linker at Ddc2 N-terminal region is connected to the rest of the complex (6.6.3). In addition, the high-resolution cryo-EM structure allowed the modelling of a complete Mec1 Kinase domain including the full-length of the activation loop (A-loop) (6.6.4), which is partially missing in the existing structure of Mec1. Finally, the generated atomic model of the complex, in combination with an existing crystal structure of Ddc2 coiled-coil (CC) domain (Deshpande

et al., 2017), provided an integrative model of the complete complex (6.6.5), improving the understanding of the complex recruitment to RPA-coated ssDNA.

6.1 The Mec1^{ATR}-Ddc2^{ATRIP} complex: plasmid DNA constructs and expression strategy

The expression vectors were designed to express full-length *S. cerevisiae* Mec1 (YBR136W) fused to 3xFLAG tag and full-length *S. cerevisiae* Ddc2 (YDR499W) fused to Protein-A, both tags located at the N-terminal region of each polypeptide (Fig. 6.1B). K. Nagai's expression system (2.3.2.1) was used to coexpress the two proteins, following the same expression scheme as for Tra1 (4.1.2). The expression vectors pRS424DQ and pRS426DQ (2.0.1, Table 3) were selected to insert Mec1 and Ddc2 sequences for coexpression, respectively. To obtain the expression vectors, the DNA for each gene was amplified by PCR (2.1.1) using primers LMDS251/LMDS252 and LMDS247/LMDS248, respectively. DNA amplification was performed from extracted *S. cerevisiae* genomic DNA (kindly provided by Alan C.M. Cheung). Vectors pRS424DQ and pRS426DQ were amplified using primers LMDS249/LMDS250 and LMDS243/LMDS244, respectively. The Protein-A (ProtA) fraction was amplified from the pBS1479-TAP plasmid (2.0.1, Table 3) using primers LMDS245/LMDS246. After DNA gel electrophoresis (2.1.2) and DNA gel extraction (2.1.3) of the PCR amplified fragments (both vectors and inserts), a two fragments GIA reaction (2.1.4) was performed to clone Mec1 and a three fragments GIA to create Ddc2 expression vector. The reaction products were independently transformed into chemically competent XL1-Blue cells (2.1.5 and 2.1.6), and plated into selective LB medium containing ampicillin (2.0.3, Table 8). The resulting colonies were screening by colony PCR (2.1.7) using primers T7-fwd and M13-rev (2.0.1, Table 4), and the positive colonies containing an insert of the expected size were selected for plasmid DNA miniprep (2.1.8). Sanger sequencing (2.1.9) of the purified plasmids was performed using primers T7-fwd, M13-rev, LMDS253-262 (Mec1 specific) and LMDS263-268 (Ddc2 specific). The resulting yeast expression vectors pRS424DQ-3xFLAG-Mec1 and pRS426DQ-ProtA-TEV-Ddc2 (2.0.1 Table 3, Fig. 1B) were cotransformed into BCY123 strain as previously described (2.3.1) for test yeast expression of the cloned polypeptides (2.3.2.2 and next section). The 3xFLAG tag of Mec1 and the Protein-A tag of Ddc2 (Fig. 6.1A and B) allows to purify the complex using a double-tag strategy (6.3 and 6.4).

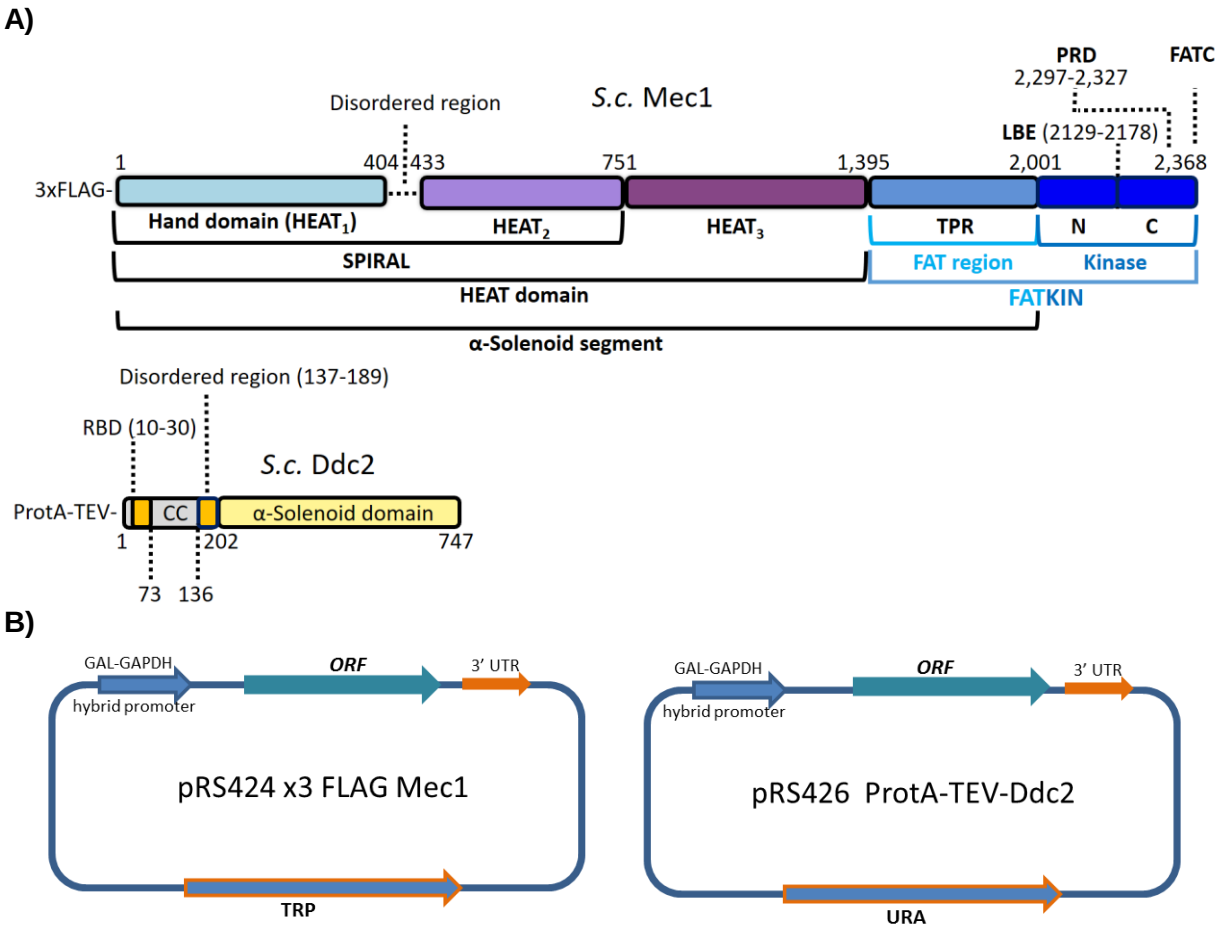


Figure 6.1 | Mec1^{ATR} and Ddc2^{ATRIP} domain organisation and expression vectors. A) Domain organisation of Mec1 and Ddc2 polypeptides in *S. cerevisiae*. RBD: RPA binding motif. CC: coiled-coil domain. **B)** Mec1 and Ddc2 were cloned into the pRS-DQ-424 and pRS-DQ-426 vectors, respectively (2.0.1, Table 3), and cotransformed into *S. cerevisiae* BCY123 strain for protein expression. Trp = *TRP* (tryptophan) 1 gene; URA= *URA* (uracil) 3 gene.

6.2 Test expression analysis

To express the complex, the yeast strain BCY123 previously co-transformed with pRS-DQ-424-3xFLAG-Mec1 and pRS-DQ-424-ProtA-TEV-Ddc2 plasmids (6.1; 2.0.1, Table 3) was plated into CSM-TRP-URA selective medium (2.0.3, Table 9) containing 2% of glucose as main carbon source. For test expression, 50 mL of CSM-TRP-URA medium containing 2% of raffinose were inoculated as a preculture for 1 L test expression (2.3.2.2). However, cell growth was severely compromised in raffinose medium, unlike its normal growth behaviour on agar plates containing glucose. Given that the complex is involved in cell-cycle control, its over-expression may result in toxicity for the cell. Following an existing expression protocol (Burgers, 1999), the presence of a low glucose concentration was enough to allow cell growth, presumably reducing the basal expression level of the complex before induction. Because upon galactose induction the expression of the complex results toxic and cell growth is impaired, the induction was carried out at slightly higher cell density than previously described

for Tra1 (OD_{600}=1.2 in this case). Yeast cells were grown in 1 L of CSM-TRP-URA medium containing either 0.05% or 0.1% of glucose (see “Materials and Methods”, 2.3.2.2), to test for growth and also for protein expression after induction with galactose. The addition of glucose at 0.05 % allowed normal growth rate of the culture and also to express the complex (Fig. 6.2A, B). 1 L of yeast culture was induced with 2% galactose over 24 hrs and protein expression was analysed by a pull-down assay and Western blot (2.4.5) of TCA precipitated material after cell lysis (2.4.4). For the pull down analysis, the cell pellet was lysed with Y-PER (2.3.2.2) and the supernatant was incubated for 2 hrs at 4°C with 50 µl of anti-FLAG M2 agarose gel (Sigma) previously equilibrated with gel filtration buffer without DTT (2.0.5, Table 16). The resin was washed three times with 500 µl of DTT-free WB[2] (2.0.5, Table 16). The supernatant and anti-FLAG M2 resin were analysed by SDS-PAGE followed by Coomassie staining (Fig. 6.2A, lane 1 and 2, respectively). A band corresponding to the 3xFLAG-Mec1 polypeptide was observed in addition to a stronger band corresponding to the ProtA-Ddc2 polypeptide (Fig. 6.2A, lane 2 and 2(b)). To analyse the effect of different expression times and pH, TCA precipitated cell material was analysed by Western blot using 3x OD_{600} units as previously defined (2.4.4). Anti-FLAG M2 antibody was used to detect 3xFLAG-Mec1 polypeptide and anti-GAPDH antibody was used for loading control purpose (2.0.6, Table 17) (Fig. 2B). The anti-FLAG M2 antibody detected Mec1 expression, and presumably the Ddc2 polypeptide is also detected due to the non-specific interaction of the ProtA tag with the Fc region of the antibody.

A) **B)**

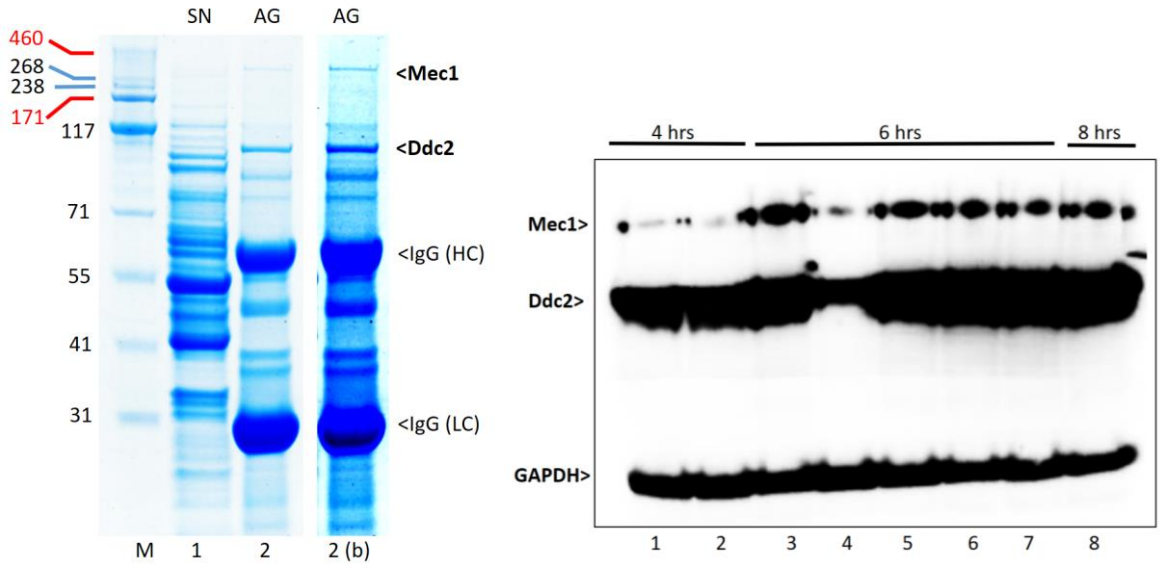


Figure 6.2 | Mec1-Ddc2 test expression. A) 4-12% SDS-PAGE and Coomassie staining of Mec1-Ddc2 test expression and pull-down analysis. SN: supernatant. AG: affinity gel anti-FLAG M2 resin. Ig HC: Ig heavy chain. Ig LC: Ig light chain. Lane 2(b): as lane 2 but with increased contrast. **M**: molecular marker or protein ladder. **B)** Western blot analysis of TCA precipitated material expressing the complex. Although mouse monoclonal anti-FLAG M2 antibody (2.0.6, Table 17) was used in the Western blot to

detect Mec1 protein, the Ddc2 subunit was also detected. Presumably, the Protein A tag present in Ddc2 interacts non-specifically with the antibody. Anti-GAPDH antibody was used as loading control (2.0.6, Table 17). **Lane 1:** 4 hrs induction, colony no. 1, pH balanced (20 mM HEPES pH 7.0). **Lane 2:** 4 hrs induction, colony no. 1, pH balanced (20 mM HEPES pH 7.0). **Lane 3:** 6 hrs induction, colony no. 1. **Lane 4:** 6 hrs induction, colony no. 2. **Lane 5:** 6 hrs induction, colony no. 2, pH balanced (20 mM HEPES pH 7.0). **Lane 6:** 6 hrs induction, mixture of colonies, 0.05% of glucose (instead of 0.1% as in the other cases). **Lane 7:** 6 hrs induction, mixture of colonies. **Lane 8:** 8 hrs induction, mixture of colonies. GAPDH: endogenous *S. cerevisiae* Glyceraldehyde-3-Phosphate Dehydrogenase.

6.3 Large scale expression, EM analysis and optimisation assays

Given the early success of the test expression and pull-down assays, the expression volume was increased to 12 litres. The resulting affinity purified material was used to perform a set of test purifications and optimisation assays (6.3.1, 6.3.2 and 6.3.4) and also to obtain a preliminary cryo-EM map of the complex (6.3.3), as will be described in the next sub-sections.

6.3.1 Large size expression and purification tests

Large scale expression was performed growing yeast cells in 12 L of casamino acid medium (2.3.4.1). After cell harvesting (2.3.2.4) and cell lysis in a freezer mill unit (2.3.2.5), affinity purification was carried out following the same procedures as for Tra1 (4.1.3) but using TEV cleavage instead in combination with 3xFLAG peptide in the elution buffer to elute the protein from anti-FLAG M2 affinity resin (see “Materials and Methods”, 2.3.4.2). Small white insoluble clumps were observed in the first eluted fraction, indicating the presence of precipitated protein, presumably due to a high concentration of Mec1-Ddc2 in this fraction, which was confirmed by SDS-PAGE and Coomassie staining (Fig. 6.3A, lane 1). The material eluted efficiently from the resin combining the 3xFLAG peptide together with TEV protease (Fig. 6.3A, lane 4). The eluted material (Fig. 6.3A, lanes 5-9) was fractionated into 0.5 mL fractions in Eppendorf tubes and stored at -80 ^{o}C.

The obtained material was initially used for test purification purpose. Given the high level of protein purity obtained after the affinity purification step (Fig. 6.3A), only a single additional purification step was needed, consisting in gel filtration chromatography on a Superpose 6 Increase 10/300 GL column (GE Healthcare). To test the behaviour of the sample into the column, 0.5 mL of affinity purified material (fraction E1, ~ 0.3 mg) were spun-down at full speed in a bench micro-centrifuge unit (Eppendorf Centriguge 5415-D) for 30 min at 4^{o}C and injected into the Superpose 6 column (Fig. 6.3B). The eluted material was analysed by SDS-PAGE (Fig. 6.3C). The chromatogram shows a first peak at ~ 8 mL corresponding to Ddc2 polypeptide aggregates (Fig. 6.3B, 3C lanes 1-3), while the Mec1-Ddc2 complex is found in the second peak at ~ 11-14 mL (Fig. 6.3B, 3C lanes 5-11), approximately matching the expected elution volume for a 720 kDa complex in this column. The last large peak in the chromatogram corresponds to the 3xFLAG peptide used for protein elution. The top band on

the gel (Fig. 6.3C, lanes 5-11) stained stronger as expected for a 1:1 stoichiometric complex where one polypeptide - Mec1 in this case - is larger than the other one - Ddc2.

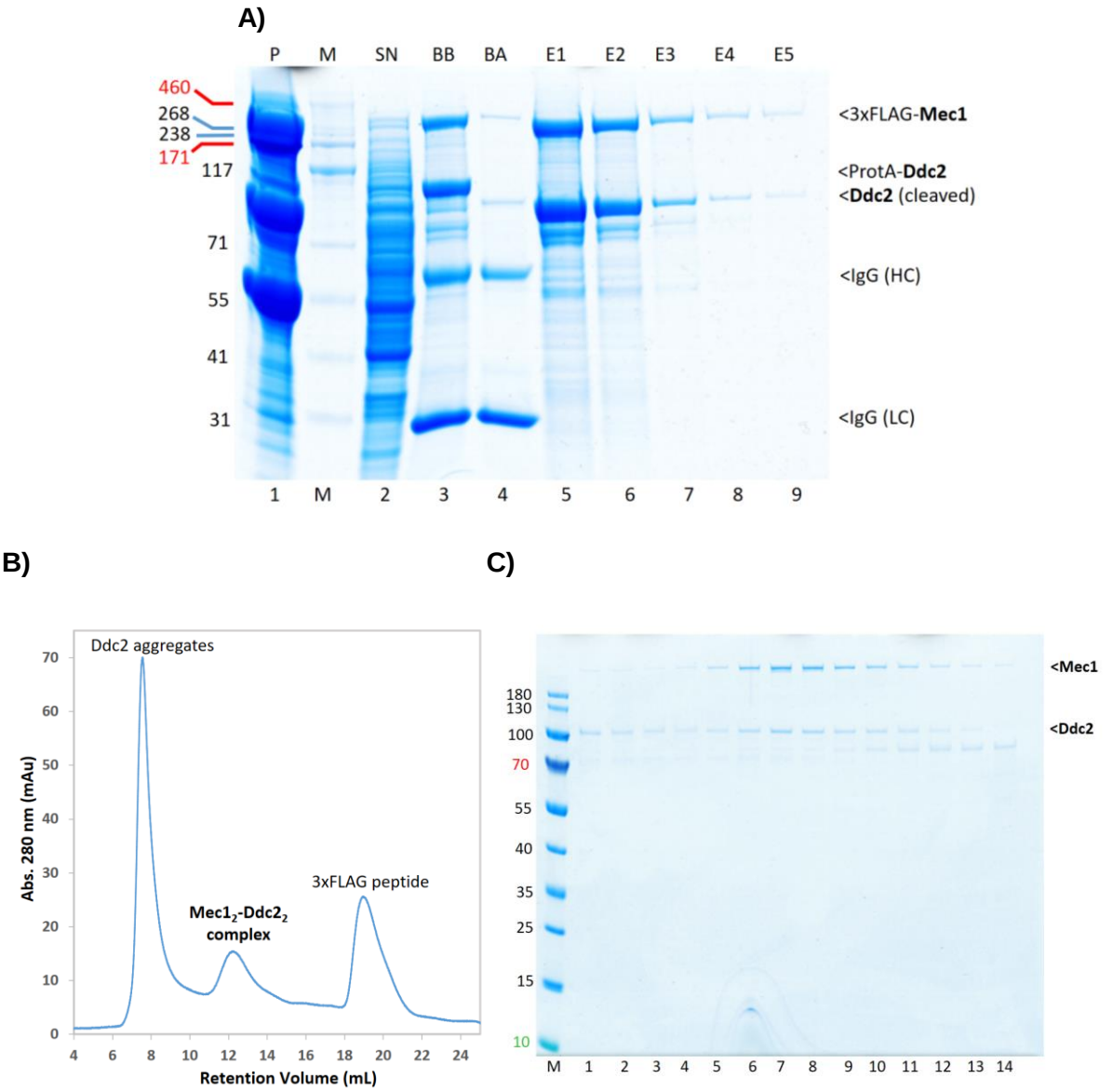


Figure 6.3 | Mec1-Ddc2 protein expression and purification. A) 4-12% SDS-PAGE analysis and Coomassie staining of affinity purified Mec1-Ddc2 complex. Instant blue was used to stain the gel (2.0.6, Table 17). **B)** Chromatography profile of the Superose 6 Increase 10/300 GL column. The first peak in the chromatogram corresponds to Ddc2 aggregates. The main peak of monodisperse Mec1-Ddc2 complex is found between 11-14 ml of elution volume and corresponds to a dimer of heterodimers architecture (Mec1$_2$-Ddc2$_2$). **C)** 4-12% SDS-PAGE analysis of **B**. Lanes 1-14: 0.5 ml fractions eluting between 9 and 16 ml.

6.3.2 BS3 cross-linking assays

To increase the complex stability during vitrification for single-particle cryo-EM analysis and also to potentially improve the range of particle orientations on the grid (Bernecky *et al.*, 2016), the sample was subjected to BS3 (Bis(sulfosuccinimidyl) suberate) crosslinking (2.3.4.3), an amine to amine water soluble crosslinker. Following manufacturer guidelines and to optimise

the crosslinking, the sample was titrated with different BS3 concentrations (Fig. 6.4A, lanes 2-7). All tested BS3-concentrations fully crosslinked the complex, including the lowest at 1.5 mM of BS3 (Fig. 6.4A, lanes 2 and 3). Thus, for an initial test to prepare cryo-EM grids, 0.5 mL of FLAG eluate (E1 fraction, ~ 0.3 mg; Fig. 6.3A, lane 5) were subjected to crosslinking at 1.5 mM of BS3 concentration and purified by gel filtration (Fig. 6.4B). No bands corresponding to free Mec1 or free Ddc2 were observed when analysed by SDS-PAGE and Coomassie staining (Fig. 6.4C), and the chromatogram of the BS3 crosslinked sample showed a similar profile to the one observed for the BS3 free complex (Fig. 6.3B and 6.4B). A very small shift to a slightly larger molecular size was detected, in agreement to the expected increase in MW after the addition of BS3 molecules to the complex. Therefore, under the tested crosslinking conditions, the population of the complex in the sample was shown to be very similar to native conditions, meaning that the complex is still monodisperse. SDS-PAGE analysis of the gel filtration chromatography shows one single band corresponding to the complex (Fig. 6.4C), and reproducing the results obtained in the titration assay (Fig. 6.4A).

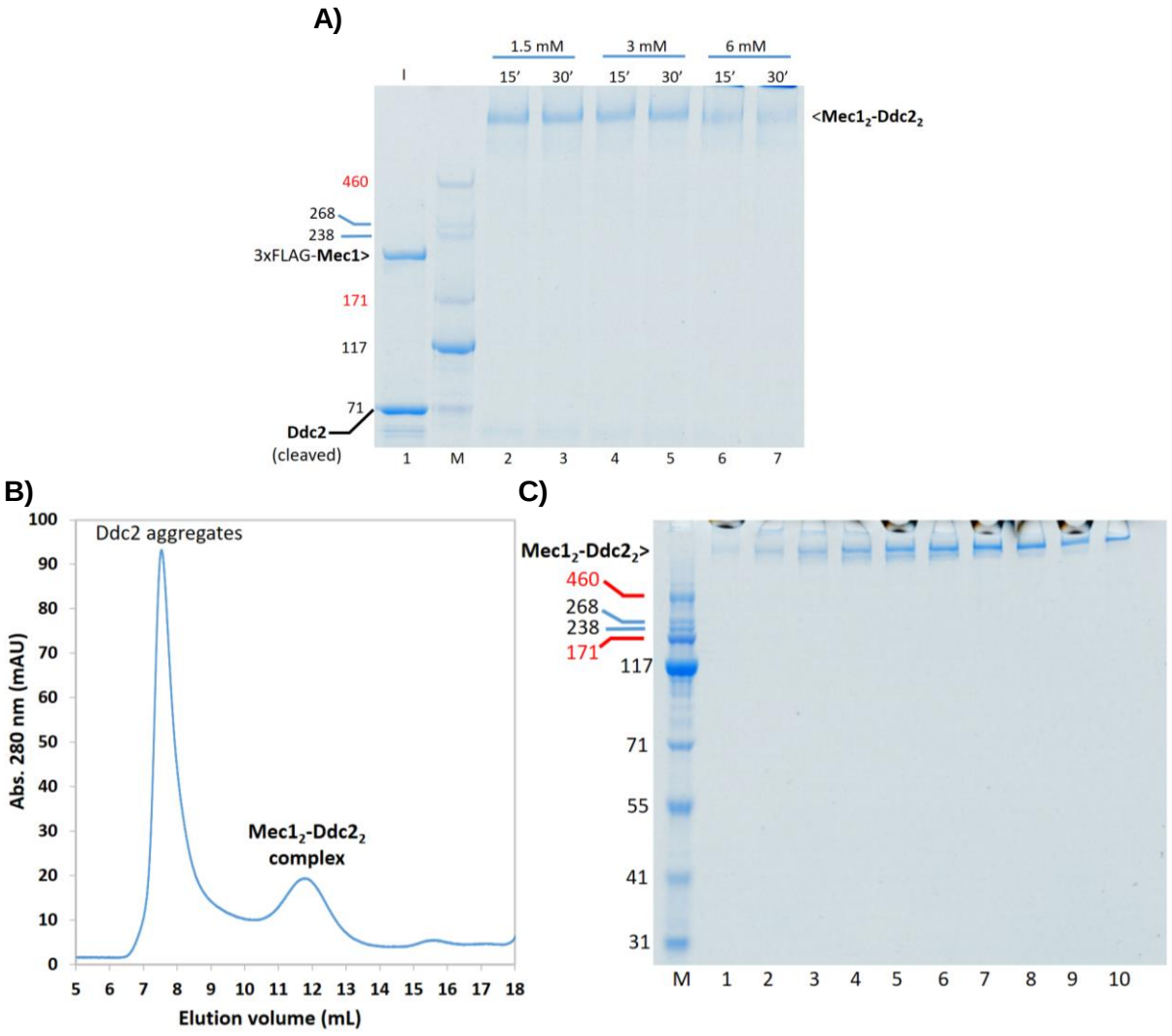


Figure 6.4 | BS3 crosslinking of the Mec1-Ddc2 complex. A) BS3 titration assay of Mec1-Ddc2. For the crosslinking of affinity purified Mec1-Ddc2, the indicated BS3 concentration was incubated with the complex for 15 min at 30 ºC using 1.5 mM of BS3 (lanes 2 and 3), 3 mM (lanes 4 and 5) and 6 mM (lanes 6 and 7). **B)** Superose 6 Increase 10/300 GL chromatography profile of BS3 crosslinked Mec1-Ddc2 sample. **C**) 4-12% SDS-PAGE and Coomassie staining of **B**. Lanes 1-10: 0.25 ml fractions eluting between 10.5 and 13 ml.

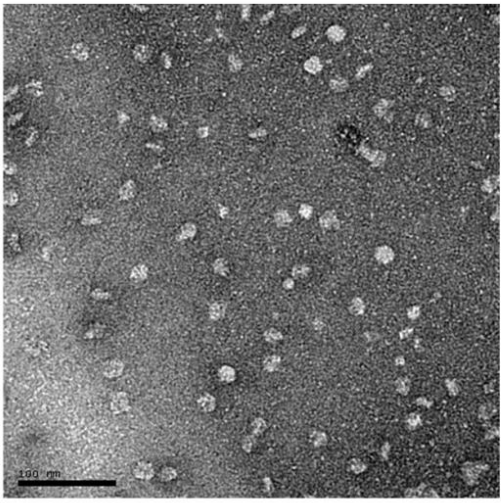
6.3.3 Preliminary EM analysis and medium resolution cryo-EM map

To initially assess protein monodispersion I performed a preliminary negative stain analysis (3.1) of the purified Mec1-Ddc2 complex that showed a similar shape when compared to a previously published reconstruction (Burgers *et al.*, 2016) (Fig. 6.5A, image on the left). Using the purified sample, the vitrification of the complex was performed by Natasha Lukoyanova on Lacey grids (Fig. 6.5A, image on the right), as well as data collection of a test cryo-EM dataset composed of 185 movies. A continuous support film was used to absorb the protein to the surface and increase the concentration of particles on the grid, so that several sample applications were performed on commercially available Lacey grids coated with a thin layer of carbon (~ 3 nm thick) (see "Materials and Methods", 3.4.3.1). Images were collected in-house at Birkbeck College (University of London) on a 300 kV FEI Tecnai G^2 F30 Polara microscope equipped with a K2 camera (Gatan) and operated in counting mode (Table 6.1).

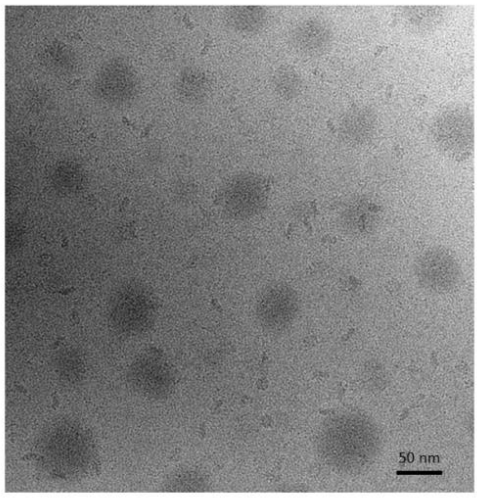
I processed the 185 cryo-EM movies resulting in a 7.2 Å cryo-EM map when C2 symmetry was applied to the reconstruction, containing a final number of 4,353 particles (Fig. 6.5B, Table 6.1).

