



Universidad Autónoma de Madrid



**Universidad Autónoma de Madrid**  
Departamento de Biología Molecular

***Structural analysis of macromolecular  
nanomachines by 3D electron  
microscopy: managing flexibility and  
heterogeneity in some defined cases***

- TESIS DOCTORAL -

**Ana Lucia Álvarez Cabrera**

Madrid, 2015





Universidad Autónoma de Madrid



**Departamento de Biología Molecular**  
**Facultad de Ciencias**

Universidad Autónoma de Madrid

***Structural analysis of macromolecular  
nanomachines by 3D electron microscopy:  
managing flexibility and heterogeneity in  
some defined cases***

Memoria presentada para optar al grado de Doctor en Bioquímica, Biología Molecular, Biomedicina y Biotecnología (Biociencias Moleculares)

**Ana Lucia Álvarez Cabrera**

**DIRECTORES DE TESIS:**

**Dr. José María Carazo García**  
**Dr. Carlos Oscar Sorzano Sánchez**

Centro Nacional de Biotecnología -CSIC





El trabajo recogido en esta memoria ha sido realizado en el Centro Nacional de Biotecnología (CNB-CSIC) bajo la dirección conjunta de los Drs. José María Carazo García y Carlos Oscar Sorzano Sánchez. Su financiación corrió a cargo de una beca predoctoral de Formación de Personal Investigador (FPI) y de los proyectos Instruct, CAM(S2010/BMD-2305) y AIC-A-2011-0638.



***A las personas más importantes en  
mi vida: mis padres, mis hermanos  
y Álvaro =^.^=***





# Agradecimientos

Quisiera agradecer de forma especial a mis directores de tesis, José María y Coss, por la oportunidad de formar parte de este maravilloso grupo de investigación y, sobre todo, por su constante apoyo a lo largo de estos años. También agradezco profundamente la amistad, enseñanzas y ayuda de todos mis compañeros del CNB y de otros centros de investigación con los que he tenido la oportunidad de trabajar durante este tiempo. Son muchas las personas que merecen, y tienen, mi más profunda gratitud por sus incontables aportes tanto a nivel profesional como personal. La amistad que me han brindado estos años en España ha hecho de este mi segundo hogar.

Mi más cariñoso e importante agradecimiento es para mis padres, mis hermanos y para Álvaro, mi mejor amigo y el amor de vida.



# Table of contents

Figure index .....	III
Table index .....	V
Abbreviations .....	VII
Summary .....	IX
Resumen .....	XI
1. Introduction .....	1
1.1. Thesis Outline .....	1
1.2. Structural biology .....	2
1.2.1. Protein structure .....	2
1.2.2. Protein dynamics and structural heterogeneity .....	4
1.2.3. Protein structure determination .....	4
1.3. Biological system: CPAP centrosomal protein .....	10
1.3.1. The centrosome .....	10
1.3.2. CPAP protein .....	11
1.4. Biological system: MCM4/6/7 helicase .....	16
1.4.1. MCMs proteins .....	16
1.4.2. MCMs as helicases .....	17
1.4.3. The eukaryotic MCMs .....	21
1.4.4. MCM4/6/7 .....	23
1.5. Biological system: TWINKLE, the human mitochondrial DNA helicase .....	26
1.5.1. The mtDNA replisome .....	27
1.5.2. Twinkle architecture: Known and proposed functions .....	28
2. Objectives .....	35
3. Materials and methods .....	36
3.1. Obtaining the protein samples .....	36
3.1.1. CPAP <sup>897-1338</sup> .....	36
3.1.2. MCM4/6/7 .....	38
3.1.3. Twinkle .....	38
3.2. <i>In silico</i> structural analysis: Sequence alignment and atomic structural modeling .....	40
3.2.1. CPAP <sup>897-1338</sup> .....	40
3.2.2. MCM4/6/7 .....	40
3.3. Transmission electron microscopy .....	41
3.3.1. Sample preparation and image acquisition .....	41

3.3.2. Image preprocessing.....	41
3.3.3. Density maps reconstruction: 3D classification and refinement.....	42
3.4. Normal Mode Analysis (NMA) of MCM4/6/7.....	44
4. Results.....	45
4.1. Human CPAP <sup>897-1338</sup> .....	45
4.1.1. Sequence analysis and structural modeling.....	45
4.1.2. Protein purification.....	47
4.1.3. Structural characterization of different CPAP <sup>897-1338</sup> homo-oligomeric complexes and fitting of atomic models.....	49
4.2. Mouse MCM4/6/7.....	63
4.2.1. Multiple sequence alignment and structure prediction modeling of mouse proteins MCM4, MCM6 and MCM7.....	63
4.2.2. EM structural characterization of the sample.....	67
4.3. Human Twinkle.....	91
4.3.1. Aggregation of helicase particles occurring at low salt concentration is reversible by increasing the ionic strength.....	91
4.3.2. Twinkle forms multiple homo-oligomeric complexes showing long and flexible arms.....	93
4.3.3. The ZBD has a role on the structural flexibility.....	99
4.3.4. Initial 3D maps: One and two floor structures.....	103
4.3.5. Establishing sample conditions for cryo-EM studies.....	105
5. Discussion.....	107
5.1. Intrinsic protein flexibility facilitates quaternary structure assembly and conformational/oligomeric fluctuations.....	107
5.2. Structural insights into CPAP conformational and oligomeric behavior: The possible hierarchical building of a scaffold.....	109
5.2.1. CPAP <sup>897-1338</sup> forms different homo-holigomeric complexes <i>in vitro</i> .....	110
5.3. Insights into the architecture and flexibility of the hexameric MCM4/6/7 ring and other lower order sub complexes.....	115
5.3.1. Structural polymorphism.....	116
5.4. Structural Characterization of TWINKLE, the human mitochondrial DNA helicase: flexibility and heterogeneity.....	124
5.4.1. Effect on the helicase sample to changes in ionic environment.....	124
5.4.2. Multiple oligomer states of the enzyme.....	125
5.4.3. Structural flexibility, possible opening of the ring and effect of the ZBD.....	126
6. Conclusions.....	130
Conclusionses.....	132
References.....	134
Appendix.....	151

# Figure index

Figure 1. Structural levels of protein organization. ....	3
Figure 2. Structure of the wild-type TCP10 domain of <i>Danio rerio</i> CPAP (PDB_4BXP) as obtained from a crystalline sample by X-ray crystallography. ....	5
Figure 3. NMR structure (PDB_2M45) of the C-terminus of the minichromosome maintenance protein MCM from <i>Sulfolobus solfataricus</i> . ....	6
Figure 4. Centrosome.....	11
Figure 5. Structural domains in human CPAP protein. ....	14
Figure 6. CPAP <sup>897-1338</sup> interacting partners. ....	14
Figure 7. General structural features of MCM.....	19
Figure 8. Proposed DNA unwinding modes by MCM helicase. ....	20
Figure 9. Active sites in AAA <sup>+</sup> domains are formed at the interface between adjacent protomers.....	21
Figure 10. Architecture of MCM helicases. ....	22
Figure 11. The human mitochondrial genome.....	27
Figure 12. General structural features of the replicative mtDNA helicase Twinkle.....	28
Figure 13. Sequence alignment of the NTD between representative Twinkle homologues and T7 gp4. ....	30
Figure 14. Ring-shape helicase 3D structures. ....	32
Figure 15. Schematic representation of Twinkle protein constructs. ....	39
Figure 16. 3D reconstruction process. ....	43
Figure 17. Secondary structural prediction of CPAP <sup>897-1338</sup> . ....	46
Figure 18. Color-coded schematic representation of the structural domains of Hs CPAP <sup>897-1338</sup> (up) and their related atomic model (down). ....	47
Figure 19. Immobilized metal affinity chromatography (IMAC) purification of CPAP <sup>897-1338</sup> . ....	48
Figure 20. SEC purification of CPAP <sup>897-1338</sup> fibers. ....	50
Figure 21. EM of CPAP <sup>897-1338</sup> fibers. ....	50
Figure 22. Structural prediction corresponding to the CPAP <sup>897-1338</sup> monomer, based on the general cane-like shape of EM images from a SEC purified sample. ....	51
Figure 23. SEC purified fraction of CPAP <sup>897-1338</sup> toroidal particles.....	52
Figure 24. EM of a SEC fraction of CPAP <sup>897-1338</sup> eluted at an apparent MW around 108kDa.....	53
Figure 25. 2D image analysis of a sample fraction eluted at an apparent MW compatible with a CPAP <sup>897-1338</sup> dimer. ....	54
Figure 26. 3D reconstruction of a putative CPAP <sup>897-1338</sup> homo-dimeric complex. ....	55
Figure 27. Purification of CPAP <sup>897-1338</sup> barrel-like particles. ....	56
Figure 28. Transmission electron microscopy of negatively stained CPAP <sup>897-1338</sup> barrel-like particles. ....	57
Figure 29. 3D reconstruction of a putative CPAP <sup>897-1338</sup> homo-tetrameric complex.....	58
Figure 30. Fitting of four copies of a triangular conformation of the proposed atomic structure of CPAP <sup>897-1338</sup> into the 3D map of the putative CPAP <sup>897-1338</sup> tetramer. ....	60
Figure 31. CPAP <sup>897-1338</sup> putative tetramers may stack together by their longer sides to form modular higher order complexes of variable length. ....	61
Figure 32. Multiple sequence alignment of mouse MCM4, MCM6 and MCM7 complete sequences with respect to their homologs in representative organisms. ....	66
Figure 33. General structure of MCM4,6 and 7.....	67
Figure 34. MCM4/6/7 sample and particles.....	69
Figure 35. Representative model of the classically described two-tier MCM helicase. ....	70

Figure 36. Individual images and reference-free class average of putative top-views from MCM4/6/7. ....	71
Figure 37. Representative reference-free class averages of MCM4/6/7 helicase sample. ....	72
Figure 38. KenDerSOM analysis. ....	73
Figure 39. Closely interacting helicase particles. ....	73
Figure 40. Putative MCM4/6/7 trimer reconstruction of a likely compact conformation. ....	76
Figure 41. Putative MCM4/6/7 trimer reconstruction showing extended densities. ....	77
Figure 42. Incipient dimerization stage of MCM trimers. ....	78
Figure 43. Class1: MCM4/6/7x2 closed ring displaying two different extra masses, a big and a small one. ...	81
Figure 44. Fitting of atomic structures within two extra masses protruding from MCM4/6/7x2 map Class 1. ....	82
Figure 45. Proposed architecture of MCM4/6/7x2 map Class1, based on structural similarities with a number of MCM2-7 EM structures featuring some comparable extra masses. ....	83
Figure 46. Class 2: MCM4/6/7x2 complex with two densities protruding from a notched AAA <sup>+</sup> region. ....	85
Figure 47. Class 3: MCM4/6/7x2 open ring with two longer subunits at the gap interface. ....	87
Figure 48. Class 4: MCM4/6/7x2 loosely closed ring with a C-terminal connecting bridge. ....	88
Figure 49. Class 5: N-terminal sealed MCM4/6/7x2 ring with a big C-terminal electron density. ....	89
Figure 50. Class 6: MCM4/6/7x2 complex with a big additional density protruding from one side of a notched AAA <sup>+</sup> region. ....	90
Figure 51. Representative NS-EM images showing the aggregation effect on a wild type Twinkle sample under three different concentrations of salt. ....	92
Figure 52. Recovery of helicase particles from an aggregated sample. ....	93
Figure 53. Reference-free class average classification (CL2D) of the complete set of 14500 particle images of the full-length Twinkle protein sample at 330 mM NaCl. ....	94
Figure 54. Lateral views of the wild type (WT) Twinkle. ....	94
Figure 55. Diagram showing representative NS-EM 2D classes of the wild type (WT) Twinkle sample at 330 mM NaCl. ....	95
Figure 56. Putative NTD. ....	98
Figure 57. Representative NS-EM micrograph of the truncate $\Delta$ ZDB protein at 250 mM NaCl. ....	99
Figure 58. Reference-free 2D class average classification (CL2D) of the complete set of 9991 particle images of the Twinkle deletion construct lacking the ZBD ( $\Delta$ ZBD) of protein. ....	100
Figure 59. Diagram showing representative NS-EM 2D classes of the $\Delta$ ZBD construct at 250 mM NaCl. ....	101
Figure 60. Lateral views of the $\Delta$ ZBD construct. ....	102
Figure 61. RCT reconstructions of both extended and compact conformations of hexameric (6 <sup>er</sup> ) and heptameric (7 <sup>er</sup> ) complexes. ....	104
Figure 62. 2D classes of particles with an apparent thin gap (pointed by a red arrow) on the ring. ....	105
Figure 63. Representative cryo-EM images of Twinkle at 330mM NaCl and in absence of glycerol. ....	105
Figure 64. Cryo-EM of the wild type Twinkle sample at 330 mM NaCl. ....	106
Figure 65. Centriolar localization of long, modular structures, proposed to be putative stacks of CPAP tetramers. ....	111
Figure 66. Tentative model for the progressive self-assembly of CPAP into gradually higher oligomeric subcomplexes until formation of a modular rope-like structure. ....	113
Figure 67. Normal mode analysis (NMA) of the MCM4/6/7x2 ring-like structures Class 1 to Class 6. ....	119
Figure 68. Proposed sequence of a putative multiple-step assembly process of MCM4/6/7x2. ....	121
Figure 69. Putative domain interactions of Twinkle that could explain the conformations observed by NS-EM. ....	128

# Table index

<i>Table 1. Primary autosomal recessive microcephaly (MCPH) associated proteins. ....</i>	<i>12</i>
<i>Table 2. Estimated resolution of the 3D maps of MCM4/6/7x2 hexamers Class 1 to 6, by the Fourier Shell Correlation (FSC) criteria at a value of 0.5.....</i>	<i>80</i>
<i>Table 3. Approximate percentage (%) of the different oligomeric state particles for the full-length wild type (WT) Twinkle sample. ....</i>	<i>96</i>
<i>Table 4. Colored dashed circles demark the approximate external diameter of the particles' central channel (blue), CDT (red) and NTD (black) rings, for each of the different oligomeric states.....</i>	<i>97</i>
<i>Table 5. Approximate percentage (%) of the different oligomeric state particles for the <math>\Delta</math>ZBD construct sample. ....</i>	<i>102</i>





# Abbreviations

$\Delta$ ZBD:	Zinc binding domain deletion
2D:	Two-dimensional
3D:	Three-dimensional
aa:	Amino acids
AAA+:	ATPase associated with diverse cellular activities
adPEO:	Autosomal dominant progressive external ophthalmoplegia
ATP $\gamma$ -S:	Adenosine-5'-(3-thiotriphosphate)
CENPJ:	Centromere Protein J
CPAP:	Centrosomal P4.1 associated protein
CL2D:	Clustering two-dimensional classification
CTD:	C-terminal domain
DNA:	deoxyribonucleic acid
dTTP:	Deoxythymidine triphosphate
dTDP:	Thymidine diphosphate
dsDNA:	Double-stranded DNA
EM:	Electron microscopy
EMMDB:	Electron Microscopy Data Bank
FSC:	Fourier Shell Correlation
hp:	hairpin
His-tag:	Histidine tag
IMAC:	Immobilized metal affinity chromatography
MCM:	Minichromosome maintenance
MCM2-7:	MCM2, MCM3, MCM4, MCM5, MCM6 and MCM7
MCM4/6/7-x2:	MCM4/6/7 hexameric ring, formed by two MCM4/6/7 heterotrimers
MCPH:	Primary autosomal recessive microcephaly
MT:	Microtubule
MTOC:	Microtubule organizing centre
mtSSB:	Mitochondrial single-stranded DNA-binding
mtRNA:	Mitochondrial ribonucleic acid
mtRNAPol:	Mitochondrial ribonucleic acid polymerase

MW:	Molecular weight
n-1:	All but one
NS:	Negative staining
NS-EM:	Negative staining electron microscopy
NTD:	N-terminal domain
OCCM:	ORC-Cdc6-Cdt1-(MCM2-7)
ORC:	Origin recognition complex
PDB:	Protein Data Bank
PCM:	Pericentriolar material
pol $\gamma$ :	Polimerase gamma
pre-RC:	Pre-replicative complex
RANSAC:	Random Sample Consensus
RCT:	Random Conical Tilt
RNA:	Ribonucleic acid
RPD:	Ribonucleic acid polymerase domain
SANS:	Small-angle neutron scattering
SAXS:	Small angle X-ray scattering
SEC:	Size exclusion chromatography
SF4:	Superfamily 4
SF6:	Superfamily 6
ssDNA:	Single-stranded DNA
T7 gp4:	Bacteriophage T7 gen 4
TEM:	Transmission electron microscopy
WT:	Wild type
ZBD:	Zinc binding domain

# Summary

Macromolecular complexes are truly nanomachines that play a crucial role in most cellular processes. Malfunctions of these biological machines are implicated in a wide variety of diseases, being the determination of their three-dimensional (3D) structures of high relevance in the medical and pharmaceutical fields. Indeed, an important facet required to understand the mechanism of a biological macromolecular assembly is the characterization of its structure and the interpretation of the potential dynamics that could represent the existence of multiple conformations.

In the present work there are presented the structural characterizations of three different biological cases: human CPAP, a centrosomal protein fundamental for centriole assembly; mouse MCM4/6/7, a group of proteins that form hetero-oligomeric ring-like complexes showing helicase activity; and human Twinkle, the mitochondrial replicative DNA helicase. All three cases share in common the fact of being highly dynamic and flexible proteins, able to form different oligomeric complexes. Through each of the presented cases of study, it was addressed, to a varying extent, the challenge of elucidating the structures of a variety of oligomeric complexes, dealing with the conformational flexibility of these macromolecular machines at the same time that it was analyzed the possible biological/functional implications of the obtained results. For this goal, it was used a structural approach employing Transmission Electron Microscopy (EM) and computational analysis techniques. The architecture characteristics of the obtained two-dimensional (2D) images and three-dimensional (3D) models, along with the already reported functional, biochemical and structural information for the specific proteins, allow us to propose putative sequences of conformational/oligomeric changes. The presented density maps, together with the suggested associated structural transition pathways, provide new insights into the possible mechanistic behavior of these machines of life.



# Resumen

Los complejos macromoleculares son verdaderas máquinas nanoscópicas que juegan un papel crucial en la mayoría de los procesos celulares. El mal funcionamiento de estas máquinas biológicas está relacionado con un gran número de enfermedades, por lo que la determinación de sus estructuras tridimensionales (3D) resulta de gran importancia con fines aplicados a los campos de la medicina y la farmacéutica. Para llegar a comprender el mecanismo de los ensamblajes moleculares biológicos se requiere, en parte importante, de su caracterización estructural y de la dinámica que puede representar la existencia de múltiples conformaciones.

En este trabajo se presentan las caracterizaciones estructurales de tres casos biológicos distintos: la proteína centrosomal de humano CPAP, esencial para el ensamblaje de los centriolos; la helicasa MCM4/6/7 de ratón, que consiste en un grupo de proteínas que forman complejos en forma de anillo, capaces de separar DNA de doble cadena; y Twinkle, la helicasa de ADN mitocondrial de humano. Todos los tres casos comparten en común la característica de ser proteínas muy flexibles y altamente dinámicas, capaces de formar distintos tipos de complejos oligoméricos. A través de cada uno de los casos presentados, se abordó, en mayor o menor grado, el reto de elucidar las estructuras de una variedad de complejos oligoméricos, lidiando con la flexibilidad conformacional de estas máquinas macromoleculares y, al mismo tiempo, proponiendo el posible significado biológico/funcional de los resultados obtenidos. Con este propósito, se empleó una aproximación estructural mediante estudios por técnicas de microscopía electrónica de transmisión (TEM) y análisis computacional. Las características estructurales de las imágenes bidimensionales (2D) y los modelos tridimensionales (3D), junto con la información funcional, bioquímica y estructural que ha sido reportada para cada proteína específica, permitió proponer secuencias putativas de cambios conformacionales/oligoméricos. Los mapas de densidad presentados y los respectivos pasos de transición estructural propuestos proporcionan nuevas perspectivas en el posible comportamiento mecánico de estas máquinas de la vida.



## Chapter 1

# 1. Introduction

## 1.1. Thesis Outline

This thesis is divided into six main sections as follows:

1. In the introductory chapter, basic concepts about protein structure are briefly summarized as well as the most common experimental techniques used in the field of structural biology to elucidate the 3D architecture of macromolecules. A special emphasis is placed into single particle electron microscopy (EM), the technique used for the resolution of the different structures obtained during the development of the three subprojects that constitute this thesis. In order to contextualize the biological relevance of each of these subprojects, the processes into which each of the related proteins/complexes are known to be involved are explained as well as what is currently known about their own properties and functional roles.
2. The objectives of the thesis are presented in the second section.
3. The third chapter describes the materials and methods.
4. The fourth section is the presentation of results. This part is subdivided into three different but related sub-sections from the perspective of managing biological samples characterized by both a high degree of conformational flexibility and oligomeric heterogeneity. Here the novel structural information obtained in all the cases under study is shown.
  - 4.1. The first project is focused on the human centrosomal protein CPAP. It encompasses the molecular cloning, expression and purification of a N-terminal truncate construct of the protein, as well as the acquisition and processing of EM images of multiple oligomeric species formed by this protein. It is proposed a putative sequence of increasing oligomerization steps, as well as the possible visual identification of the largest supramolecular CPAP structures within the centriole.

- 4.2. The mouse hexameric helicase MCM4/6/7 is the subject of the second case of study. Despite being an enormous challenge from the image analysis point of view, the structural dynamism of this sample was a unique opportunity to capture conformational and oligomeric changes that could give new insight into the assembling and action mechanisms of this complex.
- 4.3. The last project involves the work with purified samples of TWINKLE, the human mitochondrial DNA helicase. In a first instance, the conditioning and selection of the optimal sample conditions for its later use in EM studies was carried out. The 2D image analysis and the initial 3D models of different oligomeric and conformational states of the protein are presented.
5. In this part, the results of each of the projects are discussed, placing them into the context of the *state of the art* of their related topics, including an analysis of the relevance and possible contributions to the biological and biomedical fields.
6. The final section summarizes the main conclusions and findings of this thesis.

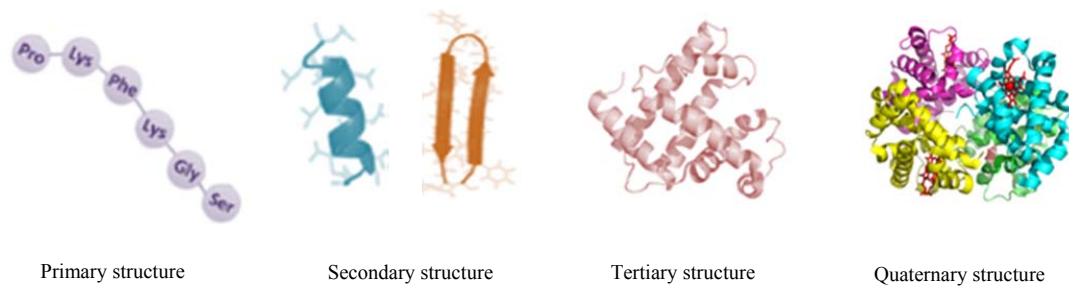
## **1.2. Structural biology**

Structural biology is a branch of the biological sciences that incorporates the principles of molecular biology, biochemistry and biophysics, among other disciplines, including bioinformatics, to study the molecular structure and dynamics of biological macromolecular assemblies, in particular proteins and nucleic acids, and how alterations in their structures affect their function.

### **1.2.1. Protein structure**

Proteins are large and complex molecules made up by a linear sequence of amino acids, whose different levels of organization in the three-dimensional (3D) space (Figure 1) give rise to structures involved in the execution and direction of virtually all the processes required to maintain living systems. Some proteins can form multimers with the same or different type of proteins (as well as other chemical entities, like nucleic acids or sugars), and such supramolecular assemblies constitute the final functional units.





**Figure 1. Structural levels of protein organization.** Primary structure (amino acid sequence), secondary structure (in blue an  $\alpha$ -helix and in orange a  $\beta$ -sheet), tertiary structure (the fold of a chain), and quaternary structure (multi-subunit complex made of several combined chains).

Proteins can be broadly classified according to their shape and solubility into three major classes:

- **Fibrous proteins** are poorly soluble, but highly flexible rod-like structures, where the secondary structure is the predominant feature and little or no tertiary structure is displayed.
- **Globular proteins** are more compact and water soluble molecules, since its structure is essentially organized in a way that most of its hydrophobic amino acids form an internal core that is sheltered by polar residues exposed to the watery surface.
- **Membrane proteins** are relatively flexible and aqueous insoluble molecules that are found in close association with lipid membranes through their hydrophobic surfaces, while their hydrophilic residues tend to protrude out.

Nevertheless, describing the structure of a protein might not be so straightforward, since a unique and static structure is not a realistic situation for most proteins. Besides, there are proteins whose sequences are predicted to be partially or completely unstructured (Romero & Obradovic 1998). Although intrinsically disordered proteins (IDPs) or regions (IDRs) can acquire a more stable folding upon interacting with a binding partner, it has been observed that some ID segments retain a disordered state in the new formed complex, which is known as a fuzzy complex (Tompa & Fuxreiter 2008).

## **1.2.2. Protein dynamics and structural heterogeneity**

For decades ago it was assumed that each protein presented a unique functional 3D structure and that its interaction with a partner followed a highly specific and rigid lock-and-key model. However, far from being static, isolated objects, proteins are dynamic entities that make use of different levels of structural flexibility (i.e. local, regional or global) to interact or associate with other molecules in order to carry out their function. While it is true that activity of many proteins is associated with a stable 3D architecture, their function must still be accomplished by a dynamic sequence of movements and biochemical interactions occurring at different timescales and ranging from few atomic fluctuations to drastic conformational rearrangements, which can include their assemble into higher order macromolecular structures. It is not unusual to find proteins or complexes that exist simultaneously in a variety of functional conformations. The conformational and oligomeric dynamic behavior shown by some proteins brings a broad and heterogeneous set of structures, whose elucidation would constitute an important pillar for the understanding of their biological function.

## **1.2.3. Protein structure determination**

A comprehensive study of protein structure and its dynamics is a determining factor for a successful and deeper understanding of their functions, so describing the conformational states of proteins and biological macromolecular assemblies is a critical point in the structural biology field. The architectural description of heterogeneous samples (due to protein flexibility and/or different oligomerization states) is a challenging, yet exciting task for structural biologists, who often employ different experimental and bioinformatics tools to handle this kind of problems.

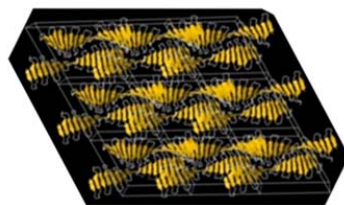
### **1.2.3.1. Experimental methods**

To date, the most used and reliable experimental approaches for structural determination of macromolecules are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, which is mainly employed for high resolution model of small proteins and single particle electron microscopy (EM) analysis to obtain medium-resolution density maps of larger proteins and complexes. Each of these methods has its own advantages and limitations, so an integrated strategy using information obtained from different techniques, when available, must be followed

in order to reach the atomic details for any biologically significant state of a macromolecular assembly.

## X-ray crystallography

This crystallographic technique employs X-rays with a wavelength around  $1\text{\AA}$ , which allows resolving structures at the atomic level. Obtaining a good protein crystal is the most crucial and difficult task of the whole process, so a number of hurdles must be overcome before getting a X-ray diffraction pattern from which the structural information of the molecule could be extracted. A first bottleneck is to have a considerable amount of a highly pure, homogenous and concentrated sample, which may be very complicated in a large number of cases. A second, and no less complicated step, is the production of a large enough (a minimum of 0.1-0.2 mm in at least 2 dimensions) and nicely organized crystal diffracting at good resolution (Egli 2010). Crystal nucleation and growth are non-trivial challenges, since those are extremely complex and poorly understood processes. Proteins presenting IDRs or very flexible domains are frequently incompatible with the X-ray crystallography technique. On the other hand, proteins forming part of a crystal have been forced to acquire the organized arrangement that allows its crystallization (Figure 2), so the resolved structure may not necessarily reflect a physiological conformation of the particle.