A)

Negative stain micrograph



Cryo-EM micrograph (Lacey)



B)

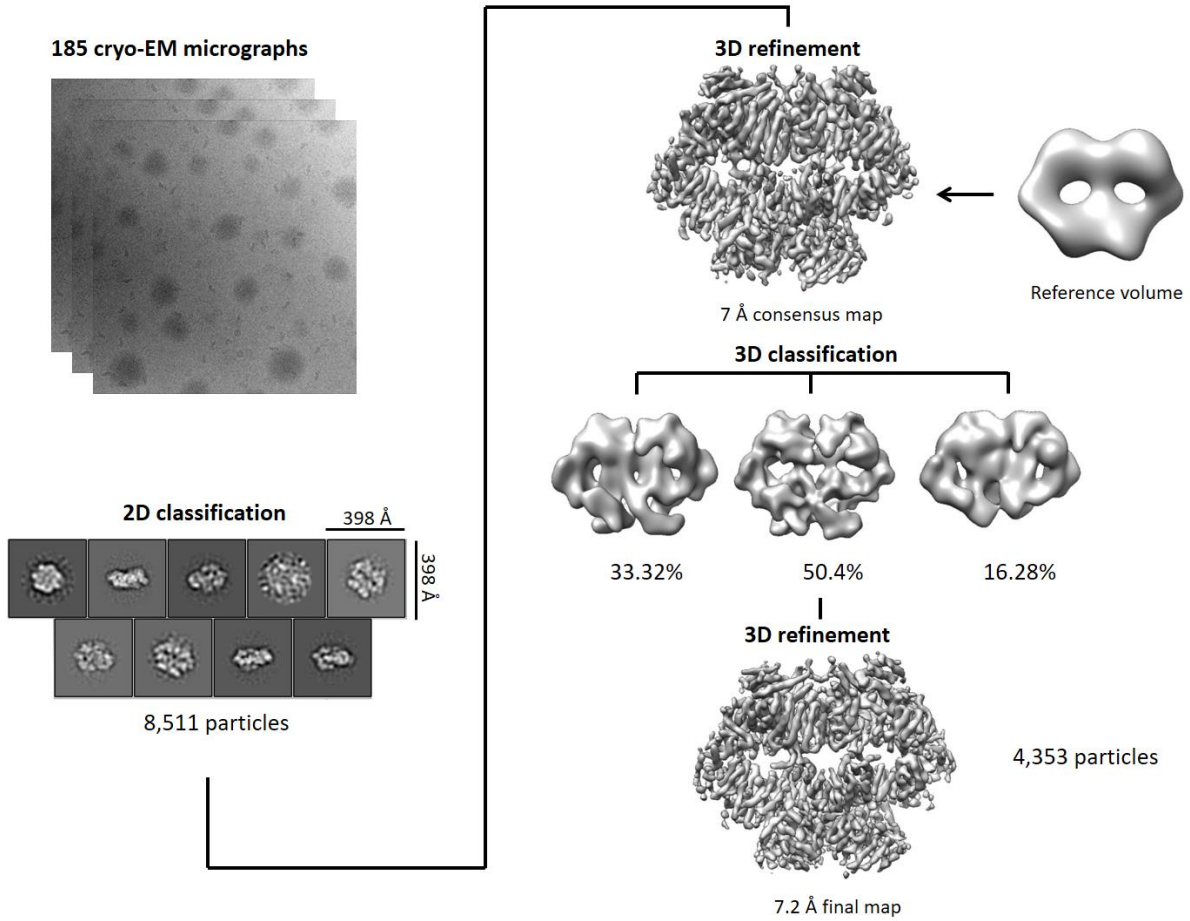


Figure 6.5 | Cryo-EM processing of the Mec1-Ddc2 complex (185 movies). A) Representative electron micrographs. Preliminary analysis of the Mec1-Ddc2 complex by negative stain (**left**); the image was colleted on a 120 kV FEI Tecnai T12 microscope operated with CCD camera. In addition, a cryo-EM micrograph of Mec1-Ddc2 is also shown (**right**), consisting in commercially available Lacey grids and collected on a direct detector device (DDD). **B)** An initial dataset composed of 185 micrographs collected on the Polara microscope generated a final cryo-EM map at 7.2 Å resolution, revealing the large α-solenoid region present in the complex, which is contributed by both the Mec1 and Ddc2 polypeptides. The 3D refinement of the reconstruction was carried out applying C2 symmetry. The reference volume for 3D auto-refinement in RELION was obtained from low-pass filtering a previously refined cryo-EM map in cryoSPARC (see "Materials and Methods", 3.4.3.2).

Table 6.1 | Cryo-EM data collection and processing of *S. cerevisiae* Mec1-Ddc2 complex (7.2 Å map, Lacey grids)	
Nominal magnification	125,000x
Voltage (kV)	300
Electron exposure (e^{-}Å^{-2})	44.25
Defocus range (μm)	-1.5 to -4
Calibrated pixel size (Å)	1.39
Symmetry imposed	C_2
Particle images after 2D (no.)	8,511

Final particle images (no.)	4,353
Map resolution (Å) at FSC=0.143	7.2

After movie alignment in MotionCor2 (Zheng *et al.*, 2017), particle picking was performed in Gautomatch (written by Kai Zhang at MRC-LMB) and both the aligned movies and particle coordinates were imported into RELION 2 (Kimanius *et al.*, 2016) for further processing. The reference volume for 3D auto-refinement in RELION was obtained from ab-initio classification in cryoSPARC v1 (Punjani *et al.*, 2017) (3.4.3.2). The resulting 7.2 Å map allowed the identification of the unique solenoid architecture present in the Mec1-Ddc2 complex (Fig. 6.5B) when compared to other PIKK proteins, revealing a dimeric architecture corresponding to an apparent C2 symmetry. An existing human ATM atomic model (PDB 5NP0) generated from a 4.7 Å map (EMD-3668) (Baretić *et al.*, 2017) matched reasonably well in our map over a long region of the Mec1 solenoid (Fig. 6.6), which was expected given the homology between $Mec1^{ATR}$ and ATM^{Tel1} (1.1.2 and 1.1.3). The Ddc2 subunit can be roughly located in the map by inference, because $Ddc2^{ATRIP}$ corresponds to the unsigned density after the fitting of ATM atomic model (Fig. 6.6, zoomed area). Ddc2 forms one of the two main dimerisation interfaces, opposite to Mec1 kinase domain region, which forms the second main dimerisation interface together with the FAT domain. Because the kinase domain is very well conserved among PIKK members, the fitting of this region can also be made unambiguously using existing atomic models for the Kinase domain (Yang *et al.*, 2013).

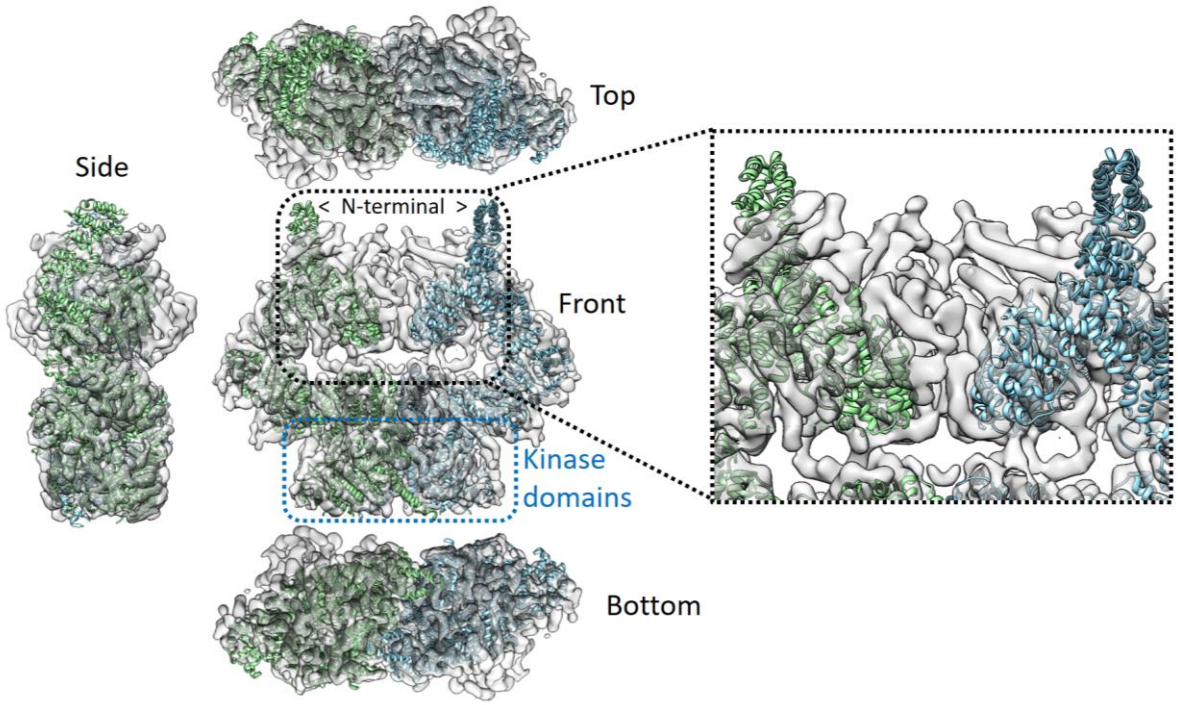


Figure 6.6 | Fitting of ATM atomic model (PDB 5NP0) into the 7.2 Å resolution cryo-EM map of the Mec1-Ddc2 complex (this study). The general architecture of the Mec1-Ddc2 complex can be defined by fitting existing atomic models into the generated cryo-EM map. The atomic model of $Mec1^{ATR}$

human paralog ATM^{Tel1} (PDB 5NP0) was fitted into the newly obtained 7.2 Å cryo-EM map of Mec1-Ddc2 (this study). The ATM Kinase domain at the C-terminal region and most of solenoid domains fit reasonably well into Mec1-Ddc2 density map. The main discrepancy is found at the N-terminal region, where ATM "Spiral" domain extends beyond the electron density at the very N-terminal end (highlighted in the figure). Likewise, we find a region within the dimeric interface opposite to the Kinase domains that was not occupied by ATM atomic model (zoomed area). This region corresponds to Ddc2^{ATRIP} density. (ATM monomers are depicted in blue and green colour each).

6.3.4 Optimisation of protein concentration for single-particle cryo-EM

The initial cryo-EM analysis suggested that the particle concentration on the grid could be further optimised (Fig. 6.5A). Particle density did not reliably increase when the sample was concentrated using a spin micro-column. Because the same behaviour was observed with Tra1, probably due to protein precipitation or aggregation, the spin concentration step was abandoned.

To test a final ion exchange chromatography step for concentration purpose, affinity purified Mec1-Ddc2 protein (Fig. 6.3A, E3 eluted fraction (lane 7)) was initially used for testing. The sample was incubated with 100 µl of bulk Q Sepharose Fast Flow (GE Healthcare) (pre-equilibrated with gel filtration buffer (2.0.5, Table 16)). After batch incubation for 2 hrs at 4 ^{o}C, the resin was recovered into a spin micro-column (Pierce micro-Spin column, Thermo Fisher) and the flow-through was applied once again to the resin. The elution was carried out in one bed volume (100 µl) by sequentially applying buffers (gel filtration buffer, 2.0.5, Table 16) with increasing NaCl concentration (from 400 to 1000 mM). Initial tests showed that the complex binds tightly to the resin, eluting at 600 mM NaCl (Fig. 6.7A, lane 4).

To analyse BS3 crosslinked sample, a gel filtration step was performed on a Superpose 6 Increase 10/300 GL column and the peak corresponding to the Mec1-Ddc2 complex (similar to 6.3.2) was recovered. Pooled fractions presented a similar behaviour as the BS3-free Mec1-Ddc2 protein i.e. similar binding efficiency to the resin and elution at the same salt concentration (Fig. 6.7B, lane 6). The result was an increased protein concentration comparing to the input. Because small amounts of protein were used for the assays - i.e. the E3 fraction for testing purpose - protein concentration measurements were very inaccurate (<0.1 mg/mL as measured by the nanodrop, 2.4.1), so that SDS-PAGE and Coomassie staining were used instead for visual assessment to confirm this concentration step.

This particular concentration strategy based on a mild concentration step at the very end of the purification and immediately before sample vitrification was combined with an increase in the amount of protein material and also a larger gel filtration column was used, as will be described in the next section (6.4). In addition, graphene oxide (GO) coated grids (3.4.3.1), which increased the contrast of the Mec1-Ddc2 particles on the collected images, replaced the previously used Lacey continuous carbon grids. All together allowed to optimise the

number of particles per micrograph and to obtain a high-resolution cryo-EM reconstruction of the Mec1-Ddc2 complex (6.4-6.6).

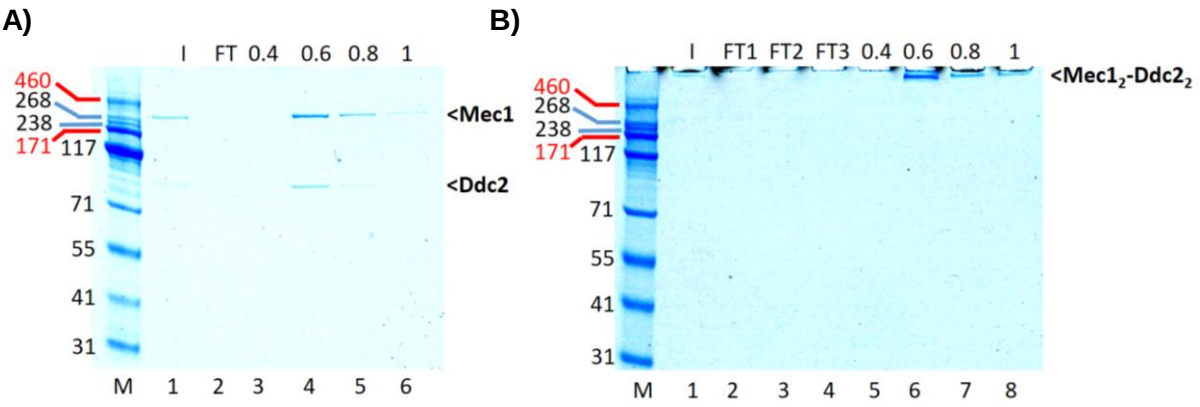
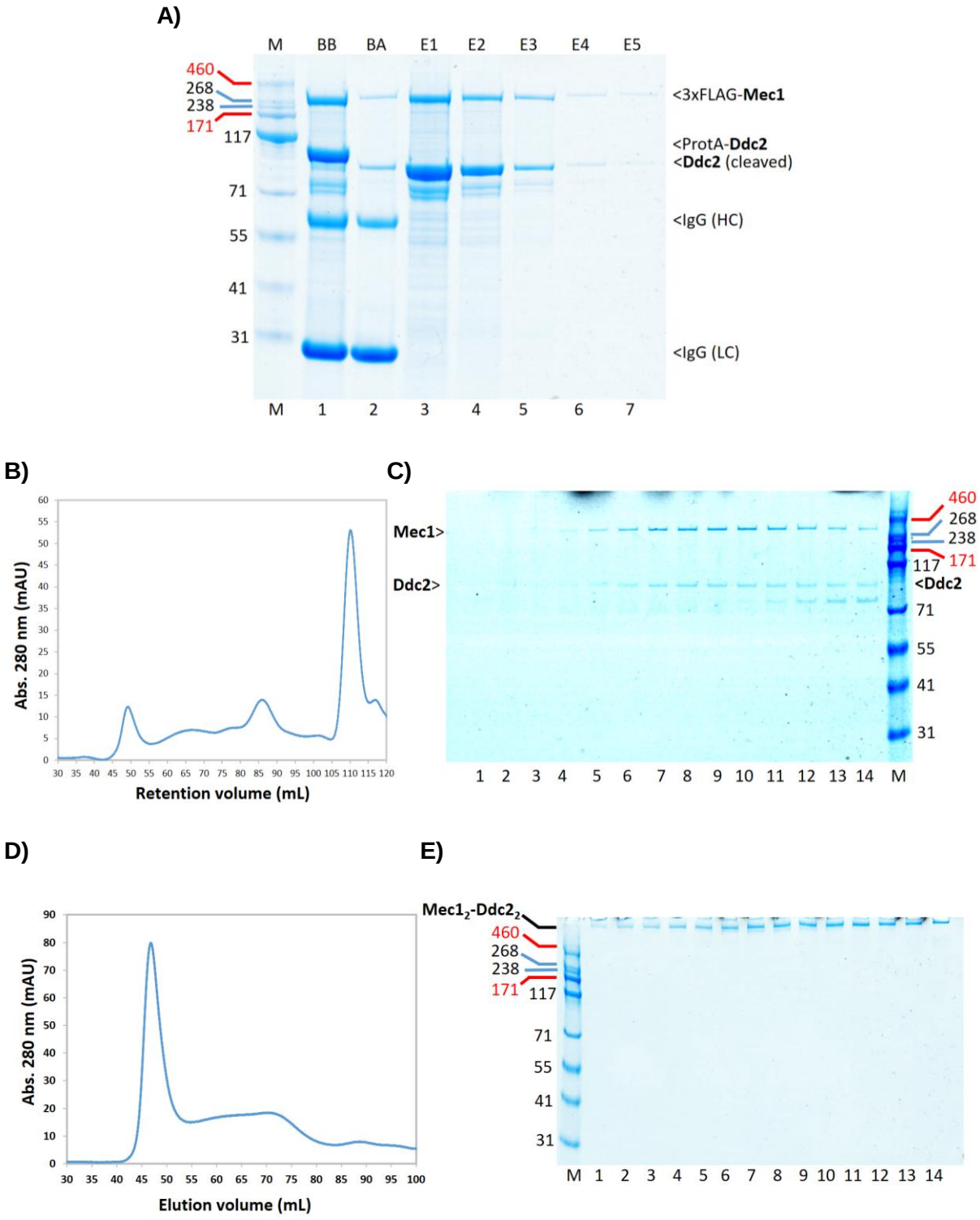


Figure 6.7 | Mec1-Ddc2 binding assays to Q-sepharose resin. A) Affinity purified BS3-free complex was bound to Q-sepharose resin and eluted with increasing NaCl concentration. I: input. FT: flow-through. The indicated NaCl concentration is in molar (0.4-1). **B)** As **A** but using BS3 crosslinked and Superpose 6 Increase 10/300 GL purified Mec1-Ddc2 complex. (The images in the figure are shown with increased contrast to facilitate visual analysis).

6.4 Optimised purification for high-resolution single-particle cryo-EM analysis

This time, the full amount of the affinity purified material was used for cryo-EM analysis. Following the optimised procedures of protein expression and purification previously described (6.3 and 2.3.4), 12 L of yeast culture were grown to affinity purify the complex using a double tag affinity purification strategy as before (6.3.1), and this time a larger volume of FLAG M2 resin was used (see “Materials and Methods”, 2.3.4.2). To avoid any potential protein precipitation as the last time, the elution volume was also increased from 1 to 2 bed volumes of resin. The affinity purification step resulted in the efficient and successful elution of most of material bound to the resin (Fig. 6.8A, lane 2). The gel filtration step was modified using the larger XK 16/70 column (GE Healthcare) to inject a higher amount of material into the column without any previous concentration step before injection. The new column was tested with 1.5 mL of BS3-free E2 fraction (~ 0.15 mg/ml) mixed with 1.5 mL of E3 (~ 0.05 mg/ml) as control (Fig. 6.8A, lanes 4 and 5, respectively). The peak for the complex in the newly tested column was found between ~ 68-78 mL (Fig. 6.8B; 6.8C, lanes 5-10). For single-particle cryo-EM analysis, 3 mL of E1 fraction (~ 0.3 mg/ml) were crosslinked with 0.75 mM of BS3. The elution peak corresponding to monodisperse Mec1-Ddc2 complex was found in a similar position as for the BS3 free sample (Fig. 6.8D; 6.8E, ~ 68-78 mL peak). After collecting the peak, the resulting volume was incubated with 100 µl of pre-equilibrated bulk Q sepharose resin, as tested before (6.3.4). The protein was eluted using “gel filtration buffer” (2.0.5, Table 16) containing 600 mM NaCl instead (Fig. 6.8F), and the eluted material was used for cryo-EM grid preparation on the Vitrobot by Natasha Lukoyanova, applying one single

time the fresh, eluted monodisperse sample to gold grids coated with a thin layer of graphene oxide (GO) (3.4.3.1).



F)

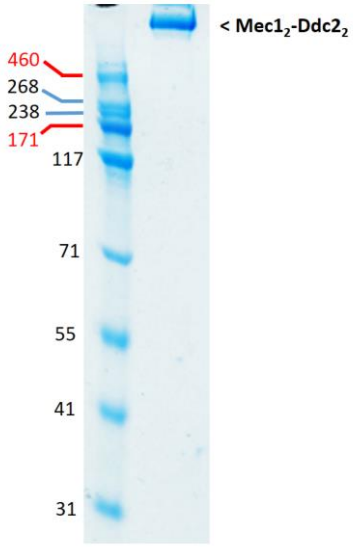


Figure 6.8 | Optimised purification of the Mec1-Ddc2 complex for high-resolution cryo-EM analysis. A) 4-12% SDS-PAGE and Coomassie staining of affinity purified Mec1-Ddc2 complex. The first lane corresponds to Mec1-Ddc2 bound to FLAG M2 resin before elution, the second lane after elution. Most of material eluted from the column, and a shift in the Ddc2 polypeptide size can be observed after TEV cleavage of the N-terminal ProtA tag. Eluted fractions E1-E5 follow (lanes 3-7). **B)** Test run of BS3-free sample on a XK 16/70 SEC column, showing a first peak corresponding to Ddc2 aggregates followed by the Mec1-Ddc2 complex. **C)** 4-12% SDS-PAGE and Coomassie staining of **B**. Lanes 1-14: 0.75 ml fractions eluting between 68.75 and 89.75 ml; one fraction every two was loaded for analysis by SDS-PAGE. The Coomassie stained gel shows the two bands corresponding to the complex. **D)** 3 mL of affinity purified Mec1-Ddc2 complex (E1 fraction, ~ 0.3 mg/ml) were BS3 crosslinked and injected into the XK 16/70 gel filtration column. **E)** 4-12% SDS-PAGE and Coomassie staining of **D.** Lanes 1-14: 0.75 ml fractions eluting between 67.5 and 78 ml, corresponding to the Mec1-Ddc2 BS3 cosslinked complex. **F)** The peak corresponding to the BS3 crosslinked Mec1-Ddc2 complex (**D**, **E**) was concentrated to 140 µl for cryo-EM analysis using Q-sepharose resin. 10 µl of the concentrated material from **D** were analysed by 4-12% SDS-PAGE and Coomassie staining. (The image in **C** is shown with increased contrast to facilitate visual analysis).

6.5 Cryo-EM structure of the Mec1^{ATR}-Ddc2^{ATRIP} complex at near-atomic resolution

6.5.1 Cryo-EM processing and 4.1 Å cryo-EM reconstruction

The optimised purification allowed to significantly increase the number of particles per micrograph (Fig. 6.9A) comparing to the previous dataset (Fig. 6.5A). The processing of the new in-house cryo-EM dataset generated high-resolution 2D classes (Fig. 6.9B) and resulted in a high-resolution cryo-EM reconstruction (Fig. 6.9C; 6.10A-C) (3.4.3.2), as will be discussed in more detail below.

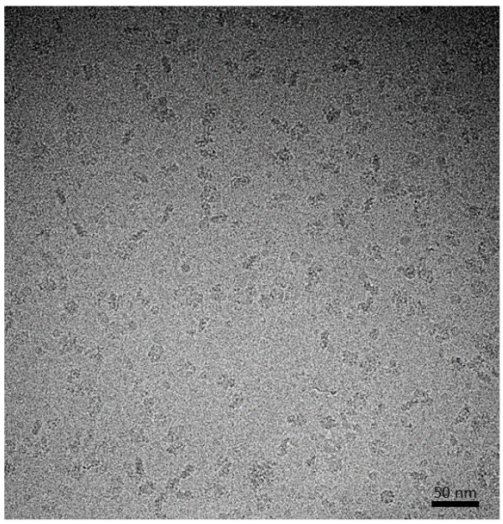
The purified and concentrated complex was applied to a grid containing graphene oxide support. Because the final concentration step before cryo-EM grid preparation consisted in the elution with 600 mM NaCl buffer (6.4), the excess of salt in the sample was washed previous to plunge-freezing in liquid ethane applying a 3 µl drop of gel filtration buffer (150 mM

NaCl, 2.0.5, Table 16) to the grid after sample application and a first blotting step. Data collection was performed in-house at Birkbeck College on a 300 kV FEI Tecnai G^2 F30 Polara microscope mounted with a K2 detector and operated in counting mode (Table 6.2). Image acquisition was performed in automated mode with SerialEM. The vitrification of the sample as well as high-resolution data collection with DDD camera were performed by Natasha Lukoyanova. I performed the cryo-EM processing of the images (3.4.3.2), generating a final reconstruction at 4.1 Å nominal resolution from 73,163 particles (Fig. 6.9C) (Table 6.2).

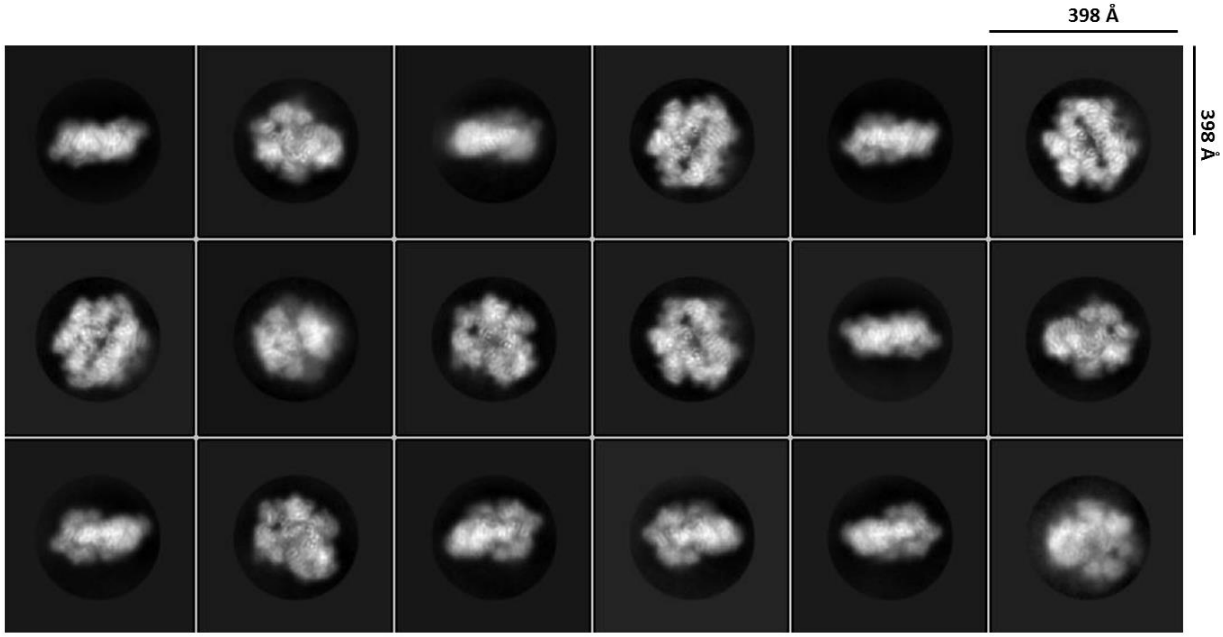
Table 6.2 | Cryo-EM data collection and processing of *S. cerevisiae* Mec1-Ddc2 complex (4.1 Å map, GO gold grids)	
Nominal magnification	125,000x
Voltage (kV)	300
Electron exposure (e^{-}Å^{-2})	49.5
Defocus range (μm)	-1.5 to -4
Calibrated pixel size (Å)	1.39
Symmetry imposed	C_2
Particle images after 2D (no.)	140,990
Final particle images (no.)	73,163
Map resolution (Å) at FSC=0.143	4.1

A)

Cryo-EM micrograph (GO support)



B)



C)

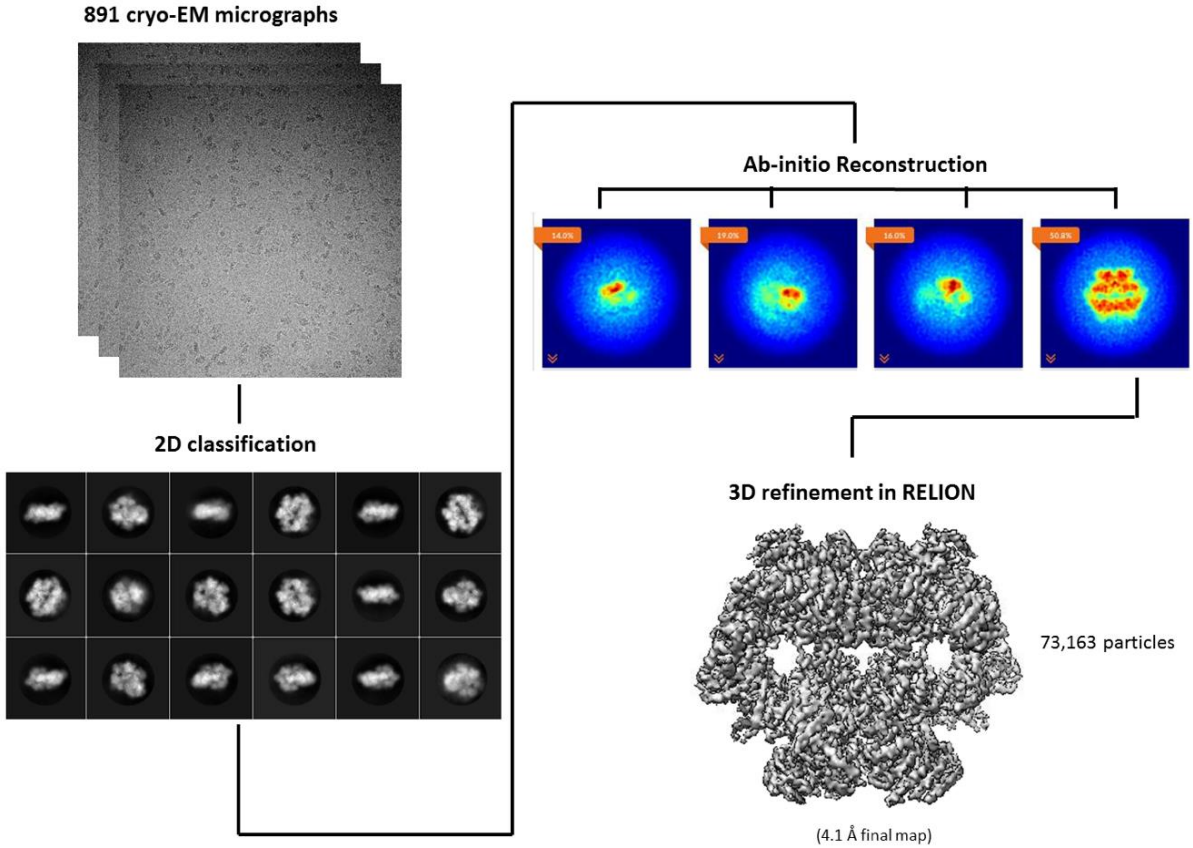
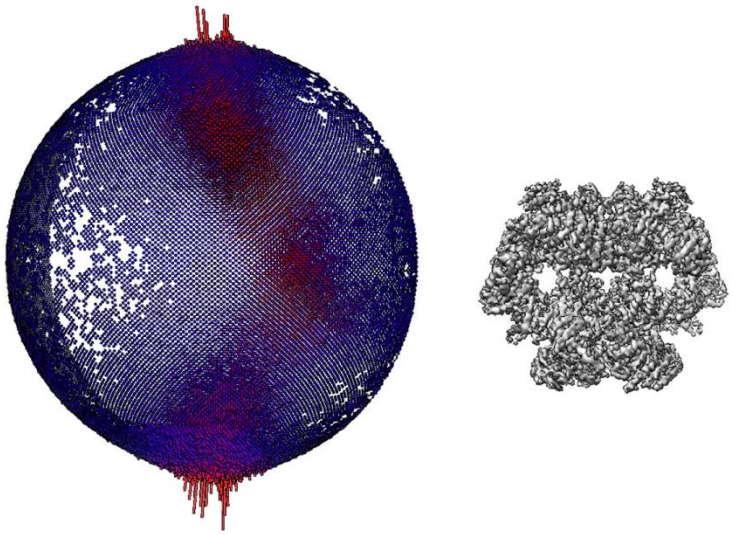
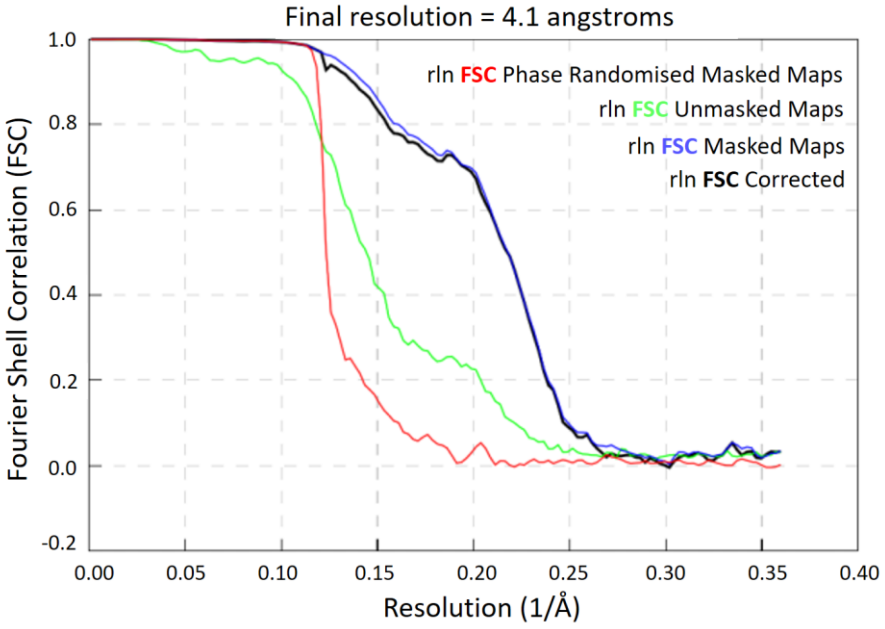


Figure 6.9 | Cryo-EM processing of the Mec1-Ddc2 complex from 891 movies. A) Representative cryo-EM micrograph. Scale bar is 50 nm size. **B)** High-resolution 2D classses of the Mec1-Ddc2 complex showing secondary structure features. **C)** Schematic representation of the cryo-EM processing from 891 micrographs. After 2D classification and ab-initio reconstruction in cryoSPARC, 73,163 particles were selected and 3D-refined in RELION, resulting in a 3D reconstruction at 4.1 Å resolution (FSC=0.143 criterion). The 3D refinement was carried out applying C2 symmetry to the reconstruction.

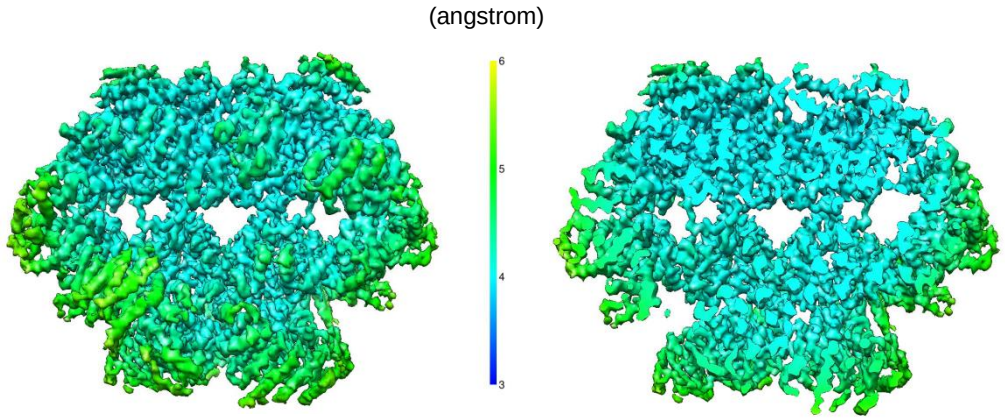
A)



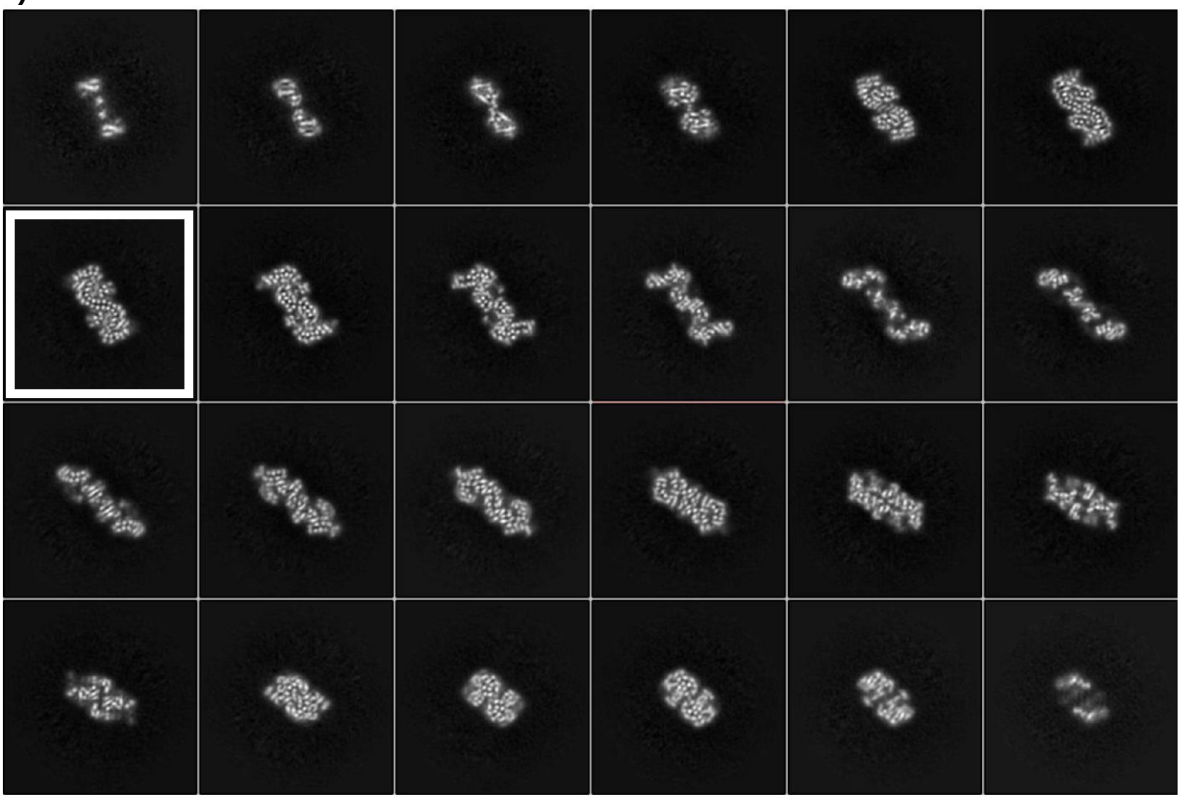
B)

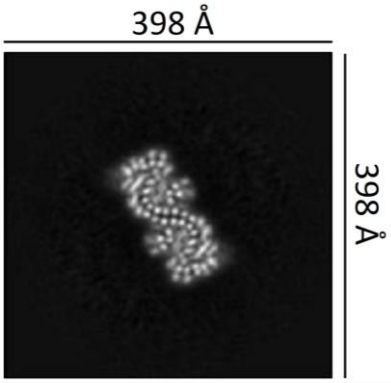

Final resolution = 4.1 angstroms
rln FSC Phase Randomised Masked Maps
rln FSC Unmasked Maps
rln FSC Masked Maps
rln FSC Corrected
Fourier Shell Correlation (FSC)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Resolution (1/Å)

C)


(angstrom)
6
5
4
3

D)




398 Å
398 Å

E)

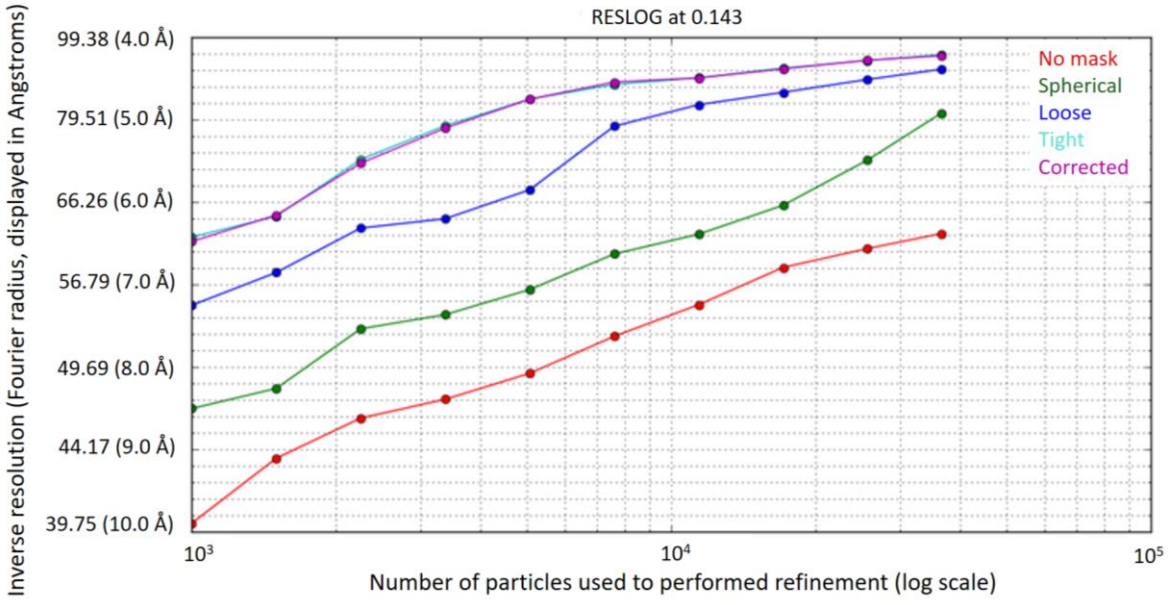

RESLOG at 0.143
99.38 (4.0 Å)
79.51 (5.0 Å)
66.26 (6.0 Å)
56.79 (7.0 Å)
49.69 (8.0 Å)
44.17 (9.0 Å)
39.75 (10.0 Å)
Inverse resolution (Fourier radius, displayed in Angstroms)
10³
10⁴
10⁵
Number of particles used to performed refinement (log scale)
No mask
Spherical
Loose
Tight
Corrected

Figure 6.10 | Analysis of Mec1-Ddc2 4.1 Å cryo-EM map. A) Angular distribution plot showing good distribution of orientations for the particles contributing to the final 3D reconstruction. **B)** FSC curve reporting 4.1 Å overall resolution at 0.143 criterion. **C)** Local resolution map. A cross-section of the map is also shown (**right**). **D)** Cross-sections of the 4.1 Å unprocessed Mec1-Ddc2 cryo-EM map (from top to bottom as rendered before in **C**) showing its large content in α-helix and the α-solenoid architecture present in the complex where pairs of α-helices are clearly visible. The S-shaped cross-section enclosed within the white squared is magnified in a closer view below; the cross-section shows the interaction of two different Ddc2 polypeptides within the dimer of heterodimers, creating a "Solenoid Blade" like structure whith an essential role in the dimerisation interface. **E)** ResLog analysis from cryoSPARC, showing the estimated resolution using different numbers of particles as well as different masks, including mask free calculations. The calculated ResLog plot indicates that more particles could significantly improve the resolution in the near-atomic range, given that a theoretical plateau is not reached yet with the actual number of particles used in the 3D refinement of the final map.

After MotionCor2 alignment of the collected movies, manual supervision and selection of micrographs was performed excluding areas with bent or absent graphene oxide. A final set of 891 movies was selected for further processing. Image processing was performed by a combination of software for single-particle analysis; the aligned movies were used for particle-picking in Gautomatch, cryoSPARC was used for 2D classification and 3D ab-initio classification, and RELION for the 3D auto-refinement of the map (see "Materials and Methods", 3.4.3.2).

There is a good distribution of particle orientations for the complex on the grid as can be deduced from the large number of different 2D classes obtained after 2D classification (Fig. 6.9B), and corroborated by the angular distribution plot analysis obtained after 3D auto-refinement in RELION (Fig. 6.10A).

No significant conformational differences were observed for the complex during 3D classification, which provided a single 3D class that refined to high-resolution (Fig. 6.9C and 10B). Sub-classification of this 3D class did not provide any additional information. Local resolution analysis shows that the core of the complex presents the highest local resolution value, as can be expected (Fig. 6.10C). Cross-section analysis of the map shows the high-content in α-helices within the complex, organised into pairs and contributing to the large solenoid region present in Mec1 and Ddc2 (Fig. 6.10D). Given that it is clearly possible to improve the resolution by increasing the number of particles, as inferred by the ResLog analysis in cryoSPARC (Fig. 6.10E), the results stimulated for the possibility of getting a standard cryo-EM dataset on a high-end 300 kV FEI Titan Krios microscope to push for the best possible resolution.

Shortly after obtaining our in-house 4.1 Å reconstruction from Polara images, a similar structure was published (Wang *et al.*, 2017), largely overlapping with the work carried out on the Mec1-Ddc2 complex. The same model organism and complex were used for the study, obtaining a similar resolution - 3.9 Å in this case. The map was published with the corresponding atomic model, showing a dimer of heterodimers architecture for the complex.

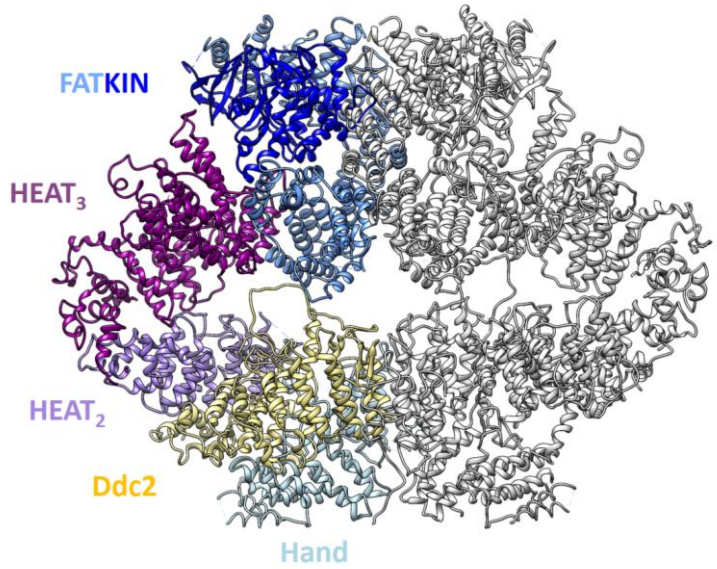
However, the comparison of our 4.1 Å cryo-EM map against the published structure shows strong discrepancy at the N-terminal region of both Mec1 and Ddc2 polypeptides, where 235 residues of Mec1 are missing in the published work. In addition, the generation of an atomic model from our map also allowed for the detailed comparison of our atomic structure against the human homolog ATR-ATRIP complex (Rao *et al.*, 2017), as will be discussed in the next section (6.5.2).

6.5.2 A Full-length Atomic Model of Mec1^{ATR}

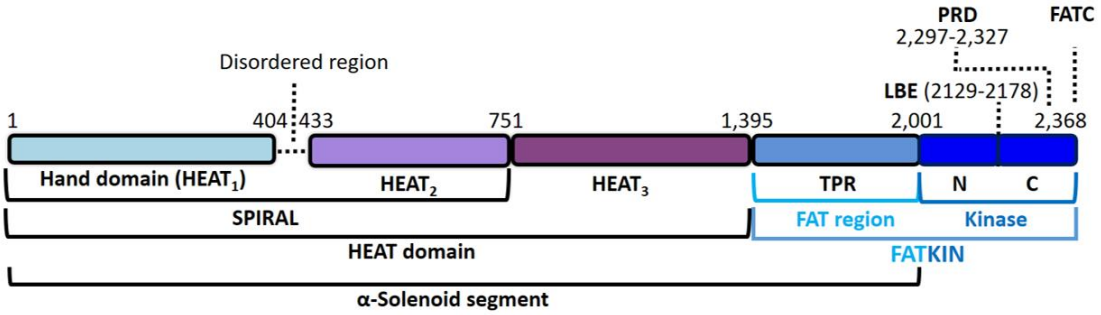
The atomic model generated from the cryo-EM map presented in this study represents the first full-length atomic structure of the Mec1 polypeptide to my knowledge, and allows to understand for the first time the architecture and assembly of the Mec1^{ATR}-Ddc2^{ATRIP} complex. The Mec1-Ddc2 complex is a dimer of heterodimers formed by the interaction of two Mec1 polypeptides with two Ddc2 polypeptides. The domain organisation of the complex is represented in Fig. 6.11, presenting a characteristic hole in the middle of the structure. Two main dimerisation interfaces can be identified in opposite sides to the mentioned hole: one is formed by the FAT and Kinase domains (FATKIN or Head globular domain) of the Mec1 polypeptides, and the other dimerisation interface is created by the interaction of the two Ddc2 subunits at the N-terminal region of Mec1.

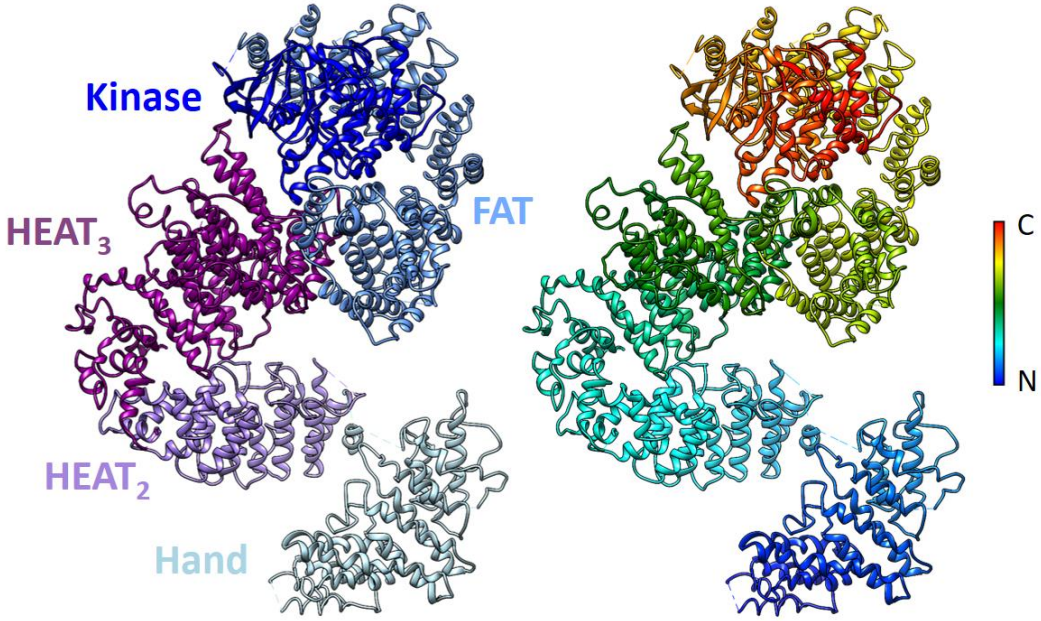
Mec1 HEAT and FAT domains constitute the very large alpha-solenoid region (Fig. 6.1A), extending from the very N-terminal region to the kinase domain, which is located at the very C-terminal region. The Ddc2 polypeptide consists of a RPA binding domain (RBD) (residues 10-30) in close proximity to an N-terminal coiled-coil (CC) domain (residues 73-136) which is connected by a long disordered region (residues 137-188) to the Ddc2 solenoid domain (residues 203-747) (Fig. 6.1A and 6.11C). Hence, Ddc2 very N-terminal region including the RBD, the CC domain and the disordered linker are missing in the cryo-EM map (residues 1-188), as will be discussed next. For that reason, the reported Ddc2 atomic model in this study corresponds to a short structured linker (residues 189-202) found after the long disordered linker together with the large solenoid domain (~ 75% of total Ddc2 residues). As will be also discussed later (6.6.5), available crystal structures of Ddc2 RBD-CC domains (Deshpande *et al.*, 2017) were used to generate an integrative model of Ddc2.

A)


FATKIN
HEAT3
HEAT2
Ddc2
Hand

B)


PRD
2,297-2,327
FATC
Disordered region
LBE (2129-2178)
1
404
433
751
1,395
2,001
2,368
Hand domain (HEAT1)
HEAT2
HEAT3
TPR
N
C
SPIRAL
FAT region
Kinase
HEAT domain
FATKIN
α-Solenoid segment


Kinase
FAT
HEAT3
HEAT2
Hand
C
N

C)

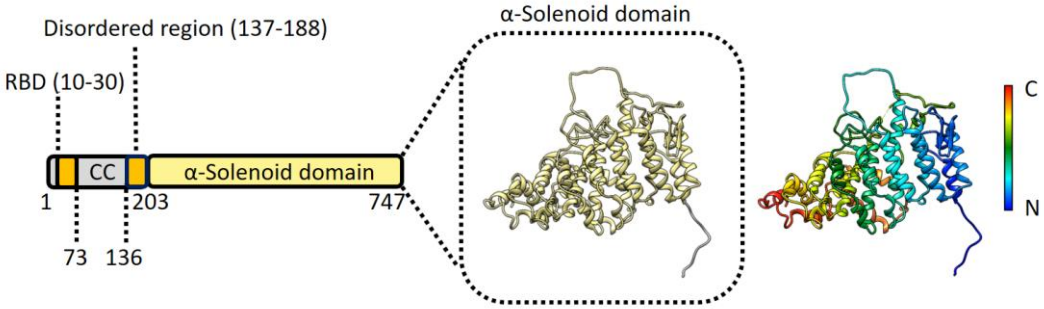


Figure 6.11 | Domain organisation of the Mec1^{ATR}-Ddc2^{ATRIP} complex. A) The Mec1-Ddc2 complex presents a dimer of heterodimers architecture. Ddc2 solenoid domain and the different Mec1 domains are labelled following the same colour code as **B** and **C** - in one polypeptide only - whereas the other heterodimer is represented in grey. **B)** Mec1 domain organisation from to N to C-terminal and its corresponding atomic model, comprising the Hand domain (HEAT region 1), HEAT region 2 and 3, the FAT domain and the Kinase domain (divided into the N and C-lobes). **C)** Ddc2 domain organisation from N to C-terminal and its corresponding atomic model, comprising the RPA binding domain (RBD) (residues 10-30), the coiled-coil (CC) domain (73-136), a disordered linker (137-188), a small structured linker (189-202) and a large α-solenoid domain (residues 203-747). The solenoid region and the small structured linker at its N-terminal end are visible in the map, whereas the density corresponding to residues 1-188 is missing because of the unstructured nature of the flexible linker connecting this region to the rest of the complex makes it invisible to cryo-EM.

As mentioned before (6.5.1), in addition to the available partial structure of Mec1 (Wang *et al.*, 2017), a second, human ATR-ATRIP complex was also published (Rao *et al.*, 2017), this time at an overall resolution of 4.7 Å, and improving to 3.9 Å in the Head region containing the FAT and Kinase domains. Due to resolution limitations at the N-terminal region of the human ATR-ATRIP complex map, the existing ATRIP^{Ddc2} atomic model was built as a poly-alanine backbone and the stoichiometry of the complex could not be unambiguously assigned, as described by the authors. However, the density corresponding to the full-length of human ATR polypeptide was visible, where the N-terminal region is composed of HEAT repeats organised into an alpha-solenoid fashion, as expected from secondary structure (SS) prediction analysis. Given the high level of homology between yeast Mec1^{ATR} and human ATR^{Mec1}, the information obtained from SS prediction and the available literature (Cortez *et al.*, 2001; Ball and Cortez, 2005; Itakura, 2005; Kim *et al.*, 2005; Sawicka *et al.*, 2016; Deshpande *et al.*, 2017), all together indicates *"a priori"* that the density corresponding to the Mec1 polypeptide should be entirely present in the map. The experimental cryo-EM map at 4.1 Å resolution presented in this study clearly shows density for the full-length Mec1 polypeptide (Fig. 6.12), which allowed to generate a complete atomic model of Mec1 (3.4.4.2). Furthermore, the local resolution in our cryo-EM map at the N-terminal region of Mec1 where Ddc2 is in close proximity (Fig. 6.10C) allows to track at side chain level each polypeptide within the complex, revealing the architecture and assembly of the complex (6.6.2).