**Figure 2. Structure of the wild-type TCP10 domain of *Danio rerio* CPAP (PDB\_4BXP) as obtained from a crystalline sample by X-ray crystallography.**

## Nuclear magnetic resonance (NMR) spectroscopy

The methodological implementation of NMR involves the use of a number of separate and specialized techniques, and is traditionally used for the determination of the atomic structure and dynamic of relatively small proteins or nucleic acids in solution. As in large molecules the NMR signals are more prone to overlap and their lower frequency Brownian motion produces a

significant reduction in sensitivity and resolution, typically the samples employed do not exceed 30-40kDa. Through NMR spectroscopy studies, a set of structures coming from an analyzed macromolecule can be obtained (Figure 3), taking advantage of the fact that the nuclei of atoms can be aligned by applying a strong magnetic field, and then the nuclei magnetic momentum can be changed back and forth from the equilibrium by using the energy of certain radio wave frequencies. The detected nuclear resonance spectrum is interpreted to obtain structural information.



**Figure 3. NMR structure (PDB\_2M45) of the C-terminus of the minichromosome maintenance protein MCM from Sulfolobus solfataricus.**

Although in recent years the use of more powerful technologies and new complementary methods has allowed the use of NMR spectroscopy in the study of some larger proteins and complexes (Frueh et al. 2013), those are still exceptional cases. As in X-ray crystallography, a crucial requirement for NMR experiments is the high purity of the protein, so this technique is limited by both, size and sample availability.

## Transmission Electron Microscopy (TEM)

EM is a versatile technique that can be employed in the structural study of a broad variety of samples types and sizes. For example, *single-particle EM* is the method of choice when working with individual particles and filaments with or without symmetry, which have not been able to be solved by other techniques. *Electron crystallography* is employed for the study of small 2D crystals (< 0.1  $\mu\text{m}$ ) because they cannot be analyzed by X-ray crystallography due to their limiting size. Finally, *electron tomography* allows obtaining 3D information of pleomorphic entities, such as cells or of sub-cellular macromolecular objects.

In an electron microscope, electrons provided by an emission source (e.g. a tungsten, lanthanum hexaboride filament, or a field emission gun -FEG-) are accelerated along a vacuum system column

by an electric field at a high voltage and are focused by electromagnetic lenses into a collimated beam. The electron beam passes through the sample interacting with the specimen, and the transmitted electrons that will finally reach a detector form a 2D image of the sample being imaged. The brightness of the different areas in the image is proportional to the phase shift between scattered and unscattered electrons transmitted through the sample at each point. Under the electron microscope, biological samples produce a very low contrast and are highly sensible to damage caused by vacuum conditions and electron dose irradiation, so the sample structure must be preserved in some way before and during its observation. Two main techniques for sample preparations can be employed, although variations of them can be done:

- Negative staining (NS) consists in the incubation of the sample with an electron dense material (usually a solution of a heavy metal salt), which covers the particles creating a sort of mask with the same shape, that is more resistant to radiation and dehydration. The stain increases the Signal-to-Noise Ratio (SNR) of the images, but as a drawback, the resolution is limited by the grain size and penetration capability of the stain; additionally, some fragile structures can be deformed. Particles as small as 53 kDa (Zhang et al. 2013) have been solved by NS, although currently, this small size is more the exception than the norm.
- Cryo-vitrification of the sample is carried out by its fast immersion in liquid ethane; in this way the specimen gets frozen into an amorphous layer of ice that keeps the particles in a state close to the physiological one. Vitrified samples do not suffer from the resolution limit of NS, but the contrast is very low, which is a fundamental issue when further processing these cryo-EM images.

Detectors for image acquisition were traditionally photographic films or charge-coupled devices (CCD), but recently, a new generation of Direct Electron Detectors (DED) has been developed which directly receive the energy of the incident electrons. Among other advantages, they greatly improve the spatial and time resolution of the image (Milazzo et al. 2011). The posterior 3D reconstruction method can be done thanks to the fact that both phase and amplitude information is present in the 2D image. Resolution of EM density maps is normally measured by random but equally dividing the image data set into two sub-sets, so two independent 3D reconstructions are obtained and a comparison between their Fourier shell correlations (FSC) can be carried out. The most commonly used criterion is the 0.5 FSC cutoff, which refers to when the correlation

coefficient of the Fourier shells drops below 0.5 (Manuscript 2011), although different criteria are also possible (Rosenthal & Henderson 2003; Scheres & Chen 2012).

### ***Single-particle EM***

Most of the activities within the cell are carried out by dynamic interactions and association between proteins. The hallmark strength of TEM, compared to X-ray crystallography and NMR, is the capacity of this technique to capture structural information of large biological macromolecules and complexes in their native environment, using just a tiny amount (nanograms) of sample. Sample purity requirements do not need to be so high, especially if the particle under study can be visually distinguished from possible contaminants. Single-particle EM analysis is a powerful tool for the structural study of macromolecular assemblies that exist in multiple conformational states, which confer to this technique an additional and unique advantage when compared with other methods. Big size, symmetry and homogeneity of the analyzed particle, are factors that boost the reachable resolution of the final 3D reconstruction, while the opposite situation makes more difficult but does not preclude obtaining volumes with valuable structural features (Burgess et al. 2004).

In single-particle analysis, thousands of preprocessed images obtained from randomly oriented particles, are computationally aligned and classified. Those images are subsequently used for constructing initial density maps (using methods like random conical tilt -RCT- (Radermacher 1988) or RANSAC (Vargas et al. 2014)) or for the refinement of preexistent volumes (Frank 1996). The above mentioned processes use to be carried out by employing a combination of different software image processing programs (e.i. Xmipp (De la Rosa-Trevín et al. 2013), Relion (Scheres 2015), EMAN (Ludtke et al. 1999), FREALIGN (Grigorieff 2007), among others).

### **1.2.3.2. Complementary bioinformatics methods**

A number of technical and sample related issues are still limiting factors for EM to reach the theoretical resolution that electron scattering would allow, especially for small, asymmetric particles (Glaeser & Hall 2011). Resolution can be partially improved using hybrid methods that involve fitting high resolution structures solved by other techniques into the EM density maps

(Rossmann et al. 2005). The level of success of this method depends on the actual resolution of the EM map and the coverage of the available atomic models to be fitted. Bioinformatics programs used for *de novo* predicting and modeling atomic structures are very helpful when there are not experimentally determine structures, or when the available ones are incomplete. Prediction of secondary structures, globular or disordered regions based on the protein sequence, also brings useful information.

## **1.3. Biological system: CPAP centrosomal protein**

### **1.3.1. The centrosome**

Centrosomes are found in most animal cells as the primary microtubule organizing centers (MTOC), carrying out an important number of cellular processes, such as the regulation of normal cell division and motility; the organization of interphasic cytoskeleton's microtubules and the spindle apparatus, and the establishment of the cellular polarity. This organelle is composed by two barrel-shape centrioles (a mature one called the mother centriole, and an immature centriole that is assembled during the previous cell cycle, called the daughter centriole) interconnected through fibers attached to their proximal base, and surrounded by the pericentriolar material (PCM), an apparently amorphous proteinaceous matrix (Bettencourt-Dias & Glover 2007) (Figure 4A).

The centrosome is normally localized at the central region of interphasic cells. However, after its replication through the S phase of the cell cycle, the centrosomes migrate to opposite poles of the cell during the prophase of mitosis, allowing the formation of the mitotic spindle between the two centrosomes (Figure 4B). The cellular bipolarity and the equal partition of chromosomes into two daughter cells are all dependent on normal centrosome function and duplication, which at the same time depends on the correct assemble of the centrioles. Centrosomes are sites of integration and activation of proteins that trigger cell division. During the last several years numerous proteins predicted to be notably rich in coiled-coil motifs and unstructured regions have been reported to be associated with the centrosome, either as permanent components, or temporally linked and concentrated at the centrosome during some stages of the cell cycle (Nogales-Cadenas et al. 2009; Alves-Cruzeiro et al. 2014).



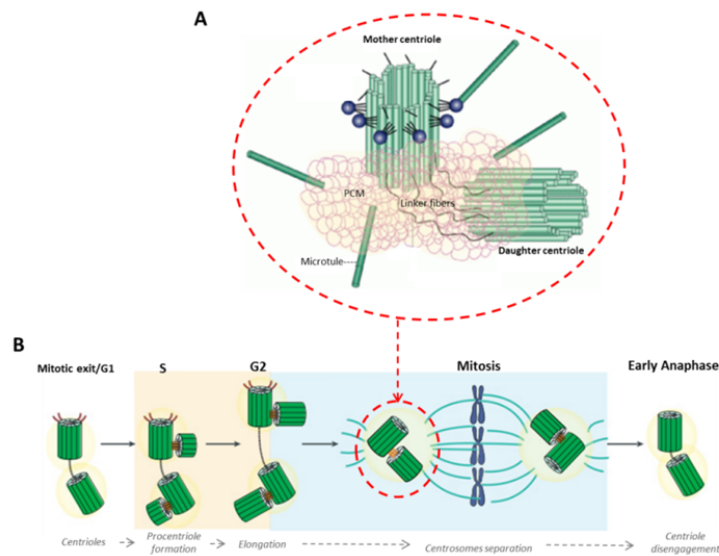


Figure 4. Centrosome. **(A)** Main structural components of centrosome. Adapted from (Doxsey 2001). **(B)** Representation of the centrioles duplication process coupled with the cell cycle. Adapted from (Debec & Sullivan 2010).

### 1.3.2. CPAP protein

CPAP (Centrosomal P4.1 associated protein, also known as Centromere Protein J –CENPJ-, or as Sas-4 in *Drosophila* and SAS-4 in *C. elegans*) is a centrosomal protein that plays a crucial role in the recruitment of microtubules (MT) during centriole formation, so it is a critical component implicated in the biogenesis of the centrosome (Cho et al. 2006; Kohlmaier et al. 2009). It has been demonstrated that CPAP and its orthologs are mainly associated with the tethering of the PCM to centrioles (X. Zheng et al. 2014; Kohlmaier et al. 2009; Tang et al. 2009; Tang et al. 2011; Gopalakrishnan et al. 2011; Hsu et al. 2008), being this a key process for the proper assembly of the centrosome which, in turn, is tightly implicated in the regulation of cell cycle progression (Doxsey et al. 2005).

CPAP is regulated during the cell cycle (Tang et al. 2009) and can be localized in different structures within the centrosome (Kleylein-sohn et al. 2007). CPAP associates with many other centriolar and PCM proteins, such as P4.1R-135 (Hung et al. 2000), CEP152 (Cizmecioglu et al. 2010), CEP135 (Y.-C. Lin et al. 2013; Vulprecht et al. 2012), CEP120 (Y.-N. Lin et al. 2013; Comartin et al. 2013), Tubuline heterodimers (Hung et al. 2004),  $\gamma$ -tubulin (Hung et al. 2000; Hung et al. 2004), STIL (Tang et al. 2011; Cottee et al. 2013), 14-3-3 (Chen et al. 2006) and , Centrobilin (Gudi et al. 2014). Furthermore, it has been reported that SAS-4 also interacts with cytoplasmic complexes,

and very interestingly, that its role in the PCM assembly (through its N-terminus) is independent to the role that it carries out along the formation of the centriole (through its C-terminus) (Gopalakrishnan et al. 2011). All the mentioned facts highlight the multifunctionality nature of CPAP, which make us realize about the potential structural complexity that this protein must have in order to be properly adapted to perform different but specific tasks in each situation when it is required. At present, the oligomeric and structural dynamics of CPAP continues poorly understood.

In CPAP, the non-conservative mutation E1235V has been described in patients diagnosed with primary microcephaly (MCPH) (Leal et al. 2003), a genetically heterogeneous disorder characterized by a significant reduction in normal brain volume and intellectual disability, whereas maintaining the general architecture of a normal brain. To date, fourteen loci, including the gene CENPJ that codes for CPAP (locus MCPH6), have been found to be related with MCPH disease (Table 1); most of these genes code for proteins located in the centrosome, either in a constitutive way or transiently associated to the centrosome during part of the cell cycle.

Locus name	Protein
MCPH1	*Microcephalin 1
MCPH2	*WDR62
MCPH3	*CDK5RAP2
MCPH4	CASC5
MCPH5	*ASPM
MCPH6	*CPAP/CENPJ
MCPH7	* <sup>#</sup> STIL
MCPH8	* <sup>#</sup> CEP135
MCPH9/SCKL5	* <sup>#</sup> <i>CEP152</i>
MCPH10	ZNF335
MCPH11	<i>PHC1</i>
MCPH12	*CDK6
---	*SAS-6
SCKL6	* <i>CEP63</i>

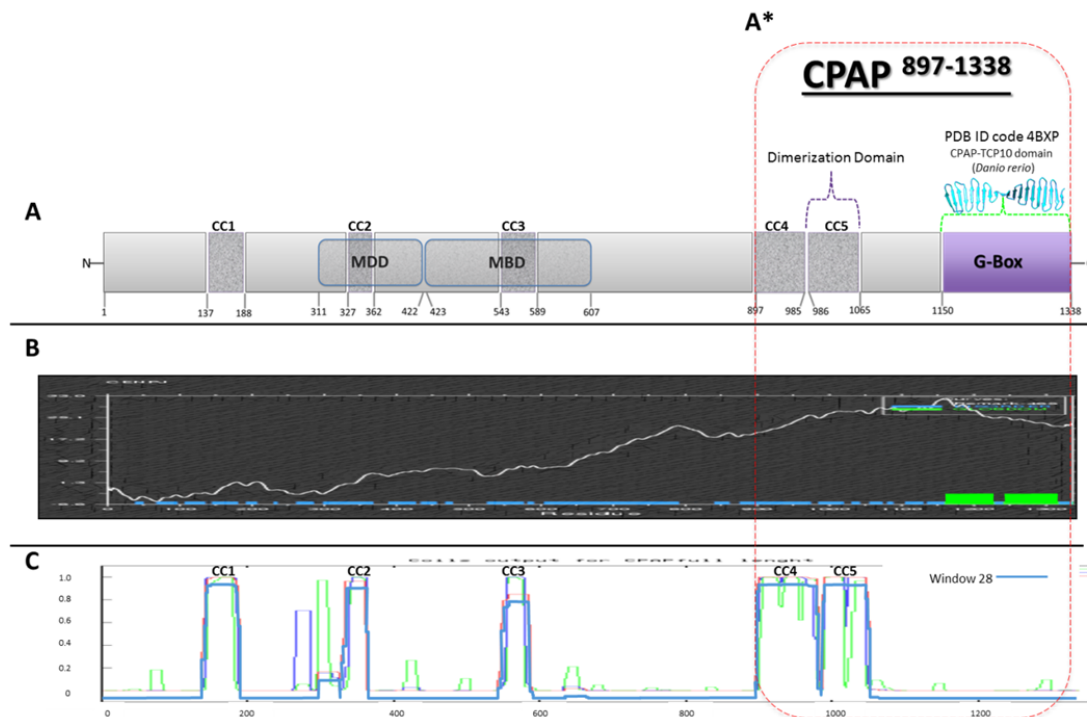
\* Proteins localized at the centrosome in a constitutive way or transiently associated to centrosome during some stage of the cell cycle. <sup>#</sup>Proteins that have been observed to directly interact with CPAP.

**Table 1. Primary autosomal recessive microcephaly (MCPH) associated proteins.**

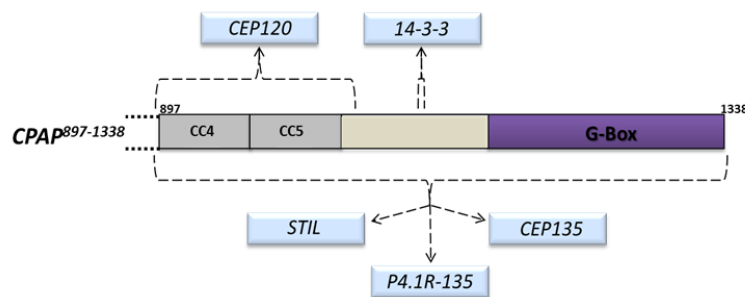
### 1.3.2.1. CPAP structure

At the structural level, CPAP presents a C-terminus with several short glycine-rich repeats that together form the so called G-Box domain (also known as TCP domain), which has a highly conserved sequence through evolution, pointing to its primordial role in centriole/centrosome development, a fact that has been widely corroborated experimentally. In recent years, a number of crystallographic structures of the G-Box/TCP domain of CPAP have been resolved, revealing a solvent-exposed  $\beta$ -sheet structure that remains stable by its own despite lacking a hydrophobic core (X. Zheng et al. 2014; Hatzopoulos et al. 2013; Cottee et al. 2013). The rest of CPAP is formed by disordered regions and coiled-coil motifs (Hung et al. 2000; Carvalho-Santos et al. 2010). The human CPAP (*Hs* CPAP) protein presents five coiled-coil domains, of which the 4<sup>th</sup> and 5<sup>th</sup> coiled-coils (CC4 and CC5, respectively) show a circular dichroism (CD) spectra consistent with an  $\alpha$ -helical secondary structure (Hatzopoulos et al. 2013), which agree with the *in silico* prediction of its amino acid (aa) sequence. CPAP also contains both a MT binding domain (MBD, 423-607aa), that binds to a polymerized microtubule, and a MT destabilizing domain (MDD, 311-422aa), that binds to an  $\alpha/\beta$ -tubulin heterodimer (Figure 5). The later interaction results in inhibiting nucleation and MT depolymerization (Hung et al. 2004; Cormier et al. 2009; Hsu et al. 2008).

The region of *Hs* CPAP between residues 897-1338 contains the G-Box (residues 1150-1338), the coiled-coils CC4 and CC5 (which together cover the residues 897-1056, and that we will call "CC5/CC4" from now on), and a predicted unstructured zone (residues 1066-1149, that will be referred from now on as the "CCGb-linker" region), that makes a linker between CC4/CC5 and G-Box (Figure 5A\*). Interestingly, it has been found that a number of regions included in this 897-1338 sequence are involved in the interaction of CPAP with other centrosomal proteins, such as P4.1R-135, CEP135, CEP120, STIL and 14-3-3 (Figure 6). The CC5 domain is responsible for the dimerization of CPAP (Zhao et al. 2010). Indeed, it is known that the formation of CPAP oligomers as well as its disruption once the protein is phosphorylated during mitosis, is essential for the correct centrosome integration along the cell cycle (Zhao et al. 2010) and for the regulation of the interactions of CPAP with other proteins (Chen et al. 2006).



**Figure 5. Structural domains in human CPAP protein.** (A) Schematic representation of the structural domains in *Hs* CPAP full-length protein. Dark gray boxes represent the five predicted coiled-coils (labelled as CC1, CC2, CC3, CC4 and CC5), being the CC5 coiled-coil the oligomerization domain of CPAP; light gray corresponds to unstructured regions and the purple box refers to the G-Box domain, whose crystallographic structure was already solved in *Danio rerio* (PDB-4BXP); MDD: Microtubule destabilization domain and MBD: Microtubule binding domain. (A\*) Red dashed lines demarcate the CPAP<sup>897-1338</sup> construct. (B) In silico prediction of globular and disordered regions of proteins *Hs* CPAP (full-length), obtained using the program GlobProt2. The green boxes show the predicted globular domains, whereas the blue boxes correspond to disordered regions of the protein. (C) Probability prediction of five coiled-coil structures in CPAP, calculated with the software COILS.



**Figure 6. CPAP<sup>897-1338</sup> interacting partners.** Approximate regions within CPAP<sup>897-1338</sup> (demarcated by the corresponding slashed curly brackets) that have been reported to interact with other centrosomal proteins (light-blue boxes).

Taking into account the current available functional experimental data of CPAP, its multiple interaction partners, and the peculiar supramolecular  $\beta$ -sheet structure of its G-Box domain (Kirkham et al. 2003; Gopalakrishnan et al. 2011), it has been proposed that CPAP works initially as a scaffold for different cytoplasmic proteins, followed by tethering of the new ensemble PCM complexes to the centriole through its C-terminal domain, allowing to continue the formation of normal and functional centrosomes (Gopalakrishnan et al. 2011; Hatzopoulos et al. 2013; X. Zheng et al. 2014; Tang et al. 2009). Despite of the biochemical and biophysical evidences of CPAP dimer formation (Hatzopoulos et al. 2013; Zhao et al. 2010), there are neither clear structural data of such complex or reports of other organized higher order oligomeric arrangements of CPAP. Based on our work on the purification and EM image analysis of CPAP<sup>897-1338</sup>, here we report, for the first time, that this protein acquires several oligomeric states, showing striking new structural features that reveal the remarkably conformational versatility of this protein, which significantly contributes to the understanding of the biological role of CPAP.

## **1.4. Biological system: MCM4/6/7 helicase**

### **1.4.1. MCMs proteins**

Minichromosome maintenance (MCM) genes were identified for the first time in cold-sensitive mutant yeasts that showed abnormal plasmid segregation (defective minichromosome maintenance) (Maine et al. 1984; Hennessy et al. 1991; Moir & Botstein 1982). Early studies placed MCM proteins as central pieces in the initiation of the DNA replication process, a role substantially confirmed by subsequent work showing the helicase activity of MCMs ring complexes in different systems (for a review see (Bochman & Schwacha 2009; Bell & Botchan 2013) ). Despite being well known for their role as replicative helicases, other observations revealed that the number of copies of MCM proteins in the cell not only exceeds by far the number of replication origins, but also that they are widely distributed on heterochromatin and that their levels seems to remain stable even after the end of DNA replication (Adachi et al. 1997; Dimitrova & Todorov 1999; Krude & Musahl 1996; Madine et al. 1995; Kinoshita & Johnson 2004; Claycomb et al. 2002; Dalton & Whitbread 1995; Forsburg et al. 1997; Kimura et al. 1995; Schulte et al. 1995; Tsuruga et al. 1997). This gives rise to the so called “MCM paradox” (Woodward et al. 2006; Ibarra et al. 2008; Hyrien et al. 2003; Takahashi et al. 2005; Das et al. 2014) and opened the discussion on whether MCMs could be involved in other cell processes besides DNA replication. Indeed, to date, a number of studies suggest a participation of MCMs proteins in transcription and remodeling of chromatin (Yankulov et al. 1999; DaFonseca et al. 2001; Dziak et al. 2003; Fitch et al. 2003; Sterner et al. 1998; Holthoff 1998; Ishimi et al. 2001). It is also known that under conditions of cellular stress, the activation of dormant origins of replication depends on this “extra” amount of MCMs (Crevel et al. 2007; Woodward et al. 2006; Ibarra et al. 2008).

In the clinical field, MCMs have begun to be recognized by their potential as cell proliferation markers. Furthermore, their expression levels seems to be associated with several pathologies, which make them useful as a diagnostic and prognostic tool for different malignancies, such as cancer (Giaginis & Vgenopoulou 2010).

It seems evident that MCM proteins contribute in different ways to maintain the integrity of the eukaryotic genome and, consequently, to cell biogenesis, but further research is needed to understand other roles and processes where MCMs could be involved.

## 1.4.2. MCMs as helicases

Helicases have been classified into six superfamilies according to their sequence motif conservation. MCM proteins are found in both archaea and eukaryotes and they belong to the functionally heterogeneous AAA+ (ATPase associated with diverse cellular activities) superfamily 6 (SF6) of helicases. All SF6 member share similar motifs at their AAA+ “core domains” (Singleton et al. 2007; Daniel et al. 2013). More specifically, MCM proteins are members of clade 7, that is characterized by the insertion of an additional  $\alpha$ -helix (helix 2 insert) when compared to the general structurally conserved ATP-binding module of the AAA+ group (Iyer et al. 2004; Erzberger & Berger 2006). The MCM canonical architecture can be summarized by dividing the protein into three general segments presenting some degree of conformational independence among each other (Costa et al. 2014) (Figure 7A):

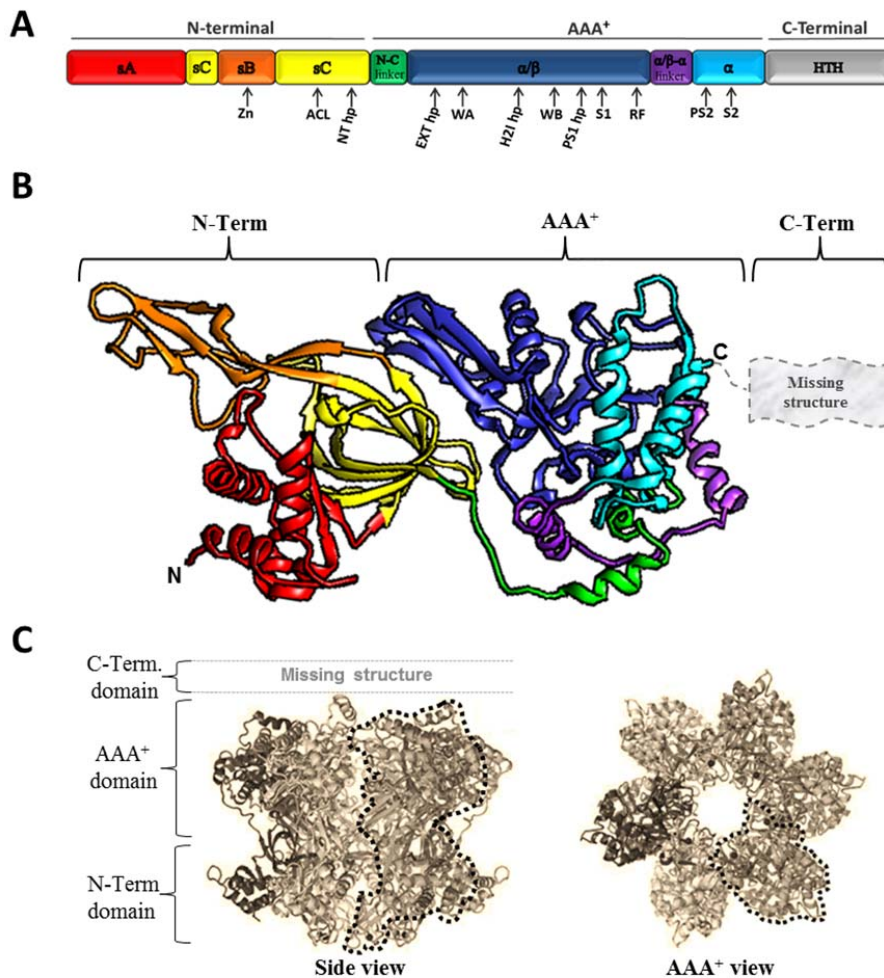
- A N-terminal of low sequence preservation among MCM homologs, that binds both ssDNA and dsDNA, and strongly affects the helicase oligomerization and processivity (Kasiviswanathan et al. 2004; Jenkinson & Chong 2006; Bochman & Schwacha 2007; Bae et al. 2009; Pucci et al. 2007; Liu et al. 2008). This region, in turn, is divided into three structural subdomains (sA, sB and sC)(Costa & Onesti 2009).
- A highly conserved AAA<sup>+</sup> core that is responsible for the catalytic ATPase and helicase activities (Tye & Sawyer 2000; Brewster et al. 2008; Bochman & Schwacha 2009). This domain has an  $\alpha/\beta$  subdomain connected via the  $\alpha/\beta$ -  $\alpha$  linker to an  $\alpha$  domain (Costa & Onesti 2009)
- A C-terminal domain that, despite of a low level of sequence conservation among MCM homologs, shares a pattern compatible with a helix-turn-helix (HTH) fold, a typical DNA-binding motif. In support of this observation, the atomic structures of the C-terminal domains of human (*Homo sapiens*, Hs) MCM6 (PDB-2KLQ), SsoMCM (PDB-22M45) and MthMCM (PDB-2MA3) (Wiedemann et al. 2015), all present a HTH fold. The high flexibility of this part is likely to be the principal factor that makes its structural characterization difficult (Costa et al. 2006b; Brewster et al. 2008; Krueger et al. 2014). This domain shows

a regulatory effect on the helicase activity of MCM archaeal homologs (Aravind & Koonin 1999; Jenkinson & Chong 2006; Pucci et al. 2007; Wei et al. 2010; Barry et al. 2007; Wiedemann et al. 2015).