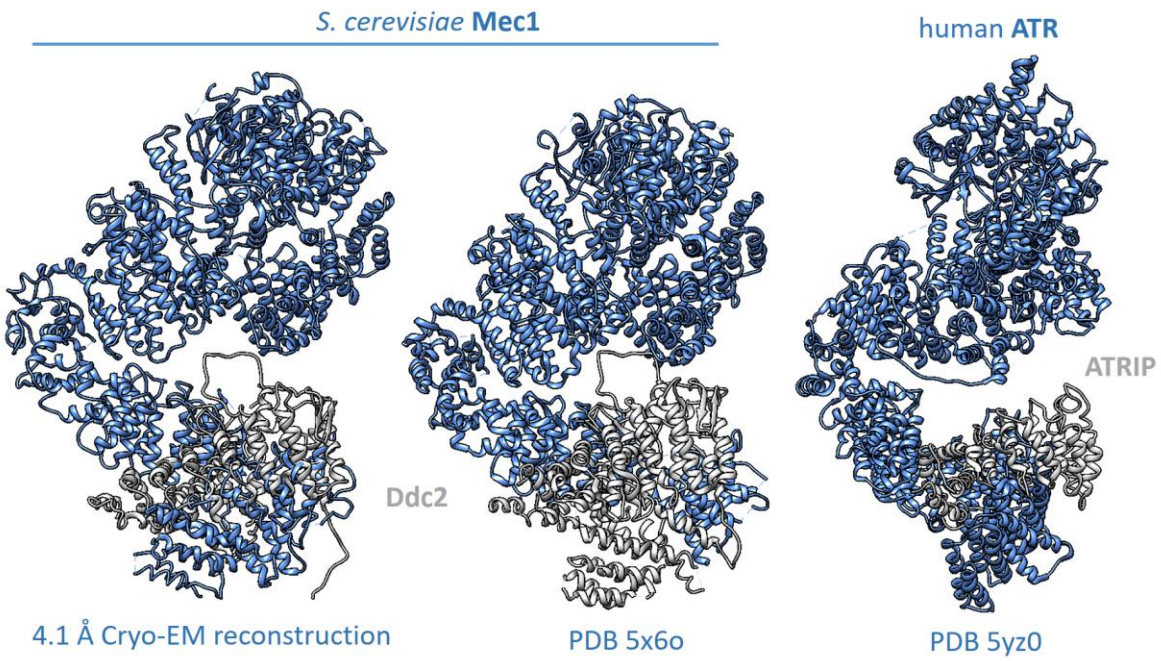
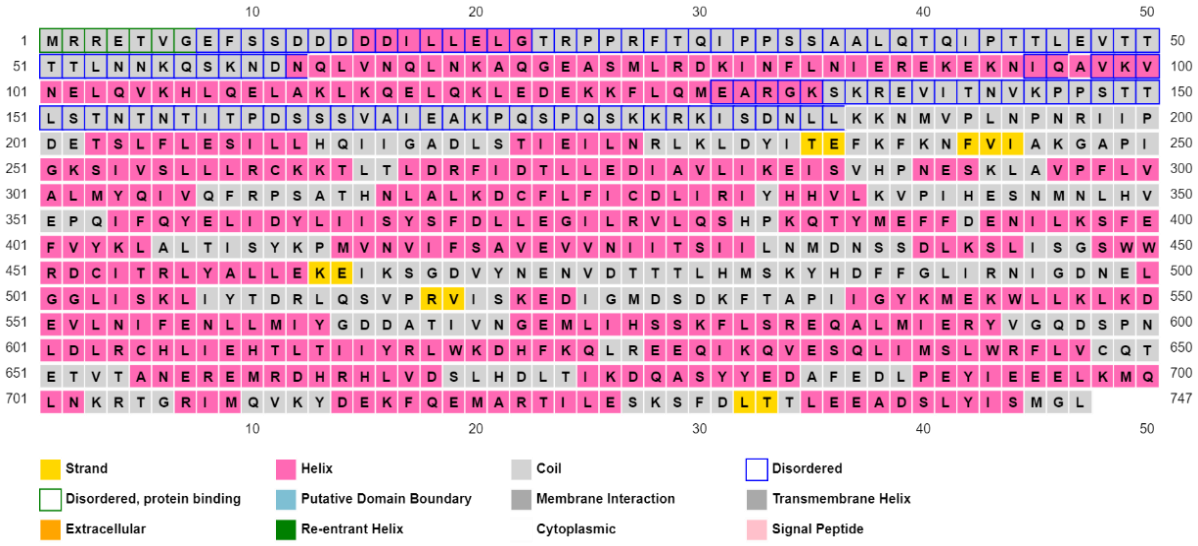


Figure 6.12 | Comparison of full-length Mec1 atomic model against available Mec1 and ATR atomic models. The generation of an atomic model from our 4.1 Å cryo-EM reconstruction is compared against Mec1 PDB 5x6o and ATR PDB 5yz0. Only one Mec1 (in blue) and one Ddc2 monomer (in grey) are rendered in the figure. While the architecture of the complex is equivalent when the Mec1-Ddc2 atomic model presented in this study and ATR-ATRIP PDB 5yz0 are compared, Mec1-Ddc2 PDB 5x6o shows discrepancy at Mec1 and Ddc2 N-terminal regions when compared to the other two atomic models of the complex.

In relation to Ddc2, PSIPRED analysis (Buchan and Jones, 2019) clearly shows that Ddc2 N-terminal region, containing the RPA binding motif and a previously reported long coiled-coil region (Fig. 6.1A), is connected to the rest of the polypeptide by a long unstructured or intrinsically disordered linker (Fig. 6.13A). Therefore, this is the most likely region to be missing in the map. On the other hand, SS analysis shows that Mec1 N-terminal region is mostly structured and folded (Fig. 6.13B), and therefore this region of Mec1 is *"a priori"* the most likely to be present in the electron density map instead of Ddc2 N-terminal region, as was confirmed by the cryo-EM map presented in this study (Fig. 6.9, 6.10 and 6.12). As previously mentioned, the analysis of the electron density map clearly shows that the FL density of the Mec1 polypeptide is present, whereas the density corresponding to Ddc2 N-terminal region – connected by an unstructured linker – is missing (residues 1-188). Thus, the visible region of Ddc2 in the electron density map corresponds to Ddc2 Solenoid domain, in addition to the short structured linker (residues 189-203) connecting Ddc2 Solenoid domain to the N-terminal unstructured region.

A)



B)

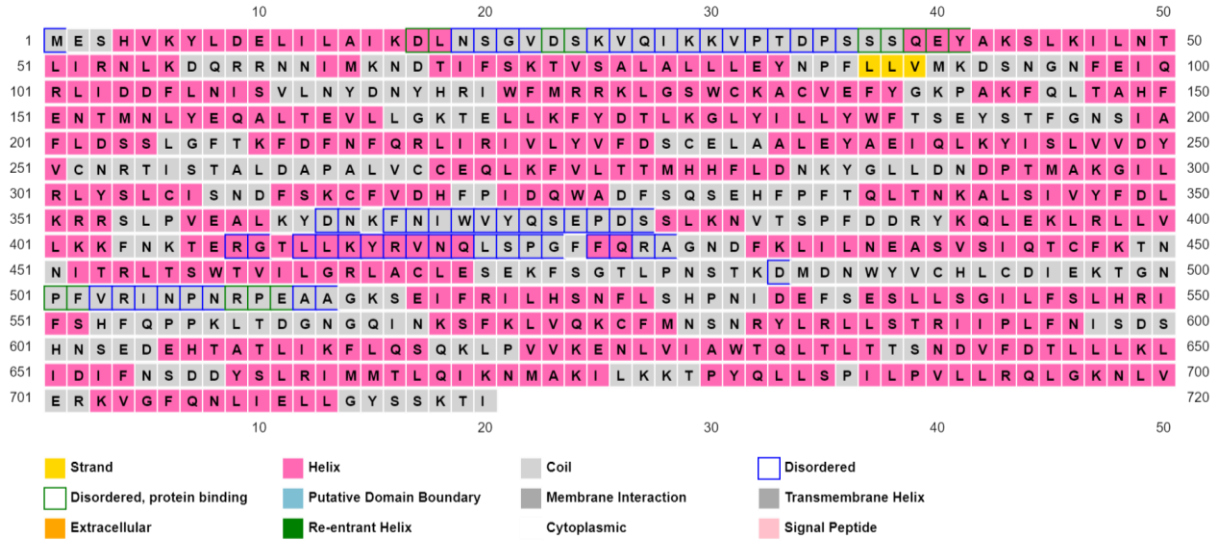


Figure 6.13 | Secondary structure prediction analysis of full-length Ddc2 (residues 1-747) and Mec1 N-terminal region (residues 1-720) by the PSIPRED server. A) Analysis of full-length Ddc2 by PSIPRED shows that after Ddc2 RPA binding domain (RBD) and the coiled-coil (CC) region (Fig. 6.1A and 6.11C), a flexible linker consisting in an intrinsically disordered region (residues 130-186 in the SS analysis) connects the Ddc2 coiled-coil to the Solenoid domain. **B)** PSIPRED analysis of Mec1 shows that the N-terminal region is mostly structured, and a flexible or unstructured linker (residues 409-427) is predicted to break the so-called “Spiral” that is found in ATR (Fig. 6.11B, 6.12 and 6.14). This linker is missing in the cryo-EM map and connects the very N-terminal Hand domain of Mec1 (residues 1-404) to the rest of the alpha-solenoid region.

The identification of the density corresponding to Ddc2 structured linker is essential to understand the complex recruitment to RPA-coated ssDNA. In this regard, a work explaining Mec1-Ddc2 recruitment to RPA-coated ssDNA was published by the Gasser lab (Deshpande *et al.*, 2017), reporting a crystal structure of Ddc2 coiled-coil domain and combining it with an available 21 Å resolution map of Mec1-Ddc2 (Sawicka *et al.*, 2016) to propose a model to

explain the complex recruitment to ssDNA. The first visible residues in our cryo-EM map for the N-terminal region of Ddc2 approximately match the beginning of the short structured region as predicted by PSIPRED (residues 189-203), and locates precisely the insertion point of Ddc2 coiled-coil domain within the complex. This short structured linker is stabilised by a hydrophobic pocket within Mec1 N-terminal region, as will be discussed next (6.6.3). An integrative atomic model of the complex generated by the combination of our atomic structure together with the existing Ddc2 coiled-coil and RPA binding domains is illustrated in the 3.3 Å cryo-EM map section (6.6.5, Fig. 6.27).

The comparison of the full-length atomic model of Mec1 presented in this study to the available atomic model of human ATR (PDB 5yz0) shows that the “Spiral” found in ATR is truncated in Mec1 by a long linker connecting the very N-terminal domain to the rest of Mec1 solenoid (Fig. 6.12 and 6.14). This flexible linker region between residues 404 to 433 is a short unstructured region as predicted by PSIPRED (Fig. 6.13B). Therefore, in Mec1, the “Spiral” region found in ATR is divided by an unstructured linker that defines a Hand-like domain (Fig. 6.1A and 6.11A, B) at the very N-terminal end of Mec1 (residues 1-404). The Hand domain clamps Ddc2 solenoid region against the next HEAT repeat segment in Mec1 (residues 433-751), which is found immediately after the disordered linker. The flexibility provided by this flexible linker between Mec1 segments (“HEAT 1” or Hand and “HEAT 2”; Fig. 6.11A, B) might help to accommodate the Ddc2 Solenoid domain within the complex (Fig. 6.12 and 6.14; 6.6.2).

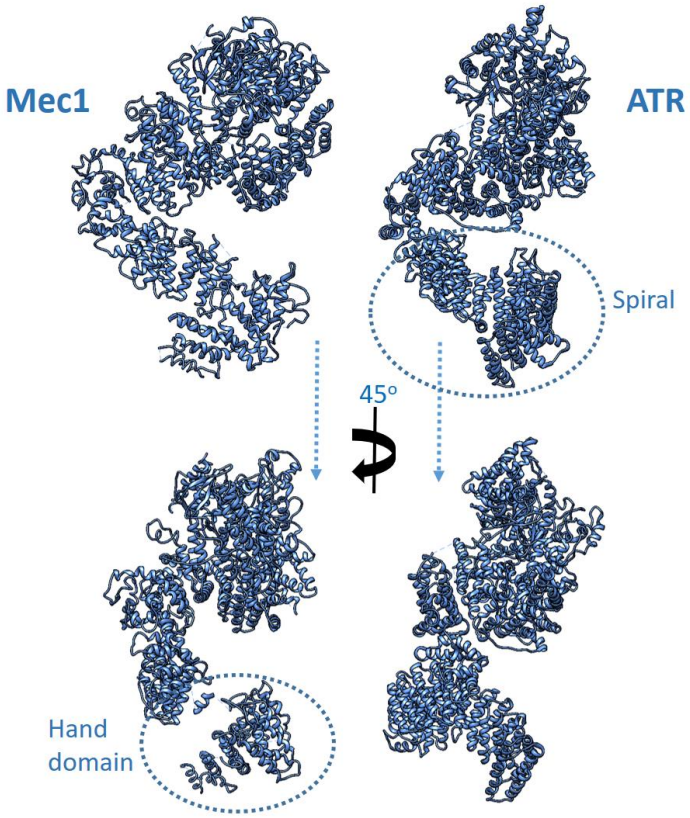


Figure 6.14 | Yeast Mec1 versus human ATR atomic models. The comparison of *S. cerevisiae* Mec1 atomic model (this study) against ATR PDB 5yz0 shows that while ATR N-terminal region is a continuous solenoid, in Mec1 the equivalent region is fragmented by a flexible linker (residues 404-433) that generates a separated N-terminal solenoid domain. In the figure, this domain is labelled as the Hand domain (also defined as "HEAT 1" in Fig. 6.11B). The rotation of the atomic model 45º in the "Y" axis facilitates the interpretation of the discontinuity in Mec1 versus ATR.

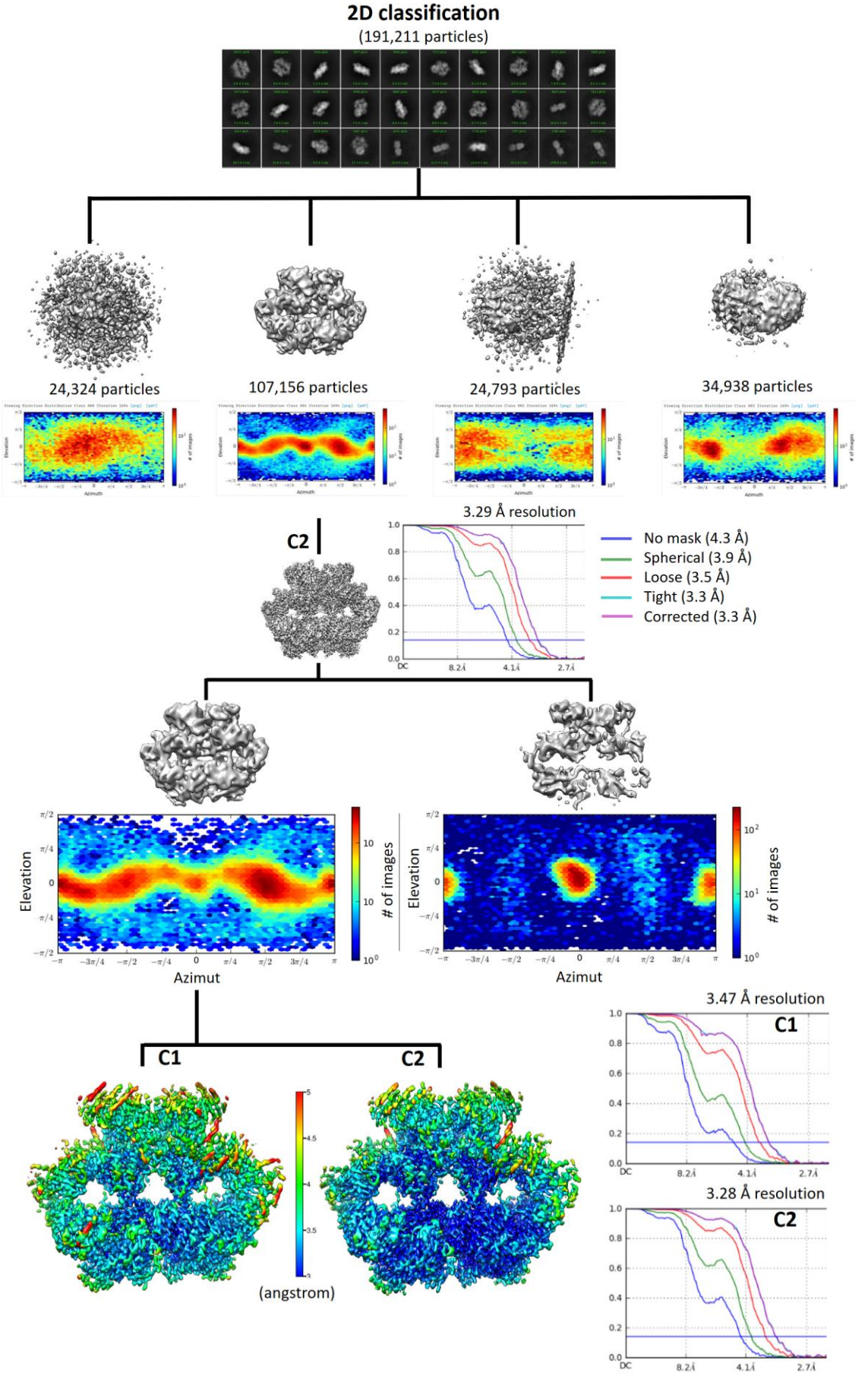
6.6 An improved cryo-EM map of the Mec1^{ATR}-Ddc2^{ATRIP} complex

6.6.1 Single-particle cryo-EM analysis

In order to generate a high-quality atomic model of the Mec1-Ddc2 complex, especially in regard to the kinase domain, cryo-EM data was collected on a 300 kV FEI Titan Krios at Birkbeck College (Table 6.3) (3.4.3). Data collection was performed on a K2 summit camera operated in counting mode, and the cryo-EM reconstruction of the complex was solved at a nominal resolution of 3.3 Å applying C2 symmetry, and 3.5 Å for the asymmetric (C1) reconstruction (Fig. 6.15). Many more side chains could be tracked this time in comparison to the previous map at 4.1 Å resolution (6.5.1), as well as missing loops in the previous structure that were built this time (3.4.4.2).

Table 6.3 | *S. cerevisiae* Mec1-Ddc2 complex (3.3 Å map, GO grids): Cryo-EM data collection, processing, refinement and validation		
Data collection and processing		
Nominal magnification	165,000x	
Voltage (kV)	300	
Electron exposure (e^{-}Å^{-2})	59.67	
Defocus range (μm)	-1 to -4	
Calibrated pixel size (Å)	0.85	
Symmetry imposed	C_1 / C_2	
Particle images after 2D (no.)	191,211	
Final particle images (no.)	92,242	
Map resolution (Å) at FSC=0.143	3.28 (C_2) / 3.47 (C_1)	
Refinement	**Mec1**	**Ddc2**
Model resolution (Å) (FSC=0.143)	3.1	3.1
Map CC (whole unit cell)	0.84	0.88
Average B-factor (Å^2)	-51.41	-32.73
RMS deviations – Bonds (Å)	0.016	0.007
RMS deviations – Angles (deg)	1.370	0.910
Validation		
Molprobity Score	1.95	1.43
Clashscore	3.52	1.82
Ramachandran plot (%)		
Favoured	84.52	91.82
Allowed	15.12	7.82
Outliers	0.37	0.36
C-beta deviations	0	0
Rotamer Outliers (%)	1.52	0.21

A)


2D classification
(191,211 particles)
24,324 particles
107,156 particles
24,793 particles
34,938 particles
C2
3.29 Å resolution
No mask (4.3 Å)
Spherical (3.9 Å)
Loose (3.5 Å)
Tight (3.3 Å)
Corrected (3.3 Å)
Elevation
Azimut
# of images
C1
C2
(angstrom)
3.47 Å resolution
C1
3.28 Å resolution
C2

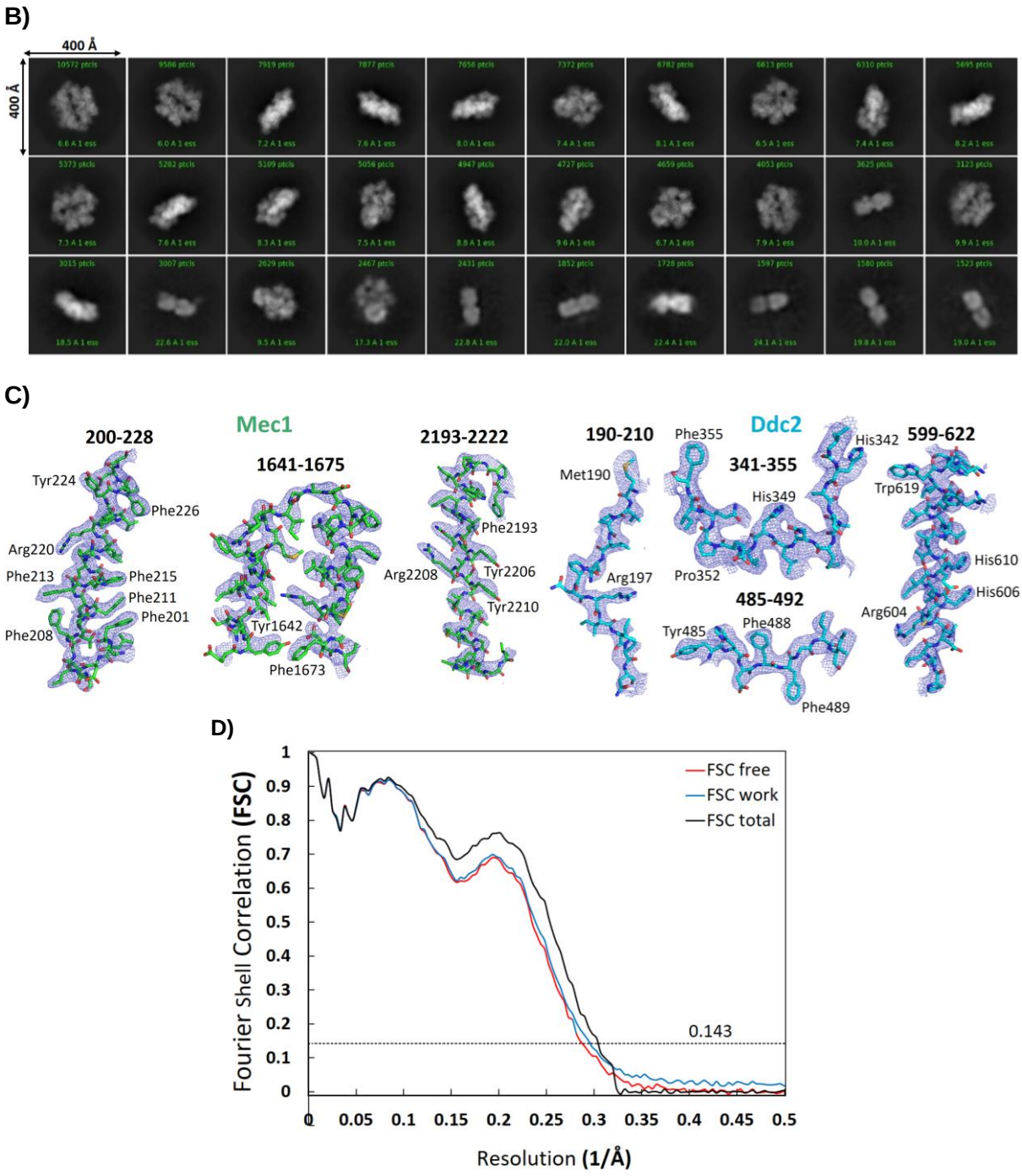


Figure 6.15 | Cryo-EM data processing of the Mec1-Ddc2 complex and atomic models. A) Overview of the cryo-EM processing in cryoSPARC v2. The resolution of each reconstruction after 3D refinement applying C1 and C2 symmetry is shown (FSC=0.143 criterion, blue line), in addition to the angular distribution plots for the maps (generated in the ab-initio classification). **B)** Enlarged view of the different 2D classes generated by cryoSPARC. **C)** Details of the atomic model fitting into the electron density map. Rendering was performed in PyMOL Molecular Graphics System (v1.8.2.1 Schrödinger, LLC). **D)** Map-vs-model FSC validation curves (detailed description in section 3.4.4).

Although many different regions benefited from the very large increase in resolution, the most relevant and significant improvement was achieved in the Kinase domain, with visible side chains for key catalytic residues within the active site of the enzyme. Likewise, it was possible

to build the full-length of the activation loop, which is incomplete in the available partial atomic model of Mec1.

After movie alignment, particle picking and extraction using the built-in processing modules in cryoSPARC v2, several rounds of 2D classification were performed, and the best 2D classes showing high-resolution features were selected (Fig. 6.15A, B) (3.4.3.2). Particles contributing to the best 2D classes were selected for ab-initio classification in cryoSPARC using four different initial classes. One “good” 3D class was refined - applying C2 symmetry - to a nominal resolution of 3.3 Å (FSC=0.143 threshold). An extra round of 3D ab-initio classification with two classes yielded a better quality 3D class, which refined to the same resolution as before. However, visual analysis of the refined map showed higher quality and better defined features comparing to the previous map, therefore this last map was selected to build the atomic model (Fig. 6.15C, D) (3.4.4.2). The local filtering tool in cryoSPARC was used to generate a map that was easier to interpret, especially in the lower resolution regions such as poorly structured loops.

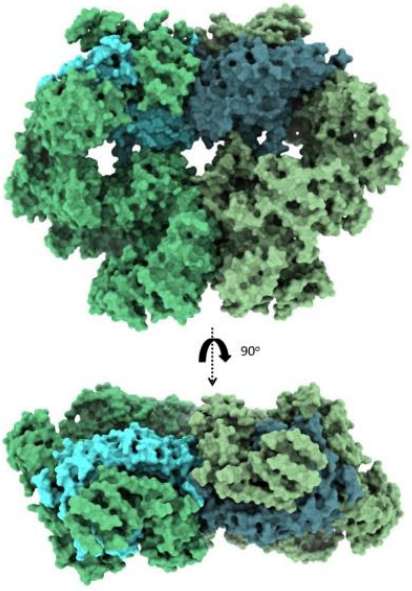
Distinct to the previously reported Mec1^{ATR} paralog protein ATM^{Tel1} (Baretić *et al.*, 2017), no open and closed conformations were observed during processing. This is also in agreement with recently published high-resolution structures of Tel1^{ATM} (Yates *et al.*, 2019), ATM^{Tel1} (Xiao *et al.*, 2019) and a 8.7 Å cryo-EM structure of Tel1^{ATM} (Wang *et al.*, 2016). However, a near-atomic resolution reconstruction of Tel1^{ATM} (Jansma *et al.*, 2019) also reports open and closed conformations; in addition, another recent reconstruction of Tel1^{ATM} (Xin *et al.*, 2019) describes different conformations for the polypeptide. One possibility is that some of these structures represent a transition state from dimer to monomer, as previously suggested (Xiao *et al.*, 2019), but in any case related to kinase activation either through monomerisation and/or binding to regulatory protein partners. In the case of the Mec1 polypeptide, no monomeric form was observed during data processing, as the Mec1-Ddc2 complex used for cryo-EM analysis was originated from a homogeneous cross-linked sample (Fig. 6.8D-F). The freshly purified complex used for cryo-EM was BS3 cross-linked previous to the injection to the gel filtration column using an optimised condition, potentially increasing the stability of the complex during sample freezing in the vitrification step and preventing the potential disassembly of the complex during the process. In addition, no monomeric form was observed during the purification process.

6.6.2 Complex architecture and assembly of subunits

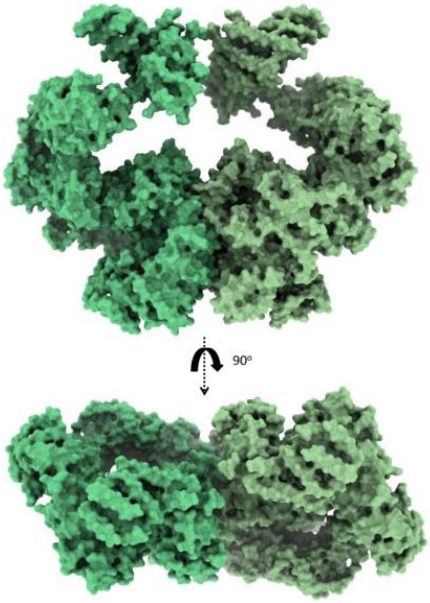
To facilitate the interpretation of the complex assembly, the Mec1-Ddc2 heterodimer was rendered in solid (Fig. 6.16), showing the fitting of Ddc2 polypeptides into the Mec1 subunits. Mec1 N-terminal region (residues 1-404) clamps Ddc2 solenoid domain like a **Hand** (Fig. 6.16A and B). Two Ddc2 solenoid domains dimerise together in a continuous fashion to form a **Solenoid Blade** (Fig. 6.16C,D). Each Mec1 Hand domain within the heterodimeric complex clamps one Ddc2 polypeptide at the ends of the Solenoid Blade (Fig. 6.16A and B), creating a compact structure as shown in the local resolution map (Fig. 6.15A), where the highest local resolution corresponds to the core of this dimerisation region.

The new cryo-EM map allowed the building of a higher quality atomic model of full-length Mec1 and Ddc2 solenoid region (see “Materials and Methods”, 3.4.4.2; Fig. 6.15C, D), providing a more accurate analysis of electrostatic interactions between them (Fig. 6.17). Surface charge analysis of the Mec1-Ddc2 complex shows that the interaction of the different polypeptides within the complex is largely driven by complementary patches of opposite charge, rather than hydrophobic interactions (Fig. 6.17). In the case of the FATKIN globular domain at the C-terminal region of Mec1, the positively charged patch in the FAT domain of one polypeptide interacts with the negatively charged interface of the Kinase domain within the other Mec1 polypeptide (Fig. 6.17A and B). This interaction generates a quadrupole-like structure on the solvent exposed interface of the Kinase region (Fig. 6.17A, Kinase side). Likewise, the interaction between Mec1 Hand domain and Ddc2 Solenoid Blade is largely driven by complementary patches of charged surfaces (Fig. 6.17C), as well as the dimerization interface between the two Ddc2 polypeptides in the Solenoid Blade (Fig. 6.17D), where there is another quadrupole-like interaction on the interface.

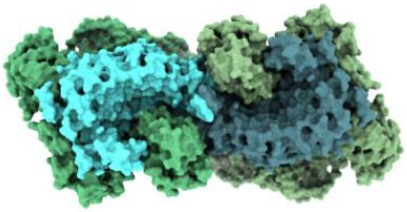
A) Mec1(FL)-**Ddc2**(195-747) heterodimer



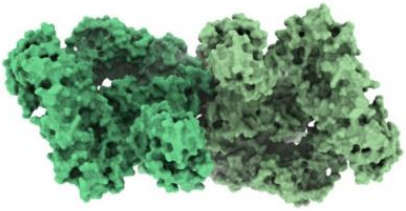
B) **Mec1** (FL) dimer



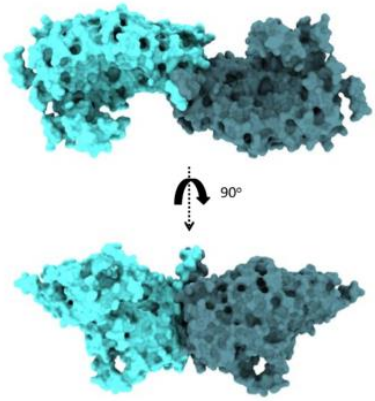
Mec1(236-2,368)-**Ddc2**(195-747) heterodimer

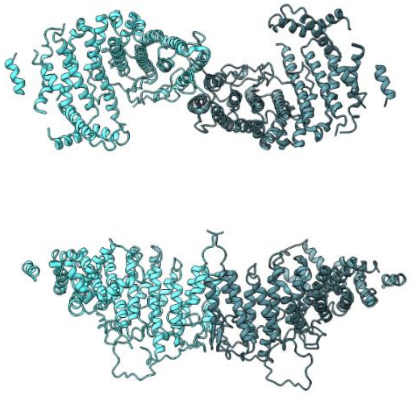


Mec1 (236-2,368) dimer



C) **Ddc2** (195 to 747) dimer





D)

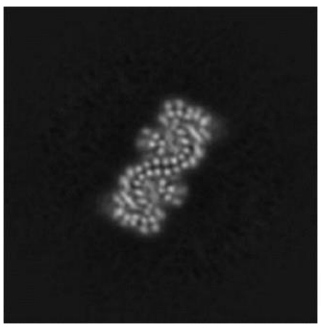
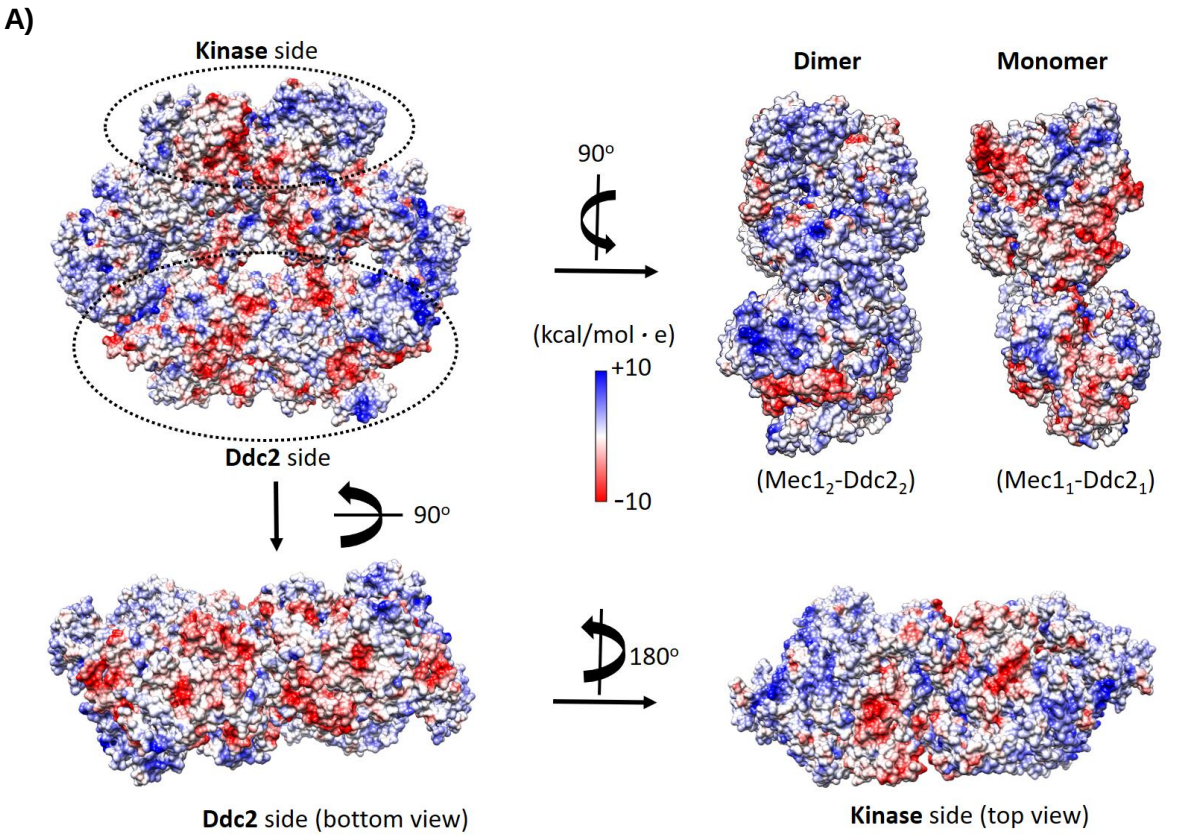
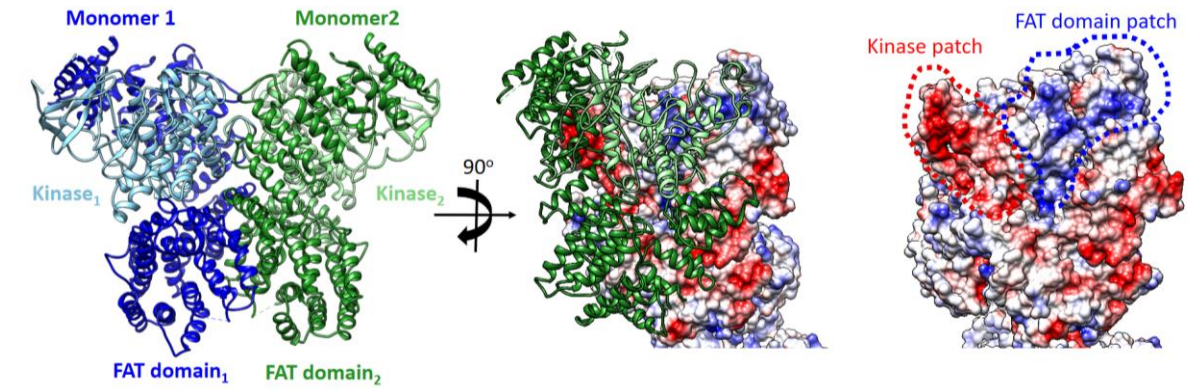


Figure 6.16 | Mec1-Ddc2 complex assembly. ChimeraX (Goddard *et al.*, 2018) was used to render the complex architecture. **A)** Mec1 polypeptides are rendered in dark and light green colour and Ddc2 polypeptides in blue. In the last image at the bottom, the first 235 residues of Mec1 are removed to facilitate the interpretation of the structure and to show Ddc2 Solenoid blade composed of two Ddc2 Solenoid domains. In the front view of the complex, the FATKIN domain is located at the bottom and Mec1 N-terminal region at the top. **B)** Same as **A** but without Ddc2. **C)** Ddc2 Solenoid Blade rendered in solid (**left**) and ribbon (**right**). **D)** A cross-section of the unprocessed Mec1-Ddc2 cryo-EM map as in Fig. 6.10D, showing Ddc2 Solenoid Blade. Next to the S-shaped cross-section, the corresponding atomic model of the Solenoid Blade is rendered at the equivalent position to the cross-section.





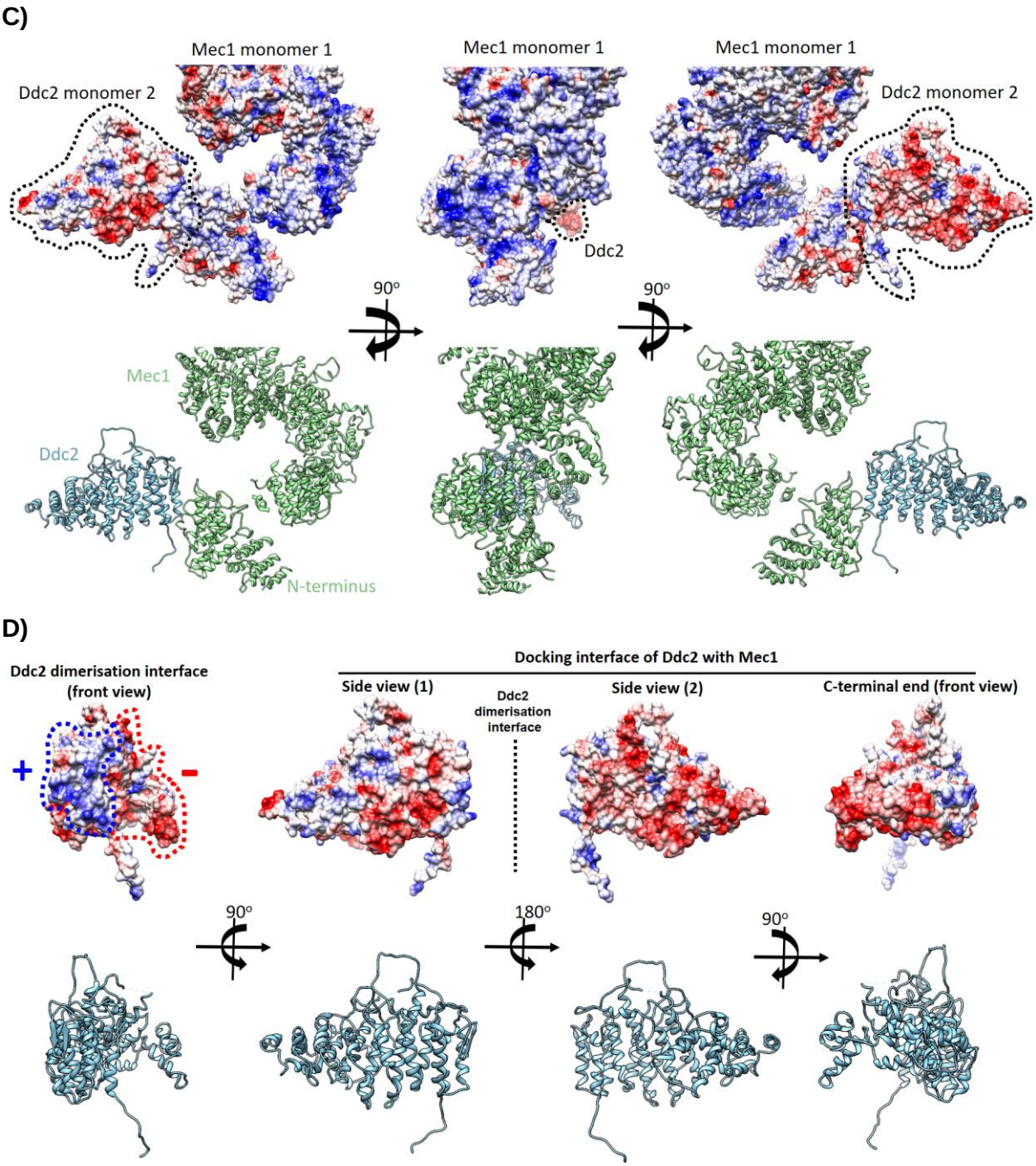


Figure 6.17 | Coulombic surface calculation and assembly of the [Mec1-Ddc2]$_2$ heterodimer complex. A) Coulombic surface rendering in Chimera (Goddard *et al.*, 2018) shows that a quadrupole like structure drives the interaction of the two Mec1 polypeptides at the FATKIN domain (FAT solenoid region together with the Kinase domain), where complementary charged surfaces interact with each other. In the front view, the FATKIN domain is located at the top and Mec1 N-terminal region at the bottom. Red colour represents negatively charged areas whereas positively charged surfaces are shown in blue (+/- kcal/mol · e). **B)** The FATKIN domains of the two Mec1 polypeptides within the complex are shown in blue and green (**left**). In the **middle** image, a 90 degrees rotation in the "Y" axis shows one of the Mec1 FATKIN domains rendered as in **A**. A closer view of the charged surfaces in the Kinase domain shows that a positively charged patch found at the FAT region of one Mec1 polypeptide interacts with the negatively charged patch found in the interface with the Kinase domain of the other Mec1 monomer. On the **right**, ribbon rendering has been removed for clarity and comparison purpose. Thus, the FAT domain of one monomer interacts with the Kinase domain of the other in a quadrupole like interaction. **C)** Mec1 and Ddc2 coulombic surface rendering is depicted to

show that the interaction along the extensive surface between Mec1 and Ddc2 is largely driven by complementary charged patches present on the interaction interface between them, where Mec1 is mainly positively charged and Ddc2 is negatively charged. **D)** A more detailed description of Ddc2 Solenoid domain is shown as in **C**, in addition to the corresponding ribbon representation. The interaction between Ddc2 solenoid domains is also driven by a quadrupole-like interaction ("Ddc2 dimerisation interface"), as is the case for the FATKIN domain.

6.6.3 The interaction of Ddc2^{ATRIP} N-terminal region with Mec1^{ATR}

The 3.3 Å cryo-EM map allowed to model at side chain resolution the first visible residues of the Ddc2 polypeptide present in the map (Fig. 6.18), which are stabilised by a hydrophobic pocket present in Mec1 (Fig. 6.19). This pocket represents a critical region for model building, considering that the two polypeptides are in contact and only a map with resolved features can be used to properly assign the density corresponding to each polypeptide (Fig. 6.12, 6.18 and 6.19).

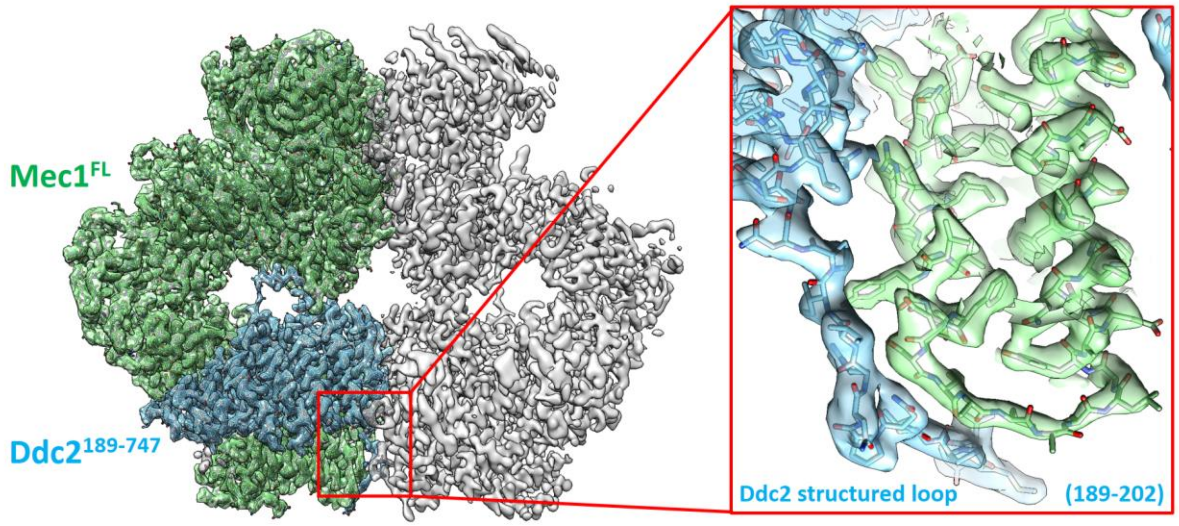


Figure 6.18 | Architecture of the Mec1^{ATR}-Ddc2^{ATRIP} complex. The atomic models of Mec1 and Ddc2, and the corresponding high-resolution cryo-EM map are shown. Full-length Mec1 is in green and Ddc2 residues 189-747 in blue. The zoomed region corresponds to the first N-terminal region of Ddc2 that is visible in the cryo-EM map, which is stabilised by Mec1 HEAT repeats 3 and 4 (Fig. 6.19).

The protein complex is recruited to RPA coated-ssDNA through the interaction of Ddc2 RBD with the Rfa1 subunit within the RPA trimeric complex (1.4.1). The accurate location of Ddc2 first visible residues is of enormous relevance to generate a reliable Ddc2 atomic model and thus to interpret the correct complex assembly and also to understand the recruitment of the complex to RPA coated-ssDNA during DNA damage or replication stress situations. The improved reconstruction of the complex very precisely shows the position of the first visible Ddc2 residue located at the N-terminal end corresponding to Asn189 (Fig. 6.19A), the first residue of the short N-terminal structured linker (Fig. 6.19A), and representing the integration point of Ddc2 RBD and CC domains into the complex. The first residues of the visible structured region are stabilised by the interaction of Ddc2 hydrophobic residues with a

hydrophobic pocket present in Mec1 located in between HEAT repeats 3 and 4 (Fig. 6.19B and C).

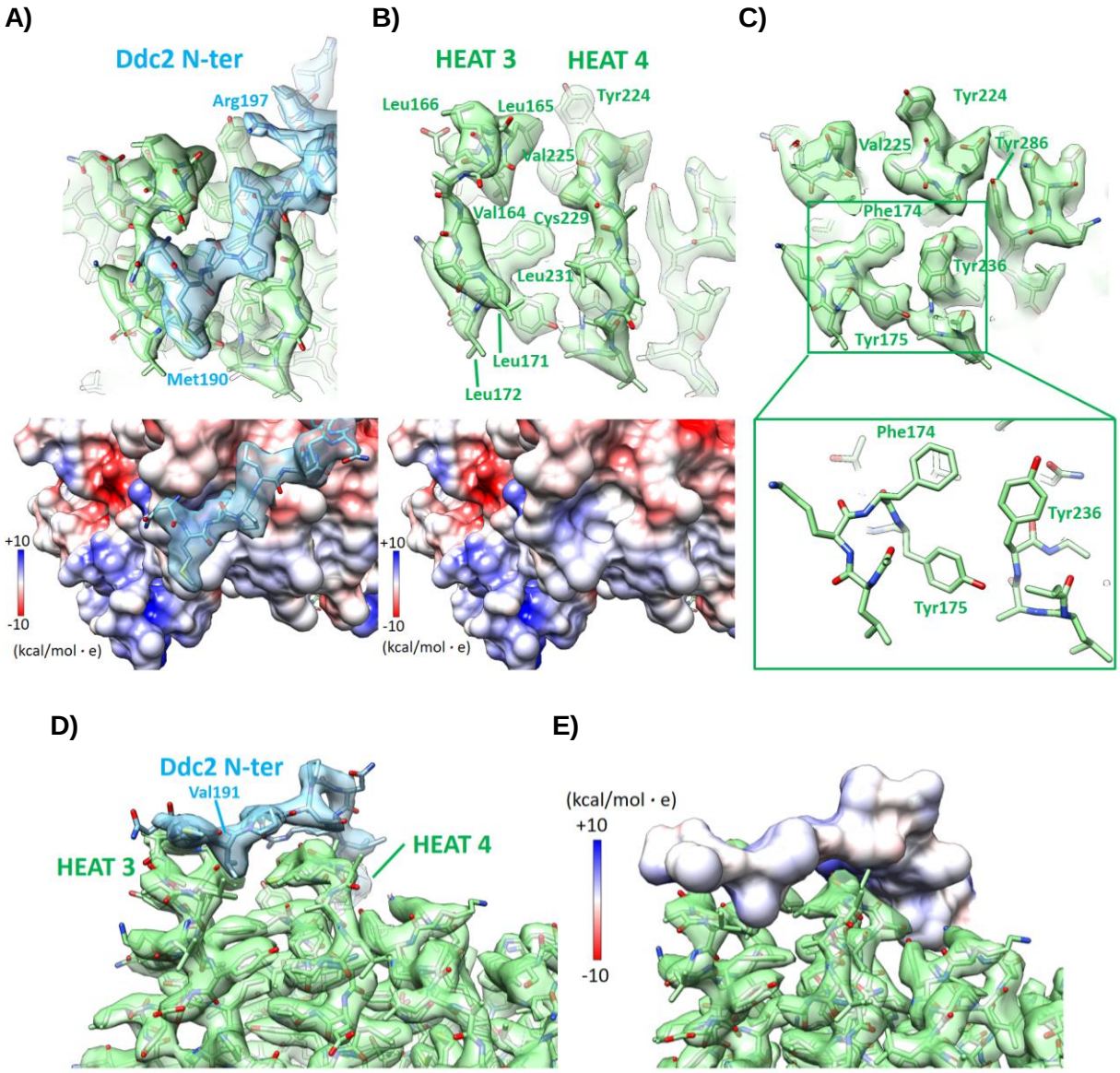


Figure 6.19 | Ddc2 N-terminal structured linker is stabilised by a hydrophobic pocket located between Mec1 HEAT repeats 3 and 4. The atomic models of Mec1 (green) and Ddc2 (blue) and the corresponding electron density map. **A)** Top view of the complex (Mec1 N-terminal region) to show the first visible N-terminal region of Ddc2 stabilised by Mec1 hydrophobic pocket. Coulombic surface calculation in Chimera for the Mec1 polypeptide is shown below. **B)** As **A** but Ddc2 is removed for clarity, to show the hydrophobic pocket where Ddc2 Val191 is inserted and the Mec1 hydrophobic residues contributing to the pocket. **C)** A cross-section of the map shown in **A** and **B** is rendered to show the cluster of hydrophobic residues contributing to the pocket where Ddc2 Val191 is inserted. **D)** A side view of the complex showing Ddc2 Val191 density within the hydrophobic pocket. **E)** As **D** but rendering coulombic surface calculation for Ddc2 to show the hydrophobic nature of the first track of Ddc2 residues present in the map.

More specifically, this hydrophobic pocket at the N-terminal region of Mec1 is formed by the proximity of two Mec1 loops connecting the α-helices of HEAT repeats 3 and 4. Ddc2 Val191 inserts into the pocket to stabilise the structured linker at Ddc2 N-terminal region (Fig. 6.19D, E).

6.6.4 Mec1 Kinase domain

6.6.4.1 General description of the Mec1 Kinase domain

The improved resolution obtained with the new cryo-EM reconstruction allowed to generate a high-quality atomic model of the kinase domain, covering most of the side chains present in this region. We can identify in Mec1 equivalent catalytic elements as found in other eukaryotic protein kinases (ePK) (Fig. 6.20), although the sequence of the catalytic motifs is not fully conserved in PIKKs when compared to canonical ePKs and therefore its definition as atypical protein kinases (1.1).

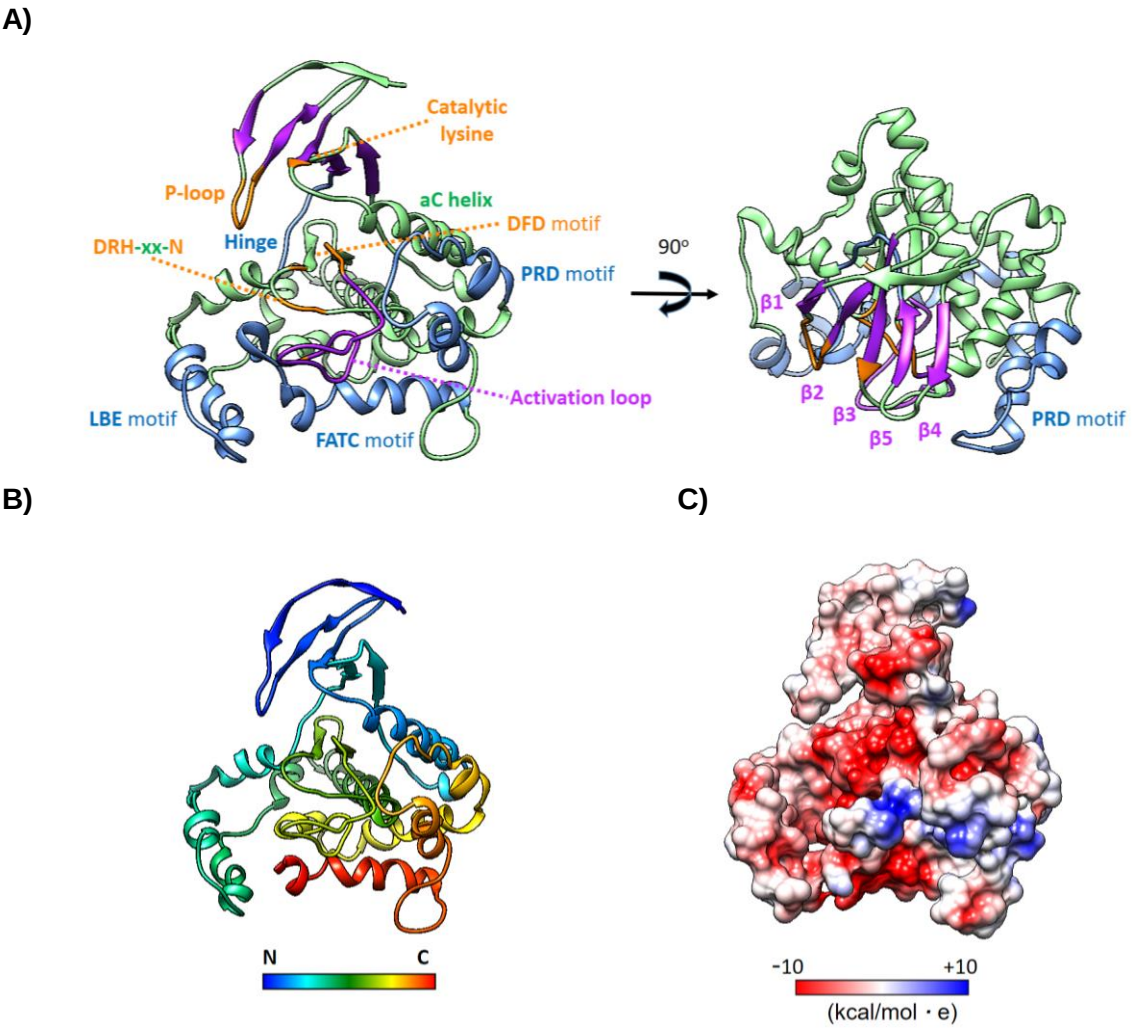


Figure 6.20 | The Mec1 Kinase domain. A) The Mec1 kinase domain is coloured according to its different motifs. The LBE, PRD and FATC structural motifs are in blue, and correspond to conserved motifs in PIKKs. The motifs or structural elements in orange are also present in ePKs. **B)** Rainbow representation of the kinase domain, in blue the N-terminal and in red the C-terminal region. **C)** Coulombic surface calculation in Chimera shows that the catalytic groove is dominated by negatively charged residues. Red indicates negatively charged and blue positively charged surfaces.

The N-lobe of the Kinase domain is mainly defined by five β-strands together with the αC helix, in addition to the P-loop and key residues that coordinate ATP binding to the active centre

such as the conserved catalytic lysine (Fig. 6.20A). In the C-lobe we find the activation and the catalytic loop (A-loop and C-loop, respectively). Previously described structural elements in PIKKs are the LBE, PRD and FATC motifs, from N to C-terminus (Fig. 6.20A, B). Interestingly, the Mec1 kinase domain is dominated by negatively charged surfaces (Fig. 6.20C), probably with a role in the interaction with its multiple substrates.

The so-called FATC domain inserts into a hydrophobic pocket where it is stabilised. Hydrophobic residues present in the LBE motif and Hinge helix are in close proximity to the aromatic residues found at the very C-terminal region of Mec1 (Fig. 6.21), and a cluster of five aromatic residues can be observed in this region.

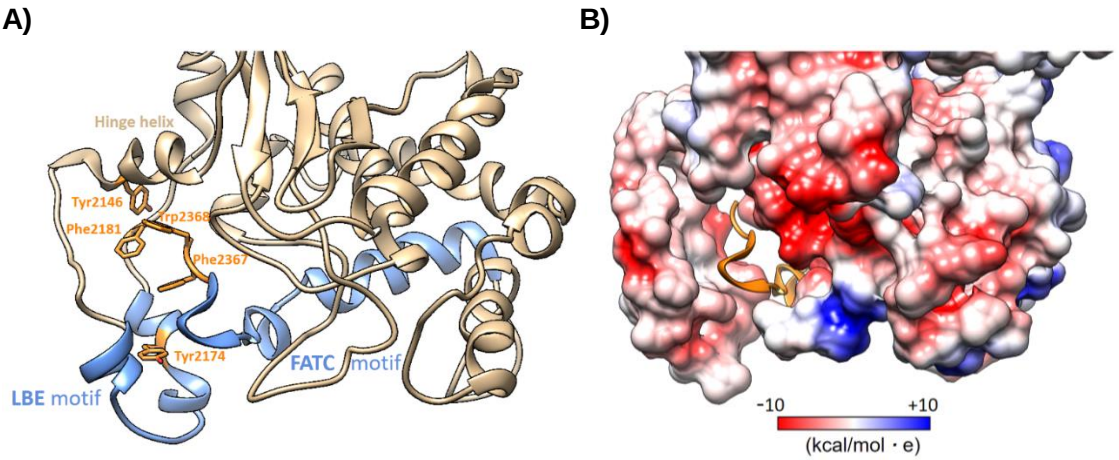


Figure 6.21 | Close up view of Mec1 C-terminal region. A) The conserved PIKK LBE and FATC motifs are rendered in blue. The very C-terminal end in Mec1 contains two aromatic residues in close proximity to other aromatic residues found in the Kinase Hinge and the LBE motifs (aromatic residues in orange). **B)** Coulombic surface colouring in Chimera of **A,** showing the hydrophobic contacts of the Mec1 C-terminal end, located at the FATC motif (orange).

6.6.4.2 Evolutionary analysis of the Mec1 Kinase domain

The first conserved structural element found in Mec1 is the P-loop (Fig. 6.22), which is called G-rich loop in ePKs. However, no glycines are found in PIKKs in this region, and presents the conserved consensus motif **S-L/A-**X**-KPK**. After the P-loop, a so-called catalytic lysine is also found in ePKs (K2080 in *S.c.* Mec1, Fig. 6.22), which is involved in ATP coordination in the active site of the enzyme (Huse and Kuriyan, 2002). In the case of PIKKs, the catalytic lysine residue is present and conserved in all members with demonstrated catalytic activity but the so-called VAIK motif containing this catalytic residue in ePKs is not conserved in PIKKs. A conserved hydrophobic residue in the so-called αC-helix is also found in evolution (Tyr2090 in *S.c.* Mec1), responsible to contact and stabilise a conserved phenylalanine in ePKs which is present in the so-called DFG motif of the activation loop. The DFG motif is replaced by a DFD motif instead in Mec1 (residues 2243-2245), and by the consensus motif HV**DF**-**D**/N when *S.c.* Mec1 sequence is compared to different organisms from yeast to humans (Fig. 6.22). At the end of the activation loop, a conserved threonine residue (Thr2265 in *S.c.* Mec1)

is found within the conserved **PFRLT** motif present in PIKKs, located in close proximity to the catalytic loop (4.3.3). For this reason, the A-loop is also sometimes called T-loop.