Some eukaryotic MCM proteins display additional low conserved sequence extensions at their N- and/or C-terminus that may play specialized roles in particular organisms (Lyubimov et al. 2012; Wei et al. 2010; Xu et al. 2013; Tye & Sawyer 2000; Sun et al. 2014; Sheu et al. 2014). In *Methanothermobacter thermautotrophicus* and *Sulfolobus solfataricus* (the most commonly studied archaea systems), there is only one class of MCM protein, *MthMCM* and *SsoMCM*, respectively (Kelman et al. 1999; McGeoch et al. 2005). In contrast with archaea, eukaryotes present several MCM homologous proteins, where each of the MCM2, MCM3, MCM4, MCM5, MCM6 and MCM7 genes defines a preserved separate family sharing significant sequence similarities to each other, mostly but not confined to the region that conforms the AAA<sup>+</sup> core domain (Koonin 1993).

Thanks to the high sequence conservation of the AAA<sup>+</sup> core domain across species and between MCM subfamilies, the analysis of the near full-length (lacking the first 6 N-terminal and the last 85 C-terminal residues) crystallographic structure of *SsoMCM* (PDB ID code 3F9V) has provided more detailed information about several characteristic and relevant elements of MCM helicases (Brewster et al. 2008) (Figure 7B). The most prominent features observed are two globular regions, a big AAA<sup>+</sup> domain and a smaller N-terminal domain, which are interconnected by a narrow linker, giving the appearance of a slim “waist” flanked by two asymmetric lobules. This broad shape can be distinguished in the side views of the 2D images and 3D maps of different MCM helicase complexes obtained to date by Transmission Electron Microscopy (EM)(Lyubimov et al. 2012; Costa et al. 2011). The C-terminal domain is a much smaller and flexible part compared with both AAA<sup>+</sup> and N-terminal domains, which makes it difficult to observe and is usually a missing part in the solved structures of most of these protein.

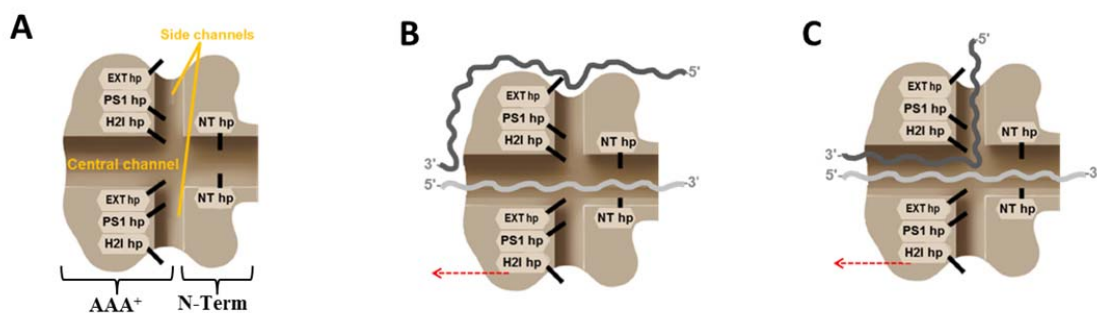




**Figure 7. General structural features of MCM.** (A) Linear schematic representation of the structural motifs of *SsoMCM*. Structural features are labeled and colored as follows: **sA** (red box), subdomain A; **sB** (orange box), subdomain B; **sC** (yellow), subdomain C; **N-C linker** (green box), amino and carboxyl domains linker;  **$\alpha/\beta$**  (blue box),  $\alpha/\beta$  domain;  **$\alpha/\beta$ - $\alpha$  linker** (purple box), linker between domains  $\alpha/\beta$  and  $\alpha$ ;  **$\alpha$**  (cyan box),  $\alpha$  domain; **HTH** (grey box), C-terminal domain helix-turn-helix motif; **Zn**, Zinc finger; **ACL**, allosteric communication loop; **NT hp**, N-terminal  $\beta$ -hairpin; **Ext hp**, external  $\beta$ -hairpin; **WA**, Walker A motif; **H2I hp**, helix 2 insert  $\beta$ -hairpin; **WB**, Walker B motif; **PS1 hp**, presensor 1  $\beta$ -hairpin; **S1**, sensor 1; **RF**, arginine finger motif; **PS2**, presensor 2 insertion; **S2**, sensor 2. (B) Ribbon diagram of the near full-length structure of *SsoMCM* (PDB-3F9V). Domains and linkers are colored with the same code-color used in panel A. (C) Side and top views of a hexameric atomic model of *SsoMCM* lacking the C-terminal domain. Black dotted lines contour one of the monomers. (Figures of panel C were adapted from (Brewster et al. 2008)).

MCM proteins can form ring-shaped complexes (Figure 7C) with ATPase and helicase activities, showing a DNA unwinding and translocation polarity in the 3'→5' direction (reviewed in (Costa & Onesti 2009; Bochman & Schwacha 2009; Brewster & Chen 2010; Bell & Botchan 2013)). Different models have been proposed to explain how MCM effects the DNA unwinding (Brewster & Chen 2010), been the steric exclusion model and the side-channel extrusion model the most accepted ones (Bochman & Schwacha 2009) (Figure 8). Fluorescence resonance energy transfer (FRET)

studies in archaea *Sso*MCMs support a steric exclusion model in which one strand of DNA (3'-tail) go all along the helicase central channel, while the second strand (5'-tail) is displaced out ahead by the helicase (McGeoch et al. 2005; Rothenberg et al. 2007), albeit maintaining a weak and dynamic interaction with the helicase surface (Graham et al. 2011). Less clear is the helicase mechanism of eukaryotic MCM complexes, and the actual path that follows the 5'-tail DNA strand remains to be clarified.

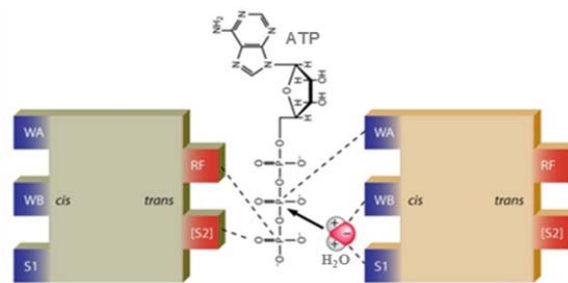


**Figure 8. Proposed DNA unwinding modes by MCM helicase.**(A) Schematic representation of a MCM hexameric helicase, showing a longitudinal cut, parallel to the central channel. The respective labeled  $\beta$ -hairpins are represented by short black bars. (B) Steric exclusion model and (C) side-channel extrusion model, proposed to explain the helicase activity of MCM. Light gray and dark gray wavy lines represent the DNA leading-strand and lagging-strand templates, respectively. Red dashed arrows indicate the movement directionality of the helicase (Figures adapted from (Brewster et al. 2008)).

Mostly found as hexameric single or double rings, the asymmetries and conformational changes displayed by some MCM helicases suggest a certain degree in both intra- and inter-domain flexibility (Costa et al. 2006a). Recently, the structure of a full-length *Mth*MCM dodecameric complex was solved by using a combination of small-angle neutron scattering (SANS) and all-atom molecular modeling techniques, where the free movement of the AAA+ core domains could be observed. The flexible loop connecting the AAA+ domain with the WH motif suggest that the C-terminus is very flexible (Krueger et al. 2014). The observed asymmetric positioning between the two tiers of the MCM2-7 ring in the EM structure of the *Dme*CMG-DNA-ATP $\gamma$ S complex reveal that the N-terminus and AAA+ regions can behave conformationally independent of each other (Costa et al. 2014).

Similar to other AAA+ proteins, the binding of ATP to MCM helicases and its subsequent hydrolysis is thought to be carried out at the interface between two consecutive protomers (Figure 9). One of the monomers provides the motifs for ATP binding in *cis* (Walker A, -WA-) and positioning of the

nucleophilic water molecule (Walker B -WB- and sensor 1 -S1-), while the adjacent monomer contributes with the *trans* motifs (arginine finger -RF- and sensor 2 -S2-) that makes contact with the  $\gamma$ -phosphate of the same nucleotide. (Iyer et al. 2004; Erzberger & Berger 2006; Brewster et al. 2008). In this way, the RF from one subunit catalyzes the hydrolysis of the ATP bound to the next subunit. The way S2 function in other AAA+ proteins is in *cis*, but in MCM proteins the insertion of presensor 2 (PS2) allows S2 to act in *trans* (Brewster et al. 2008; Erzberger & Berger 2006). The sequence of PS2 notably diverges among the six eukaryotic MCM proteins (Bochman & Schwacha 2009). A conserved region, called the allosteric communication loop (ACL), facilitates the coordination of the N- and C-terminal motifs of adjacent subunits that are involved in the binding and hydrolysis of ATP (Barry et al. 2009). Near the N-terminus of the protein, there is a zinc finger motif (Zn finger) that is involved in the folding stabilization of the N-terminal domain. Another remarkable feature is the presence of four  $\beta$ -hairpins that protrude from the surface (Brewster et al. 2008). Starting from the amino terminus of the sequence, the first hairpin (hp) is the N-terminal hp (NT-hp), followed by the exterior-hp (EXT-hp), the helix 2 insert hp (H2I-hp) and, finally, the presensor 1 hp (PS1-hp). The  $\beta$ -hairpins are involved in DNA interactions (Brewster & Chen 2010; Bell & Botchan 2013).



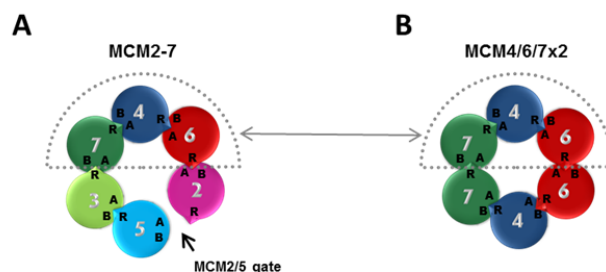
**Figure 9. Active sites in AAA<sup>+</sup> domains are formed at the interface between adjacent protomers.** In MCM proteins, the motifs WA, S1 and WB act in *cis*, on the other hand RF and S2 motifs act in *trans* to hydrolyze an ATP molecule (Figure adapted from (Bochman & Schwacha 2009)).

### 1.4.3. The eukaryotic MCMs

In eukaryotes, all six MCM paralogous proteins, MCM2 to MCM7 (MCMs 2-7), are required for replication of DNA. The heterohexameric MCM2-7 helicase has been purified isolated or as part of bigger complexes (the pre-replicative complex, pre-RC (Sun et al. 2014); the intermediate ORC-Cdc6-Cdt1\_MCM2-7 complex, OCCM (Sun et al. 2013); or as part of the holo-helicase complex Cdc45-(MCM2-7)-GINS, CMG (Costa et al. 2011; Costa et al. 2014)). The previously mentioned

evidences pointed to the MCM2-7 heterohexamer as the replicative helicase in eukaryotes (Chong et al. 1996; Kearsley & Labib 1998; Kearsley et al. 1996; Mcm 1994). However, the activation of the MCM2-7 helicase depends on its phosphorylation by DDK and CDK kinases (Remus & Diffley 2009), followed by its association with GINS and Cdc45 (Ilves et al. 2010). *In vitro* helicase activity of the isolated MCM2-7 complex has been observed in budding yeast and, more recently, in human, but only under specific buffer anion conditions (requiring the addition of glutamate and/or acetate) that seems to induce the closure of the MCM2/5 gate (Bochman & Schwacha 2008; Hesketh et al. 2015).

The subunit organization within the MCM2-7 ring complex was experimentally established as MCM5→3→7→4→6→2 (within each interacting pair, the subunits to the left of the arrow provide the motifs in *trans*, while the subunits to the right contribute with the motifs in *cis*) (Ishimi 1997a; Schwacha & Bell 2001; Davey et al. 2003; Costa et al. 2011; Sun et al. 2013; Sun et al. 2014), where the interaction of MCM2 with MCM5 forms the close/opening ring interface that has been observed as a gap by EM (Figure 10A) (Costa et al. 2011; Lyubimov et al. 2012). This gap could work as an ATP-modulated gate for ssDNA to pass through, so it can access the central channel of the ring (Bochman & Schwacha 2008; Costa et al. 2011; Sun et al. 2013).



**Figure 10. Architecture of MCM helicases.** Diagrams for the protomer organization of (A) MCM2-7 and (B) the proposed putative protomer arrangement in the MCM4/6/7x2 hexamer, based on the known physical interactions between the different MCM proteins. Grey dotted semicircle figure highlights the MCM trimer composed by MCM4, 6 and 7, which would putatively share a similar organization in both helicases. The position of the motifs: Walker A, Walker B, and arginine finger, are indicated by capital letters “A”, “B”, and “R”, respectively.

In addition to the MCM2-7 heterohexamer, it has been found eukaryote MCMs proteins can be associated in several other combinations, giving rise to dimers (Tye 1996; Burkhart et al. 1995; Sherman & Forsburg 1998), trimers (Musahl et al. 1995), tetramers (Ichinose 1996), hexamers (Tye 1996; Richter & Knippers 1997) and double-hexamers (Evryn et al. 2009; Remus et al. 2009). Some of the MCM polypeptides are also able to self-interact (Tye 1996; Yabuta et al. 2003; Tran et al. 2010). Deletion and mutational studies were carried out to better understand the role of each of

the six MCMs proteins within the different subcomplexes. It was shown that each of the six types of MCM proteins contributes differentially to the biophysical and biochemical properties of the complex, involving the interaction with other different proteins/complexes (You & Masai 2008; Wei et al. 2010; Liu et al. 2012; Costa et al. 2011; Sun et al. 2013; Nguyen et al. 2012), oligomerization and stabilization of the hexameric complex, ATP hydrolysis and, DNA binding and helicase activity (Lee & Hurwitz 2000; Ma et al. 2010; Xu et al. 2013; Schwacha & Bell 2001; Bochman et al. 2008; Davey et al. 2003; Bochman & Schwacha 2007; Ying & Gautier 2005). From the above, it was concluded that the MCM4/6/7 heterotrimer forms the catalytic core of the MCM2-7 helicase, whereas MCM2, 3 and 5 play a negative regulatory role (Ishimi 1997a; Lee & Hurwitz 2000; Sato et al. 2000; Schwacha & Bell 2001). In agreement with all these results, the different oligomers present particular properties and levels of activity.

#### **1.4.4. MCM4/6/7**

Aside from the MCM2-7 helicase, significant effort has been invested in the study of the biochemical properties of the MCM4/6/7 helicase, an hexameric ring complex formed by two MCM4/6/7 heterotrimers (hereinafter alternatively referred to as MCM4/6/7x2) which, in contrast to the MCM2-7 hexamer, shows an intrinsically helicase activity in a persistent manner (You et al. 1999). MCM4/6/7x2 presents limited processivity (Ishimi 1997a; Lee & Hurwitz 2000; Lee & Hurwitz 2001), but a double heterohexameric complex (a dimer of MCM4/6/7x2), that assemble on forked DNA substrate, has been shown to be a processive helicase (Lee & Hurwitz 2001). MCM4/6/7x2 has been isolated from diverse systems, and both its ATPase and DNA helicase activities are conserved across species (Sherman & Forsburg 1998; Prokhorova & Blow 2000; Holthoff 1998; Davey et al. 2003; Coué et al. 1998; Ichinose 1996; Ishimi 1997b; You et al. 1999; Lee & Hurwitz 2000; Sato et al. 2000; Lee & Hurwitz 2001; You et al. 2002; Biswas-Fiss et al. 2005; You & Masai 2005; Bochman & Schwacha 2007; You & Masai 2008; Xu et al. 2013), suggesting that this complex may be implicated in important cellular processes. Although currently the biological role of MCM4/6/7x2 remains unknown, it is not surprising that dysregulation or mutations of its components are associated with several clinical conditions, including cancer (Watanabe et al. 2012; Gineau et al. 2012; Hughes et al. 2012; T. Zheng et al. 2014).

Regarding the catalytic role of the MCM4/6/7 trimer subcomplex and the regulatory role of the MCM2, 3 and 5 proteins, it was observed that MCM4/6/7x2 showed both a better dsDNA binding performance and a higher ssDNA association rate when compared to MCM2-7 (Bochman & Schwacha 2007). The authors of that work proposed that these differences are a consequence of conformational changes in a putative regulatory site in the MCM2/5 gate. Whether there could exist a similar gate in MCM4/6/7x2 remains an important question to be answer. Another observation also supporting the idea of the regulatory action of MCM2, 3 and 5, was that incubation of either MCM2 or MCM3/5 with MCM4/6/7x2, causes a disassembly of the dimeric complex of MCM4/6/7 along with the loss of DNA helicase activity (Lee & Hurwitz 2000).

All protomers surely contribute in some general way to the proper function of the helicase, but some principal roles have been associated to each of the three types of subunits, MCM4, 6 and 7. Phylogenetic analysis suggest that MCM4 is the most ancient form of eukaryotic MCMs (Kearsey & Labib 1998). WA and WB motifs of MCM4 play key roles in the ssDNA binding activity of the helicase (Bochman & Schwacha 2007; Gómez et al. 2002; You et al. 1999) It has been suggested that in fission yeast, the C-terminus of MCM4 blocks the DNA unwinding activity of MCM2-7 once there is enough ssDNA at the stalled replication fork for checkpoint activation. On the other hand, the MCM4 C-terminal part is also required for efficient resumption of fork progression during recovery from the replication block (Nitani et al. 2008). Mutations in the Zn-finger of MCM4 promote the dissociation of MCM4/6/7x2 into its trimers (You et al. 2002), suggesting that dimerization of MCM4/6/7 is somehow directed by the Zn-finger of the MCM4 subunits. The above observation suggest that MCM4 should presents at least two different domains with particular oligomerization roles, one for the formation of the MCM4/6/7 trimer and the other, within the N-terminus, involved in the dimerization of the trimers themselves (You et al. 2002). In turn, MCM6 plays a major role in binding of ATP, and MCM7 is required for hydrolysis of ATP and helicase activity (You et al. 1999). Based on these considerations, a simplified model for the DNA unwinding mechanism of MCM4/6/7x2 was proposed (You et al. 2002), in which there is a sequential participation of the protomers and each protein class (e.i. MCM4, 6 or 7) makes a major-contribution to a different task. First, DNA binds to the helicase by interacting with MCM4, then, an ATP molecule binds to MCM6 causing a conformational change in the complex that would facilitate the subsequent hydrolysis of the nucleotide by MCM7. A tight communication and coordinated work between all subunits in MCM4/6/7x2 is essential in order to fully carry out its helicase activity.

Usually isolated as the hexameric complex MCM4/6/7x2 (Ishimi 1997b; Lee & Hurwitz 2000; Yabuta et al. 2003; Ishimi 1997a; You et al. 1999; Sato et al. 2000; Ma et al. 2010; Su et al. 1996), proteins MCM4, MCM2 and MCM7 have been also purified as a stable single MCM4/6/7 heterotrimer (Musahl et al. 1995; Sherman et al. 1998; Ma et al. 2010; Ichinose 1996; Ishimi 1997b). Although it is not clear which are all the factors directing the oligomerization state of MCM4/6/7, it seems that ATP promotes the formation of higher order oligomers (You & Masai 2008). A study with *S.cerevisiae* proteins, in which MCM4, 6 and 7 were individually purified and then mixed together, showed a shift from the trimer to the hexamer form after incubation with ATP, a cofactor required for helicase activity. Incubation with ATP $\gamma$ S had a similar oligomerization effect to ATP, and it was also found that ADP fosters the formation of the hexamers but to a lesser extent (Ma et al. 2010). In agreement with the idea that the active complex is an hexamer and not a single trimer, a mutational study showed that the RF motifs of the three proteins are required for a full ATPase activity (Ma et al. 2010).

Despite certain disagreement in the literature regarding some of the direct pair interactions between eukaryote MCM subunits (including some self-interactions), most of the experimental reports support an architecture model for the MCM4/6/7x2 hexameric ring in which the two MCM4/6/7 trimers face each other in a way that the two MCM6 subunits binds together (MCM6-MCM6) as well as the two MCM7 proteins (MCM7-MCM7), and each of the MCM4 is located between the two different homotypic pairs, that is, MCM4-6-6-4-7-7 (Yabuta et al. 2003; Yu et al. 2004; Xu et al. 2013), (Figure 10. B). To the best of our knowledge, there is only one study where an alternative model is proposed, suggesting a direct binding of MCM6 and MCM7 (Ma et al. 2010).

To date there are plenty of studies focused on the biochemical properties of the hexameric helicase MCM4/6/7x2. Contrarily, at the structural level there are only a few works showing some 2D images of the closed ring-shaped complex, so many questions remain yet to be answered. In order to better understand the structural and functional organization of MCM4/6/7x2, we have taken advantage of the similarities that exist between this helicase and MCM2-7. In this work we present the structural study of diverse MCM4/6/7 complexes as an important contribution to continue with the process of elucidating the functional mechanism of this macromolecular motor. Our data suggests how MCM4/6/7 trimers dimerize, and we propose, based on a number of novel

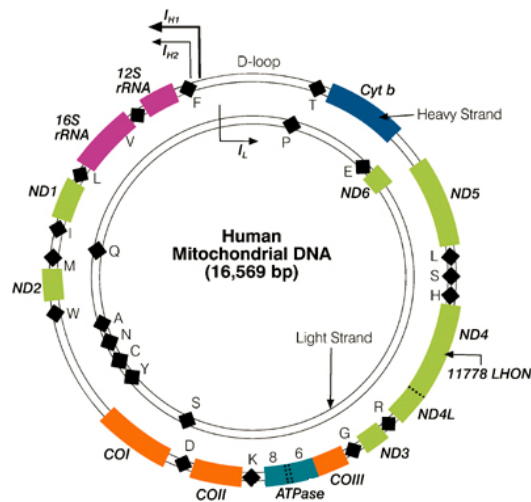
and revealing conformational changes of the hexameric complex, a sequence that would give rise to a putatively active MCM4/6/7x2 structure, able to perform DNA binding and helicase activities.

## **1.5. Biological system: TWINKLE, the human mitochondrial DNA helicase**

Twinkle, the mitochondrial replicative DNA helicase, is a ring-shaped enzyme (Ziebarth et al. 2010; Fernández-Millán et al. 2015) required for strand separation at the replication fork of mitochondrial DNA (mtDNA), making it essential for mtDNA maintenance and regulation of its copy number (Tynismaa et al. 2004). The gene C10orf2 on chromosome 10q24 encoding the Twinkle protein was originally identified through its link with a neuromuscular disorder associated with mtDNA damage (Spelbrink et al. 2001), a clinical syndrome known as autosomal dominant progressive external ophthalmoplegia (adPEO), which is associated with mutations in the Twinkle gene (Fratteer et al. 2010) and multiple mtDNA deletions (Suomalainen et al. 1997).

Mitochondria are popularly known as the “powerhouse of the cell” because they are responsible for producing most of the ATP used as source of chemical energy for the cells (Scheffler 2007), which requires an intact and functional mitochondrial genome. These organelles had also been found to be tightly involved in the execution of diverse cellular events, such as the regulation of metabolism, cell-cycle control, development, antiviral responses and cell death (Mcbride & Neuspiel 2006; Tower 2015). The homeostasis and adaptability of the cell depend on the qualitative and quantitative status of the mitochondrial genome. Mitochondria contain their own genome. Normally, mtDNA is organized as a circular, covalently closed, double-stranded DNA (dsDNA), which is thought to be derived from the  $\alpha$ -proteobacterium that eventually became the mitochondrion (Gray et al. 1999; Burger et al. 2003). The human mtDNA molecule (and that of mammals in general) is around 15.5 kb, and its two strands are differentiated by their nucleotide content, with a guanine-rich strand referred to as the heavy strand (H-strand) and a cytosine-rich strand referred to as the light strand (L-strand). Together, both H-strand and L-strand contain 37 genes (Figure 11), all of them essential for mitochondrial function (Kyriakouli et al. 2008). Alterations in the mtDNA molecule or its copy number are associated with a number of clinical conditions (Mao & Holt 2009; Nogueira et al. 2014; Spelbrink et al. 2001; Suomalainen et al. 1997).





**Figure 11. The human mitochondrial genome.** This small (16,569 bp) genome is almost completely transcribed from both strands, initiating from either one of two promoters (IH1, IH2) on the Heavy (H-) strand, or the single promoter (IL) on the Light (L-) strand. All of these promoters and elements involved in replication initiation are found in the displacement (D-) loop, the only major non-coding region. The genome encodes 22 mt-tRNA (black diamonds), 2 mt-rRNA genes (fuchsia) and 13 protein-coding genes (olive, ND1-6 encoding members of NADH:ubiquinone oxido-reductase; blue, Cytb encoding apocytochrome *b* of ubiquinol:cytochrome *c* oxido-reductase; orange, COI-III encoding members of cytochrome *c* oxidase; aquamarine, ATPase 6, 8 encoding two members of Fo-F1 ATP synthase). Single letter code is given for each mt-tRNA-encoding gene (Figure and legend modified from (Kyriakouli et al. 2008)).

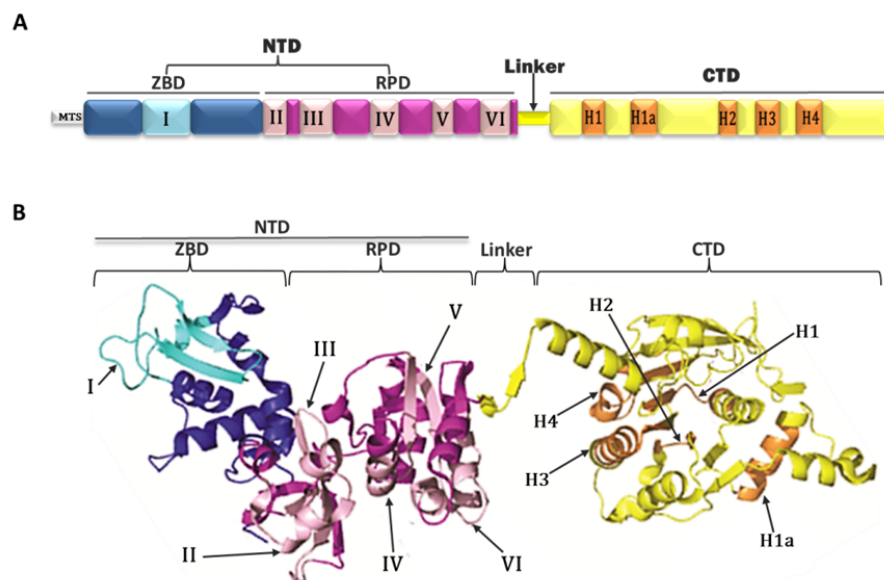
### 1.5.1. The mtDNA replisome

The proteins that constitute the minimal mtDNA replisome, a protein complex required for the maintenance of mtDNA (Korhonen et al. 2004; Spelbrink et al. 2001; Wanrooij & Falkenberg 2010; Tynismaa et al. 2004), are all encoded by nuclear genes and form a small set that includes Twinkle helicase, both the catalytic ( $\alpha$ ) and accessory ( $\beta$ ) subunits of DNA polymerase  $\gamma$  (pol  $\gamma$ ), the mitochondrial single-stranded DNA-binding protein (mtSSB) (Korhonen et al. 2004), and the mitochondrial RNA polymerase (mtRNAPol), which functions as the mtDNA primase (Wanrooij et al. 2010; Shi et al. 2008; Reyes et al. 2013; Pham et al. 2006). Interestingly, it seems that frequently along mitochondrial history, functional redirection of ancestral proteins has happened, and such is the case for mtDNA replisome components, where Twinkle, the catalytic subunit pol  $\gamma$ - $\alpha$  and mtRNAPol share ancestry with the gp4 primase-helicase, the gp5 DNA polymerase and the gp1 RNA polymerase from the bacteriophage T7, respectively; whereas mtSSB is a homologue of the homotetrameric SSBs from eubacteria (Shutt & Gray 2006a), and the accessory subunit pol  $\gamma$ - $\beta$  evolved from the class II aminoacyl-tRNA synthetases of eubacteria (Uhn 1999; Fan et al. 2006).

## 1.5.2. Twinkle architecture: Known and proposed functions

Sequence similarities between the mitochondrial replicative DNA helicase and the bifunctional primase-helicase of bacteriophage T7 gen 4 (T7 gp4) (Spelbrink et al. 2001; Shutt & Gray 2006b), together with additional studies showing related physical features of their architectures (Ziebarth et al. 2007), assigned Twinkle to the superfamily 4 (SF4) of helicases. Both proteins present a general structure where a flexible linker connects distinct N-terminal and C-terminal domains (

Figure 12).



**Figure 12. General structural features of the replicative mtDNA helicase Twinkle. (A)** Linear schematic representation of the structural motifs, which are labeled and colored as follows: The N-terminal domain (NTD) is formed by two smaller domains, the zinc binding domain (ZBD, in cyan and blue) and the RNA polymerase domain (RPD, in pink and fuchsia); Motifs from the NTD (I, II, III, IV, V and VI) and those from the CTD (H1, H1a, H2, H3 and H4). **(B)** Ribbon diagram of an homology model of Twinkle, showing the approximate localization of the structural motifs, where the colors and nomenclature showed in panel A are maintained (Figure from panel B has been adapted from (Fernández-Millán et al. 2015)).

Twinkle has a NTPase-dependent unwinding activity with a 5' to 3' translocation directionality (Korhonen et al. 2003), as well as a tendency to oligomerize into toroidal complexes (Ziebarth et

al. 2010; Fernández-Millán et al. 2015), which are features showed by the SF4 members (Singleton et al. 2007). While it is true that Twinkle can oligomerize in a NTP-independent way, its oligomeric status (hexamerization and/or heptamerization) is affected by different ionic conditions and cofactors (Ziebarth et al. 2010; Farge et al. 2008; Fernández-Millán et al. 2015); for example, it has been observed that high salt conditions (Ziebarth et al. 2010; Fernández-Millán et al. 2015) or addition of MgUTP (Sen et al. 2012) favors the presence of hexamers over higher oligomers. In a similar way, T7 gp4 forms both hexamers and heptamers (Toth et al. 2003; Crampton et al. 2006), where the first oligomeric state predominates in presence of triphosphate, while the use of diphosphate produces a higher proportion of heptamers over hexamers (Crampton et al. 2006). In the presence of DNA and nucleotide (both dTTP and dTDP), T7 gp4 is detected as hexamers that bind ssDNA. On the other hand, once formed, T7 gp4 heptamers are not able to efficiently bind either ssDNA or dsDNA (Crampton et al. 2006). Twinkle interacts with a variety of DNA substrates, even in the absence of cofactors, including both linear and circular ssDNA and linear dsDNA (Jemt et al. 2011), and shows higher binding affinity for dsDNA than for ssDNA (Sen et al. 2012; Ziebarth et al. 2010; Farge et al. 2008; Longley et al. 2010), while T7 gp4 binds preferably to ssDNA over dsDNA, requiring cofactors for binding ssDNA (Matson et al. 1985; Hingorani & Patel 1993). Currently, three different DNA binding sites in Twinkle have been detected; one for dsDNA and two for ssDNA, where the latter sites are putatively localized one in the central channel and the other on the external surface of the helicase ring (Sen et al. 2012). Similarly to T7gp4, Twinkle requires substrates with both a single-stranded 5'-DNA loading site and a short 3'-tail (which resembles the structure of a replication fork) to efficiently initiate the unwinding of the DNA duplex (Korhonen et al. 2003; Patel & Picha 2000).

**The C-terminal domain (CTD) of Twinkle shows high homology with the equivalent region in T7gp4 (Spelbrink et al. 2001; Shutt & Gray 2006b), corresponding in both cases to their helicase domain and an important part for oligomerization of these enzymes, where the conserved catalytic motifs Walker A and Walker B (within SF4 are known as motifs H1 and H2, respectively) found in helicases and translocases are housed (Singleton et al. 2007). The three additional motifs H1a, H3 and H4 present at the CTD of Twinkle, are characteristic signatures of SF4 (Singleton et al. 2007) (**

Figure 12). In Twinkle, the linker region is important for oligomerization (Guo et al. 1999), and several adPEO disease-related mutations reside within this region (Ji et al. 2014; Dündar et al. 2012). Regarding the N-terminal domain (NTD) of the mtDNA helicase, in mammals, this region of the protein has diverged dramatically from that of other metazoans, losing most of the residues present in T7gp4 that are essential for the primase activity so that, in mammals, Twinkle doesn't show this function (Shutt & Gray 2006b). In particular, human Twinkle lacks three of the four



et al. 2003; Keck et al. 2000; Podobnik et al. 2000). The ZBD of the primase binds to ssDNA and recognizes specific nucleotide triplet patterns from priming sites in the ssDNA template (Lee et al. 2012), which subsequently passes to the RPD, where binding and the catalytic condensation of NTPs occurs; RPD is required for the template-directed coupling of the NTPs (Kuchta & Stengel 2010).

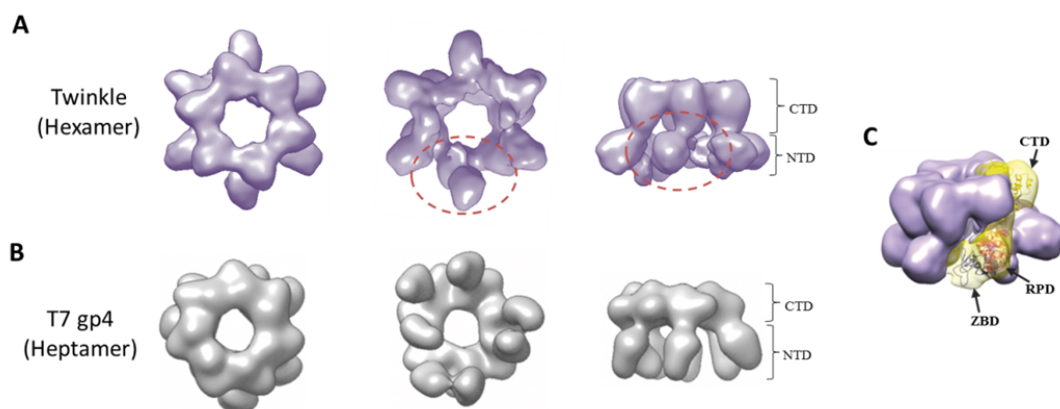
Twinkle mutants lacking its NTD can only support the synthesis of short DNA products (~ 1 kb) (Farge et al. 2008), so despite the fact that the NTD of Twinkle is not absolutely required to carry out the most basic function of the mtDNA replisome, it is important for the efficient ssDNA binding and dsDNA unwinding activities of the enzyme, which are required for formation of long DNA products (Farge et al. 2008). Furthermore, it seems that the NTD also contributes to the proper oligomerization of the helicase ring complex (Farge et al. 2008). Other authors have proposed possible additional roles of the NTD, such as stimulation of primer utilization (Shutt & Gray 2006b). Removal of ZBD in human Twinkle produces a decrement in ssDNA binding compared with that of the full-length protein, affecting also the ssDNA-stimulation of the ATPase activity of the enzyme (Farge et al. 2008).

On the other hand, the lack of the ZBD did not affect the protein binding to dsDNA, suggesting that the ZBD contributes specifically to ssDNA binding. It has been proposed that, in T7gp4, the initial contact of a DNA binding site at its primase domain with a DNA molecule would produce a transient opening of the ring, allowing the access of DNA to the central channel, after which the gap is closed again (Ahnert et al. 2000; Farge et al. 2008; O'Shea, V.L. and Berger 2014). An alternative mechanism for DNA loading of T7 gp4 has been proposed based on the observation that it forms both hexameric and heptameric complexes (Patelt 1995; Toth et al. 2003; Crampton et al. 2006). It is suggested that T7 gp4 heptamers could lose one protomer, which would lead to an opening in the resulting hexamer (Crampton et al. 2006). Since Twinkle forms both hexameric and heptameric toroidal complexes (Ziebarth et al. 2010; Fernández-Millán et al. 2015) and is able to load by its own on the DNA, similar mechanism to the ones proposed for T7 gp4 could be also considered in the case of Twinkle as possible strategies to create a gap that allows the helicase ring to load onto DNA without assistance from additional factors. The assembly into different oligomeric ring-shaped complexes is a feature showed by a number of AAA<sup>+</sup> ATPases, including some members of superfamily 6 (SF6) of helicases, as for example MCM helicases from different

archaea and eukaryotic organisms (Yu et al. 2002; Fletcher et al. 2003; Adachi et al. 1997; Lee & Hurwitz 2001).

Interestingly, in addition to the already known NTPase-dependent helicase activity of Twinkle, the antagonistic function, an efficient ability of annealing two complementary ssDNAs, has been detected to occur both with and without cofactors (Sen et al. 2012). Indeed, it was observed that under certain conditions (absence of gp2.5 or ssDNA trap) the annealing activity competes with the unwinding activity. The annealing activity of Twinkle is sensitive to the ionic strength of the environment, being inhibited at concentrations higher than 150mM NaCl (Sen et al. 2012), which can be partially explained by the fact that protein-DNA interactions start being disrupted at high salt conditions, and are abolished at 420mM NaCl (Yakovchuk et al. 2006).

Recently, the 3D structure of a hexameric complex of Twinkle that was isolated under high salt conditions (1M NaCl) and stabilized by chemical fixation with the GraFix method, was solved by cryo-electron microscopy (EM) (the density map is still not available on the Electron Microscopy Data Bank, EMD) (Fernández-Millán et al. 2015) (Figure 14A). The 3D map shows a compact, two-tier ring conformation, which resembles the general shape of a T7 gp4 heptamer complex which naturally lacks the ZBDs (PDB-1Q57) (Toth et al. 2003) (Figure 14B).



**Figure 14. Ring-shape helicase 3D structures.** (A) EM map of a Twinkle hexamer (Fernández-Millán et al. 2015) and (B) the structure of an heptamer of T7 gp4 (PDB-1Q57, filtered at 30Å). From left to right the CTD-, NTD- and lateral-views are shown, respectively. A red dashed line highlights a particular NTD in the structure of Twinkle that is isolated from the rest of NTDs, which are interconnected through contacts with the ZBD of the neighbouring monomer. (C) Tilted view from Twinkle EM map, with one segmented subunit and a homology model fitted in (Figure from panel C was adapted from (Fernández-Millán et al. 2015).

The proposed flexible fitting of an atomic homology model of the full-length Twinkle protein (a chimeric model based on predicted secondary structure similarities between Twinkle and both the CTD of T7 gp4 and the NTD of DnaG) into the cryo-EM map (Figure 14C) accommodates the CTDs into the density of a closed six fold symmetry (C6) ring, showing similar interfaces than the ones observed in both the T7 gp4 hexamer (PDB-1E0J and 1E0K) (Singleton et al. 2000) and heptamer (PDB-1Q57) structures. The RPDs are fitted in the second floor of the map, which unlike the CTDs layer, shows a slightly asymmetric ring. Four smaller extra densities, located in a level slightly lower than the RPD layer, are assigned to some of the ZBDs, being asymmetrically positioned in a way that interconnects five of the RPDs, while the last RPD remains disconnected from the rest. The lack of two ZBDs in the density map is attributed to a higher flexibility of these domains. The different interface contacts within this hexamer of Twinkle are: *in cis*, ZBD/RPD and *in trans*, both ZBD/RPD and RPD/CTD. The above is in agreement with studies on the related SF4 proteins DnaG (Corn et al. 2005) and T7 gp4 (Lee & Richardson 2002), where ZBD/RPD interactions, both *in cis* and *in trans*, have been detected; additionally, contacts of RPD and RPD-CTD linker with the CTD of the neighbouring protomer have been observed in T7 gp4 (Singleton et al. 2000). In the same study of Twinkle protein (Fernández-Millán et al. 2015), negative staining (NS) EM examination of the high-salt, unfixed sample, showed both hexamers and heptamers with a radial arrangement of the subunits where a central pore was formed by the interacting edge of the protomers, while the opposite side of all the subunits remained extended to the outside with variable orientations. Small angle X-ray scattering (SAXS) analysis of the particles in solution at 1.5M NaCl showed a dynamic sample with a mixture of extended conformations of both hexamers and heptamers (with a calculated ratio 3/2), where the protein presented high flexibility at the inter-domain connectors.

Aside from observing that Twinkle forms both flexible hexamers and heptamers, It is also known that low salt concentrations result in protein aggregation (Ziebarth et al. 2010; Fernández-Millán et al. 2015), which have made difficult its structural analysis under more physiological, lower ionic strength conditions. Obviously, much more work is still required to elucidate the structural mechanisms that control the different functions of this essential enzyme for cell biogenesis. Here, we present an EM analysis of the human Twinkle, by comparing the full-length protein (mature Twinkle includes residues 43-684) with a truncate construct lacking the ZBD ( $\Delta$ ZBD protein

includes residues 147-684), under salt concentrations between 250-330mM NaCl. Our comparative analysis between the full-length and N-terminal truncate proteins provides direct visual evidence of the fundamental role of the ZBD on the conformational behavior of the ring complexes. Indeed, deletion of the ZBD significantly restricts the capability of the enzyme to acquire the whole range of additional conformational movements showed by the WT, which we suggest would be very likely required for the correct function and interactions that must be normally carried out by Twinkle. We also report, for the first time, images of octameric and pentameric ring-like complexes of Twinkle which coexist in a small proportion with a majority population of the already reported hexameric and heptameric species. Our findings of structural details of Twinkle flexibility and the influence of the ZBD provide new insights to understand the mechanistic features displayed by the mitochondrial replicative DNA helicase in mammals.



## Chapter 2

# 2. Objectives

The general aim of this thesis is the structural characterization by transmission electron microscopy of several macromolecular complexes, such as the centrosomal human protein construct CPAP<sup>897-1338</sup> and the mouse and the human helicases MCM4/6/7 and Twinkle, respectively, being the three of them, highly dynamic and flexible complexes, able to form different oligomeric assemblies. In particular, the following objectives have been pursued:

1. Purification of human CPAP<sup>897-1338</sup> and isolation of different oligomeric complexes formed by this protein. This task has been followed by the structural characterization and the three dimensional reconstructions of the most stable assemblies.
2. Three dimensional reconstruction of different oligomeric and conformational states of mouse MCM4/6/7, and fitting of atomic structures into the obtained density maps for the putative localization of structural domains.
3. Two dimensional characterization of the structure and flexibility of multiple oligomeric complexes formed by human Twinkle, determining the effect of the zinc binding domain on the structural dynamics of the helicase by comparing the full-length protein with a N-terminal deletion mutant. Additionally, initial low resolution three dimensional maps of the most representative oligomeric and conformational states have been obtained.

## Chapter 3

# 3. Materials and methods

## 3.1. Obtaining the protein samples

### 3.1.1. CPAP<sup>897-1338</sup>

#### 3.1.1.1. Plasmid construct

The full-length human cDNA of CPAP (1338 residues) was kindly provided by Dr. Tang T. K., from the Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan (Hung et al. 2000). The C-terminal residues 897 to 1338 of CPAP were amplified by PCR using the sense primer STO3203 that incorporates the BamHI restriction site (5'-GCGGATCCCCTGGTGACAATGCTCG -3'), and the STO3204 antisense primer containing BsrGI restriction site (5'-CGGTGTACATTACAGCTCCGTGTCCATTG -3'). To create a N-terminal 6xHis-tagged CPAP<sup>897-1338</sup> recombinant DNA construct (His-CPAP<sup>897-1338</sup>), the amplified fragment was cloned into the BamHI-BsrGI site of the pST66Trc2-His expression vector (this is a pET3a modified plasmid created and kindly provided by Dr. Song Tan, from the Center for Gene Regulation, Pennsylvania State University, State College, PA. EE. UU.), which allowed the incorporation of a six histidine tag (His-tag) to the N-terminus of the CPAP<sup>897-1338</sup> protein; in this way the recombinant expression plasmid pST66Tcr2-HisCPAP<sup>897-1338</sup> was obtained. The strain *E. coli* DH5 $\alpha$  was used as host for the cloning work. The isolation of the recombinant plasmid was performed using standard methods (Green & Sambrook 2012).

#### 3.1.1.2. Protein expression

*E. coli* BL21(DE3)pLysS cells were transformed with the pST66Tcr2-HisCPAP<sup>897-1338</sup> recombinant plasmid, and protein expression was induced adding 0.2mM IPTG followed by an incubation at 23°C/16h. Solubility tests showed that CPAP<sup>897-1338</sup> protein is highly insoluble in low salt buffers, so

it was purified and maintained most of the time in presence of 300mM NaCl, with the exception of a part of the protein sample that was dialyzed in buffer with 150mM NaCl to be posteriorly used for an ion exchange chromatography step. The cell pellet was resuspended in cold lysis buffer pH 7.4 (30mM Na<sub>2</sub>HPO<sub>4</sub>, 20mM NaH<sub>2</sub>PO<sub>4</sub>, 300mM NaCl, 10% v/v Glycerol, 0.1% Triton X-100, 25mM Imidazole, 10µg/ml DNase A, 10µg/ml RNase I, 0.5mg/ml Lysozyme, 1X Protease Inhibition Cocktail EDTA-free. Sigma) and lysed by French Press.

### **3.1.1.3. Protein purification by immobilized metal affinity chromatography (IMAC)**

After centrifugation (40.000rpm/45min) of the cell lysate, the soluble His-tagged protein in the clarified supernatant was purified by immobilized metal affinity chromatography (IMAC) with a Talon resin (Talon superflow. GE Healthcare) following the recommendations of the manufacturer. The collected fractions were analyzed by SDS-PAGE (Any KD Mini-PROTEAN TGX Gel. BIO-RAD), followed by staining with SimplyBlue SafeStain (Invitrogene); the CPAP identity of the observed protein bands was confirmed by Western Blot (against the His-tag) and Mass spectrometry (MALDI TOF/TOF MS) analysis. The IMAC purified fractions were pooled to be used in further chromatographic steps.

### **3.1.1.4. Size exclusion chromatography (SEC) analysis**

The IMAC purified protein was loaded into either preparative or analytical Superdex 200 columns equilibrated with cold buffer 20mM Hepes pH7.4, 300mM NaCl, 0.25mM TCEP and 10% Glycerol. The collected fractions were analyzed by SDS-PAGE (Any KD Mini-PROTEAN TGX Gel. BIO-RAD) stained with SimplyBlue SafeStain (Invitrogene). Some of the fractions were analyzed by negative staining transmission electron microscopy (NS-EM).

### **3.1.1.5. Ion exchange chromatography**

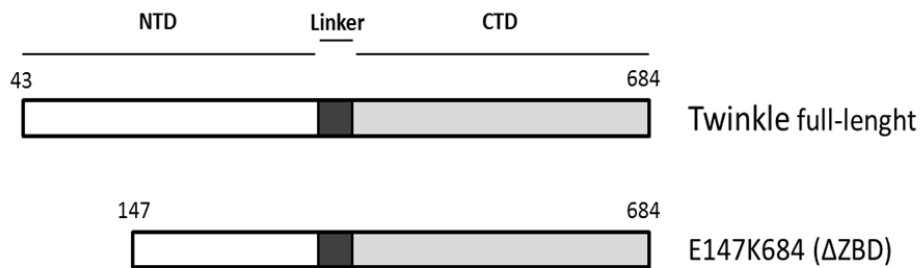
Part of the IMAC purified protein was pooled and dialyzed for 3h at 4°C against buffer 20mM Hepes pH7.4, 150mM NaCl, 0.25mM TCEP and 10% Glycerol, and then applied on an anion exchange HiTrap Q HP column (GE Healthcare Life Sciences). The collected fractions were analyzed by NS-EM

### **3.1.2. MCM4/6/7**

Mouse MCM4/6/7 protein sample was kindly provided by Professor Hisao Masai, from the Department of Genome Medicine, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo, 156-8506, Japan. The genes of MCM4-6xHis + MCM6 and MCM7-FLAG were coexpressed in a baculovirus system. MCM4/6/7 complexes were produced in High 5 insect cells coinfecting with recombinant baculoviruses, and purified by successive steps of metal affinity chromatography, anti-FLAG M2 antibody-agarose affinity chromatography and, finally, glycerol gradient sedimentation, as described in (You & Masai 2008). The sample buffer was 50mM Hepes-Na, pH 7.5, 50mM Na-Acetate, 1mM DTT, 0.5mM EDTA, 2mM Mg-Acetate, 1mM ATP, 0.01% Triton X-100 and ~1.5% Glycerol.

### **3.1.3. Twinkle**

Human mitochondrial replicative DNA helicase protein (Twinkle) samples were kindly provided by Prof. Laurie S. Kaguni, Distinguished Professor from Michigan State University, U.S.A and the Tampere University, Finland. The human full-length (residues 43-684) Twinkle-6xHis protein and an N-terminal deletion variant of Twinkle (6xHis-Del.E147K684) lacking the first 146 residues, where it is included the ZBD ( $\Delta$ ZBD) (Figure 15), were coexpressed in a baculovirus system. Both samples were produced in *Spodoptera frugiperda* cells coinfecting with recombinant baculoviruses, and purified by successive steps of metal affinity chromatography, heparin-sepharose affinity chromatography and ultracentrifugation in 12–30% glycerol gradients, as described in (Ziebarth et al. 2007). The purified samples were chromatographed on a Superdex 200 HR 10/30 column and analyzed by SDS-PAGE and Coomassie staining. Fractions corresponding with a mix of both hexameric and heptameric oligomers, as determined by size exclusion chromatography (SEC) analysis, were used for our EM analysis. The buffer composition of the protein samples was 35 mM Tris-HCl, pH 7.5, 2 mM  $\beta$ -mercaptoethanol, ~5-20% glycerol, with 330 mM NaCl or 250 mM NaCl for the full-length protein and the  $\Delta$ ZBD, respectively. For some of the experiments conducted along this work, we decrease the salt concentration in a range between 100-150 mM NaCl, in order to allow for the binding of cofactors (4 mM  $MgCl_2$  and 2 mM ATP $\gamma$ -S).



**Figure 15.** Schematic representation of Twinkle protein constructs. Mature full-length TWINKLE protein and Del.E147K684, a deletion construct lacking the ZBD ( $\Delta$ ZBD), where it is shown the position of the NTD (white), the NTD-CTD linker (Linker, in black) and the CTD (grey).

### 3.1.3.1. Solubility assay

**Decreasing salt concentration:** An initial sample reaction mixture containing 35 mM Tris-HCl, pH 7.5, 2 mM  $\beta$ -mercaptoethanol, 5% glycerol and 330 mM NaCl, was analyzed by negative staining electron microscopy (NS-EM). Separately, two other aliquots of the same reaction were directly diluted until reaching either 150 mM NaCl or 100 mM NaCl concentrations, while maintaining the same concentration of the rest of reagents and also adding 4 mM  $MgCl_2$  and 2 mM ATP $\gamma$ -S in both cases. The reaction samples were incubated for 10 min at 10 °C and subsequently analyzed by NS-EM.

**Increasing salt concentration:** An initial sample reaction mixture containing 35 mM Tris-HCl, pH 7.5, 2 mM  $\beta$ -mercaptoethanol, 5% glycerol, 4 mM  $MgCl_2$ , 2 mM ATP $\gamma$ -S and 100 mM NaCl, was analyzed by NS-EM. Separately, the salt concentration of another aliquot of the same reaction was increased until 330 mM NaCl. The reaction was incubated for 25 min at 25 °C and subsequently analyzed by NS-EM.

## **3.2. *In silico* structural analysis: Sequence alignment and atomic structural modeling**

### **3.2.1. CPAP<sup>897-1338</sup>**

Predictions of order/globularity and disorder regions of *Hs* CPAP (full-length) were done using the GlobProt2 server (Linding 2003). For CPAP<sup>897-1338</sup>, the secondary structure was predicted with Jpred3 (Cole et al. 2008); for the prediction of coiled-coil regions the web server COILS (Lupas et al. 1991) was used; and, finally, the web servers ProCoils (Mahrenholz et al. 2011) and LOGICOIL (Vincent et al. 2013) were employed to predict the oligomeric state probability of the coiled-coil sequence. The atomic modeling of the CPAP coiled-coil CC4/CC5, as well as the CCGb-linker region, was carried out with I-TASSER, a protein structure/function prediction server that incorporates multiple-threading alignments, *ab initio* modeling and refinement by iterative assembly simulation of template fragments (Zhang 2008; Roy et al. 2010). The crystallographic structure of the G-Box/TCP domain (PDB-4BXP), together with the structures modeled for the CC4/CC5 and CCGb-linker regions, were used for both, proposing a model of a CPAP<sup>897-1338</sup> monomer and making a tentative atomic fitting in the EM reconstructed volume of a CPAP<sup>897-1338</sup> oligomeric complex.

### **3.2.2. MCM4/6/7**

The sequences of a number of MCM homologs from different organism from both eukaryote and archaea were compared with the mouse proteins by making multiple sequence alignments with the program ClustalX2 (Larkin et al. 2007). Atomic models for each of the mouse proteins MCM4, MCM6 and MCM7 were generated using the structural prediction server I-TASSER (Roy et al. 2010; Zhang 2008).

### **3.3. Transmission electron microscopy**

Small, non-symmetrical flexible particles are difficult specimens for structural studies, although this heterogeneity may reflect the natural state required for their function, and so, study of these “hard” specimens is of high biological relevance.

#### **3.3.1. Sample preparation and image acquisition**

For negative staining (NS) of the samples, a drop of the protein solution was applied directly onto a glow-discharged EM grid (QUANTIFOIL. Formvar/Carbon. Cu 400 mesh grids), and allowed to be adsorbed on the grid surface (for few seconds or minutes, depending on protein concentration); then the drop was blotted with filter paper (Whatman grade No. 1) and the grid was washed by touching the surface with two consecutive drops of 0.75% (w/v) uranyl formate, blotting each time, and stained for 1min with one more drop of the same staining agent. Finally, the grid was blotted again and allowed to air dry before observation. Grids were examined in a JEOL JEM-1230 (accelerating voltage 100kV) electron microscope, and images recorded with a CCD camera ORIUS SC100 (4k x 2.7k pixel) at 40000x magnification, resulting in an image pixel size of 2.28 Å/pixel.

#### **3.3.2. Image preprocessing**

All image preprocessing, particle selection and two-dimensional (2D) analysis steps were carried out following the general workflow of the image processing package Xmipp3.1 (De la Rosa-Trevín et al. 2013). EM single particles were either manually or semi automatically selected.

Phase contrast is the dominant image formation mechanism for both stained and unstained specimens, and arises from the phase shift between scattered and unscattered electron waves (Amos et al. 1982). During EM data acquisition, the ideal projection image is convoluted with a point spread function (PSF) within the objective lens system. The PSF is a direct consequence of the lens aberrations, such as spherical (Cs), chromatic aberrations (Cc) and, astigmatism (as result of imperfect construction of the lens), among others, plus certain degree of defocus combined with a small percentage of amplitude contrast. The Fourier Transform of the PSF is the CTF (Contrast Transfer Function), which has a very characteristic pattern of cylindrical rings known as Thon-rings. A more in-depth discussion of the CTF and its analytical expression can be found in (Zhu & Frank 1997) and (Amos et al. 1982). Naturally, the blurring effect of the CTF must be

corrected, at least partially, before doing any further processing of the images. CTF estimation and correction was done with Xmipp3.1. CTF corrected images were downsampled by a factor of 2, reaching a final sampling rate of 4.56 Å/pixel. The selected particles were extracted and the images were normalized for its further processing.