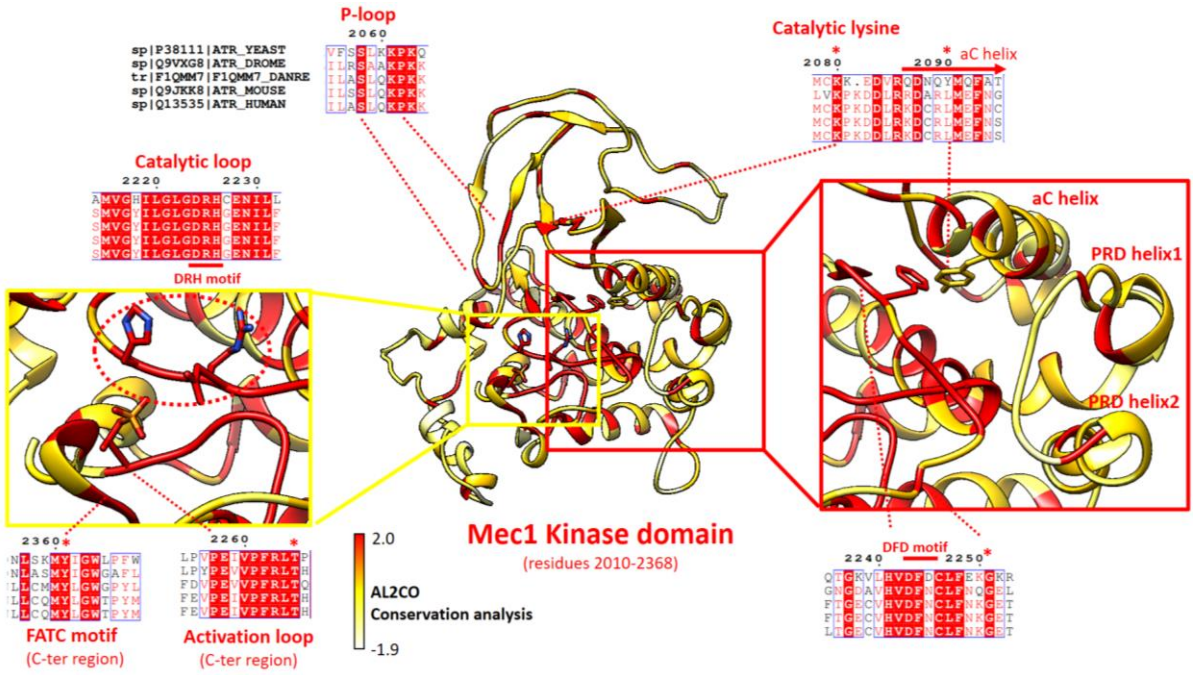


Figure 6.22 | Residue conservation in Mec1^{ATR} kinase domain from yeast to humans. *S. cerevisiae* Mec1, *D. melanogaster* ATR, *Danio rerio* ATR, *Mus musculus* ATR and *Homo sapiens* ATR protein sequences were aligned in Clustal Omega (Sievers and Higgins, 2018) and rendered in Chimera. The AL2CO algorithm (Pei and Grishin, 2001) was used for rendering, where values ranged from -1.891 (in white) for the less conserved regions to 1.985 (in red colour) for fully conserved residue positions; yellow colour represents a value of 0.047, as calculated by AL2CO. Sequence alignment is also shown for key catalytic residues present in the active site. The rendering of the aligned sequences was performed in ESPript 3.0 (Gouet, Robert and Courcelle, 2003).

6.6.4.3 Mec1 Kinase domain versus the other PIKK members

Sequence analysis of PIKKs with demonstrated catalytic activity (i.e. all of them but Tra1^{TRAAP}) shows that the only conserved residue in the P-loop is a proline (Pro2062 in *S.c.* Mec1) (Fig. 6.23). This proline is in the β2 strand of the kinase domain and is aligned in space to the conserved catalytic Lysine 2080 (β3), Gly2127 (β5) and Ser2121 (β4), the last one being a proline in the other PIKKs, including the Mec1 human homolog ATR. Mec1 Asp2083 and Asp 2087 are also remarkably well conserved among catalytic PIKKs, and both are facing towards the ATP/Mg++ biding pocket. Interestingly, the only conserved position in the so-called DFG motif found in ePKs is aspartic acid (Fig. 6.24), which is the first residue of the DFD motif in Mec1 and the DFN motif in ATR. The second position of the DFG motif in PIKKs can be occupied by a Phe, Tyr or Leu, and the residue in this position will very likely interact with a Val, Met, Leu or Iso present in the αC-helix which is located in close proximity (Fig. 6.24), positioning for catalysis the aspartic acid residue in the DFG motif present in the activation loop as previously described in ePKs (Kornev *et al.*, 2006). Mec1 is the only PIKK with two aromatic residues in these two closely located positions (a D**F**D motif in the activation loop

and a tyrosine in close proximity present in the αC-helix). The C-terminal region of the activation loop (Pro2257-Thr2265 in Mec1) is also extremely well conserved in PIKKs, where a **PFRLT** motif can always be identified (4.3.3; 6.6.4.2). This is also the case for the catalytic loop (Ile2219-Asn2229), where the catalytic residues are present within a different consensus motif comparing to ePKs; instead of a HRD motif, in PIKKs we find an inverted DRH motif, so that the catalytic motif in PIKKs is the conserved **GDRH-XX-N** sequence (4.3.3; Fig. 6.23). Stickily speaking, we can identify the conserved **ILGL**-**GDRH**-XX-**N** motif in the activation loop in all catalytic PIKKs, with only two exceptions where a single position is substituted by a functionally equivalent residue.

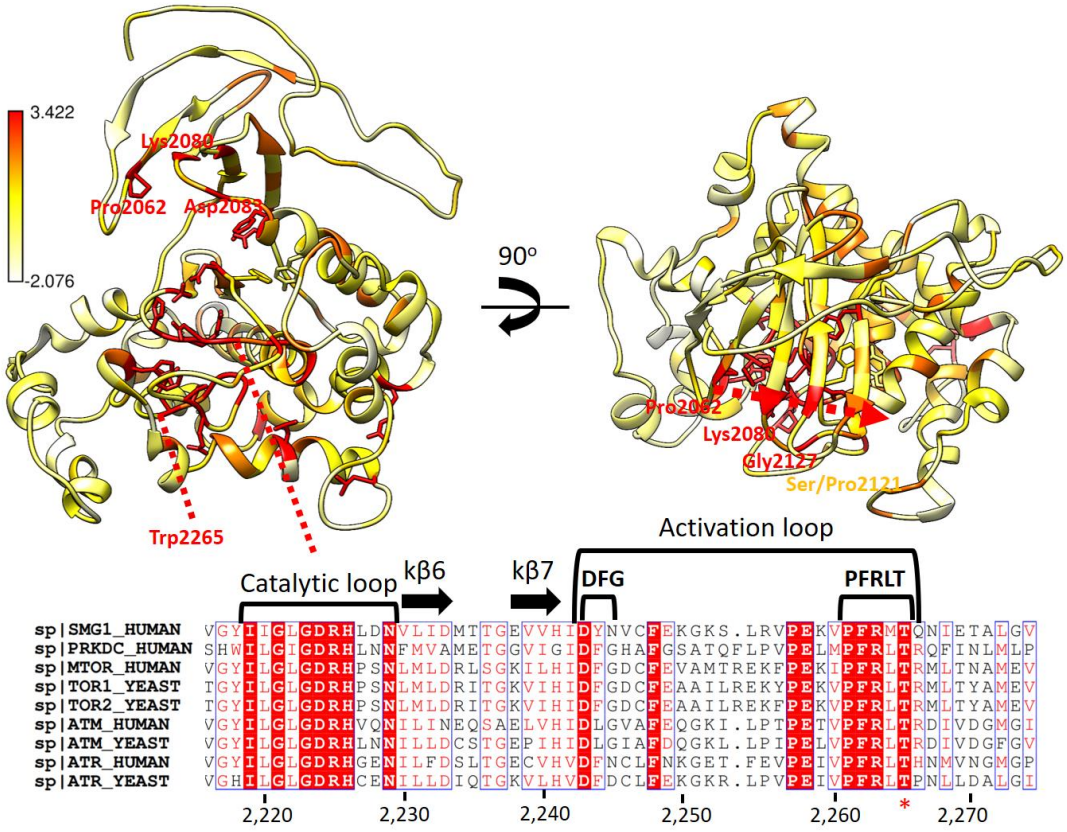


Figure 6.23 | Residue conservation in the PIKK protein family. The protein sequences corresponding to human DNA-PKcs, mTOR, ATR, ATM and SMG1, and the yeast PIKKs Tor1, Tor2, Mec1 and Tel1 were aligned in Clustal Omega and rendered in Chimera using the AL2CO algorithm and Mec1 atomic model (this study) as previously described in Fig. 6.22. In the N-terminal lobe of the kinase, residues Pro2062, Lys2080, Gly2127 and Pro2121 (Ser in Mec1) are conserved and aligned in space. Asp2083, in close proximity to the catalytic lysine 2080, is also fully conserved in all aligned PIKKs. In the C-terminal lobe of the kinase, sequence alignment shows a conserved ILGL-GDRH-XX-N motif in the catalytic loop and a conserved PFRLT motif at the end of the activation loop, with the only exception of SMG1 where the leucine residue is substituted by methionine. Residue conservation values as calculated by AL2CO algorithm are shown; maximum conservation values in red (3.422) and minimum conservation values in white (-2.076), whereas intermediate values are rendered in yellow colour (0.673). PRKDC: human DNA-PKcs. (The secondary structure elements in the alignment correspond to Mec1).

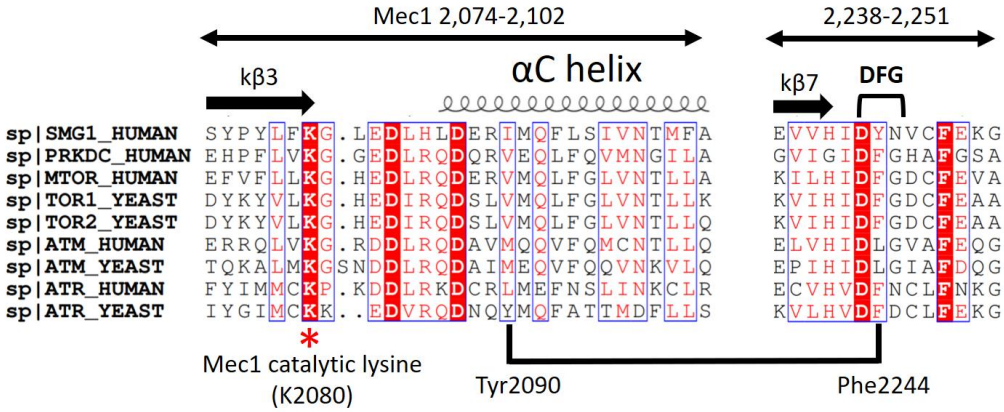


Figure 6.24 | The DFG motif in the PIKK protein family. Sequence alignment of yeast and human PIKK protein sequences in Clustal Omega as in Fig. 6.23 shows that the DFG motif found in ePKs is present in all members, with variations. The first position in the DFG motif is always occupied by an aspartic residue in all PIKKS. In the αC-helix we find a hydrophobic residue that very likely interacts with and stabilises the activation loop through the interaction with the DFG motif, orienting the aspartic acid residue for catalysis as previously described in ePKs. In the case of Mec1, Phe2244 in the DFD motif interacts with Tyr2090 located in the αC-helix. The catalytic lysine (K2080 in Mec1) is conserved in all PIKKs, in addition to two aspartic residues (D2083 and D2087 in Mec1) in close proximity, likely involved in the coordination of ATP. Therefore, a consensus motif can be defined for the PIKK protein family, where instead of the catalytic lysine containing VAIK motif in ePKs we find in PIKKs the conserved **K**G-X-D/E-**D**-I/L/V-RQ**D** motif.

Finally, a tryptophan located close to the very C-terminal region of the polypeptide is fully conserved in all catalytic PIKKs, including SMG1. This position corresponds to Trp2364 in *S.c.* Mec1, and is in close proximity in space to the catalytic loop or T-loop where Thr2265 is found, which is the last fully conserved residue within the activation loop as the last residue of the conserved PFRLT motif. The equivalent threonine position is auto-phosphorylated in PKA (Thr197) and is considered a landmark of protein kinase activation (Taylor *et al.*, 2013) given that is a requirement for activation, as will be discussed in the next section (6.6.4.4). Thus, following a similar mechanism, the arginine residue present in the **GDRH-XX-N** motif would interact with the phosphorylated threonine position in the **PFRLT** motif (Fig. 6.23). Hence, the phosphate group at this position would stabilise the active configuration of the catalytic loop as previously described for other ePKs (Taylor and Kornev, 2011), so that the phosphorylation of this threonine is a requirement previous to kinase activation. Therefore, in Mec1 this conserved threonine position (Thr2265) is very likely to be phosphorylated, as is suggested by our cryo-EM map, although MS analysis confirmation is still needed. In addition, the reconstruction in C1 symmetry suggests that Thr2265 might be found in the phosphorylated state only in one monomer at a time within the complex, but it also needs to be fully confirmed.

The catalytic loop is extremely well-conserved in evolution (Fig. 6.23), and the catalytic aspartic residue within the GDRH-XX-N motif plays a fundamental role in the transfer of the phosphate group from the ATP molecule coordinated in the active centre to the acceptor residue in the substrate protein (Kornev *et al.*, 2006). In addition, the conserved asparagine

residue found at the end of the catalytic loop in the GDRH-XX-**N** motif (Asn2229 in Mec1) has been shown to play a role in ATP coordination in mTOR (Yang *et al.*, 2017).

6.6.4.4 Understanding Mec1 and the PIKK Kinase domain by the comparison to PKA as a canonical ePK

Because PKA is one of the best studied protein kinases, with an extensive amount of literature accumulated over the years, the direct comparison of Mec1 atomic model against PKA provides valuable information regarding to Mec1 mechanism of action, as well as for the other PIKKs.

Sequence alignment of the P-loop in Mec1 indicates that this segment is poorly conserved in comparison to other ePKs, as the so-called G-rich motif present in ePKs is missing (Fig. 6.22). However, the P-loop in Mec1 also participates in the formation of a hydrophobic pocket, which is involved in ATP binding in ePKs like PKA (Kornev *et al.*, 2006), suggesting that there is structural and functional conservation (6.6.4.5).

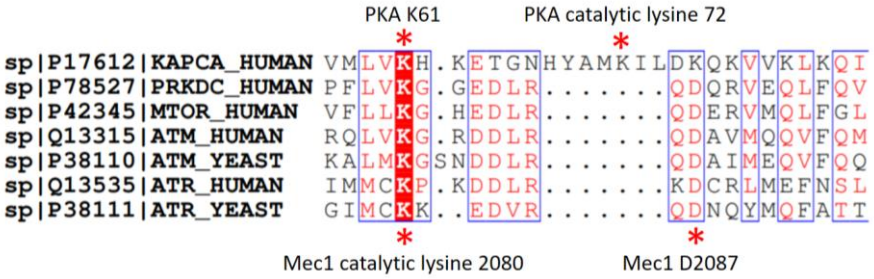
In PKA, K72 is the so-called catalytic lysine; it is found in the β3 of the kinase domain, and sequence alignment shows that this position is within a track of seven residues which is absent in the PIKK family (Fig. 6.25A). However, 3D alignment of PKA atomic model to Mec1 kinase domain shows that a conserved lysine in PIKKs occupies in space the equivalent position to the catalytic K72 (K2080 in Mec1) (Fig. 6.25A-C), showing structural conservation. Interestingly, this lysine position in Mec1 (K2080) is also conserved in the PKA sequence, and corresponds to K61 (Fig. 6.25A), but in this case the three dimensional position of this residue in PKA is different when compared to the PIKK architecture (PKA K61 is found at the end of β2 and Mec1 K2080 at the end of β3). Therefore, the conserved PKA K61 position became the catalytic lysine in PIKKs, as sequence and structural analysis strongly suggest.

Sequence alignment shows that Mec1 D2087 is also highly conserved in PIKKs (Fig. 6.25A). When Mec1 atomic model is compared to PKA, Mec1 D2087 seems to occupy the equivalent position to PKA E91, which is involved in the interaction with the catalytic lysine (Fig. 6.25B, C).

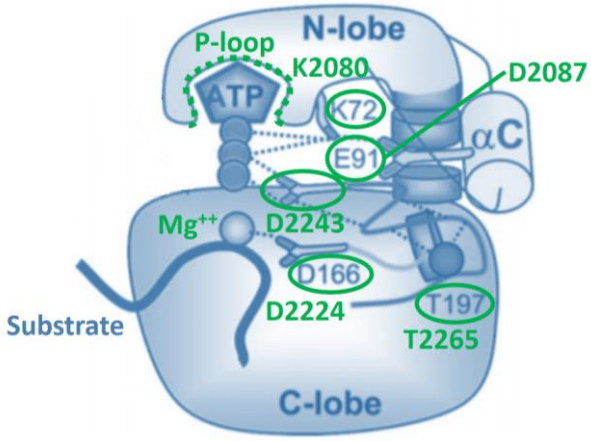
In Mec1, the activation loop is found in the so-called “IN” position in both monomers within the complex (Treiber and Shah, 2013), which is a landmark of protein activation in PKA where the activation loop can be found also in the “OUT” position. In the “IN” position, the activation loop is stabilised by hydrophobic interactions against the αC-helix (6.6.4.3). In this regard, Mec1 Phe2244 within the DFD motif is stabilised by the proximity of Tyr2090 in the αC-helix (Fig. 6.22-6.24). Although the active conformations of protein kinases have been shown to be essentially the same in all ePKs, the inactive conformations are structurally diverse (Jura *et al.*, 2011), so that this is probably one of the main differences when compared to PKA.

Distinct to PKA, where we find three threonine residues in the activation loop, the only phosphorylable position in Mec1 activation loop is Thr2265, preventing ambiguity in respect of the target residue that is auto-phosphorylated. In PKA, Thr197 becomes the phospho-threonine residue in the activation loop upon kinase activation (Fig. 6.25B), and its role is to interact with the arginine present in the HRD motif in the catalytic loop, modified to the inverted DRH motif or GDRH-XX-N in PIKKs. In addition, the phosphorylation of the activation loop is essential for the correct positing of key catalytic residues previous to phosphoryl transfer (Steichen *et al.*, 2012). Therefore, this event is a requirement in protein kinase activation as mentioned before (6.6.4.3) and can be used as a signature to assess the activation state of a kinase.

A)



B)



C)

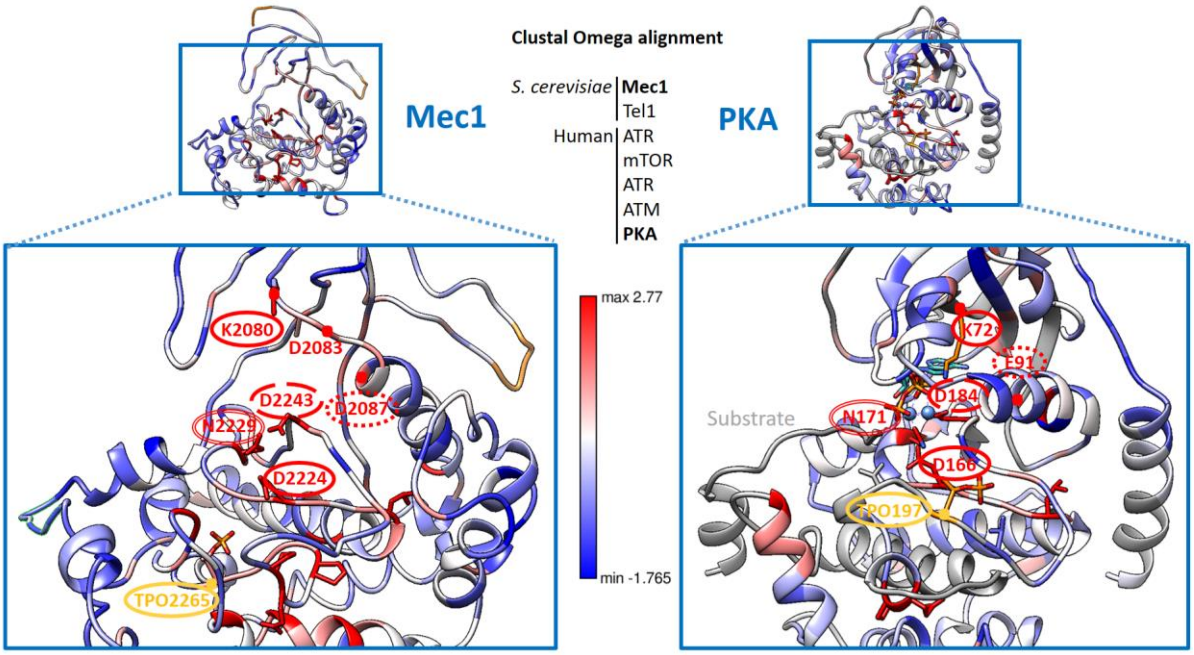


Figure 6.25 | Key catalytic residues in PKA are conserved in all PIKKs. The comparison of Mec1 and other PIKKs against PKA shows residue and structural conservation of the same catalytic elements, suggesting a universal mechanism of action in ePKs including the PIKK protein family. **A)** Sequence alignment in Clustal Omega of the catalytic lysine in PIKKs versus PKA. Mec1 K2080 is conserved in all aligned protein kinases, including PKA. **B)** Schematic representation of PKA catalytic residues (in blue), together with the equivalent catalytic residues in Mec1 (in green). (Adapted from Kornev *et al.*, 2006). **C)** Protein sequences corresponding to *S. cerevisiae* Mec1 and Tel1, together with human ATR, mTOR, ATM and PKA were aligned in Clustal Omega and plotted into Mec1 (this study) and PKA (PDB 1ATP) atomic models using AL2CO algorithm in Chimera. The key catalytic residues are conserved in space, occupying equivalent positions in the structures. Equivalent residue positions: Mec1 K2080 and PKA K72 (catalytic lysine), Mec1 D2087 and PKA E91 (involved in the interaction with the catalytic lysine), Mec1 D2224 and PKA D166 (present in the catalytic loop), Mec1 N2229 and PKA N171 (catalytic loop), Mec1 D2243 and PKA D184 (first reside of the DFG motif in the activation loop), and finally Mec1 T2265 and PKA T197 (PIKK "PFRLT" motif in the activation loop). In PKA, T197 is auto-phosphorylated previous to kinase activation; in Mec1, this position corresponds to T2265, the only Ser/Thr residue within the activation loop. In the figure, Thr2265 is illustrated with a putative phosphate group for explanation and clarity purposes. Equivalent residue positions in Mec1 and PKA atomic models are enclosed using the same line style to facilitate their identification in the figure. In addition, a conserved aspartic residue corresponding to Mec1 D2083 is found in all aligned PIKKs, and absent in PKA. Red dots are used to indicate with more clarity the position of labelled residues.

Sequence alignment of PIKKs with demonstrated catalytic activity against PKA was performed in Clustal Omega and shows that the key catalytic residues in the active site are fully conserved. The rendering of these conserved positions into the corresponding atomic models (Mec1, this study; PKA, PDB 1ATP) also shows structural conservation in addition to residue conservation (Fig. 6.25C), strongly suggesting that the mechanism of action is also conserved. Therefore, although PIKKs have been described as atypical protein kinases due to sequence deviations when compared to consensus catalytic motifs in ePKs, the analysis of the data

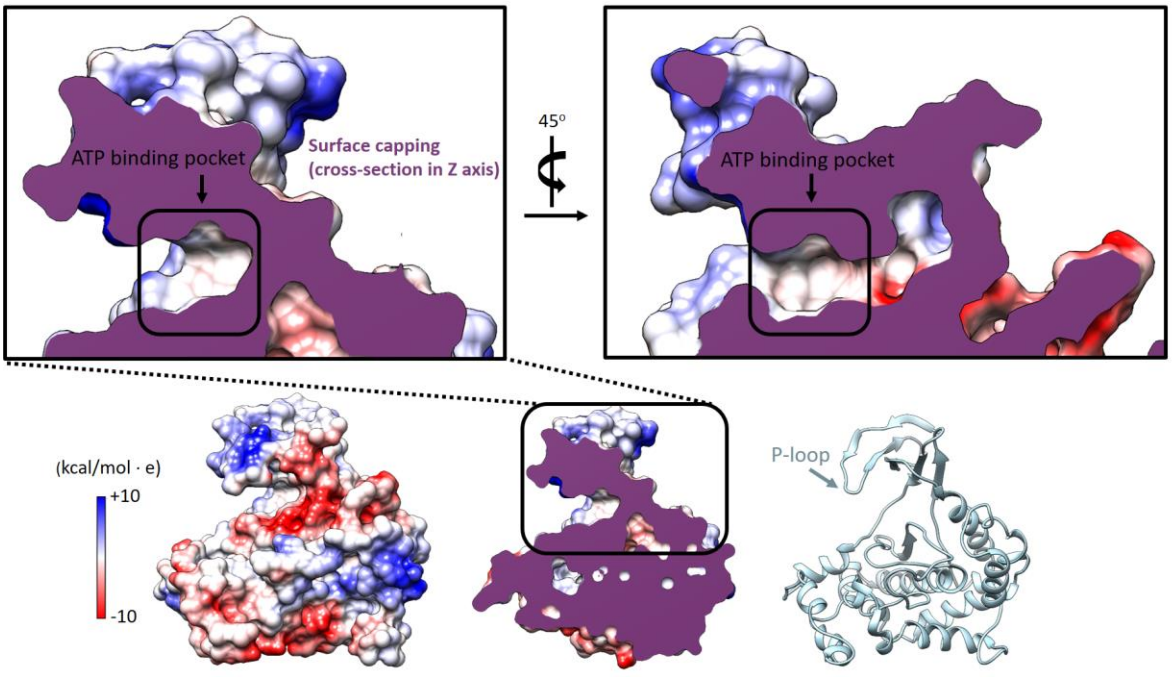
strongly suggests that in addition to structural conservation, the mechanism of action is probably also very well conserved.

Finally, although the reason behind the fact that the complex is found in the cell as a dimer still remains a mystery with the available information, it has been previously shown in some protein kinases like the EGF receptor that one kinase domain is activated by another kinase domain (Jura *et al.*, 2009). Therefore, one of the main keys to decipher Mec1 mechanism of action will be to understand the communication between the two Mec1 kinase domains. In addition to potential allosteric regulation at distant regions from the kinase domain, the direct contact of the two kinase domains opens the possibility of regulation through the direct communication between the catalytic domains within the dimer of heterodimers.

6.6.4.5 Mec1 Kinase domain: the ATP binding pocket and pharmacological implications

In ePKs, we find a structurally conserved hydrophobic pocket that accommodates the adenosine fraction of ATP (Fig. 6.26). This hydrophobic pocket is located under the P-loop and next to the Hinge connecting the N and C-lobes of the kinase, and is also present in Mec1 (Fig. 6.26A). One specific residue, located in proximity to the Hinge region, has been found to be critical for the binding of ATP analogs and therefore presents a role in the resistance to ATP competitive kinase inhibitors. This residue is the so-called gate-keeper (Noble, Endicott and Johnson, 2004), and its name is given because protein kinases usually have a bulky amino acid in this position, so that mutations to Gly or Ala facilitates the binding of ATP analogs creating a deeper hydrophobic pocket (Taylor and Kornev, 2011).

A)



B)

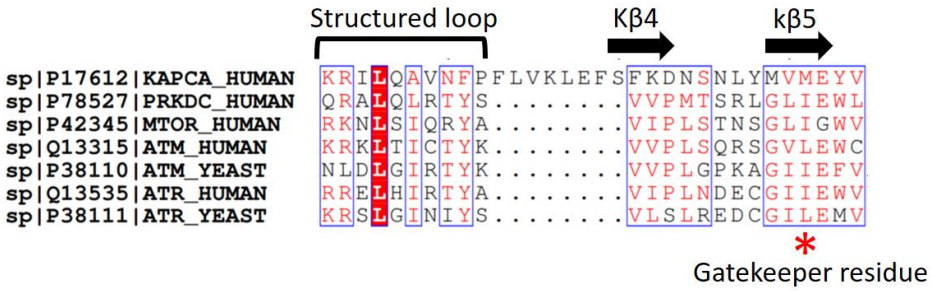


Figure 6.26 | The ATP binding pocket in Mec1. A) Coulombic surface representation of Mec1 kinase domain and its ATP binding pocket. Although the sequence of the P-loop is not conserved when compared to the consensus sequence present in other ePKs, the hydrophobic cavity or pocket that accommodates ATP is structurally conserved. **B)** Sequence alignment of different PIKKs together with human PKA (KAPCA) showing the gatekeeper position ("star"). The ATP binding pocket in PKA is formed by residues found in the P-loop and also the Hinge linker connecting the N and C-terminal lobes of the kinase. The hydrophobic pocket that accommodates the adenosine fraction of ATP is influenced by the gatekeeper residue, typically consisting in a leucine or isoleucine in PIKKs (L2129 in Mec1 and I2237 in mTOR); in human PKA we find a methionine residue at position M120. (The secondary structure elements shown in the alignment correspond to *S. cerevisiae* Mec1).