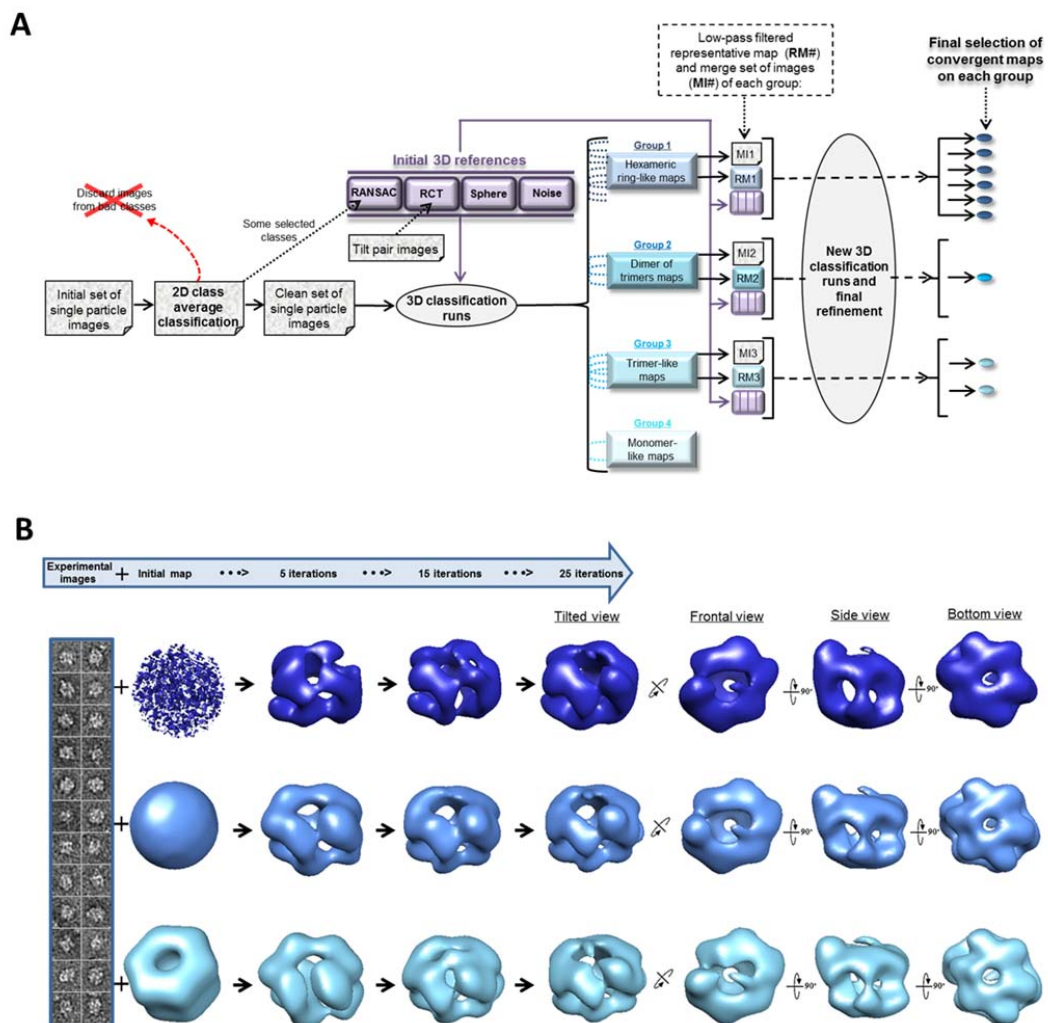
Since electron micrograph of biological particles are very noisy, the signal-to-noise ratio must be improved by performing the classification and averaging of large numbers of similar individual projections. Here, 2D class averages of particles were obtained with a clustering two-dimensional classification (CL2D) alignment and classification method (Sorzano et al. 2010), as implemented in Xmipp3.1 (De la Rosa-Trevín et al. 2013). CL2D is a multi-reference refinement approach based on correntropy (a nonlinear similarity measure between two random variables), instead of the more traditionally used correlation, as similarity criterion. Particle classification results from the comparison of the set of correntropies of all the images assigned to one class, and the set of correntropies of all the images that do not belong to the same class. 2D classification was used for several purposes, such as: cleaning the image data set by eliminating particles of the worst 2D classes, dealing with heterogeneous population of particles and generating initial 3D reference models.

### **3.3.3. Density maps reconstruction: 3D classification and refinement**

Single Particle Analysis of EM images have proved to be a powerful technique to elucidate the 3D structure of biological specimens within a wide range of different sizes and shapes, being the most appropriate choice when dealing with dynamic and heterogeneous samples. To obtain the 3D-EM map of a particle, an initial estimation/guess of its 3D structure is iteratively refined with a set of experimental 2D images (usually, several thousands). However, a phenomenon known as the “initial volume problem”, consisting in the dependence of the final result on the initial input volume (promoted by a tendency of refinement algorithms to fall in local minimum) (Pascual-Montano et al. 2006; Henderson et al. 2012), makes necessary to implement some kind of validation test. To handle the above mentioned problem, we used different initial volumes which were then refined with the same data set, so that only convergent maps, regardless of the initial volume, were finally selected. The initial volumes used included a blob of noise, an electron dense



sphere, as well as several volumes previously generated from our own experimental images with RANSAC (Vargas et al. 2014) or Random Conical Tilt (Radermacher 1988) (as implemented in Xmipp 3.1). These volumes were subjected to iterative refinements without applied symmetry and, in most cases, additional 2D and 3D classification runs were performed until obtaining convergent maps (Figure 16). A kind of validation for each of the obtained models was done by comparing projections of the 3D maps with experimental reference-free 2D class average classes.



**Figure 16. 3D reconstruction process. (A)** Diagram of the general steps followed for 2D and 3D processing of EM images of single particles, using the data of the MCM4/6/7 subproject from the present work. **(B)** Example of an iterative volume refinement process and final map consistency test: A number of independent refinement runs are performed using the same data set of experimental images (*first* column), but starting from a different initial volume (*second* column). The obtaining of a convergent structure validates the selected final map. In this example, the starting volumes are, from top to bottom, a noise, a sphere and a 150Å filtered model obtained with RANSAC (from experimental 2D average class images).

The resolution for each of the final 3D maps was determined using the Fourier shell correlation (FSC) 0.5 cutoff criterion (Heel & Schatz 2005). 3D maps were visualized with UCSF Chimera (Pettersen et al. 2004). The threshold level isosurface at which each density map was displayed to approximately represent the theoretical molecular mass calculated with the program EMAN1 (Ludtke et al. 1999) using the following command (<http://blake.bcm.edu/emanwiki/Volume>): volume <input file> <A/pix> [calc=<mass kDa>]. Here, the volume/mass conversions assume a protein density of 1.35 g/ml (0.81 Da/A<sup>3</sup>).

### **3.4. Normal Mode Analysis (NMA) of MCM4/6/7**

Normal mode analysis (NMA) of EM structures has been extensively used in the study of flexible macromolecules and complexes for predicting their functional motions (Jin et al. 2014; Tama et al. 2002; Tama et al. 2003; Ming et al. 2002), where putative dynamics is calculated directly from the structure. Using the NMA tool included in Xmipp3.1, six EM maps representing different conformational states of the ring-like MCM4/6/7x2 complex were analyzed. The distance matrix of the maps was calculated, for which it was employed the correlation among all the structures (similarity matrix) after doing the elastic alignment (deformations that one density map require to reach the shape of the next most similar 3D map) of the maps. A multidimensional scaling of the distance matrix was done in order to project all the structures into a 3D space, allowing a graphical visualization of the relative similarity distances among the six different maps of MCM4/6/7x2.

## Chapter 4

# 4. Results

## 4.1. Human CPAP<sup>897-1338</sup>

### 4.1.1. Sequence analysis and structural modeling

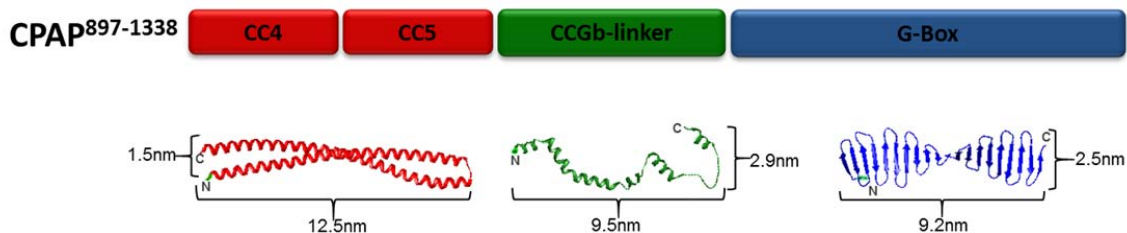
The *in silico* structural analysis of the *Hs* CPAP<sup>897-1338</sup> sequence (Figure 5A\* and Figure 17) shows that there are three structurally different regions distributed as follows:

- 38% corresponds to the CC4/CC5 coiled-coil region and is predicted to be mostly formed by  $\alpha$ -helices.
- 19% is formed by the CCGb-linker and is predicted to be unstructured.
- 43% forms the G-Box domain and is predicted to have 21 short  $\beta$ -strands, which is in agreement with the solved structure (PDB-4BXP) of the equivalent part of the protein in *D. rerio*.



**Figure 17. Secondary structural prediction of CPAP<sup>897-1338</sup>.** (A) The sequence includes the CC4/CC5 coiled-coil region (residues 897-1065), the CCGb-linker (residues 1066-1149) and the G-Box (residues 1150-1338). Red letter H refers to  $\alpha$ -helices. Yellow letter E refers to  $\beta$ -sheets. (B) Sequence alignment of the G-Box domain from *H. sapiens* (*Hs*) and *D. rerio* (*Dr*), showing a comparison of the  $\beta$ -sheet positions for both the prediction for *Hs* (labelled in orange) and the already known structure from the PDB-4BXP for *Dr* (labelled in blue). The part corresponding to the *Dr* sequence, underlined in red, is not resolved in the crystallographic structure.

Frequently, the interpretation of a 3D-EM density map involves building a molecular model. The structural modeling of the CC4/CC5 obtained with I-TASSER, shows two ~12.5nm long  $\alpha$ -helices organized in an antiparallel disposition, which is in agreement with the most probable state predicted by the LOGICOIL algorithm. Although the CCGb-linker region is predicted to be intrinsically disordered, meaning that is flexible and there is not a unique/fixed structure, we still have used some of the structures modeled by I-TASSER as a guide to have a rough idea of the general dimensions for both potentially contracted or extended states of this zone of the protein. Finally, we have used the crystallographic structure of the G-Box/TCP domain from *D. rerio*, whose sequence presents ~68% identity with the G-Box from human (Figure 18).

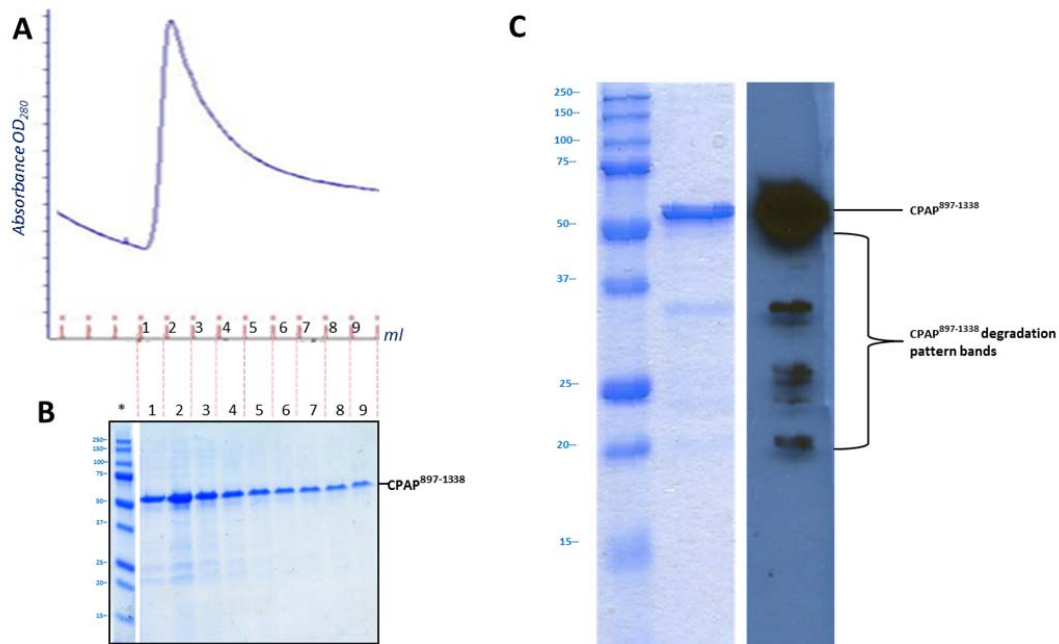


**Figure 18. Color-coded schematic representation of the structural domains of *Hs* CPAP<sup>897-1338</sup> (up) and their related atomic model (down).** From left to right: I-TASSER predicted atomic models for both the CPAP coiled-coils CC4/CC5 (in red) and an extended state version of the flexible CCGb-linker (in green), together with the crystallographic structure of the G-Box/TCP domain (PDB-4BXP) (in blue).

### 4.1.2. Protein purification

Most parts of CPAP are predicted to be disordered and coiled-coil, which are two hallmark characteristics of many centrosomal proteins, and are commonly associated with samples highly insoluble and difficult to handle and purify (Dos Santos et al. 2013; Treviño et al. 2014), which unfortunately is the case of the full-length CPAP. The CPAP<sup>897-1338</sup> sequence codes for a protein with a theoretical molecular weight (MW) around 52 kDa. This construct includes the only globular domain of CPAP (the G-Box), and the rest of the fragment, which covers around 57% of its sequence, is composed by the coiled-coils CC4/CC5 and the predicted non-structured region CCGb-linker (which connects CC4/CC5 with the G-Box).

Our N-terminal 6xHis-tagged CPAP<sup>897-1338</sup> construct is partially soluble under defined optimized expression and purification conditions (more details can be found in Materials and Methods). A one-step IMAC purification gave rise to a highly pure protein sample (Figure 19 A and B). SDS-PAGE analysis of the purified samples showed a small amount of a discrete degradation pattern of the protein; the CPAP identity of the observed lower molecular weight bands was confirmed by Mass spectrometry analysis and Western blot against the N-terminal His-tag of the protein, which revealed that the degradation occurs by the C-terminus side (Figure 19 C).



**Figure 19. Immobilized metal affinity chromatography (IMAC) purification of CPAP<sup>897-1338</sup>.** (A) IMAC purification chromatographic elution profile and (B) SDS-PAGE of the purified fractions. The lane labelled with the asterisk (\*) corresponds to the protein molecular weight markers; the lanes 1 to 9 are the pooled fractions of the purified CPAP<sup>897-1338</sup>. Dashed lines show the fraction sample correspondence between the chromatogram and SDS-PAGE. (C) CPAP<sup>897-1338</sup> protein degradation pattern. SDS-PAGE (*left*) and Western Blot against the N-terminal His-CPAP<sup>897-1338</sup> (*right*) showing a characteristic but minoritarian protein degradation pattern observed along the purification.

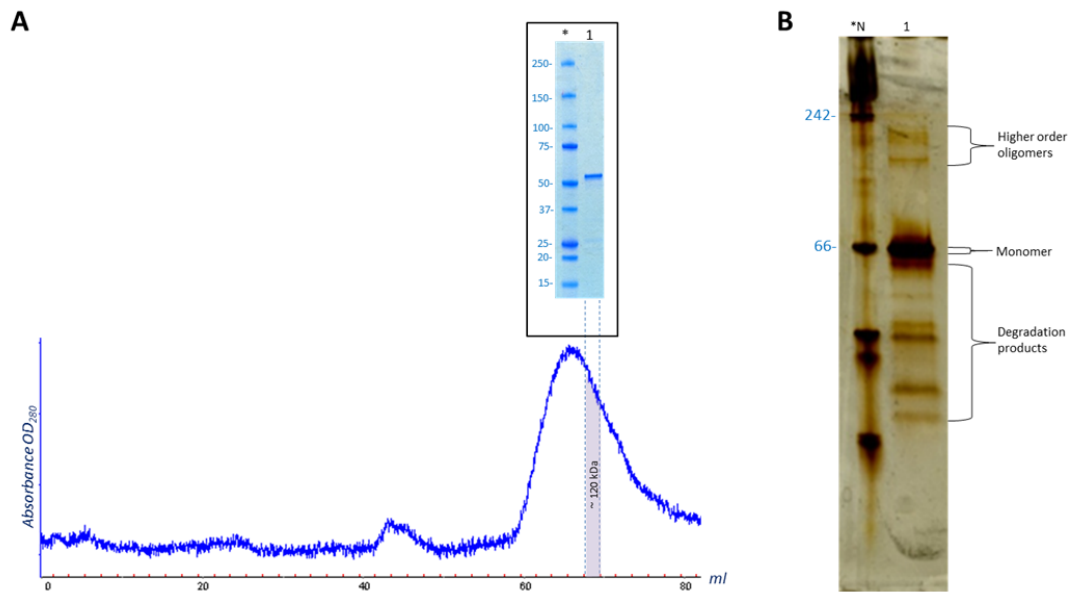
The variable SEC profiles obtained in a number of purifications of CPAP<sup>897-1338</sup> performed under the same buffer and temperature conditions indicated differences in size and abundance of the macromolecular specimens that were present. A possible explanation to this phenomenon is the change in concentration of the initial input samples passed through each of the SEC runs, which would strongly suggest that the concentration of the protein must be a determinant factor on the oligomerization state of CPAP<sup>897-1338</sup>, thus causing variations on the abundance and hydrodynamic radius of the particles. The relatively broad elution profile obtained in some cases is explained by the coexistence of particles with different shape/size in the sample; this was confirmed by EM analysis of the eluted fractions. An active exchange between different-sized complexes could also explain the relatively broad elution profile.

### **4.1.3. Structural characterization of different CPAP<sup>897-1338</sup> homo-oligomeric complexes and fitting of atomic models**

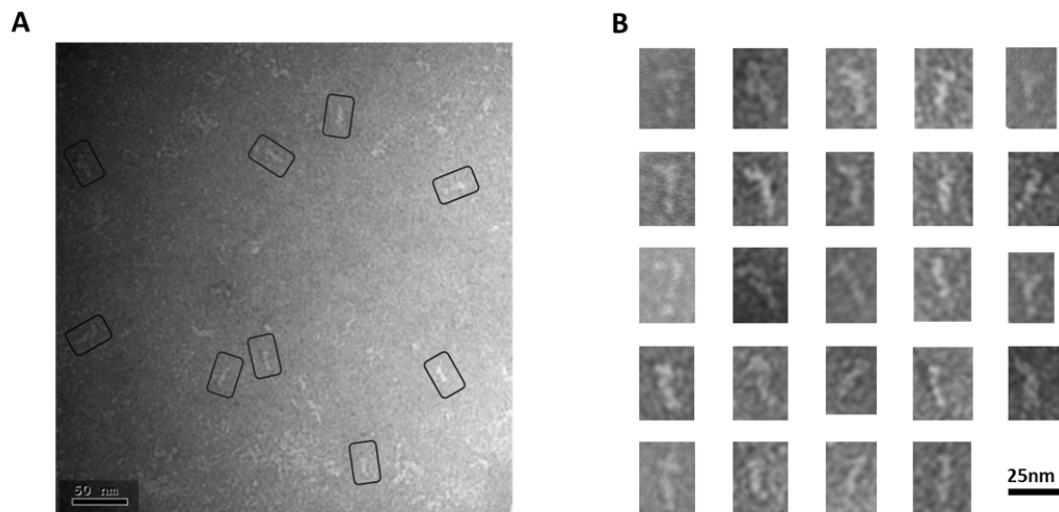
Visualization of NS-EM images of a number of CPAP<sup>897-1338</sup> purified samples allowed us to identify and characterize the structure of different oligomeric complexes formed by this protein. It is important to note that in the following we will perform putative assignment of oligomeric states based on both our biochemical data and the fitting of atomic models. Note that our fittings are only intended to show how the different maps are perfectly compatible with the oligomeric state assigned to them, and not to depict precise atomic arrangements, although some general structural features provided by the EM data allowed us to propose a tentative disposition of the different protein domains.

#### **4.1.3.1. CPAP<sup>897-1338</sup> flexible, fibrillar structures**

EM of a SEC (Superdex 200 16/60) purified CPAP<sup>897-1338</sup> protein fraction eluted around an apparent MW of 120kDa (Figure 20A), showed a population of flexible and elongated rope-like structures (Figure 21). Due to the fibrillar shape of these particles, it is expected that their hydrodynamic behavior significantly differ from that of the globular molecular weight protein markers commonly used for SEC columns calibration. A Native-PAGE was run to determine the oligomeric state of the protein in this fraction (Figure 20 B), showing that the predominant state was the monomer and only a negligible part forms some higher order oligomers (CPAP identity of the bands was confirmed by Mass spectrometry analysis). Protein degradation products similar to those previously observed (Figure 19C) were also present in this fraction.



**Figure 20. SEC purification of CPAP<sup>897-1338</sup> fibers.** (A) Size-exclusion chromatography (SEC) profile (Superdex 200 16/60) of purified CPAP<sup>897-1338</sup>. Black box shows an SDS-PAGE image, where the lane labeled with the asterisk (\*) corresponds to the protein molecular weight markers and, the lane labeled with the number 1 represents a fraction that elutes at an apparent MW of ~120kDa. EM images of this fraction show flexible fiber particles (See Figure 21). (B) Native-PAGE (silver staining) of the same fraction showed in panel A (labeled with the number 1). Most of the protein is in a monomeric state (thick band close to the 66kDa marker), although there are some very soft bands corresponding to a number of oligomers (Over the 66kDa band) and some others bands showing the previously observed protein degradation products (under the 66kDa band). \*N, shows the lane of the native protein ladder.

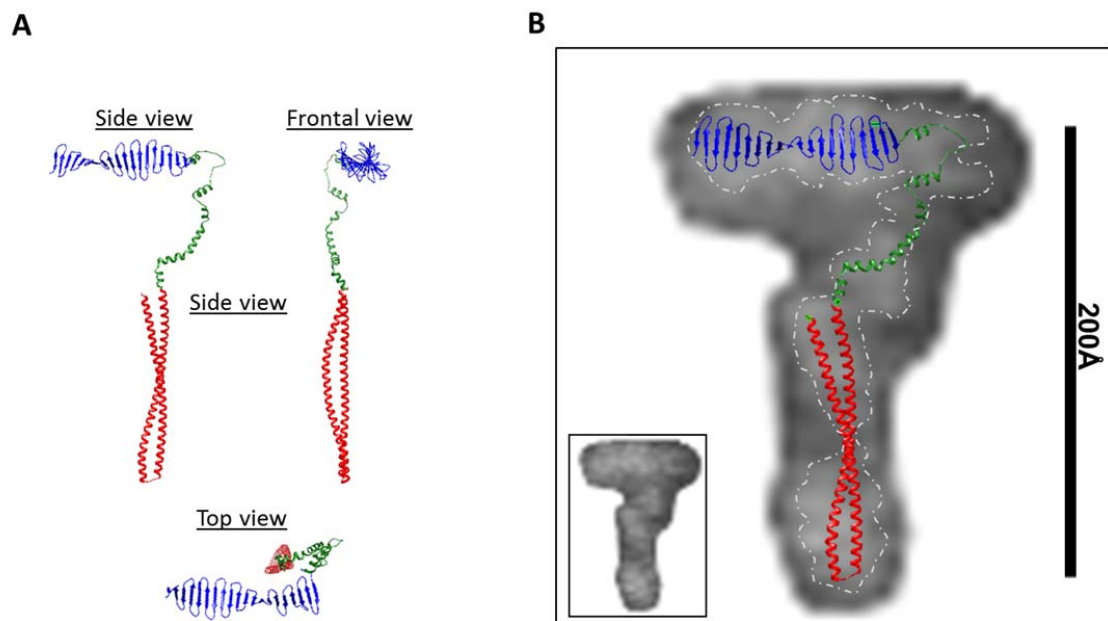


**Figure 21. EM of CPAP<sup>897-1338</sup> fibers.** (A) A representative negative stain electron micrograph of a SEC purified CPAP<sup>897-1338</sup> fraction eluted at an apparent MW of ~120kDa. It shows a large population of flexible fiber-like particles. Some selected particles are highlighted by black rectangles. (B) Windowed raw images of some side views show a general elongate cane-like structure.



Although the high flexibility of the particles made unfeasible to obtain well-defined 2D average classes of the EM images to make a 3D reconstruction, it was still possible to carry out a partial structural analysis by measuring the general dimensions of the most representative images, which presented a cane-like shape with a “shaft” of ~15-17nm long and a flexible “handle” of ~10nm long; the apparent thickness of the strings varied in a range between 2-4nm.

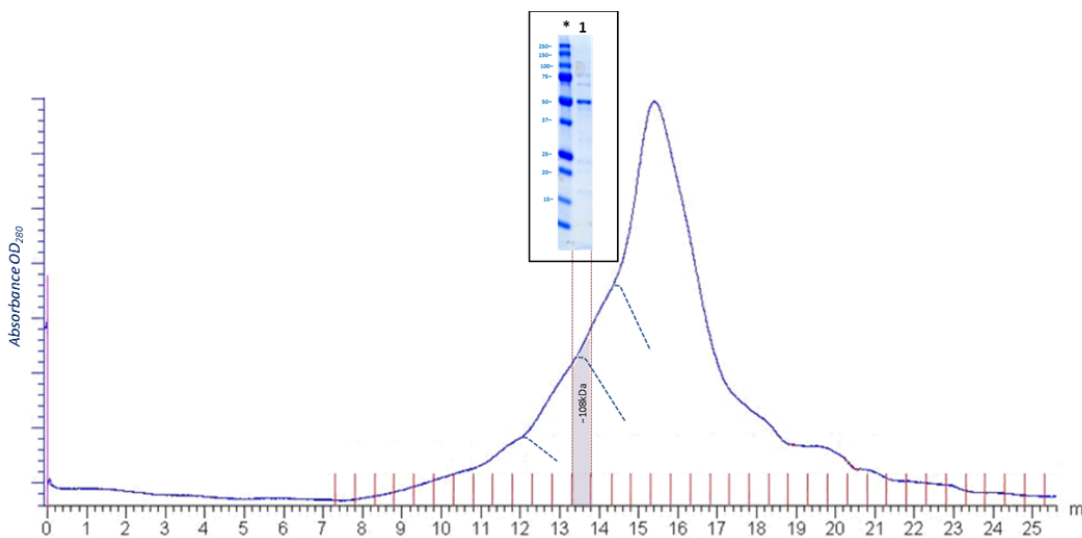
To complete the assembly of an atomic model for a possible 3D structure of the CPAP<sup>897-1338</sup> monomer, we put together the crystallographic structure of the G-Box/TCP domain (PDB-4BXP) and the models of the CC4/CC5 and the CCGb-linker (an extended version) regions, as predicted by I-TASSER; here, the CCGb-linker acts as a very flexible hinge that connects the G-Box and CC4/CC5 domains, which respectively correspond to the “shaft” and to the “handle” of the model (Figure 22 A). We observed that the dimensions and the shape of some of the fiber-like EM images fitted reasonably well with the model proposed before (Figure 22 B).



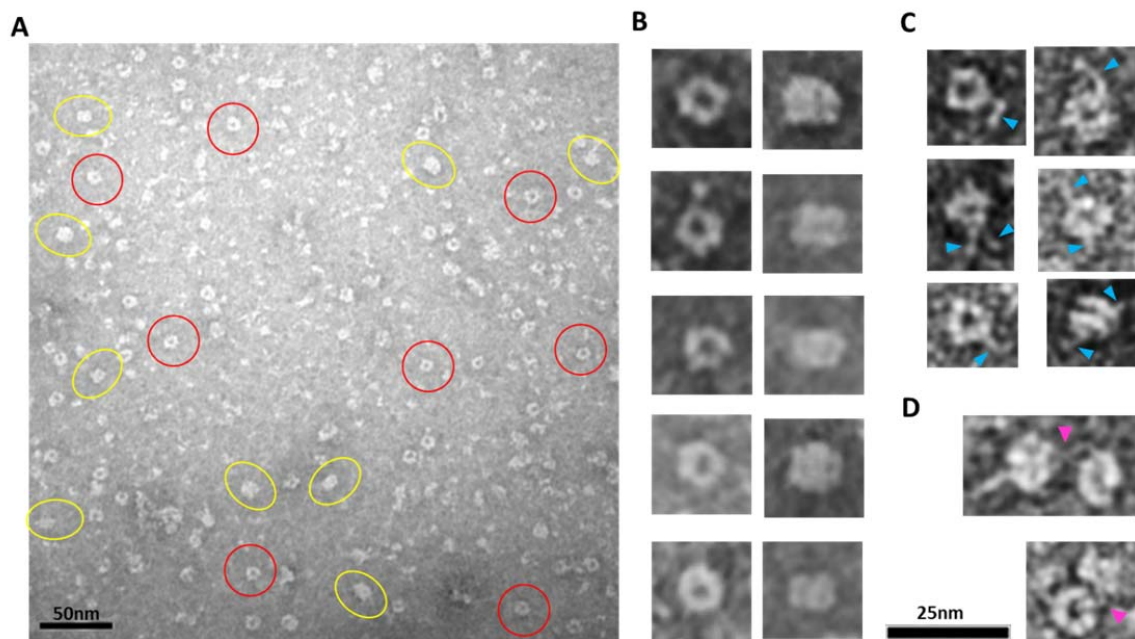
**Figure 22. Structural prediction corresponding to the CPAP<sup>897-1338</sup> monomer, based on the general cane-like shape of EM images from a SEC purified sample. (A)** Model of a possible CPAP<sup>897-1338</sup> monomer assembled using structural predictions of the CC4/CC5 and CCGb-linker domains and the X-ray structure of the G-Box (PDB-4BXP). **(B)** Superimposition of the atomic models of the CC4/CC5, CCGb-linker and G-Box domains on a representative single particle image from a NS-EM sample of CPAP<sup>897-1338</sup>. A discontinuous line demarks the contour of the EM particle. A smaller version of the same particle is shown in the inset at the lower left corner.

#### 4.1.3.2. CPAP<sup>897-1338</sup> toroidal complex

EM micrographs of a SEC (Superdex 200 10/300) purified sample eluted at an apparent MW of ~108kDa (Figure 23), which is compatible with the size of a CPAP<sup>897-1338</sup> dimer, was mostly populated by both images of toroids with an external diameter of ~10nm, and a more rectangular kind of particles with a size of ~7.5nm (*height*) x 10nm (*width*), which look like two wavy threads bound by their wider sides (Figure 24 A and B). These images could be interpreted as different projections of a structure similar to two interweaved rings, where the rectangular-like and the ring-like particles could be respectively attributed to the lateral and top views of the proposed model. Some of these particles showed thin flexible strings protruding to the outside (Figure 24 C) and, interestingly, we observed pairs of particles interacting through these flexible extensions (Figure 24 D).

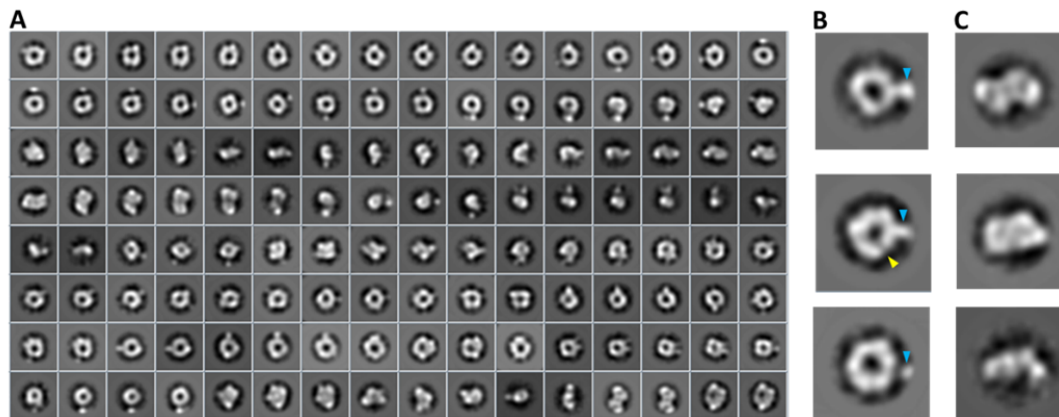


**Figure 23. SEC purified fraction of CPAP<sup>897-1338</sup> toroidal particles.** Size-exclusion chromatography (SEC) profile (Superdex 200 10/300) of purified CPAP<sup>897-1338</sup> where blue dotted lines mark three fused peaks under the graphic. The black box shows an SDS-PAGE image, where the lane labeled with the asterisk (\*) corresponds to the protein molecular weight markers and the lane labeled with the number 1 is a fraction that elutes at an apparent MW of ~108kDa, compatible with a dimer of CPAP<sup>897-1338</sup>. EM images of this fraction show some toroidal particles (See Figure 24).



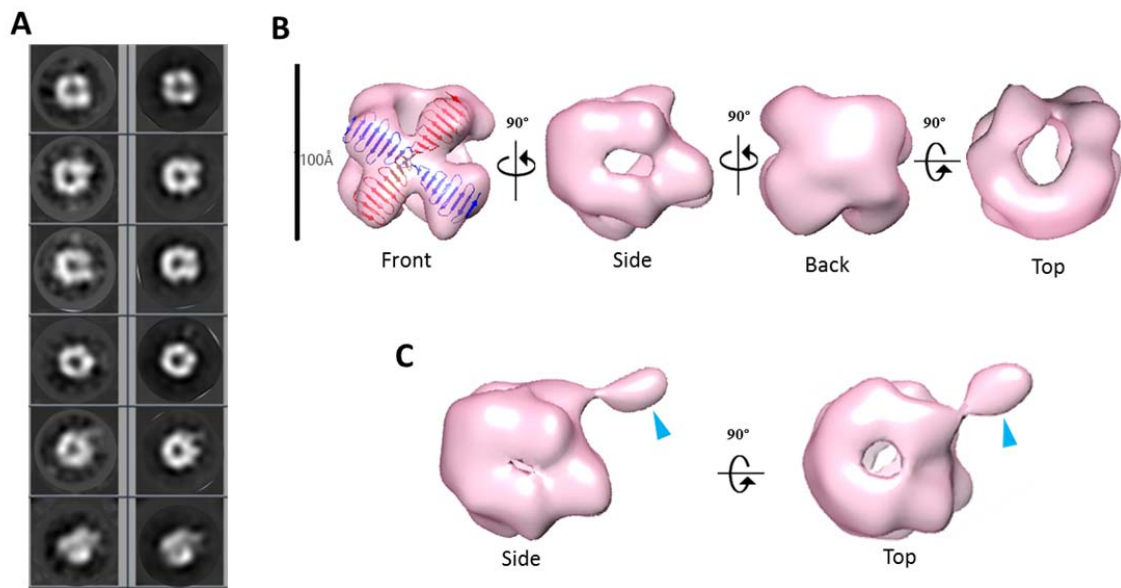
**Figure 24. EM of a SEC fraction of CPAP<sup>897-1338</sup> eluted at an apparent MW around 108kDa.** (A) A representative negative stain electron micrograph of a SEC purified CPAP<sup>897-1338</sup> fraction eluted at ~108kDa, compatible with being a dimer of the protein. Some toroid-like views are highlighted with red circles, while the more rectangular particles are highlighted by yellow ovals. (B) Windowed raw images of representative particles that could correspond with top views (*left panel*) and lateral views (*right panel*) of the same type of particle. (C) Single particles showing extended flexible projections (pointed by blue arrowheads). (D) Pair of particles interacting through flexible strings (pointed with a fuchsia arrowhead).

It is expected that during the averaging of the images, highly flexible structures get reduced, at best, to the base part, for being it the most stable point. Indeed, in agreement with the aforementioned phenomena, the 2D class average classification of 36,699 single particle images (Figure 25 A) showed some toroidal particles presenting a punctual, strong extra density at the periphery of the ring (Figure 25 B), which could be attributed to the average of the base part of the flexible strings observed in the single particle images (Figure 24 C), although it cannot be excluded that more compact conformations of the flexible projection exist. A small discontinuity is observed in some of the ring-like images, suggesting the presence of a gap (Figure 25 B). 2D image classification revealed the presence of other more elongated and flat particles (Figure 25 C), some of which looked like two small consecutive rings or an infinity symbol ( $\infty$ ), suggesting that other intermediate conformations of the dimer could also exist.



**Figure 25. 2D image analysis of a sample fraction eluted at an apparent MW compatible with a CPAP<sup>897-1338</sup> dimer.** (A) Reference-free class average image classification (CL2D). (B) Toroidal class images showing an extra mass at the periphery (pointed with a blue arrowhead), which is not always clearly connected to the ring. A weak density area (pointed with a yellow arrowhead) suggests a gap in the ring. (C) 2D class images of more elongated and flat particles that could correspond to intermediate conformation of the final dimer.

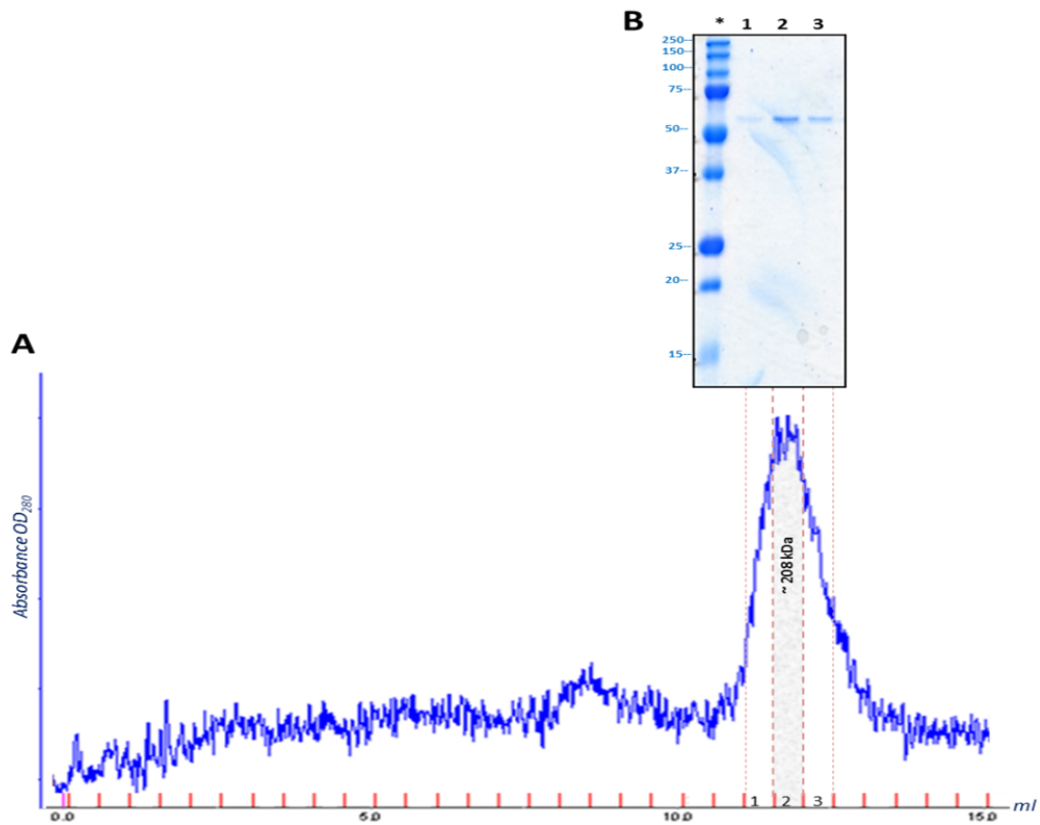
The combination of both 2D and 3D classification image processing procedures allowed to obtain a final 24 Å resolution map of an asymmetrical toroid-like structure (Figure 26) containing 4,587 images out from the initial data set. This structure represents the most frequent kind of particles (Figure 24 B) of the putative dimer of CPAP<sup>897-1338</sup>, although the reconstruction of other similar maps revealed some degree of global flexibility of the structure. The top and bottom openings of the central channel have different shape and size. The frontal view of the complex shows a slightly concave cross brace structure where two copies of the crystallographic structure of the G-Box/TCP domain (PDB-4BXP) can be reasonably fitted (Figure 26 B). The opposite side (back view) also has an X-like shape but is thicker and convex. Finally, each of the two remaining opposite sides shows a kind of two-tier like shape with an elongated opening in the middle. When the 3D map is displayed at a lower contour level (Figure 26 C) it appears an additional density emerging from a corner of the frontal face of the structure. This extra mass may correspond to part of one of the flexible strings previously described in some of the single particle images (Figure 24 C).



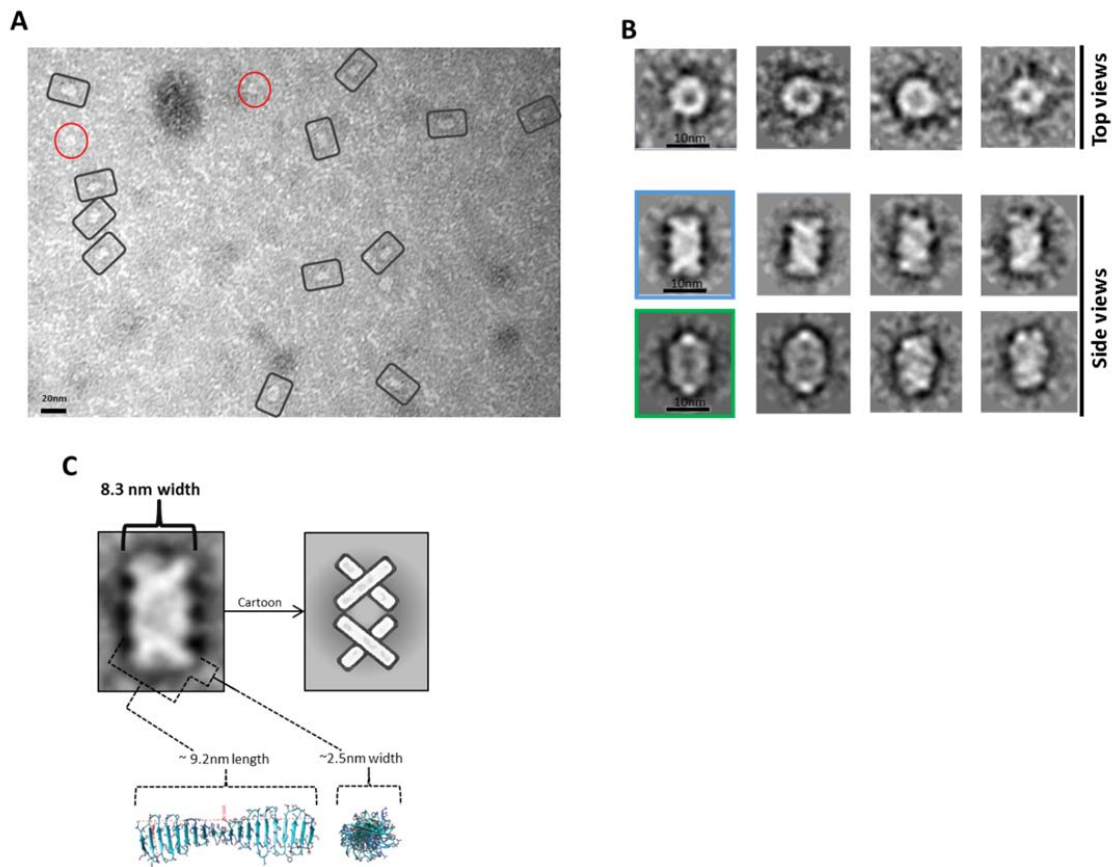
**Figure 26. 3D reconstruction of a putative CPAP<sup>897-1338</sup> homo-dimeric complex.** (A) Reference-free class averages (*left column*) and the corresponding forward projections (*right column*) of the 3D reconstruction. (B) Different views of the putative CPAP<sup>897-1338</sup> dimer with two copies of the atomic coordinates of G-Box/TCP domain (PDB-4BXP, represented in blue and red) fitted inside. (C) Top and one side view of the density map displayed at a low contour level where it is observed an extra mass (pointed by a blue arrowhead).

#### 4.1.3.3. CPAP<sup>897-1338</sup> barrel-meshed like structure

A gel Filtration (Superdex 200 10/300) purified protein fraction eluted at an apparent MW of ~208kDa (Figure 27), compatible with the size of a tetramer of CPAP<sup>897-1338</sup>, was analyzed by NS-EM (Figure 28 A). Reference free average classification of 3,048 singles images showed a meshed structure around 14.5nm length, that presents two different types of side views, each around 9.3nm and 8.3nm maximum width. The wider side view has an ellipsoid shape traversed by two wavy bridges (Figure 28 B). The narrowest side view looks like a lattice formed by strings with a diameter around 2.5nm, which agrees with the diameter of the crystallographic structure of the G-Box/TCP domain (PDB-4BXP). We also noted that the ~9.2nm length of the atomic structure of the G-Box matched surprisingly well with the longitude of the diagonal lines observed in some of the 2D average class (Figure 28 C).

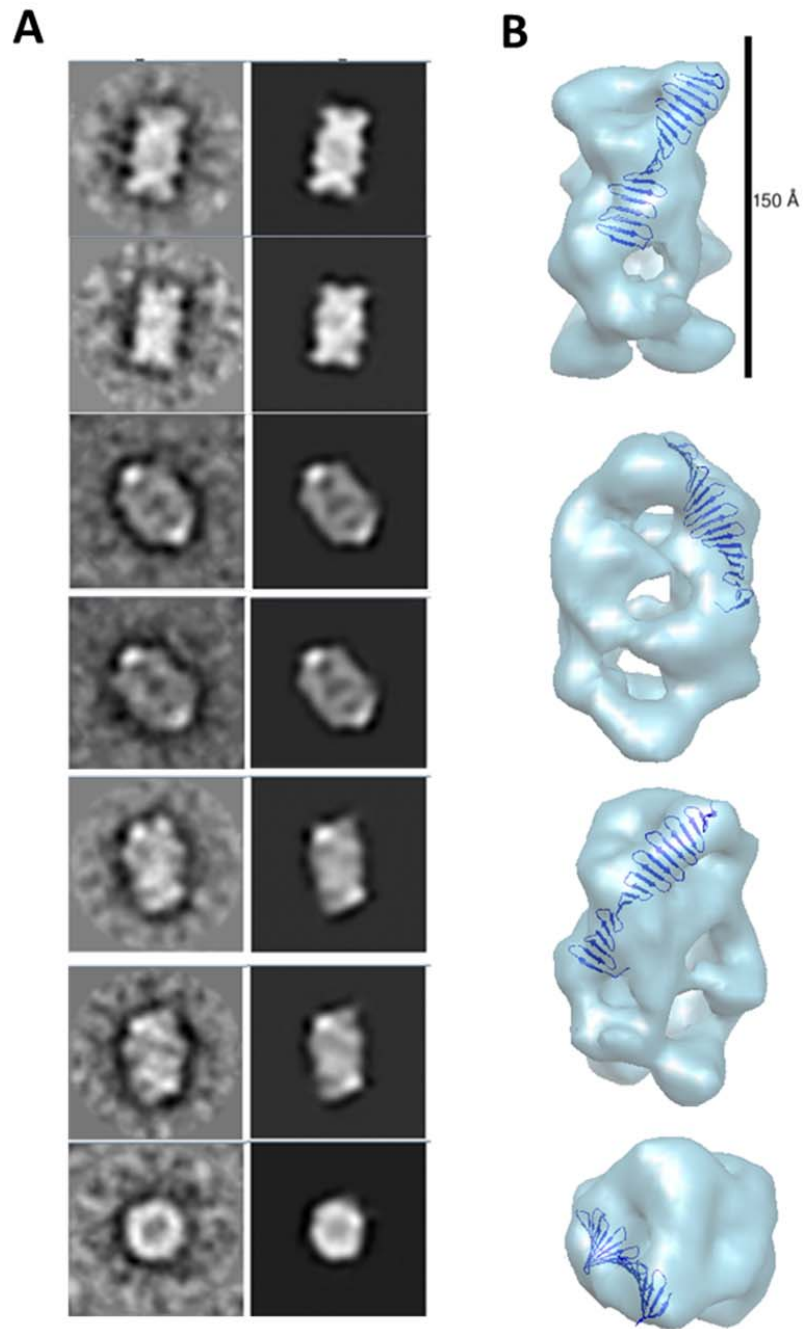


**Figure 27. Purification of CPAP<sup>897-1338</sup> barrel-like particles.** Chromatogram and associated SDS-PAGE are shown. **(A)** SEC profile (Superdex 200 10/300) of a purified CPAP<sup>897-1338</sup> sample. The center of the peak (fraction 2) elutes at an apparent MW of ~208kDa, compatible with a tetramer of CPAP<sup>897-1338</sup>. This fraction was analyzed by EM (See Figure 28). **(B)** SDS-PAGE. Lane labelled with the asterisk (\*) corresponds to the protein molecular weight markers; lanes 1-3, fractions corresponding to the pick of the purified CPAP<sup>897-1338</sup>.



**Figure 28. Transmission electron microscopy of negatively stained CPAP<sup>897-1338</sup> barrel-like particles.** (A) A representative negative stain electron micrograph of a SEC purified CPAP<sup>897-1338</sup> fraction eluted at ~208kDa, which is compatible with being a tetramer of the protein. Top views and side views are highlighted by gray rectangles and red circles, respectively. (B) Reference-free class averages of top views (panel of the first row) and side views (panels of the last two rows) of the putative CPAP<sup>897-1338</sup> tetramer complex. Some of the side view images are almost rectangular and narrower (image highlighted in blue) than other kind of images that have a more oval-shape and are wider (image highlighted in green). (C) 2D class average of a rectangular side view of 8.3nm width, which resembles two stacked crossbars (see the schematic representation to the right of the image), where the match between the dimensions of its diagonal strings (~2.5nm x ~9.2nm) and the ones of the crystallographic structure of *D. rerio* G-Box/TCP domain (PDB-4BXP) is shown.

The 24 Å resolution 3D map of the putative CPAP<sup>897-1338</sup> tetramer reveals the structure of an asymmetrical barrel-like complex made of intertwined strings. A tentative fitting of the atomic model of the G-Box (PDB-4BXP) shows how the general dimensions of this crystallographic structure fits in our EM volume (Figure 29).



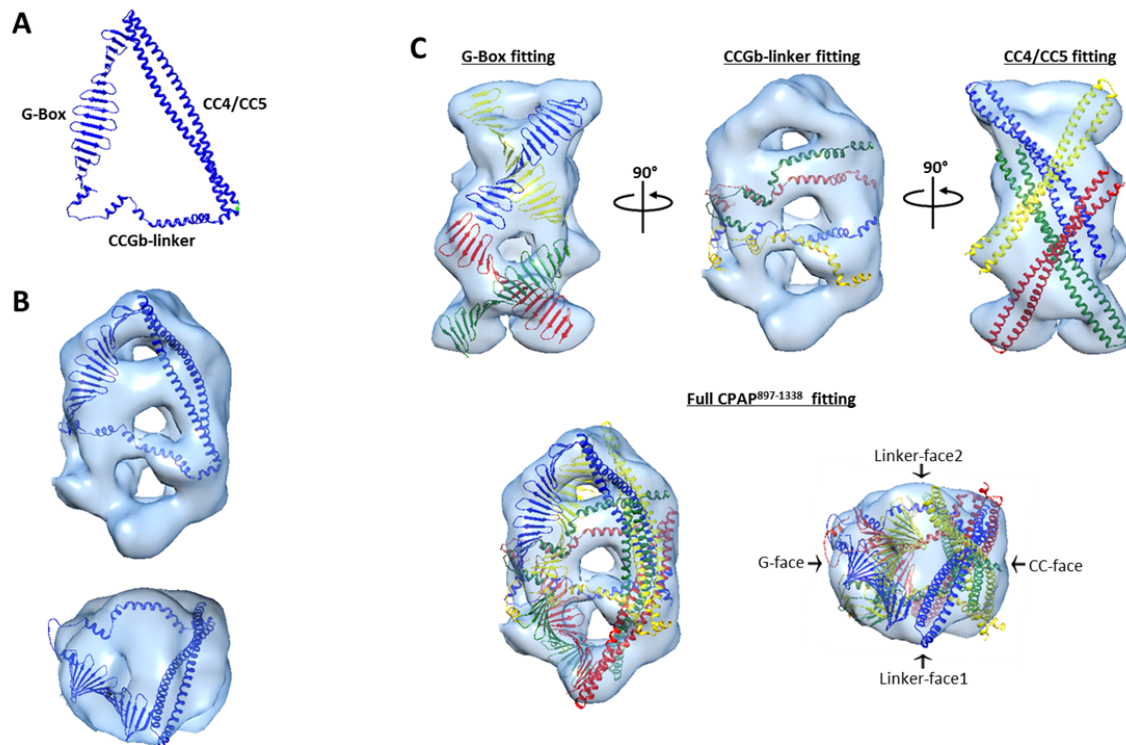
**Figure 29. 3D reconstruction of a putative CPAP<sup>897-1338</sup> homo-tetrameric complex. (A)** Reference-free class averages (*left column*) and the corresponding forward projections (*right column*) of the 3D reconstruction. **(B)** Different orientations of the putative CPAP<sup>897-1338</sup> tetramer volume, together with a tentative fitting of the atomic structure of the G-Box/TCP domain showing how the general dimensions of this structure fit into the EM volume.



Disordered regions, as it is predicted to be the CCGb-linker of CPAP (Figure 17 A), are often found as flexible linkers, allowing the protein to acquire different conformations. An alternatively proposed conformation of the CPAP<sup>897-1338</sup> atomic model presented in Figure 22, where the CC4/CC5 domain bends toward the G-Box producing a triangular-shape structural arrangement (Figure 30 A), allowed a nice fitting of four copies of the monomer inside the volume of our putative CPAP<sup>897-1338</sup> tetramer (Figure 30 B and C). Step by step, the fitting has been done by distributing the structural domains of the protein monomers into the four long lateral faces of the density map, in the following way:

- G-face: Four copies of the crystallographic structure of the G-Box domain have been fitted in one of the narrow lateral views of our map.
- CC-face: Four copies of the atomic model of CC4/CC5 have been fitted on the opposite side.
- Linker-face1 and linker-face2: Two copies of an extended version model for the CCGb-linker were accommodated at each of the two wider lateral sides of the volume.

The fitting described above gives additional support to the suggested tetramer oligomeric state for this complex. It is worth mentioning that the proposed distribution of the domains within the density map creates a clear structural polarity.