In PKA, the gatekeeper is Met120, whereas in PIKKs this position can be occupied by a Leu or Ile (Fig. 6.26B). In Mec1, the equivalent position is L2129 and is found in the β5 of the kinase catalytic domain (β8 of the polypeptide). In mTOR, this position is occupied by an isoleucine and corresponds to residue I2237 (Wu *et al.*, 2015) (Fig. 6.26B).

6.6.5 An integrative atomic model of the Mec1^{ATR}-Ddc2^{ATRIP} complex

The crystal structure of Ddc2 coiled-coil domain published by the Gasser lab (Deshpande *et al.*, 2017) perfectly explains how the complex is recruited to RPA-coated ssDNA through the interaction of Ddc2 RPA binding domain with Rfa1 N-terminal OB domain - that belongs to one of the subunits in the RPA trimeric complex (Fig. 6.27). The published work shows how the N-terminal region of two different Ddc2 polypeptides assembles to form the CC domain. However, the integration point of Ddc2 coiled-coil into Mec1 through the unstructured linker has remained unresolved until now. The structure presented in this study shows at atomic detail the insertion point of Ddc2 N-terminal region to the rest of the complex (Fig. 6.18 and 6.19). The combination of our cryo-EM map together with Gasser lab's crystal structure provides a structural framework to better understand the recruitment of the Mec1-Ddc2 complex to RPA-coated ssDNA (Fig. 6.27). Structural studies focused on an upstream step related to the binding of the RPA trimeric complex to ssDNA further expand this model (Yates *et al.*, 2018).

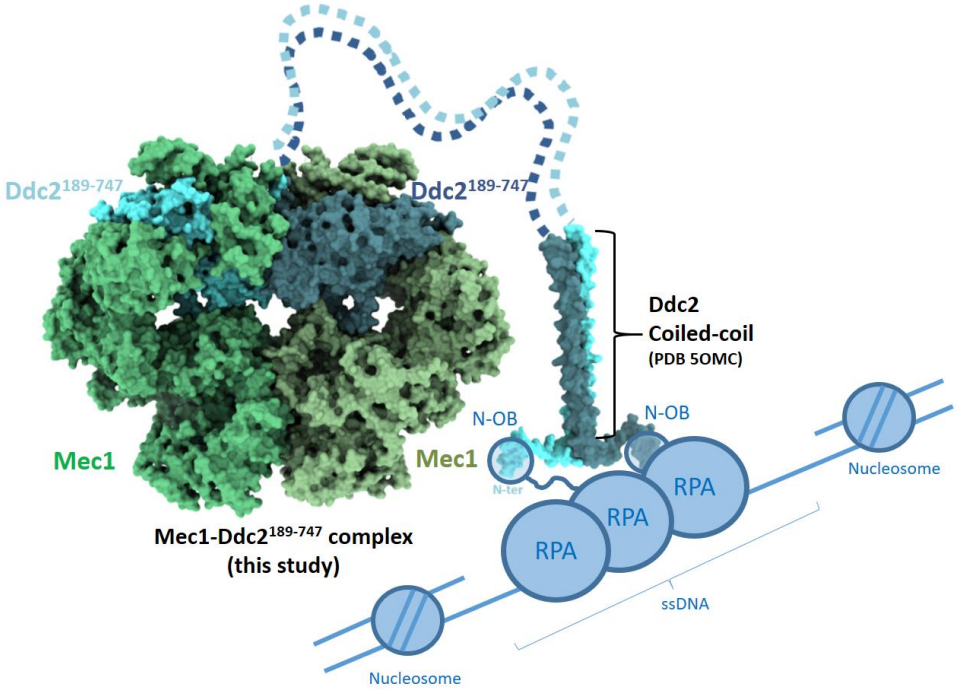


Figure 6.27 | An integrative atomic model for the Mec1-Ddc2 complex. The combination of Mec1^{FL}-Ddc2$^{189-747}$ atomic model - built from the high-resolution cryo-EM reconstructions presented in this study (6.5 and 6.6) - and an available crystal structure of Ddc2 N-terminal region from *K. lactis* (PDB 5OMC) provides a more complete model for the understanding of the complex recruitment to RPA-coated ssDNA. Dotted lines represent the disordered linker of each Ddc2 subunit within the complex. N-OB: *S. cerevisiae* Rpa1 N-OB domain. RPA: trimeric complex composed of Rpa70 (Rpa1), Rpa32 (Rpa2) and Rpa14 (Rpa3) in *S. cerevisiae*.

7 Final conclusions

7.1 Synopsis

Previous to the presented research work, no structural information was available for Tra1^{TRAAP}, which is an essential protein in all eukaryotes and the largest transcription coactivator (1.2.4), and only low resolution maps were publicly available for Mec1^{ATR}-Ddc2^{ATRIP}, an essential complex traditionally described in the DNA damage response (1.4). The presented atomic structures shed light into the key biological processes where they are involved as the highly conserved and essential eukaryotic genome regulators they are. In addition, the presented structural information greatly contributes to the understanding of the structural and functional diversification of the conserved PIKK protein family.

7.2 Main contributions of the present work

In this regard, a high-resolution cryo-EM map of *S. cerevisiae* Tra1^{TRRAP} was obtained (4.2), allowing to generate a complete *"de novo"* atomic model (4.3.1). The structure shows a very large alpha-solenoid region organised into an intricate network of HEAT repeats, which define a Finger and a Cradle or Ring region (4.3.2). The large N-terminal HEAT repeat region connects to a large globular domain that comprises the Kinase domain at the C-terminal preceded by a short solenoid segment of 15 TPR repeats (FAT domain) (4.3.3). The atomic model of Tra1 was fitted into existing low and medium resolution maps of SAGA (4.3.6) and NuA4 (4.3.7), revealing the contacts with the rest of the subunits within the complexes (4.3.8). The atomic model allowed to map previously reported mutations affecting the interaction of Tra1 with transcription activators (4.3.5); in the case of the yeast Gal4 activator, these mutations cluster together in an evolutionary conserved area from yeast to humans (4.3.4), suggesting that the same interaction region in Tra1^{TRRAP} might be used by human transcription activators. Isolated Tra1 was incubated with Gal4 transcription activation domain (Gal4 TAD) in this study (5.3), and the corresponding cryo-EM reconstruction revealed a density located at the Gal4 interacting region present in Tra1 (5.4).

Regarding to the Mec1-Ddc2 complex, the obtained high-resolution cryo-EM map allowed to generate the first complete atomic model of Mec1 (6.5.2), revealing the complex architecture and assembly. The complex adopts a stable dimer of heterodimers configuration composed of two Mec1 and two Ddc2 polypeptides (6.6.2). Two Ddc2 solenoid domains interact together to form a dimeric solenoid structure that binds to the N-terminal solenoid region in Mec1. Furthermore, the integration point of Ddc2 coiled-coil domain to the rest of the complex is reported, identifying a hydrophobic pocket that stabilises the first residues present in the short structured linker found before Ddc2 solenoid domain (6.6.3). The high-resolution

reconstruction also allowed to generate a high-quality atomic model of the kinase active site, where the full-length density of the activation loop was entirely visible and allowing to generate a complete atomic model (6.6.4). Distinct to Tra1, predicted to be a pseudo-kinase, the Mec1 active site shows structural conservation of key catalytic residues when compared to other eukaryotic protein kinases, suggesting that it is reasonable to expect also mechanistic and functional conservation. In addition, the combination of an existing structure of Ddc2 coiled-coil domain together with the Mec1-Ddc2 atomic model presented in this study allowed to generate an integrative atomic model of the complex (6.6.5), completing the existing model of complex recruitment to RPA-coated ssDNA during the DDR and RSR.

Finally, the presented atomic structures allowed to compare at the structural level the PIKK protein family (1.1.3), showing three pairs of paralog members composed of $Tra1^{TRRAP}$ (this study) and DNA-PKcs, $Mec1^{ATR}$ (this study) and $Tel1^{ATM}$, mTOR and SMG1. That might suggest that a duplication event occurred in evolution, where three initial and essential PIKKs duplicated to generate its non-essential homolog counterpart that specialised in a closely related but at the same time different biological function and becoming a paralog protein. Sequence and structural analysis allowed to define new catalytic motifs for the PIKK protein family (4.3.3; 6.6.4.2; 6.6.4.3), including the conserved **GDRH-**XX**-N** (or **ILGL**-**GDRH**-XX-**N**) motif in the catalytic loop and the conserved **PFRLT** motif at the end of the activation loop.

7.3 Future directions

The direct interaction between transcription coactivators and activators is an essential upstream process in transcription initiation and still needs to be understood in terms of affinity and specificity, and it is also essential to understand how transcription initiation is simultaneously modulated by the action of different activators on its target coactivator. In this regard, Tra1 represents a magnificent model to study this interaction, given the multitude of activators that interact with it and its high conservation level in all eukaryotes from yeast to humans. In addition, the overlapping and also unique functions of the SAGA and NuA4 complexes still need to be dissected, and the understanding of Tra1 plays a key role in this context, as the only shared subunit between the two complexes. The precise analysis of the assembly pathway of these complexes will help to understand how Tra1 is specifically loaded into one complex or the other to perform its corresponding and specific biological function. Finally, since the SAGA and NuA4 complexes participate in the DDR (1.3), the role of the putative pseudokinase $Tra1^{TRRAP}$ in this process needs to be independently understood for each complex.

Concerning to the Mec1-Ddc2 complex, still fundamental questions remain to be answered. It includes the understanding of how the different described activators and adaptor proteins

regulate its kinase activity and how the substrate is specifically recognised and phosphorylated. The identification and understanding of the different phosphorylation positions in Mec1 is key to understand how the kinase activity is regulated. Distinct to other PIKKs like DNA-PKcs, Mec1 has been always described as a dimer, and the mechanistic reason remains to be discovered yet. For that reason, the analysis of the Mec1-Ddc2 complex will also benefit the understanding of other dimeric PIKKs like mTOR and Tel1^{ATM}.

Additional functions have been described to the canonical or originally described biological functions for some PIKK members. In the case of Tra1, its role in DNA repair still needs to be fully understood (1.3). For the Mec1-Ddc2 complex there is increasing literature regarding to its role during meiosis (1.4.2), in addition to its canonical role in the DNA damage response. In addition, different PIKKs have been previously described as telomere regulators (1.1.1 and 1.4.2). Therefore, the role of the PIKK protein family in telomere regulation needs to be further analysed.

Different to other proteins, PIKKs require a specific chaperone system to properly fold this family of colossal proteins, where the R2TP complex has been described to play a central role (4.1). For that reason, Tra1 – as a huge monomer part of large protein complexes - and Mec1-Ddc2 - as a dimeric complex - represent excellent models to understand the folding mechanism in the PIKK protein family.

In addition, a deeper understanding of the PIKK family will also increase the knowledge we have so far respecting to ePKs in terms of substrate recognition, phosphorylation specificity and mechanism of action.

Finally, the pharmacological implications of a deeper biological understanding of the PIKK family are enormous (1.5), and in this regard, structural knowledge is essential. The PIKK family is present in all eukaryotes and is highly conserved in evolution from yeast to humans. Many of these proteins function as genome regulators, and its inhibition create genome instability (1.3 and 1.5). In other cases such as mTOR, the metabolic control on cell growth can also be modulated. In addition, the essential nature of some members can be pharmacologically exploited. Therefore, they are promising targets in the therapeutic field, and especially as potential anti-tumor agents. A deeper structural knowledge regarding to the mechanisms of kinase activity regulation will allow to develop more specific and potent drugs to be used in cancer therapy.

References

Allard, S. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. The EMBO Journal, 18(18), 5108–5119.

Amunts, A., Brown, A., Bai, X. -c., Llacer, J. L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S. H. W., & Ramakrishnan, V. (2014). Structure of the Yeast Mitochondrial Large Ribosomal Subunit. Science, 343(6178), 1485-1489.

Andrs, M., Korabecny, J., Jun, D., Hodny, Z., Bartek, J., & Kuca, K. (2015). Phosphatidylinositol 3-kinase (PI3K) and phosphatidylinositol 3-kinase-related kinase (PIKK) inhibitors: Importance of the morpholine ring. Journal of Medicinal Chemistry, 58(1), 41–71.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research, 44(W1), W344–W350.

Askree, S. H., Yehuda, T., Smolikov, S., Gurevich, R., Hawk, J., Coker, C., Krauskopf, A., Kupiec, M., & McEachern, M. J. (2004). A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proceedings of the National Academy of Sciences, 101(23), 8658–8663.

Ball, H. L., & Cortez, D. (2005). ATRIP oligomerization is required for ATR-dependent checkpoint signaling. Journal of Biological Chemistry, 280(36), 31390-31396.

Baptista, T., Grünberg, S., Minoungou, N., Koster, M. J. E., Timmers, H. T. M., Hahn, S., Devys, D., & Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Molecular Cell, 68(1), 130-143.e5.

Bareti, D., & Williams, R. L. (2014). PIKKs — the solenoid nest where partners and kinases meet This review comes from a themed issue on Multi-protein assemblies in signalling. Current Opinion in Structural Biology, 29, 134–142.

Baretić, D., Pollard, H. K., Fisher, D. I., Johnson, C. M., Santhanam, B., Truman, C. M., Kouba, T., Fersht, A. R., Phillips, C., & Williams, R. L. (2017). Structures of closed and open conformations of dimeric human ATM. Science Advances, 3(5), e1700933.

Barley, N. A., Candau, R., Wang, L., Darpino, P., Silverman, N., & Berger, S. L. (1995). Characterization of physical interactions of the putative transcriptional adaptor, ADA2, with acidic activation domains and TATA-binding protein. Journal of Biological Chemistry, 270(33), 19337–19344.

Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell, 116(5), 699–709.

Baxter, W. T., Leith, A., & Frank, J. (2007). SPIRE: The SPIDER Reconstruction Engine. Journal of Structural Biology, 157(1), 56–63.

Bennett, G., & Peterson, C. L. (2015). SWI/SNF recruitment to a DNA double-strand break by the NuA4 and Gcn5 histone acetyltransferases. DNA Repair, 30, 38–45.

Bentley, N. J., Holtzman, D. A., Flaggs, G., Keegan, K. S., DeMaggio, A., Ford, J. C., Hoekstra, M., & Carr, A. M. (1996). The Schizosaccharomyces pombe rad3 checkpoint gene. The EMBO Journal, 15(23), 6641–6651.

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., & Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature, 529(7587), 551–554.

Bhaumik, S. R., Raha, T., Aiello, D. P., & Green, M. R. (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes and Development, 18(3), 333–343.

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., & Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. Nature, 493(7434), 699-702.

Blackford, A. N., & Jackson, S. P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. Molecular Cell, 66(6), 801–817.

Blatch, G. L., & Lässle, M. (1999). The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. BioEssays, 21(11), 932–939.

Boudeau, J., Miranda-Saavedra, D., Barton, G. J., & Alessi, D. R. (2006). Emerging roles of pseudokinases. Trends in cell biology, 16(9), 443-452.

Bromberg, R., Guo, Y., Borek, D., & Otwinowski, Z. (2019). High-resolution cryo-EM reconstructions in the presence of substantial aberrations. BioRxiv, 798280.

Brown, C. E., Howe, L., Sousa, K., Alley, S. C., Carrozza, M. J., Tan, S., & Workman, J. L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science, 292(5525), 2333–2337.

Brzovic, P. S., Heikaus, C. C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R. E., & Hahn, S. (2011). The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Molecular Cell, 44(6), 942–953.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research, 47(W1), W402–W407.

Burgers, P. M. J. (1999). Overexpression of Multisubunit Replication Factors in Yeast. Methods, 18(3), 349–355.

Burgers, P. M., Makarova, A. V., Hailemariam, S., Wanrooij, P. H., Tannous, E., Zhang, X., Darbari, V. C., Bose, D., & Sawicka, M. (2016). The Dimeric Architecture of Checkpoint Kinases Mec1 ATR and Tel1 ATM Reveal a Common Structural Organization . Journal of Biological Chemistry, 291(26), 13436–13447.

Burma, S., & Chen, D. J. (2004). Role of DNA–PK in the cellular response to DNA double-strand breaks. DNA Repair, 3(8–9), 909–918.

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., & Grigorieff, N. (2012). Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. Structure, 20(11), 1823–1828.

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics, 8(2), 93-103.

Chen, V. B., Wedell, J. R., Wenger, R. K., Ulrich, E. L., & Markley, J. L. (2015). MolProbity for the masses-of data. Journal of Biomolecular NMR, 63(1), 77–83.

Cheng, X., Jobin-Robitaille, O., Billon, P., Buisson, R., Niu, H., Lacoste, N., Abshiru, N., Côté, V., Thibault, P., Kron, S. J., Sung, P., Brandl, C. J., Masson, J. Y., & Côté, J. (2018). Phospho-dependent recruitment of the yeast NuA4 acetyltransferase complex by MRX at DNA breaks regulates RPA dynamics during resection. Proceedings of the National Academy of Sciences of the United States of America, 115(40), 10028–10033.

Cheung, A. C. M., & Díaz-Santín, L. M. (2019). Share and share alike: the role of Tra1 from the SAGA and NuA4 coactivator complexes. Transcription, 10(1), 37–43.

Chittuluru, J. R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M. J., Sapountzi, V., Selleck, W., Huang, J., Utley, R. T., Cramet, M., Allard, S., Cai, G., Workman, J. L., Fried, M. G., Tan, S., Côté, J., & Asturias, F. J. (2011). Structure and nucleosome interaction of the yeast NuA4 and Piccolo–NuA4 histone acetyltransferase complexes. Nature Structural & Molecular Biology, 18(11), 1196–1203.

Cianfrocco, M. A., Kassavetis, G. A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J. T., & Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. Cell, 152(1–2), 120–131.

Cimprich, K A, Shin, T. B., & Keith, C. T. (1996). cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. Proceedings of the National Academy of Sciences, 93(7), 2850-2855.

Cimprich, Karlene A., & Cortez, D. (2008). ATR: An essential regulator of genome integrity. Nature reviews Molecular cell biology, 9(8), 616-627.

Clough, R. N., Moldovan, G., & Kirkland, A. I. (2014). Direct Detectors for Electron Microscopy. Journal of Physics: Conference Series, 522(1), 012046.

Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. Nucleic acids research, 36(suppl_2), W197-W201.

Cooper, T. J., Wardell, K., Garcia, V., & Neale, M. J. (2014). Homeostatic regulation of meiotic DSB formation by ATM/ATR. Experimental cell research, 329(1), 124-131.

Cortez, D., Guntuku, S., Qin, J., & Elledge, S. J. (2001). ATR and ATRIP: Partners in checkpoint signaling. Science, 294(5547), 1713–1716.

Cramer, P. (2019). Eukaryotic Transcription Turns 50. Cell, 179(4), 808-812.

Dames, S. A. (2010). Structural basis for the association of the redox-sensitive target of rapamycin FATC domain with membrane-mimetic micelles. Journal of Biological Chemistry, 285(10), 7766–7775.

Dames, S. A., Mulet, J. M., Rathgeb-Szabo, K., Hall, M. N., & Grzesiek, S. (2005). The solution structure of the FATC domain of the protein kinase target of rapamycin suggests a role for redox-dependent structural and cellular stability. Journal of Biological Chemistry, 280(21), 20558–20564.

Dang, C. V., Le, A., & Gao, P. (2009). MYC-induced cancer cell energy metabolism and therapeutic opportunities. Clinical Cancer Research, 15(21), 6479–6483.

Deshpande, I., Seeber, A., Shimada, K., Keusch, J. J., Gut, H., & Gasser, S. M. (2017). Structural Basis of Mec1-Ddc2-RPA Assembly and Activation on Single-Stranded DNA at Sites of Damage. Molecular Cell, 68(2), 431-445.e5.

Díaz-Santín, L. M., Lukoyanova, N., Aciyan, E., & Cheung, A. C. M. (2017). Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. ELife, 6, 1–20.

Diep, C. Q., Tao, X., Pilauri, V., Losiewicz, M., Blank, T. E., & Hopper, J. E. (2008). Genetic evidence for sites of interaction between the Gal3 and Gal80 proteins of the Saccharomyces cerevisiae GAL gene switch. Genetics, 178(2), 725–736.

DiMaio, F., Echols, N., Headd, J. J., Terwilliger, T. C., Adams, P. D., & Baker, D. (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. Nature Methods, 10(11), 1102–1104.

Doerr, A. (2015). Single-particle cryo-electron microscopy. Nature methods, 13(1), 23.

Doyon, Y., & Côté, J. (2004). The highly conserved and multifunctional NuA4 HAT complex. Current Opinion in Genetics & Development, 14(2), 147–154.

Doyon, Y., Selleck, W., Lane, W. S., Tan, S., & Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Molecular and Cellular Biology, 24(5), 1884–1896.

Durand, A., Bonnet, J., Fournier, M., Chavant, V., & Schultz, P. (2014). Mapping the deubiquitination module within the saga complex. Structure, 22(11), 1553–1559.

Eisenstein, M. (2016). The field that came in from the cold. Nature Methods, 13(1), 19–22.

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.-R., Park, S., Wang, X., Krebs, J., Tripathi, S., Sakabe, N., Sobreira, D. R., Huang, S.-C., Rao, S. S. P., Pruett, N., Chauss, D., Sadler, E., … Casellas, R. (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell, 178(5), 1145-1158.e20.

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

Fernandez-Leiro, R., & Scheres, S. H. W. (2016). Unravelling biological macromolecules with cryo-electron microscopy. Nature, 537(7620), 339–346.

Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. W., & Ramakrishnan, V. (2014). Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. Cell, 157(4), 823-831.

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., & Eddy, S. R. (2015). HMMER web server: 2015 update. Nucleic Acids Research, 43(W1), W30–W38.

Fishburn, J., Mohibullah, N., & Hahn, S. (2005). Function of a Eukaryotic Transcription Activator during the Transcription Cycle. Molecular Cell, 18(3), 369–378.

Forget, D., Robert, F., Grondin, G., Burton, Z. F., Greenblatt, J., & Coulombe, B. (1997). RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. Proceedings of the National Academy of Sciences of the United States of America, 94(14), 7150–7155.

Forment, J. V., & O'Connor, M. J. (2018). Targeting the replication stress response in cancer. Pharmacology & Therapeutics, 188, 155–167.

Forwood, J. K., Lange, A., Zachariae, U., Marfori, M., Preast, C., Grubmüller, H., Stewart, M., Corbett, A. H., & Kobe, B. (2010). Quantitative Structural Analysis of Importin-β Flexibility: Paradigm for Solenoid Protein Structures. Structure, 18(9), 1171–1183.

Friedel, A. M., Pike, B. L., & Gasser, S. M. (2009). ATR/Mec1: coordinating fork stability and repair. Current opinion in cell biology, 21(2), 237-244.

Galarneau, L., Nourani, A., Boudreault, A. A., Zhang, Y., Hé, L., Allard, S. P., Savard, J., Lane, W. S., Stillman, D. J., & Côté, J. (2000). Multiple Links between the NuA4 Histone Acetyltransferase Complex and Epigenetic Control of Transcription. Molecular cell, 5(6), 927-937.

Gat, Y., Schuller, J. M., Lingaraju, M., Weyher, E., Bonneau, F., Strauss, M., Murray, P. J., & Conti, E. (2019). InsP6 binding to PIKK kinases revealed by the cryo-EM structure of an SMG1–SMG8–SMG9 complex. Nature Structural & Molecular Biology, 26(12), 1089-1093.

Gibson, D G. (2011). Enzymatic assembly of overlapping DNA fragments. Method Enzymol, 498, 349–361.

Gibson, Daniel G, Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods, 6(5), 343–345.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Science, 27(1), 14–25.

Gouet, P., Robert, X., & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Research, 31(13), 3320–3323.

Grant, P. a, Schieltz, D., Pray-Grant, M. G., Yates, J. R., Workman, J. L., Amati, B., Dalton,

S., Brooks, M. W., Littlewood, T. D., Evan, G. I., Land, H., Barlev, N. a., Candau, R., Wang, L., Darpino, P., Silverman, N., Berger, S. L., Berger, S. L., Piña, B., … Zakian, V. a. (1998). The ATM-related cofactor Tra1 is a component of the purified SAGA complex. Molecular Cell, 2(6), 863–867.

Greber, B. J., Nguyen, T. H. D., Fang, J., Afonine, P. V., Adams, P. D., & Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature, 549(7672), 414–417.

Greber, B. J., Toso, D. B., Fang, J., & Nogales, E. (2019). The complete structure of the human TFIIH core complex. Elife, 8, e44771.

Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A., & Barford, D. (1999). The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell, 96(1), 99–110.

Gubbens, A., Barfels, M., Trevor, C., Twesten, R., Mooney, P., Thomas, P., Menon, N., Kraus, B., Mao, C., & McGinn, B. (2010). The GIF Quantum, a next generation post-column imaging energy filter. Ultramicroscopy, 110(8), 962–970.

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nature Reviews Molecular Cell Biology, 19(10), 621-637.

Hahn, S., & Young, E. T. (2011). Transcriptional Regulation in Saccharomyces cerevisiae: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. Genetics, 189(3), 705-736.

Hammarén, H. M., Ungureanu, D., Grisouard, J., Skoda, R. C., Hubbard, S. R., & Silvennoinen, O. (2015). ATP binding to the pseudokinase domain of JAK2 is critical for pathogenic activation. Proceedings of the National Academy of Sciences of the United States of America, 112(15), 4642–4647.

Han, Yan, Luo, J., Ranish, J., & Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. The EMBO Journal, 33(21), 2534–2546.

He, yuan, yan, chunli, Fang, J., inouye, carla, Tjian, R., ivanov, ivaylo, & Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature, *533*(7603), 359-365.

Helmlinger, D., Marguerat, S., Villén, J., Swaney, D. L., Gygi, S. P., Bähler, J., & Winston, F. (2011). Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. EMBO Journal, 30(14), 2843–2852.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences of the United States of America, 110(45), 18037–18041.

Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Lleonart, M., Jackson, S., & Wang, Z.-Q. (2001). Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. Nature Genetics, 29(2), 206–211.

Hiom, K. (2005). DNA repair: How to PIKK a partner. Current biology, 15(12), R473-R475.

Hoke, S. M. T., Guzzo, J., Andrews, B., & Brandl, C. J. (2008). Systematic genetic array analysis links the Saccharomyces cerevisiae SAGA/SLIK and NuA4 component Tra1 to multiple cellular processes. BMC genetics, 9(1), 46.

Hoke, S. M. T., Irina Mutiu, A., Genereaux, J., Kvas, S., Buck, M., Yu, M., Gloor, G. B., & Brandl, C. J. (2010). Mutational analysis of the C-terminal FATC domain of Saccharomyces cerevisiae Tra1. Current Genetics, 56(5), 447–465.

Huisinga, K. L., & Pugh, B. F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Molecular Cell,

13(4), 573–585.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., … Yeats, C. (2008). InterPro: the integrative protein signature database. Nucleic Acids Research, 37, 211–215.

Huse, M., & Kuriyan, J. (2002). The Conformational Plasticity of Protein Kinases. Cell, 109(3), 275–282.

Ikura, T., Ogryzko, V. V, Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., & Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell, 102(4), 463–473.

Imseng, S., & Aylett, C. H. (2018). Architecture and activation of phosphatidylinositol 3-kinase related kinases. Current Opinion in Structural Biology, 49, 177–189.

Itakura, E. (2005). Dimerization of the ATRIP Protein through the Coiled-Coil Motif and Its Implication to the Maintenance of Stalled Replication Forks. Molecular biology of the cell, 16(12), 5551-5562.

Jansma, M., Linke-Winnebeck, C., Eustermann, S., Litz, C., Kessler, B., & Hopfner, K.-P. (2019). Near-Complete Structure and Model of Tel1ATM from Chaetomium thermophilum Reveals a Robust Autoinhibited ATP State. Structure, 28(1), 83-95.

Jiang, X., Sun, Y., Chen, S., Roy, K., & Price, B. D. (2006). The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. Journal of Biological Chemistry, 281(23), 15741–15746.

Jínek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., & Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. Nature Structural and Molecular Biology, 11(10), 1001–1007.

Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., Wemmer, D. E., Zhang, X., & Kuriyan, J. (2009). Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. Cell, 137(7), 1293–1307.

Jura, N., Zhang, X., Endres, N. F., Seeliger, M. A., Schindler, T., & Kuriyan, J. (2011). Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. Molecular Cell, 42(1), 9–22.

Juven-Gershon, T., & Kadonaga, J. T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology, 339(2), 225–229.

Kakihara, Y., & Houry, W. A. (2012). The R2TP complex: Discovery and functions. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1823(1), 101–107.

Kannan, N., & Taylor, S. S. (2008). Rethinking Pseudokinases. Cell, 133(2), 204-205.

Kappel, C., Zachariae, U., Dölker, N., & Grubmüller, H. (2010). An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. Biophysical Journal, 99(5), 1596–1603.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10(6), 845–858.

Kim, S. M., Kumagai, A., Lee, J., & Dunphy, W. G. (2005). Phosphorylation of Chk1 by ATM- and Rad3-related (ATR) in Xenopus egg extracts requires binding of ATRIP to ATR but not the stable DNA-binding or coiled-coil domains of ATRIP. Journal of Biological Chemistry, 280(46), 38355-38364.

Kimanius, D., Forsberg, B. O., Scheres, S. H. W., & Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. Elife, 5, e18722..

Klein, J., Nolden, M., Sanders, S. L., Kirchner, J., Weil, P. A., & Melcher, K. (2003). Use of a genetically introduced cross-linker to identify interaction sites of acidic activators within native transcription factor IID and SAGA. The Journal of Biological Chemistry, 278(9), 6779–6786.