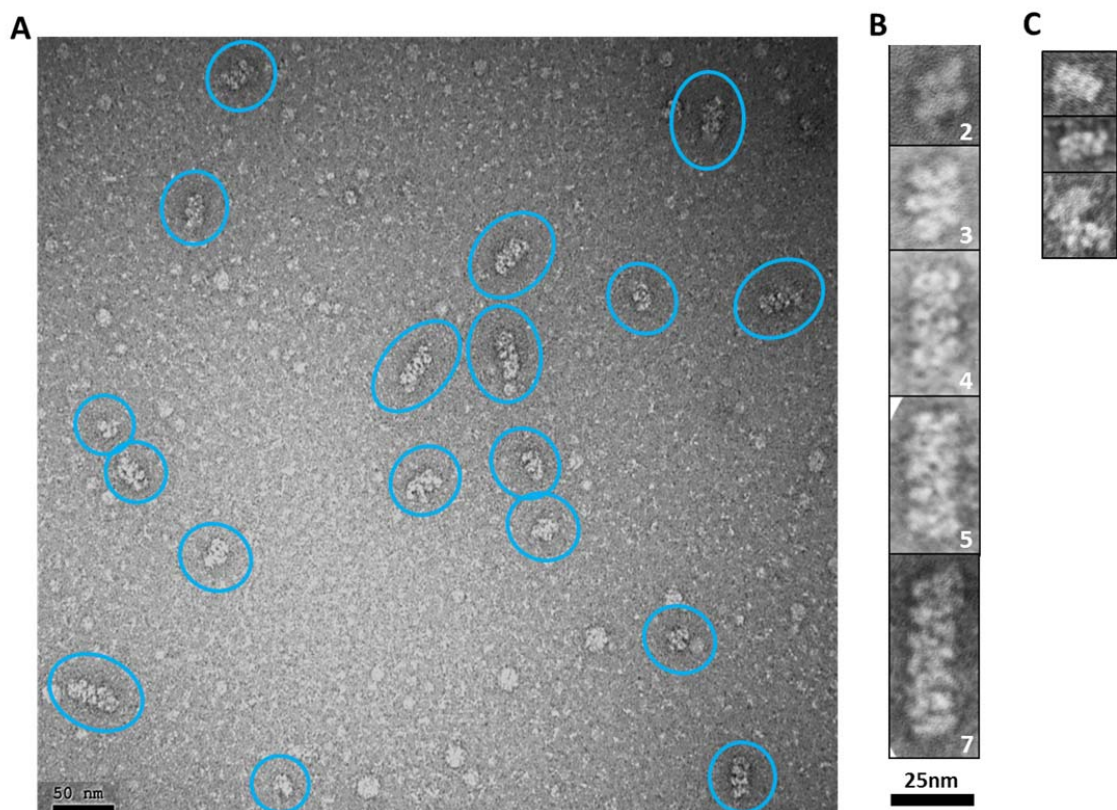


**Figure 30. Fitting of four copies of a triangular conformation of the proposed atomic structure of CPAP<sup>897-1338</sup> into the 3D map of the putative CPAP<sup>897-1338</sup> tetramer.** (A) Proposed triangular conformation of CPAP<sup>897-1338</sup> monomer, using the obtained atomic models of the CC4/CC5 and CCGb-linker domains and the X-ray crystallographic structure of the G-Box (PDB-4BXP). (B) Fitting of one copy of the complete monomer model into the 3D volume, where two different views of the EM density map of the complex are shown in translucent blue. (C) For clarity, in these figures the three structural domains of each of four fitted CPAP<sup>897-1338</sup> monomers are painted with a different color. In the first row there are shown different views of the density map with the fitting of four copies of each of the structural domains, and in the second row it is shown the fitting of four copies of the complete monomer model into the 3D map. Abbreviations used for labeling different faces of the figure in the panel at the bottom right corner are: G, G-Box; Linker, CCGb-linker and CC, CC4/CC5.

Interestingly, the obtained 3D map of particles in the fraction eluted at a MW ~108 kDa, (which we proposed previously to be a dimer), presents both dimensions and shape that could be compatible with half of our putative tetramer 3D map. The above mentioned consideration, together with the observed images of two molecules of the putative dimers interacting with each other (Figure 24C), suggest that the tetramers may be formed by the dimerization of the ~108kDa toroidal complexes.

#### 4.1.3.4. Modular higher order supramolecular rope-like structures

NS-EM of protein fractions eluted between 350-450mM NaCl from an ion exchange column, showed modular rope-like structures of different lengths. At first glance, these structures seem to be formed by linear stacks of a variable number of rectangular modules with dimensions varying between  $\sim 7\text{-}8.5$  nm (*length*) x  $\sim 15\text{-}16$  nm (*width*) (Figure 31A and B), whose shape and size resemble the narrow side view of our previously described putative CPAP<sup>897-1338</sup> tetramer. Interestingly, the dimensions of completely or partially isolated blocks found in this sample (Figure 31C), match with those of the putative tetrameric complex. Therefore, it seems reasonable to propose that the putative tetramers self-assembly into organized, modular rope-like structures. A possible explanation for the slightly variable internal organization of the elongated, modular structures is that their building blocks have some flexibility, causing the small differences in length and width observed among modules.



**Figure 31.** CPAP<sup>897-1338</sup> putative tetramers may stack together by their longer sides to form modular higher order complexes of variable length. (A) Representative negative stain electron micrograph of an anion exchange purified CPAP<sup>897-1338</sup> fraction, where there are observed higher order modular-like complexes of different length (highlighted by

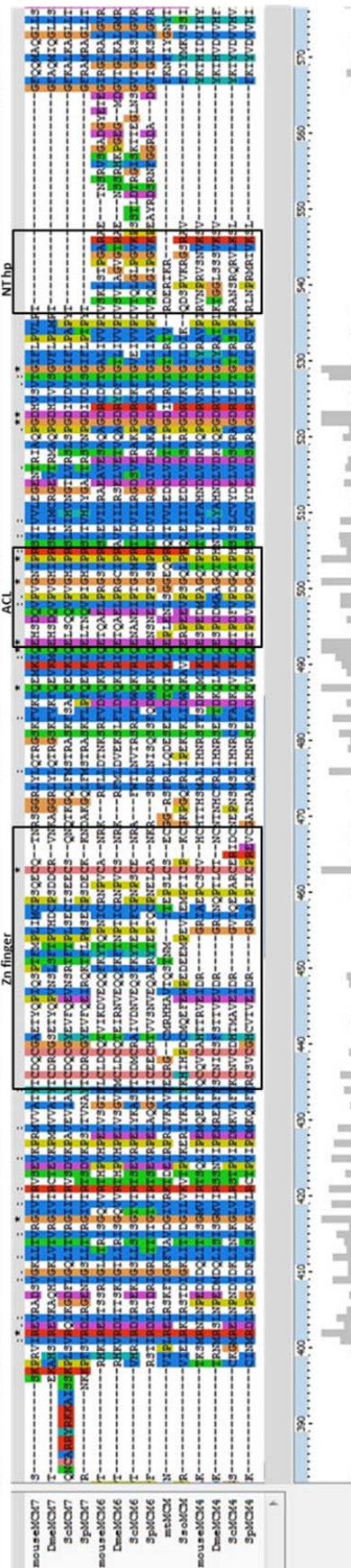
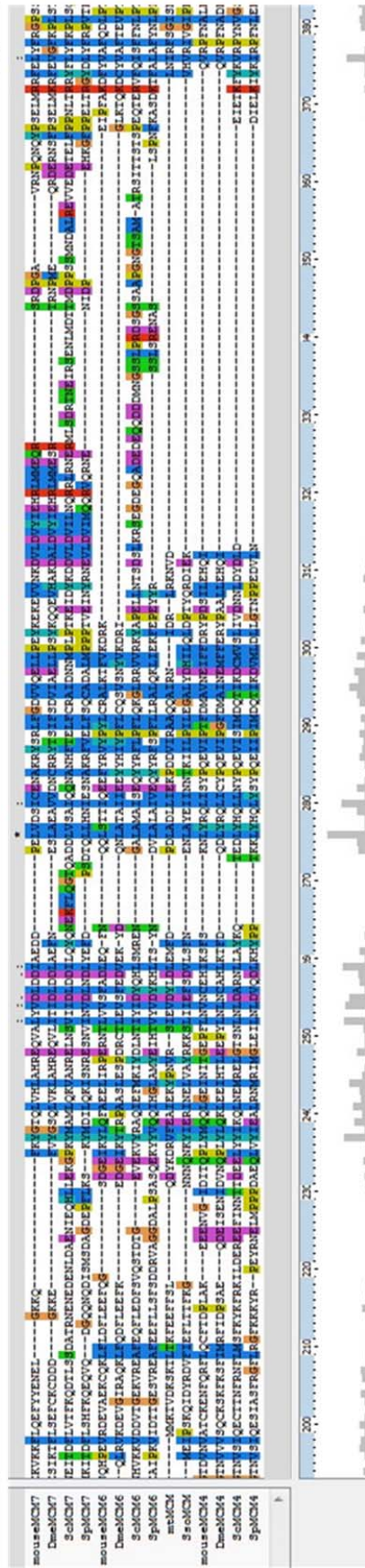
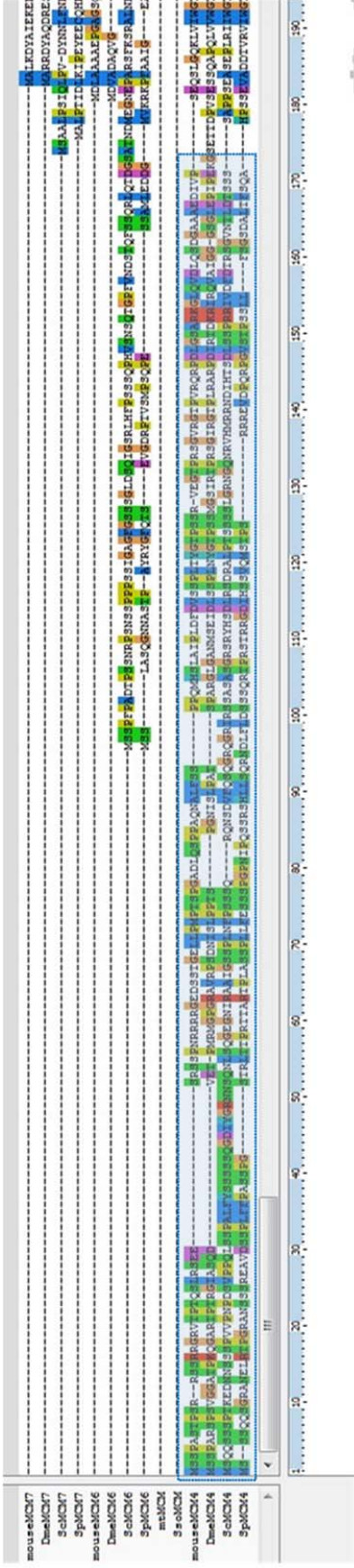
blue ovals). **(B)** Windowed zoomed images of some selected particles where the number of stacked modules is indicated by the number written on the lower right corner of each image. **(C)** Images of individual modules (first two rows) or partially interacting modules (last row).

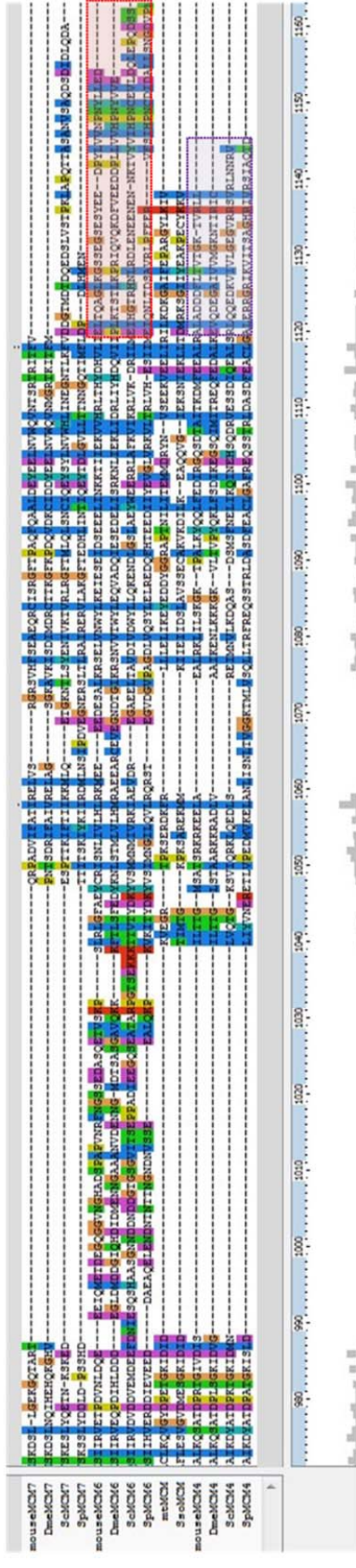
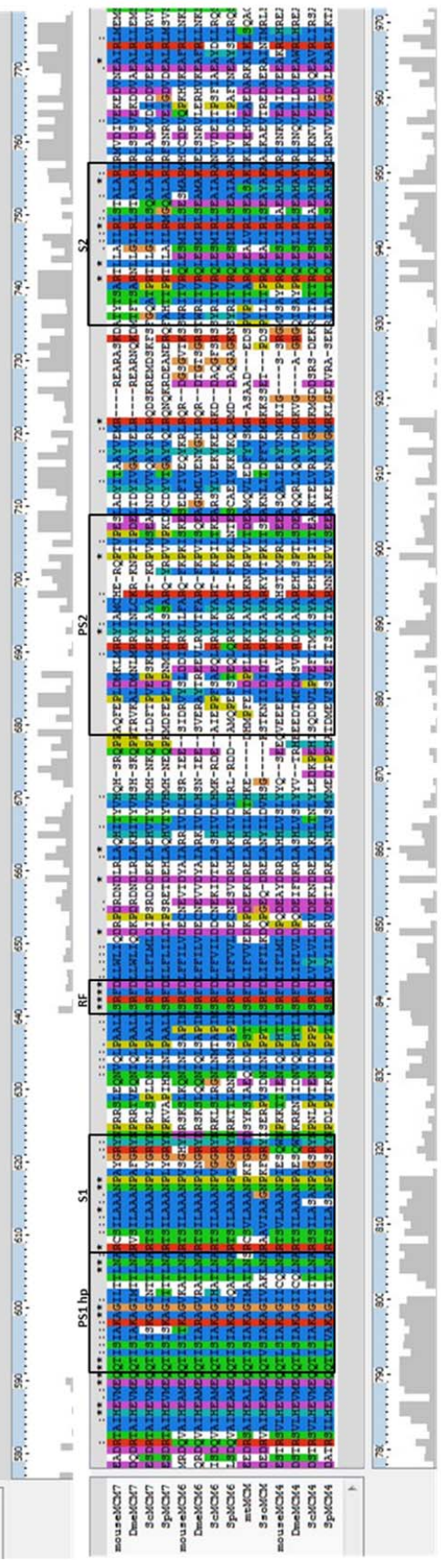
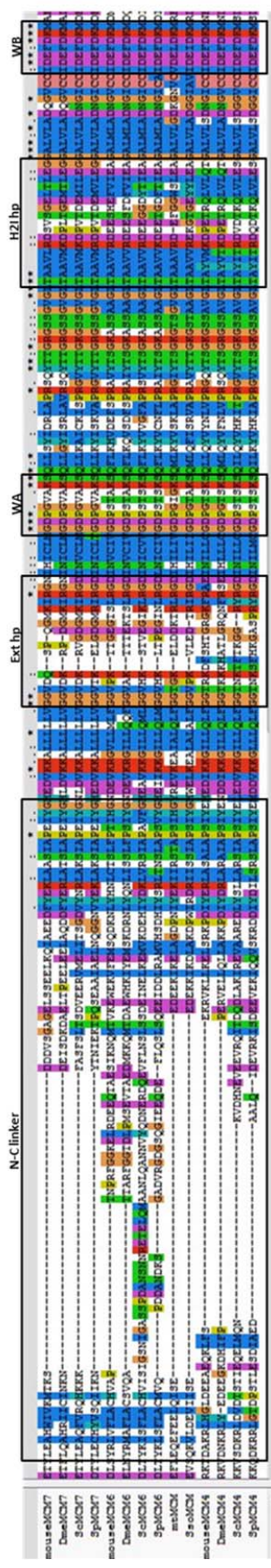
It has been proposed that higher order assemblies of CPAP could act in a synergistic way as a platform to provide an interface mediating the tethering of PCM to the centriole (Gopalakrishnan et al. 2011; Hatzopoulos et al. 2013; X. Zheng et al. 2014). Our 3D-EM reconstruction of the CPAP<sup>897-1338</sup> tetramer and the observed higher order extended and modular organized structures with an ~8 nm axial periodicity, are compatible with the last mentioned idea.

## **4.2. Mouse MCM4/6/7**

### **4.2.1. Multiple sequence alignment and structure prediction modeling of mouse proteins MCM4, MCM6 and MCM7**

To date, multiple biochemical and structural studies on MCM proteins from different organisms are available. We performed a multiple sequence alignment of mouse MCM4, MCM6 and MCM7 sequences against those from other species (Figure 32). The results of this analysis are highly informative and useful to get a better picture of the level of structural and functional similarity that can be expected between mouse proteins and the available information from other MCM homologs. It is observed that most of the structural motifs within the AAA<sup>+</sup> domain are well conserved, while the N-terminal and C-terminal regions are quite variable. When compared against most of MCM7 sequences (and with the *SsoMCM*), it become noticeable that MCM6 proteins present a C-terminal extension, as it happens with some of the MCM4 sequences, although to a much lesser extent. Furthermore, the MCM4 proteins display a long extension of residues toward the N-terminus.

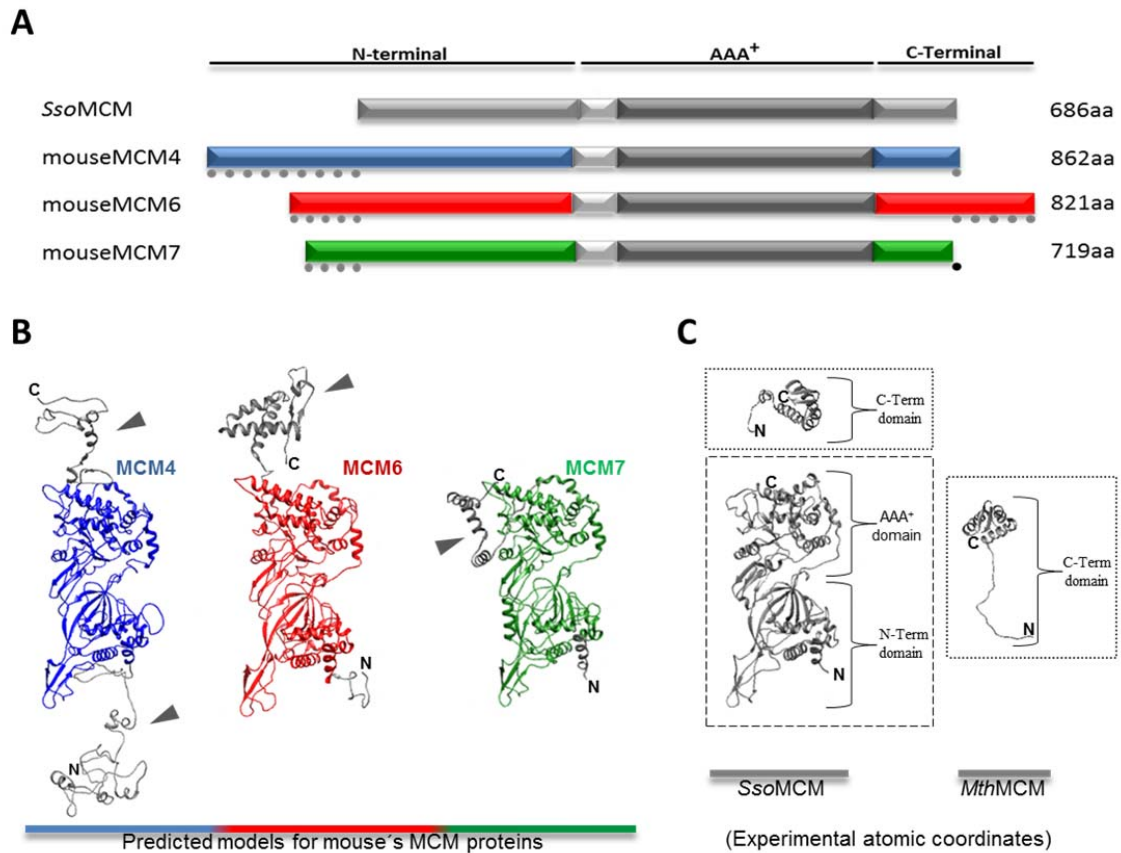




**Figure 32. Multiple sequence alignment of mouse MCM4, MCM6 and MCM7 complete sequences with respect to their homologs in representative organisms.** *D. melanogaster* (*Dme*), *S. cerevisiae* (*Sc*), *S. pombe* (*Sp*), and the MCM protein in *S. solfataricus* (*Sso*) and in *M. thermautotrophicus* (*mt*). Shared motifs among MCM proteins are highlighted into black boxes labeled with the respective motif abbreviation: Zn finger, Zinc finger; ACL, allosteric communication loop; NT hp, N-terminal  $\beta$ -hairpin; N-C linker, amino and carboxyl domains linker; Ext hp, external  $\beta$ -hairpin; WA, Walker A motif; H2I hp, helix 2 insert  $\beta$ -hairpin; WB, Walker B motif; PS1 hp, presensor 1  $\beta$ -hairpin; S1, sensor 1; RF, arginine finger motif; PS2, presensor 2 insertion; S2, sensor 2. Residues conservation is highlighted by ClustalX color code and the light gray columns under the alignment are an indicator of the conservation level at each point (Chenna 2003; Larkin et al. 2007). Blue dashed box highlights a N-terminal MCM4 residue extension zone; red dashed box highlights a C-terminal MCM6 residue extension zone; purple dashed box highlights a C-terminal MCM4 residue extension zone.

Consistently with the conserved domains and variable regions observed in the multiple sequence alignment of MCM homologs, the I-TASSER modeled structures of mouse MCM4, MCM6 and MCM7 maintain a core globular architecture similar to that of the crystal model of the near full-length *Sso*MCM (PDB-3F9V), which is lacking its C-terminal domain. On the other hand, length differences between proteins are observed at both their N-terminal and C-terminal parts. Mouse MCM6 has a long C-terminal extension and shares 95% of sequence identity with the structure of the C-terminus of *hs*MCM6 (PDB-3F9V), which is a winged helix (WH) motif. MCM4 features both a very long and unstructured N-terminal extension and a flexible but much shorter C-terminus, which is similar in length to that of archaea MCM. Sequence identity between the C-terminal region of mouse MCM4 and the solved C-terminals structures of *Sso*MCM (PDB-2M45) and *Mth*MCM (PDB-2MA3) are around 16% and 20%, respectively. Compared with the other proteins, the C-terminus of MCM7 is the shortest one; the structural prediction modeling of this part adopts a HTH conformation that folds down against the AAA<sup>+</sup> domain of the protein (Figure 33).





**Figure 33. General structure of MCM4,6 and 7. (A)** Comparative schematics of *SsoMCM* and mouse MCM4,6 and 7 proteins. Gray dots demark the terminal extensions of the mouse proteins when compared against the archaea MCM, while black dots represent the length difference between the C-terminal region from both MCM7 and *SsoMCM*, being the latter the longest one. **(B)** I-TASSER predicted structures for mouse MCM4 (blue), MCM6 (red) and MCM7 (green) full-length proteins. Arrowheads point to the grey-colored long N-terminal extension of MCM4 and the C-terminal domains of all the three mouse proteins, being the one of MCM6 the crystallographic structure PDB-2KLQ. **(C)** The near full-length structure of *SsoMCM* (PDB-3F9V) is highlighted by a dashed black box, and the C-terminal structures of *SsoMCM* (PDB-2MA3) and *MthMCM* (PDB-2M45) are highlighted by dotted black boxes.

## 4.2.2. EM structural characterization of the sample

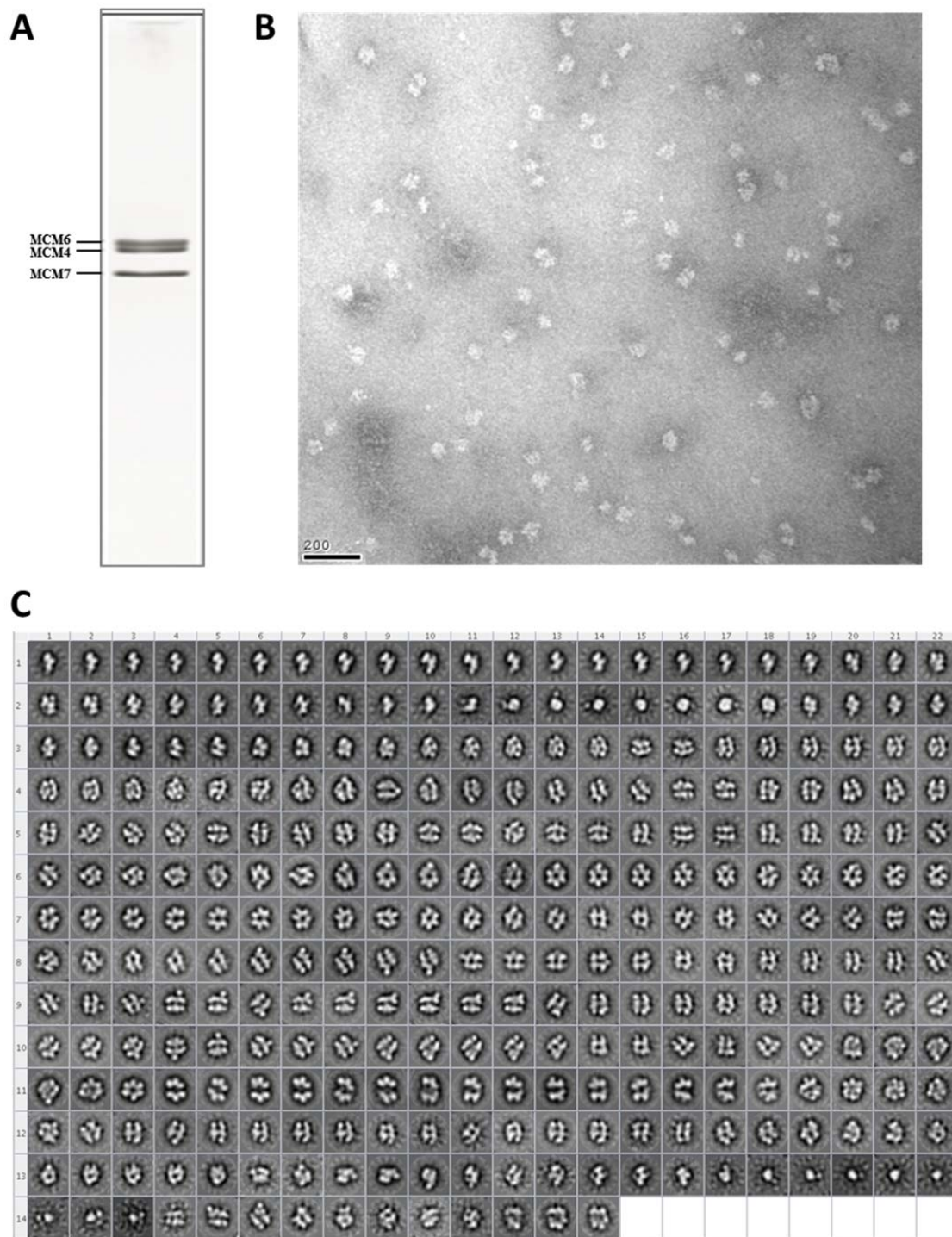
Helicases are macromolecular motors that employ the energy obtained from ATP hydrolysis to carry out its functions. Conformational changes induced by ATP binding and hydrolysis produce a highly dynamic behavior. It has been reported that in the absence of nucleotide, the MCM4/6/7 trimeric state is prevalent. On the other hand, ATP (and a non-hydrolyzable ATP analog) favors the dimerization of the trimer to form the MCM4/6/7x2 hexamer, and an intermediate state of trimers and hexamers is observed with ADP (Ma et al. 2010). By analyzing the MCM4/6/7 sample

in presence of ATP, we were able to capture and study a number of putative ATP binding/hydrolysis-dependent intermediate complexes. Several of the presented architecture features in this work reveal novel structural details never reported before.