Knutson, B. a., & Hahn, S. (2011). Domains of Tra1 Important for Activator Recruitment and Transcription Coactivator Functions of SAGA and NuA4 Complexes. Molecular and Cellular Biology, 31(4), 818–831.

Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., & Cramer, P. (2019). Structural basis of TFIIH activation for nucleotide excision repair. Nature Communications, 10(1), 2885.

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

Kornev, A. P., Haste, N. M., Taylor, S. S., & Eyck, L. F. Ten. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proceedings of the National Academy of Sciences of the United States of America, 103(47), 17783–17788.

Koutelou, E., Hirsch, C. L., & Dent, S. Y. R. (2010). Multiple faces of the SAGA complex. In Current Opinion in Cell Biology. Current opinion in cell biology, 22(3), 374-382.

Kvas, S., Gloor, G. B., & Brandl, C. J. (2012). Loss of nonsense mediated decay suppresses mutations in Saccharomyces cerevisiae TRA1. BMC genetics, 13(1), 19.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. Cell, 172(4), 650-665.

Lecona, E., & Fernandez-Capetillo, O. (2018). Targeting ATR in cancer. Nature Reviews Cancer, 18(9), 586–595.

Lee, T. I., Causton, H. C., Holstege, F. C. P., Shen, W. C., Hannett, N., Jennings, E. G., Winston, F., Green, M. R., & Young, R. A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature, 405(6787), 701–704.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., & Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods, 10(6), 584–590.

Lin, L., Chamberlain, L., Zhu, L. J., & Green, M. R. (2012). Analysis of Gal4-directed transcription activation using Tra1 mutants selectively defective for interaction with Gal4. Proceedings of the National Academy of Sciences, 109(6), 1997–2002.

Lis, J. T. (2019). A 50 year history of technologies that drove discovery in eukaryotic transcription regulation. Nature structural & molecular biology, 26(9), 777-782.

Liu, G., Zheng, X., Guan, H., Cao, Y., Qu, H., Kang, J., Ren, X., Lei, J., Dong, M.-Q., Li, X., & Li, H. (2019). Architecture of Saccharomyces cerevisiae SAGA complex. Cell Discovery, 5(1), 25.

Louder, R. K., He, Y., Ramón López-Blanco, J., Fang, J., Chacón, P., & Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature, 531(7596), 604-609.

Lovejoy, C. A., & Cortez, D. (2009). Common mechanisms of PIKK regulation. DNA Repair, 8(9), 1004–1008.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. Science, 298(5600), 1912–1934.

Maréchal, A., & Zou, L. (2013). DNA damage sensing by the ATM and ATR kinases. Cold

Spring Harbor Perspectives in Biology, 5(9).

Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., Bou-Assaf, G. M., Engen, J. R., Englander, S. W., Faber, J., Garlish, R., Griffin, P. R., Gross, M. L., Guttman, M., Hamuro, Y., Heck, A. J. R., Houde, D., Iacob, R. E., Jørgensen, T. J. D., … Rand, K. D. (2019). Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nature Methods, 16(7), 595–602.

McMahon, S. B., Van Buskirk, H. A., Dugan, K. A., Copeland, T. D., & Cole, M. D. (1998). The novel ATM-related protein TRRAP is an essential cofactor for the c- Myc and E2F oncoproteins. Cell, 94(3), 363–374.

McMullan, G., Faruqi, A. R., & Henderson, R. (2016). Direct Electron Detectors. Methods in enzymology (Vol. 579, pp. 1-17).

Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of Structural Biology, 142(3), 334–347.

Monté, D., Clantin, B., Dewitte, F., Lens, Z., Rucktooa, P., Pardon, E., Steyaert, J., Verger, A., & Villeret, V. (2018). Crystal structure of human Mediator subunit MED23. Nature Communications, 9(1), 1–7.

Mordes, D. A., Glick, G. G., Zhao, R., & Cortez, D. (2008). TopBP1 activates ATR through ATRIP and a PIKK regulatory domain. Genes and Development, 22(11), 1478–1489.

Mordes, D. a, Nam, E. a, & Cortez, D. (2008). Dpb11 activates the Mec1-Ddc2 complex. Proceedings of the National Academy of Sciences of the United States of America, 105(48), 18730–18734.

Murr, R., Vaissière, T., Sawan, C., Shukla, V., & Herceg, Z. (2007). Orchestration of chromatin-based processes: Mind the TRRAP. Oncogene, 26(37), 5358-5372.

Mutiu, a. I., Hoke, S. M. T., Genereaux, J., Hannam, C., MacKenzie, K., Jobin-Robitaille, O., Guzzo, J., Côté, J., Andrews, B., Haniford, D. B., & Brandl, C. J. (2007). Structure/function analysis of the phosphatidylinositol-3-kinase domain of yeast Tra1. Genetics, 177(1), 151–166.

Nakane, T., Kimanius, D., Lindahl, E., & Scheres, S. H. W. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. Elife, 7, e36861.

Neely, K. E., Hassan, A. H., Brown, C. E., Howe, L., & Workman, J. L. (2002). Transcription Activator Interactions with Multiple SWI/SNF Subunits. Molecular and Cellular Biology, 22(6), 1615–1625.

Noble, M. E. M., Endicott, J. A., & Johnson, L. N. (2004). Protein kinase inhibitors: insights into drug design from structure. Science (New York, N.Y.), 303(5665), 1800–1805.

Nogales, E., Patel, A. B., & Louder, R. K. (2017). Towards a mechanistic understanding of core promoter recognition from cryo-EM studies of human TFIID. Current opinion in structural biology, 47, 60-66.

Nozawa, K., Schneider, T. R., & Cramer, P. (2017). Core Mediator structure at 3.4 Å extends model of transcription initiation complex. Nature, 545(7653), 248-251.

O'Connor, M. J. (2015). Targeting the DNA Damage Response in Cancer. Molecular Cell, 60(4), 547–560.

Ogi, H., Goto, G. H., Ghosh, A., Zencir, S., Henry, E., & Sugimoto, K. (2015). Requirement of the FATC domain of protein kinase Tel1 for localization to DNA ends and target protein recognition. Molecular Biology of the Cell, 26(19), 3480–3488.

Orlova, E. V., & Saibil, H. R. (2011). Structural Analysis of Macromolecular Assemblies by Electron Microscopy. Chemical Reviews, 111(12), 7710–7748.

Pacheco, S., Maldonado-Linares, A., Marcet-Ortega, M., Rojas, C., Martínez-Marchal, A., Fuentes-Lazaro, J., Lange, J., Jasin, M., Keeney, S., Fernández-Capetillo, O., Garcia-Caldés, M., & Roig, I. (2018). ATR is required to complete meiotic recombination in mice. Nature Communications, 9(1), 2622.

Papai, G., Weil, P. A., & Schultz, P. (2011). New insights into the function of transcription factor TFIID from recent structural studies. In Current Opinion in Genetics and Development (Vol. 21, Issue 2, pp. 219–224).

Park, J., Kunjibettu, S., McMahon, S. B., & Cole, M. D. (2001). The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis. Genes and Development, 15(13), 1619–1624.

Patel, A. B., Louder, R. K., Greber, B. J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., & Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science, 362(6421).

Pei, J., & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics, 17(8), 700–712.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: Five essential questions. Nature Reviews Genetics, 14(4), 288-295.

Peterson, C. L., & Workman, J. L. (2000). Promoter targeting and chromatin remodeling by the SWI/SNF complex. Current opinion in genetics & development, 10(2), 187-192.

Piskacek, S., Gregor, M., Nemethova, M., Grabner, M., Kovarik, P., & Piskacek, M. (2007). Nine-amino-acid transactivation domain: Establishment and prediction utilities. Genomics, 89(6), 756–768.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., & Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature, 533, 353–358.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V, Borchers, C. H., Baumeister, W., Herzog, F., Villa, E., & Cramer, P. (2015). Architecture of the RNA polymerase II– Mediator core initiation complex. Nature, 518(7539), 376-380.

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods, 14(3), 290–296.

Rajakulendran, T., & Sicheri, F. (2010). Allosteric Protein kinase regulation by pseudokinases: Insights from STRAD. Sci. Signal., 3(111), pe8-pe8.

Ramachandran, S., Haddad, D., Li, C., Le, M. X., Ling, A. K., So, C. C., Nepal, R. M., Gommerman, J. L., Yu, K., Ketela, T., Moffat, J., & Martin, A. (2016). The SAGA Deubiquitination Module Promotes DNA Repair and Class Switch Recombination through ATM and DNAPK-Mediated γH2AX Formation. Cell Reports, 15(7), 1554–1565.

Rao, Q., Liu, M., Tian, Y., Wu, Z., Hao, Y., Song, L., Qin, Z., Ding, C., Wang, H.-W., Wang, J., & Xu, Y. (2017). Cryo-EM structure of human ATR-ATRIP complex. Cell research, 28(2), 143-156.

Reeves, W. M., & Hahn, S. (2005). Targets of the Gal4 transcription activator in functional transcription complexes. Molecular and Cellular Biology, 25(20), 9092–9102.

Rivera-Calzada, A., López-Perrote, A., Melero, R., Boskovic, J., Muñoz-Hernández, H., Martino, F., & Llorca, O. (2015). Structure and Assembly of the PI3K-like Protein Kinases (PIKKs) Revealed by Electron Microscopy. AIMS Biophysics, 2(2), 36–57.

Robert, F., Hardy, S., Nagy, Z., Baldeyron, C., Murr, R., Dery, U., Masson, J.-Y., Papadopoulo, D., Herceg, Z., & Tora, L. (2006). The Transcriptional Histone Acetyltransferase Cofactor TRRAP Associates with the MRN Repair Complex and Plays a Role in DNA Double-

Strand Break Repair. Molecular and Cellular Biology, 26(2), 402–412.

Robinson, P. J., Trnka, M. J., Bushnell, D. A., Davis, R. E., Mattei, P.-J., Burlingame, A. L., Kornberg Correspondence, R. D., & Kornberg, R. D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell, 166, 1411–1422.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci, 21(9), 327–335.

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. Journal of Structural Biology, 192(2), 216–221.

Rosenthal, P. B., & Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. Journal of Molecular Biology, 333(4), 721–745.

Rundle, S., Bradbury, A., Drew, Y., & Curtin, N. J. (2017). Targeting the ATR-CHK1 axis in cancer therapy. Cancers, 9(5), 1–25.

Russo, C. J., & Passmore, L. A. (2014). Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. Nature Methods, 11(6), 649–652.

Russo, C. J., & Passmore, L. A. (2014). Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. Science (New York, N.Y.), 346(6215), 1377–1380.

Sabatini, D. M. (2006). mTOR and cancer: insights into a complex relationship. Nature Reviews Cancer, 6(9), 729–734.

Saleh, A., Schieltz, D., Ting, N., McMahon, S. B., Litchfield, D. W., Yates, J. R., Lees-Miller, S. P., Cole, M. D., & Brandl, C. J. (1998). Tra1p is a component of the yeast Ada·Spt transcriptional regulatory complexes. Journal of Biological Chemistry, 273(41), 26559–26565.

Sawicka, M., Wanrooij, P. H., Darbari, V. C., Tannous, E., Hailemariam, S., Bose, D., Makarova, A. V., Burgers, P. M., & Zhang, X. (2016). The dimeric architecture of checkpoint kinases Mec1^{ATR}and Tel1^{ATM}reveal a common structural organization. Journal of Biological Chemistry, 291(26), 13436-13447.

Scheeff, E. D., Eswaran, J., Bunkoczi, G., Knapp, S., & Manning, G. (2009). Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site. Structure, 17(1), 128–138.

Scheres, S. H. w. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. ELife, 3, e03665.

Scheres, S.H.W. (2016). Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods in Enzymology, 579, 125–157.

Scheres, Sjors H W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology, 180(3), 519–530.

Scheres, Sjors H W, & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nature methods, 9(9), 853.

Schilbach, S., Tegunov, D., Urlaub, H., & cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature, 551(7679), 204-209.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods, 9(7), 671-675.

Setiaputra, D., Ahmad, S., Dalwadi, U., Steunou, A.-L., Lu, S., Ross, J. D., Dong, M.-Q., Côté, J., & Yip, C. K. (2018). Molecular architecture of the essential yeast histone acetyltransferase complex NuA4 redefines its multi-modularity. Molecular and cellular biology, 38(9), e00570-17.

Setiaputra, D., Ross, J. D., Lu, S., Cheng, D. T., Dong, M. Q., & Yip, C. K. (2015).

Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. Journal of Biological Chemistry, 290(16), 10057–10070.

Sharov, G., Voltz, K., Durand, A., Kolesnikova, O., Papai, G., Myasnikov, A. G., Dejaegere, A., Shem, A. Ben, & Schultz, P. (2017). Structure of the transcription activator target Tra1 within the chromatin modifying complex SAGA. Nature Communications, 8(1).

Shigechi, T., Tomida, J., Sato, K., Kobayashi, M., Eykelenboom, J. K., Pessina, F., Zhang, Y., Uchida, E., Ishiai, M., Lowndes, N. F., Yamamoto, K., Kurumizaka, H., Maehara, Y., & Takata, M. (2012). ATR-ATRIP kinase complex triggers activation of the fanconi anemia DNA repair pathway. Cancer Research, 72(5), 1149–1156.

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. Nature Reviews Genetics, 15(4), 272-286.

Sibanda, B. L., Chirgadze, D. Y., Ascher, D. B., & Blundell, T. L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. Science, 355(6324), 520–524.

Sibanda, B. L., Chirgadze, D. Y., & Blundell, T. L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature, 463(7277), 118–121.

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Science, 27(1), 135–145.

Smith, G. C. M., & Jackson, S. P. (2010). Chapter 77 – The PIKK Family of Protein Kinases. Handbook of Cell Signaling (pp. 575-580).

Smits, V. A. J., Reaper, P. M., & Jackson, S. P. (2006). Rapid PIKK-dependent release of Chk1 from chromatin promotes the DNA-damage checkpoint response. Current Biology, 16(2), 150–159.

Sommer, L. A. M., Schaad, M., & Dames, S. A. (2013). NMR- and circular dichroism-monitored lipid binding studies suggest a general role for the FATC domain as membrane anchor of phosphatidylinositol 3-kinase-related kinases (PIKK). Journal of Biological Chemistry, 288(27), 20046–20063.

Steichen, J. M., Kuchinskas, M., Keshwani, M. M., Yang, J., Adams, J. A., & Taylor, S. S. (2012). Structural Basis for the Regulation of Protein Kinase A by Activation Loop Phosphorylation. Journal of Biological Chemistry, 287(18), 14672–14680.

Struhl, K. (1989). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. In Trends in Biochemical Sciences. Trends in biochemical sciences, 14(4), 137-140.

Sun, Y., Jiang, X., & Price, B. D. (2010). Tip60: Connecting chromatin to DNA damage signaling. Cell cycle, 9(5), 930-936.

Taylor, S. S., & Kornev, A. P. (2011). Protein kinases: Evolution of dynamic regulatory proteins. In Trends in Biochemical Sciences. Trends in biochemical sciences, 36(2), 65-77.

Taylor, S. S., Zhang, P., Steichen, J. M., Keshwani, M. M., & Kornev, A. P. (2013). PKA: Lessons learned after twenty years. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1834(7), 1271–1278.

Tegunov, D., & Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nature Methods, 1–7.

Tompa, P., & Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends in Biochemical Sciences, 33(1), 2–8.

Tong, A. S., Stern, J. L., Sfeir, A., Kartawinata, M., de Lange, T., Zhu, X.-D., & Bryan, T. M. (2015). ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to

Telomeres. Cell Reports, 13(8), 1633–1646.

Traven, A., Jelicic, B., & Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. EMBO Reports, 7(5), 496–499.

Treiber, D. K., & Shah, N. P. (2013). Ins and Outs of Kinase DFG Motifs. Chemistry & Biology, 20(6), 745–746.

Tsai, K.-L., Yu, X., Gopalan, S., Chao, T.-C., Zhang, Y., Florens, L., Washburn, M. P., Murakami, K., Conaway, R. C., Conaway, J. W., & Asturias, F. J. (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature, 544(7649), 196–201.

Tuttle, L. M., Pacheco, D., Warfield, L., Hahn, S., & Klevit, R. E. (2019). Mediator subunit Med15 dictates the conserved "fuzzy" binding mechanism of yeast transcription activators Gal4 and Gcn4. BioRxiv, 840348.

Tuttle, L. M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., & Klevit, R. E. (2018). Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Reports, 22(12), 3251–3264.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Björling, L., & Ponten, F. (2010). Towards a knowledge-based Human Protein Atlas. Nature biotechnology, 28(12), 1248-1250.

Urnavicius, L., Zhang, K., Diamant, A. G., Motz, C., Schlager, M. A., Yu, M., Patel, N. A., Robinson, C. V., & Carter, A. P. (2015). The structure of the dynactin complex and its interaction with dynein. Science, 347(6229), 1441–1446.

Vadas, O., & Burke, J. E. (2015). Probing the dynamic regulation of peripheral membrane proteins using hydrogen deuterium exchange-MS (HDX-MS). Biochemical Society Transactions, 43(5), 773–786.

van Attikum, H., & Gasser, S. M. (2005). The histone code at DNA breaks: a guide to repair? Nature Reviews Molecular Cell Biology, 6(10), 757–765.

Vignard, J., Mirey, G., & Salles, B. (2013). Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up. Radiotherapy and Oncology, 108(3), 362–369.

von Morgen, P., Hořejší, Z., & Macurek, L. (2015). Substrate recognition and function of the R2TP complex in response to cellular stress. Frontiers in Genetics, 6, 69.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C., & Raunser, S. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Communications Biology, 2(1), 218.

Wang, X., Ahmad, S., Zhang, Z., Côté, J., & Cai, G. (2018). Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex. Nature communications, 9(1), 1-11.

Wang, X., Chu, H., Lv, M., Zhang, Z., Qiu, S., Liu, H., Shen, X., Wang, W., & Cai, G. (2016). Structure of the intact ATM/Tel1 kinase. Nature communications, 7(1), 1-8.

Wang, X., Ran, T., Zhang, X., Xin, J., Zhang, Z., Wu, T., Wang, W., & Cai, G. (2017). 3.9 Å structure of the yeast Mec1-Ddc2 complex, a homolog of human ATR-ATRIP. Science, 358(6367), 1206–1209.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., & Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Molecular Cell, 68(1), 118-129.e5.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics,

25(9), 1189–1191.

Weinert, T. A., Kiser, G. L., & Hartwell, L. H. (1994). Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. Genes and Development, 8(6), 652–665.

Widger, A., Mahadevaiah, S. K., Lange, J., ElInati, E., Zohren, J., Hirota, T., Pacheco, S., Maldonado-Linares, A., Stanzione, M., Ojarikre, O., Maciulyte, V., de Rooij, D. G., Tóth, A., Roig, I., Keeney, S., & Turner, J. M. A. (2018). ATR is a multifunctional regulator of male mouse meiosis. Nature Communications, 9(1), 2621.

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. Journal of Molecular Biology, 293(2), 321–331.

Wu, P.-Y. J., Ruhlmann, C., Winston, F., & Schultz, P. (2004). Molecular Architecture of the S. cerevisiae SAGA Complex. Molecular Cell, 15(2), 199–208.

Wu, T. J., Wang, X., Zhang, Y., Meng, L., Kerrigan, J. E., Burley, S. K., & Zheng, X. F. S. (2015). Identification of a non-Gatekeeper hot spot for drug-resistant mutations in mTOR kinase. Cell Reports, 11(3), 446–459.

Xiao, J., Liu, M., Qi, Y., Chaban, Y., Gao, C., Pan, B., Tian, Y., Yu, Z., Li, J., Zhang, P., & Xu, Y. (2019). Structural insights into the activation of ATM kinase. Cell research, 29(8), 683-685.

Xin, J., Xu, Z., Wang, X., Tian, Y., Zhang, Z., & Cai, G. (2019). Structural basis of allosteric regulation of Tel1/ATM kinase. Cell Research, 29(8), 655–665.

Xu, P., Li, C., Chen, Z., Jiang, S., Fan, S., Wang, J., Dai, J., Zhu, P., & Chen, Z. (2016). The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism. Molecular Cell, 63(6), 965–975.

Yang, H., Jiang, X., Li, B., Yang, H. J., Miller, M., Yang, A., Dhar, A., & Pavletich, N. P. (2017). Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. Nature, 552(7685), 368–373.

Yang, H., Rudge, D. G., Koos, J. D., Vaidialingam, B., Yang, H. J., & Pavletich, N. P. (2013). mTOR kinase structure, mechanism and regulation. Nature, 497(7448), 217–223.

Yates, L. A., Aramayo, R. J., Pokhrel, N., Caldwell, C. C., Kaplan, J. A., Perera, R. L., Spies, M., Antony, E., & Zhang, X. (2018). A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. Nature Communications, 9(1), 5447.

Yates, L. A., Williams, R. M., Hailemariam, S., Ayala, R., Burgers, P., & Zhang, X. (2019). Cryo-EM Structure of Nucleotide-Bound Tel1ATM Unravels the Molecular Basis of Inhibition and Structural Rationale for Disease-Associated Mutations. Structure, 28(1), 96-104.

Yin, X., Liu, M., Tian, Y., Wang, J., & Xu, Y. (2017). Cryo-EM structure of human DNA-PK holoenzyme. Cell Research, 27(11), 1341–1350.

Yoshimura, S. H., & Hirano, T. (2016). HEAT repeats - versatile arrays of amphiphilic helices working in crowded environments?. Journal of cell science, 129(21), 3963-3970.

Zeqiraj, E., & van Aalten, D. M. F. (2010). Pseudokinases-remnants of evolution or key allosteric regulators?. Current opinion in structural biology, 20(6), 772-781.

Zhang, H., Photiou, A., Grothey, A., Stebbing, J., & Giamas, G. (2012). The role of pseudokinases in cancer. Cellular signalling, 24(6), 1173-1184.

Zhang, N., Ichikawa, W., Faiola, F., Lo, S.-Y., Liu, X., & Martinez, E. (2014). MYC interacts with the human STAGA coactivator complex via multivalent contacts with the GCN5 and TRRAP subunits. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1839(5), 395-405.

Zhang, X., Wu, J., & Luan, Y. (2017). Tip60: Main Functions and Its Inhibitors. Mini-Reviews

in Medicinal Chemistry, 17(8), 675–682.

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

Zhou, J., Lim, C. U., Li, J. J., Cai, L., & Zhang, Y. (2006). The role of NBS1 in the modulation of PIKK family proteins ATM and ATR in the cellular response to DNA damage. Cancer letters, 243(1), 9-15.

Zhu, L., Li, L., Qi, Y., Yu, Z., & Xu, Y. (2019). Cryo-EM structure of SMG1–SMG8–SMG9 complex. Cell research, 29(12), 1027-1034.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife, 7, e42166.

Zivanov, J., Nakane, T., & Scheres, S. (2019). Estimation of high-order aberrations and anisotropic magnification from cryo-EM datasets in RELION-3.1. BioRxiv, 798066.

Abbreviations and symbols

°C	Degree Celsius
3C	Human Rhinovirus 3C protease
3xFLAG	Asp-Tyr-Lys-Asp-His-Asp-**Gly**-Asp-Tyr-Lys-Asp-His-Asp-**Ile**-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys
Å	*ångström*
AAD	ATR activation domain (referred to DNA repair)
AAD	Acidic activation domain (referred to transcription)
ABD	Activator binding domain
Ade	Adenine
A-loop	Activation loop
Amp	Ampicillin
APS	Ammonium peroxodisulfate
ATM	Ataxia telangiectasia mutated
ATP	Adenosine 5'-triphosphate
ATR	ATM and RAD3-related protein kinase
ATRIP	ATR-interacting protein
AU	Absorbance units
BA	Beads after elution
BB	Beads before elution
bp	Base pair
BSA	Bovine serum albumin
Cam	Chloramphenicol
CC	Coiled coil domain
CCD	Charge-coupled device
CDK	Cyclin-dependent kinase
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
C-loop	Catalytic loop
CMOS	Complementary metal-oxide semiconductors
CSM	Complete supplement mixture
CTD	C-terminal domain
CTF	Contrast transfer function

CV	Column volume
Da	dalton
DBD	DNA binding domain
Ddc1	DNA damage checkpoint protein 1
Ddc2	DNA damage checkpoint protein 2 (Standard name: Lcd1; alias: Pie1)
DDD	Direct electron detector device
DDR	DNA damage response
DLS	Diamond Light Source
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dNTP	Deoxynucleoside triphosphate
Dpb11	DNA polymerase B 11
DQE	Detective quantum efficiency
DSB	DNA double-strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	Escherichia coli
e.g.	For example, from Latin “exempli gratia”
EB	Elution buffer
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
et al.	And others [Latin: et alii]
EtOH	Ethanol
FA	Fanconi anemia
FAT	FRAP-ATM-TRRAP domain
FATC	FRAP-ATM-TRRAP C-terminal domain
FEG	Field emission gun
FL	Full-length
FRB	FKBP12-rapamycin-binding domain
FSC	Fourier shell correlation
fwd	Forward (primer)
Gal/GAL	Galactose
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase

GFP	Green fluorescent protein
GIA	Gibson isothermal assembly
GO	Graphene oxide
GST	Glutathione S-transferase (~ 26 kDa)
h/hrs	Hour
HEAT	Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2*A*), yeast kinase TOR1 domain
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
His	Histidine
HR	Homologous recombination
HRP	Horseradish peroxidase
HU	Hydroxyurea
i.e.	That is [Latin: id est]
ICL	DNA interstrand crosslink
IEX	Ion exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Ionizing radiation
Kam	Kanamycin
l/L	Litre
LB	Lysogeny broth, Luria broth or Luria-Bertani medium
LB	Lysis buffer
LBE	mLST8 protein binding site motif
Lcd1	Lethal, checkpoint-defective, DNA damage-sensitive protein 1
Leu	Leucine
M	Molar (mole/litre)
MALDI-TOF	Matrix-assisted laser desorption/ionization, time of flight mass spectrometer
Mec1	Mitosis entry checkpoint protein 1 (Aliases: Esr1, Sad3, Rad31)
MES	2-(N-morpholino)ethanesulfonic acid
min	Minute
MOPS	3-(N-morpholino)propanesulfonic acid
Mrc1	Mediator of the replication checkpoint 1
MRN	Mre11-Rad50-Nbs1 complex
mRNA	Messenger RNA
MRX	Mre11-Rad50-Xrs2 complex

MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
MW	Molecular weight
MWCO	Molecular weight cut off
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NMD	non-sense mediated mRNA decay
NS	Negative staining
N-ter	N-terminal
NTP	Nucleoside triphosphate
NuA4	Nucleosome acetyltransferase of H4
O/N	Overnight
OD_N	Optical density at N nm
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
PEG	poly(ethylene glycol)
PI	protease inhibitor
PI3K	Phosphatidylinositol-3 lipid kinase
PIC	Pre-initiation complex
PIKK	Phosphatidylinositol/Phosphoinositide 3-kinase-related kinase
P-loop	Phosphate binding loop
PMSF	Phenylmethylsulfonyl fluoride
Pol	Polymerase
PRD	PIKK Regulatory Domain
ProtA	Protein A tag (~ 42 kDa protein from *Staphylococcus aureus*)
rcf	Relative centrifugal force
RELION	REgularised LIkelihood OptimisatioN
rev	Reverse (primer)
Rfa1	Replication factor A protein 1
Rfa2	Replication factor A protein 2
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RPA	Replication protein A

Rpa1	Replication protein A subunit 1
Rpa2	Replication protein A subunit 2
rpm	Revolutions per minute
RSR	Replication stress response
RT	Room temperature
S c.	*Saccharomyces cerevisiae*
S.p.	*Schizosaccharomyces pombe*
s/sec	Second
SAGA	Spt-Ada-Gcn5 acetyltransferase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SNR	Signal-to-noise ratio
ssDNA	Single-stranded DNA
TA	Transcription activator
TAD	Transcription(al) activation, trans-activating or trans-activation domain
TAE	Tris, acetic acid, EDTA
Taq	Thermus aquaticus
TBP	TATA-binding protein
TCA	Trichloroacetic acid
Tel1	Telomere maintenance 1
TEM	Transmission electron microscope
TEMED	N,N,N',N'-tetramethylethylenediamine
Tet	Tetracycline
TEV	Tobacco etch virus derived protease
TF	Transcription factor
TFB-I, TFB-II	Transformation buffer I and II
TFIID	Transcription Factor II D
T-loop	Activation loop (named because of the phosphorylated Thr at the C-terminal)
TopBP1	DNA topoisomerase 2 binding protein 1
Tor1	Target of rapamycin kinase 1
Tor2	Target of rapamycin kinase 2
TPR	Tetratricopeptide repeat
Tra1	Transcription-associated protein 1

Tris	Tris-(hydroxymethyl)-aminomethane
Trp	Tryptophane
TRRAP	Transformation/transcription domain-associated protein
Trx/TrxA	Thioredoxin solubility tag (~ 12 kDa)
U	Unit
UAS	Upstream activating (or activation) sequence
Ura	Uracil
UTR	Untranslated region
UV	Ultraviolet light
v/v	Volume per volume
w/o	Without
w/v	Weight per volume
WB	Wash buffer
WT	Wild type
x *g*	Times gravity
YNB	Yeast nitrogen base without amino acids
YPD	Yeast extract peptone dextrose

Bases in nucleic acids (single letter abbreviations):

A adenine

C cytosine

G guanine

T thymine

U uracil

Protein amino acid residues (single letter and three letter abbreviations):

A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.