#### **4.2.2.1. Description of the sample**

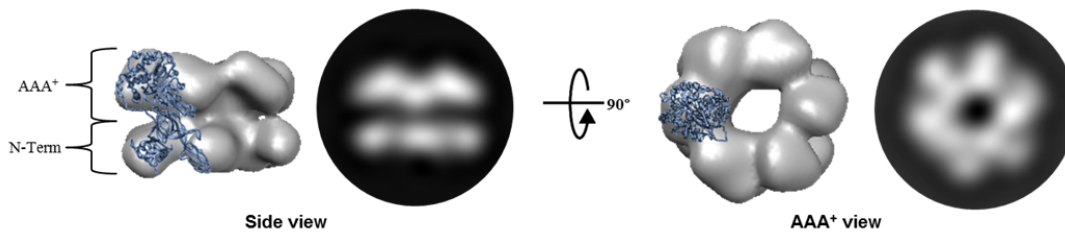
A -80°C preserved sample purified in the presence of ATP, containing a homogenous complex with a molecular weight agreeing with a hexameric assembly (You & Masai 2008) and formed by stoichiometric quantities of the proteins MCM4, MCM6 and MCM7 (Figure 34 A), was analyzed by negative-stain EM (Figure 34 B). Reference free 2D image classification on a set of 58,010 manually selected particles revealed a highly dynamic sample, mostly populated by flexible hexameric complexes, although lower oligomers were also observed (Figure 34 C). The above mentioned, as well as other more detailed structural features that will be discussed later in this section, becomes more evident after comparing our experimental 2D classes with projections of a canonical MCM hexameric two-tier ring model (Figure 35). The side view of this canonical model presents a two-tier shape, where the N-terminal domains from all the protomers form a first layer relatively flat, which is topped by a thicker layer formed by their AAA<sup>+</sup> domains. The more flexible regions, corresponding to the C-terminus and the N-terminal extensions present in some eukaryotic MCMs, are not visible in the canonical model. The top view has a ring-like shape with six interconnected lobes around a central pore.

It is known that the presence of nucleotide stimulates the formation of the hexameric complex, while lower oligomeric forms of the protein are obtained in the apo state (Ma et al. 2010). On the other hand, eukaryote MCM proteins have DNA-dependent ATPase motifs in their AAA<sup>+</sup> central domains (Ishimi 1997a; Hu et al. 1993; Koonin 1993), and ssDNA importantly increases the ATPase activity of MCM4/6/7x2 (Kanter et al. 2008). Studies in *Mth*MCM showed that substantial conformational changes occur during ATP binding and hydrolysis (Sakakibara et al. 2009), and in *Ecu*MCM2-7 it has been described a compaction of the ring upon ATPγS binding (Lyubimov et al. 2012). On the basis of the foregoing, we speculate that different states of the helicase (apo, ATP-bound and ATP-hydrolysis intermediates) may give rise to the observed conformational changes.



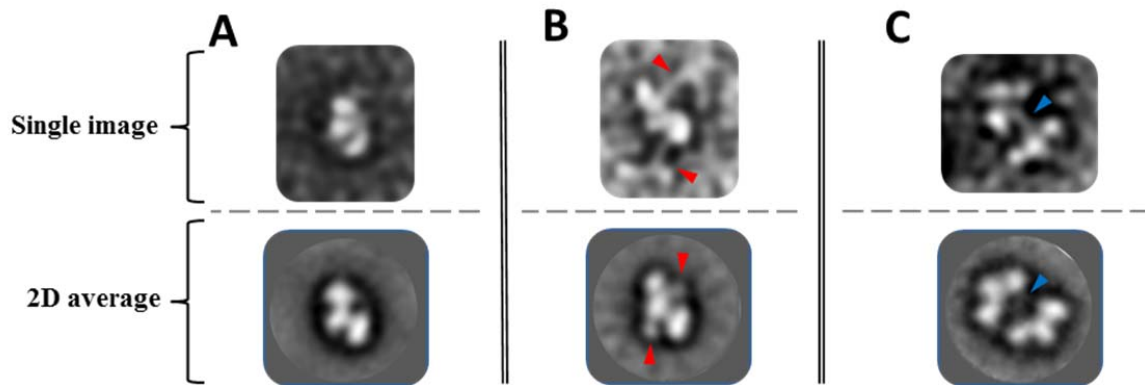
**Figure 34. MCM4/6/7 sample and particles.** (A) SDS-PAGE gel of a glycerol gradient purified fraction of mouse MCM4/6/7x2 complex. (B) Representative negative stain electron micrograph of the sample. (C) Reference-free class average classification (CL2D) of the complete set of 58,010 particles images of the sample.

### MCM hexamer canonical model



**Figure 35. Representative model of the classically described two-tier MCM helicase.** Both the AAA<sup>+</sup> and side views of the 3D map (*left*) and their corresponding 2D projections (*right*) are shown. A monomer of the atomic structure of *Sso*MCM (PDB-3F9V), lacking its C-terminal domain, is fitted into the density model.

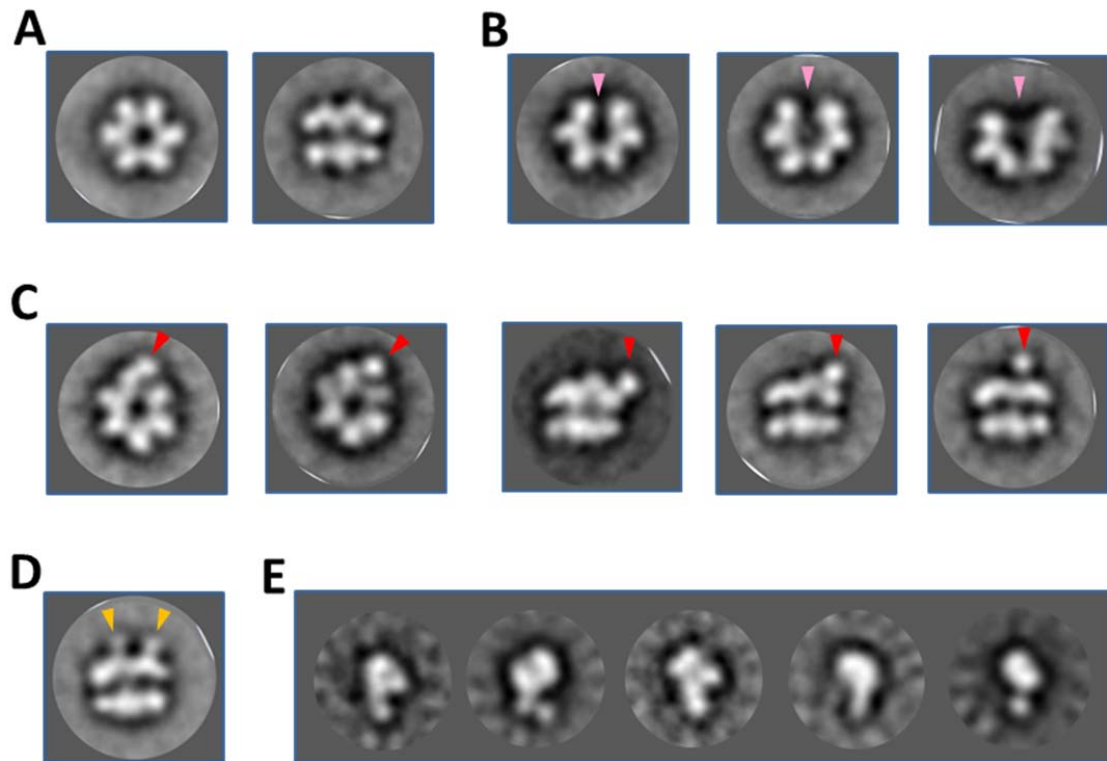
2D image analysis showed some novel classes with size and shape compatible with being top/bottom-views of one half of a hexamer only (i.e. a trimer) (Figure 36 A). Other similar classes, but displaying two smaller additional densities located at opposite sides respect to each other, are also observed. Interestingly, a number of images from individual particles revealed the presence of thin, flexible strings emerging from the body of the proposed MCM trimer (Figure 36 B). It is expected that during the averaging of the images, these highly flexible structures get reduced, at best, to the base part, for being the most stable point. Considering the above observations, and the existence of long extensions in some of the MCMs subunits (Figure 33), the additional electron densities in the 2D classes may correspond to the N-terminal and C-terminal sequence extension of MCM4 and MCM6, respectively. Another interesting 2D class average showed what seems to be two separated and confronted trimers. A careful view of the area between the two putative trimers reveals a soft density, and the observation of the single particle images make evident a thin bridge connecting the two trimers (Figure 36 C).



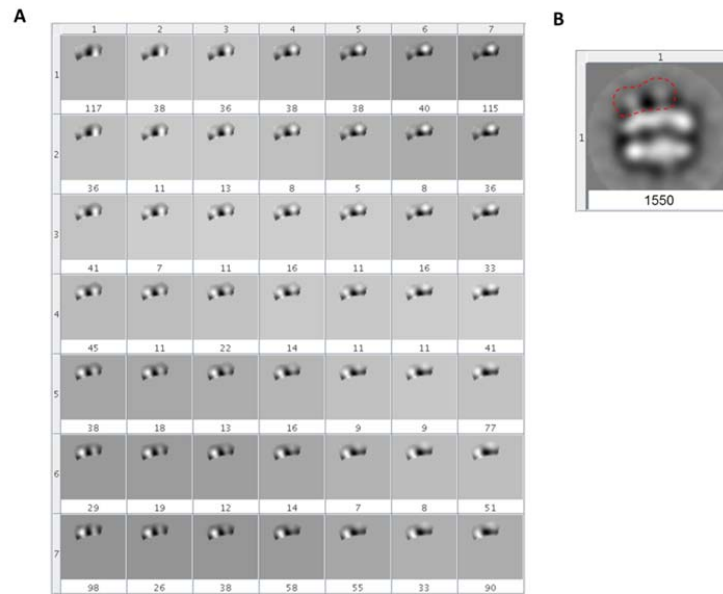
**Figure 36. Individual images and reference-free class average of putative top-views from MCM4/6/7. (A)** simple trimer, **(B)** trimer with two protruding flexible strings (pointed by red arrowheads) and **(C)** two facing trimers looking as if bound through a thin central bridge (pointed by a blue arrowhead).

In addition to classes similar to the more classically described side and top-view projections of the two-floor ring-like hexameric helicase (Figure 37 A, compared with the projections in Figure 35), and a gapped or even some more widely opened rings (Figure 37 B), we also captured top- and side-views of hexamer conformations showing, quite sharply, the presence of one strong extra density, which, by comparison with the canonical side view of an MCM helicase (Figure 35), seems to be located at the C-terminal region of the ring (Figure 37. C). A different novel conformation is a side view presenting two similar additional electron densities; once again, by comparison with the canonical side view of an MCM helicase (Figure 35), both densities could correspond to the C-terminal domain of some of the protomers (Figure 37 D). These two densities are less strong than the single one observed in the aforementioned 2D classes, reflecting their smaller size. KenDerSOM (Smoothly Distributed Kernel Probability Density Estimator Self Organizing Map) analysis was performed in order to confirm the presence of the two different densities, by discarding the possibility that average of particles with only one extra mass, but in different positions, were aligned, creating an artifactual class that appeared to have two extra densities (Figure 38). Two copies of MCM6 form part of the hexameric MCM4/6/7x2, and this protein has a substantial C-terminal extension when compared with both MCM4 and MCM7. Then, following this reasoning, it seems very plausible that each of the two extra masses protruding from the AAA<sup>+</sup> region of the helicase may correspond to the HTH domain of both MCM6 protomers. In the same line, and considering that in the classes showing only one extra mass emerging from the AAA<sup>+</sup> this is much more electron dense than the ones in class with two masses, one might think that the

strong electron density could be formed by the partial or complete interaction of the HTH motifs of both MCM6s, which in turn would support the proposed model where the two MCM6s are contiguous (Figure 10 B). Finally, we observed a minority group of small particles that could, tentatively, correspond with views of free monomers (Figure 37 E), which are around 80-95 kDa.

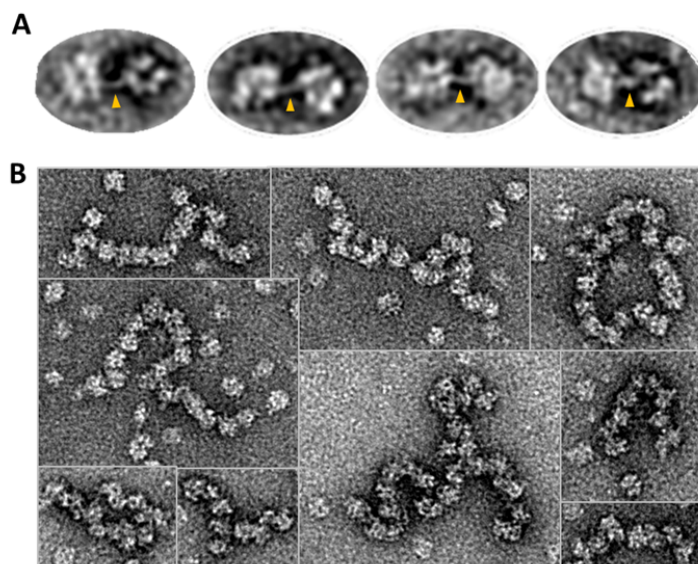


**Figure 37. Representative reference-free class averages of MCM4/6/7 helicase sample. (A)** Classical ring-shape and side views of and MCM helicase; **(B)** open rings with different level of aperture (pink arrowhead points to the gap); **(C)** ring-shaped (first two *left* panels) and side views (last three *right* panels) of the helicase, displaying a sharp extra electron density (pointed by a red arrowhead); **(D)** a side view with two light extra electron densities (pointed by yellow arrowheads) and **(E)** putative side views of monomers.



**Figure 38. KenDerSOM analysis.** (A) Analysis of a zone showing two extra densities (demarcated by a red dashed contour) in (B) a reference-free class average. Numbers refer to the particles on each class.

Interestingly, some helicase pairs can be found bound to each other through a thin cord (Figure 39 A), suggesting that protruding extensions are involved in the interaction between helicase particles. Additionally, some string-like structures formed by a variable number of helicase particles were also present (Figure 39 B).



**Figure 39. Closely interacting helicase particles.** Representative images of (A) helicase pairs interconnected by a fine thread in between (pointed by orange arrowheads), and (B) strings formed by helicase particles.

#### 4.2.2.2. 3D reconstructions and fitting of atomic structures

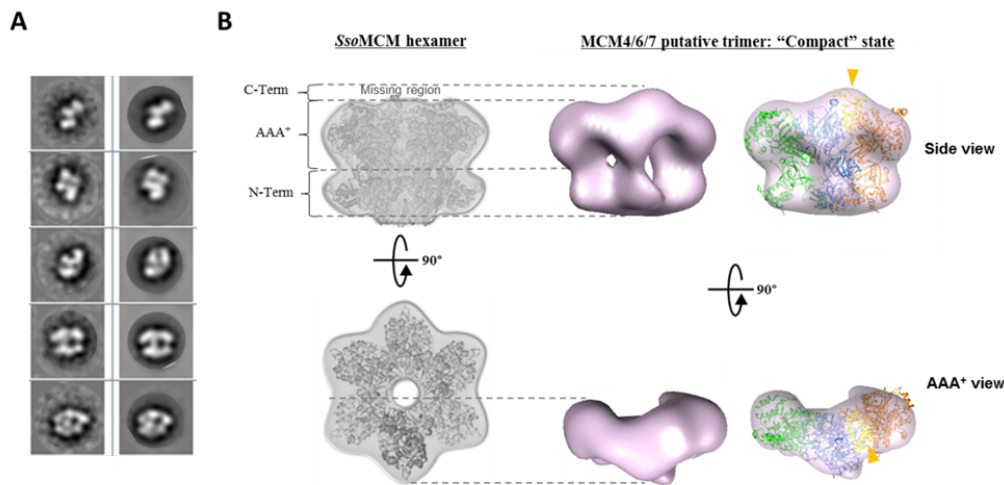
To construct active ATPase sites within the helicase, a proper alignment between subunits must occur; in this way, the correct *cis* and *trans* interactions of the catalytic motifs at the interfaces of the proteins are formed (Figure 9) (Brewster et al. 2008; Miller et al. 2014). The intersubunit nature of the active sites, in conjunction with the proposed protomer arrangement of MCM4/6/7x2 (Figure 10 B), imply that, despite having the same protomer composition, the two trimers do not have symmetric intersubunit interactions. That is, in the hexamer there would be both a MCM7→4→6 and a MCM6'→4'→7' trimer (hereinafter an apostrophe sign will be used to differentiate the two subunits of the same type of protein for each of the two trimers), where the subunits to the left of the arrow provide the motifs in *trans*, while the subunits to the right contribute with the motifs in *cis*. The organization of the MCM7→4→6 trimer has normally been considered to be the same than the one that is contained in the helicase MCM2-7(Figure 10). However, because the two trimers are not equally organized, their structures do not necessarily be exactly the same, neither their functions within the complex. Indeed, all MCM4/6/7x2 structures presented in this work show a clear symmetry mismatch. An important additional consideration is the fact that ATP binding promotes the stabilization of the hexamer, while ATP hydrolysis has a destabilization effect that shift the equilibrium toward the trimer (Ma et al. 2010).

Using negative-stain EM, single-particle classification and reconstruction methods, we obtained multiple 3D maps corresponding to different oligomeric and conformational states of mouse MCM4/6/7. In order to facilitate the reader to recognize the side and front views within our 3D maps, as well as the three principal structural regions (N-terminus, AAA<sup>+</sup> and C-terminus), we have adopted a similar lay out as the one in Figure 7 C. Importantly, it must be noted that fitting of atomic structures within the presented low resolution density maps is only a rough guide of their likely relative positions, and does not intend to represent the absolute orientation in any of the cases.



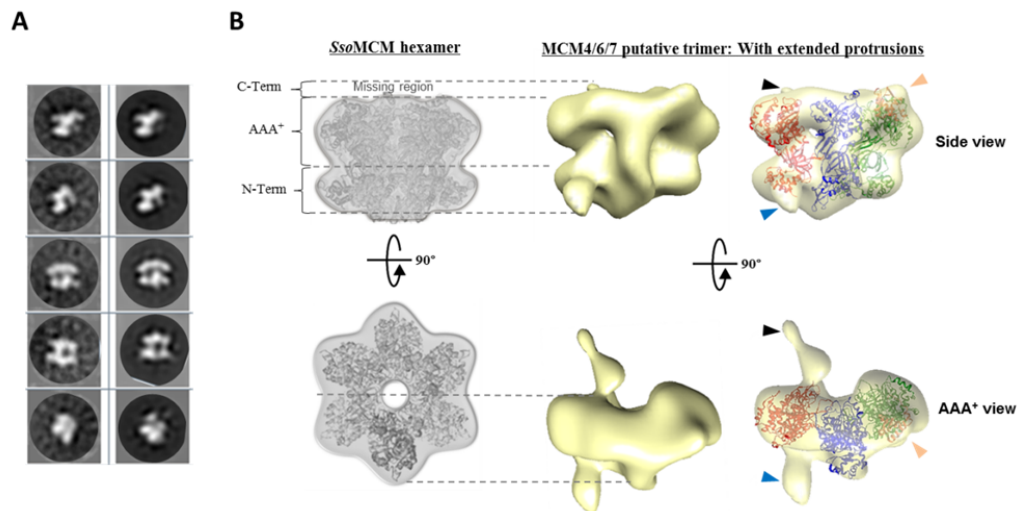
## Identification of isolated MCM trimer configurations

To date, a number of biochemical works have reported the formation of a MCM4/6/7 heterotrimer (Musahl et al. 1995; Sherman et al. 1998; Ma et al. 2010; Ichinose 1996; Ishimi 1997b), but there is not further information at the structural level. Along the process of classification and refinement of the different 3D maps reconstructed from our analyzed EM sample data, some of the obtained structures presented a principal body with shape and size compatible with the mass of a MCM trimer. Some structural variants of this putative trimer were observed; an apparently more compact state as well as conformations showing two extended and flexible protrusions emerging from the principal body. Hence, we present a representative structure of each of these putative states. A first 28 Å estimated resolution 3D map shows three column-like structures bound together by their top and bottom parts, creating a complex that resembles the side-view of the hexameric helicase (Figure 40). Three copies of the crystallographic structure of the near full-length *Sso*MCM (PDB-3F9V) fits well on this volume, which shows, unequivocally, that this is a MCM trimer. An additional feature is an extra density located at the interface between two of the subunits, protruding above what could correspond to their AAA<sup>+</sup> region. Indeed, it is known that MCM6 contains a C-terminal domain much bigger than the ones of MCM4 and MCM7, and we noted that the atomic coordinates of the HTH C-terminal structure of *Hs*MCM6 (PDB-2KLQ) can be nicely accommodated in this extra space.



**Figure 40. Putative MCM4/6/7 trimer reconstruction of a likely compact conformation.** Designation of the C-terminal AAA<sup>+</sup> and N-terminal domain regions has been done by comparison with a *SsoMCM* hexamer model. **(A)** Reference-free class averages (*left* column) and the corresponding forward projections (*right* column) of the 3D model. **(B)** From left to right there are shown the side views (*top* panels) and AAA<sup>+</sup> views (*bottom* panels) of: an *in silico* assembled hexamer model of *SsoMCM* lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of a putative MCM4/6/7 trimer and the fitting within the same map of three copies of the *SsoMCM* monomer (PDB-3F9V, in light-green, light-blue and orange) and the C-terminal domain of *hsMCM6* (PDK-2KLQ, in yellow). Yellow arrowhead points to the extra density where the C-terminal structure has been fitted.

The second map, resolved at an estimated 26 Å resolution, also shows dimensions that are congruent with a possible MCM trimer, although at lower threshold levels features two additional, extended electron densities (Figure 41). One of these densities extrudes from the putative AAA<sup>+</sup> region of one of the side protomers and projects to the inner part of the trimer. A second extra density is projected to the external side of the trimer and departs from the putative N-terminal region, at the interface between the subunit with the putative AAA<sup>+</sup> extension and the protomer at the middle of the trimer. By their position, these protrusions could correspond to extended versions of the C-terminal of MCM6 and the N-terminal extension of MCM4. 2D averages and single particle images displaying small and flexible extra densities, respectively, suggest that these projections are not artifacts of the 3D reconstruction. Finally, there is a small extra mass protruding from the interface between the central protomer and the one opposite to the firstly described two extended densities; in this space the C-terminal structure of *MthMCM* (PDB-2MA3) can be fitted. Due to the expected flexibility of the projections and the limited resolution of the maps, it is not possible to ensure if the differences between the maps of the two trimers are just variations of the same complex, or if they really point to structural characteristics inherent to each of the two putative trimers, MCM7→4→6 and MCM6'→4'→7'.

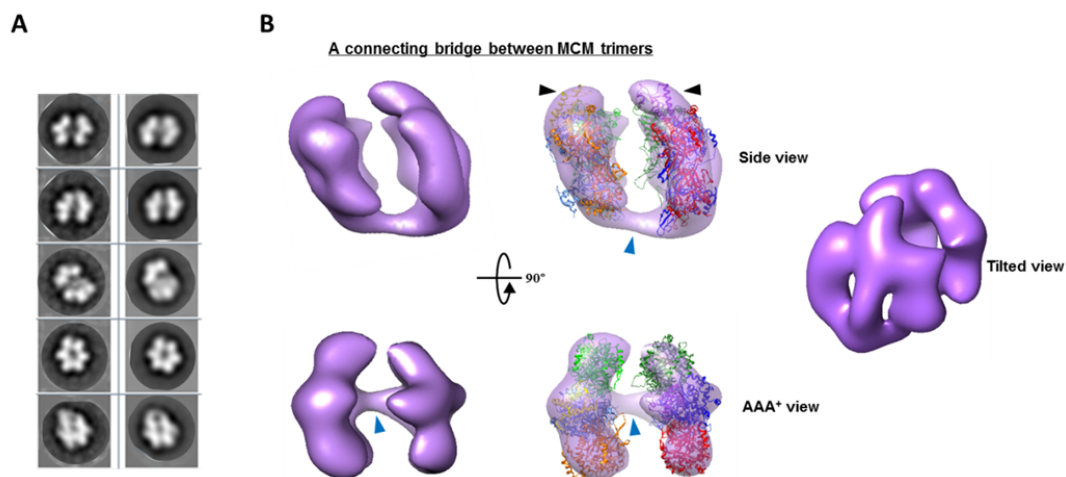


**Figure 41. Putative MCM4/6/7 trimer reconstruction showing extended densities.** Designation of the C-terminal AAA<sup>+</sup> and N-terminal domain regions has been done by comparison with a *Sso*MCM hexamer model. **(A)** Reference-free class averages (*left* column) and the corresponding forward projections (*right* column) of the 3D model. **(B)** From left to right there are shown the side views (*top* panels) and AAA<sup>+</sup> views (*bottom* panels) of: an *in silico* assembled hexamer model of *Sso*MCM lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of a putative MCM4/6/7 trimer and the fitting within the same map of three copies of the *Sso*MCM monomer (PDB-3F9V, in red, blue and green) and the C-terminal of *Sso*MCM (PDB-2M45, in pale-pink). Pale-pink arrowhead points to the area where the C-terminal atomic coordinates have been fitted. Extra densities that would putatively corresponded to HTH motif of MCM6 and N-terminal extension of MCM4, are pointed with black and blue arrowheads, respectively.

## Incipient dimerization of MCM trimers

Mutational and biochemical studies on mouse MCM4 have implicated residues at its N-terminal domain (more specifically in the zinc finger) as essential for dimerization of MCM4/6/7 trimers (You et al. 2002). Along this line, one of our 3D maps, resolved at an estimated resolution of 30 Å, presents two principal bodies, each one likely to be a MCM trimer, that are connected through a bridge emerging at one side of the two putative trimers, near the protomer of the middle (Figure 42), which could correspond to MCM4 (Figure 10). After fitting six copies of *Sso*MCM monomer (PDB-3F9V), it remains empty space within the bridge, which would be putatively filled by the interacting N-terminal extensions of the MCM4 subunits. Additionally, both trimers had an extra density extruding at the opposite site of the bridge, near to the subunit of the middle. Each of these densities has enough room to accommodate the atomic coordinates of the C-terminal structure of *hs*MCM6 (PDB-2KLQ).

This 3D reconstruction would provide direct visual evidence in support of the oligomerization role of the N-terminal of MCM4, which is involved in the assembly of two MCM trimers into the heterohexameric complex MCM4/6/7x2. However, further labeling studies for mapping the N-terminus of MCM4 are required to confirm the identity of the observed bridge.



**Figure 42. Incipient dimerization stage of MCM trimers.** (A) Reference-free class averages (*left* column) and matching forward projections (*right* column) of the density map. (B) Putative side (*top* panels) and AAA<sup>+</sup> (*bottom* panels) views of the 3D map and the respective fitting of six copies of the SsoMCM monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Blue arrowheads point to an extra density that connects the two trimers. Black arrowheads point to two extra densities at the putative C-terminal region, each with a copy of the HTH motif of HsMCM6 (PDB-2KLO, in light-orange and purple) fitted inside. An intermediate tilted view of the map is shown to get a better idea of its overall shape.

## Architecture of the heterohexamer MCM4/6/7x2: Structures of multiple conformational classes exemplifying the versatility of the complex

Hexameric MCM helicases have been shown to be highly dynamic molecular machines, able to make conformational changes implying different movements of their N-terminal, AAA<sup>+</sup> and C-terminal regions, as well as partial or complete disruption of the binding interface between some of the subunits.

A number of 3D maps, with estimated resolutions between 28-33 Å (Table 2), corresponding to different structural classes adopted by the MCM4/6/7x2 heterohexamer, were classified and refined as described in Materials and Methods (Figure 16). Common to all classes is an almost planar configuration that can be divided into three regions, one on top of the other, containing the N-terminal (at the bottom), the AAA<sup>+</sup> (at the middle), and the protruding C-terminal (at the top) domains of some of the protomers. Six copies of the near full-length crystal structure of *Sso*MCM (PDB-3F9V) can be fitted as rigid bodies into each of the 3D EM maps; some densities left unoccupied, may correspond to N-terminal or C-terminal extensions present in some of the mouse MCM subunits. Consistently with the dimensions showed by an archaea MCM hexamer model lacking its C-terminal region (side view in Figure 7 C and Figure 35), the N-terminal region of our maps is less voluminous than the AAA<sup>+</sup> region, however, additional masses at the putative C-terminal region of some protomers are systematically observed. In all cases, an asymmetrical top and bottom opening of the central helicase channel is observed, being the one at the AAA<sup>+</sup> region the wider. Up to now, works on MCM2-7, the proposed replicative helicase, suggest that it depends on additional proteins or special buffer conditions for both helping to seal its MCM2/5 gate and acquire a planar configuration which, respectively, seem to be possible mechanisms for avoiding DNA to slide out through the ring gap and to get the appropriate intersubunit alignment required for an initial activation of the complex.

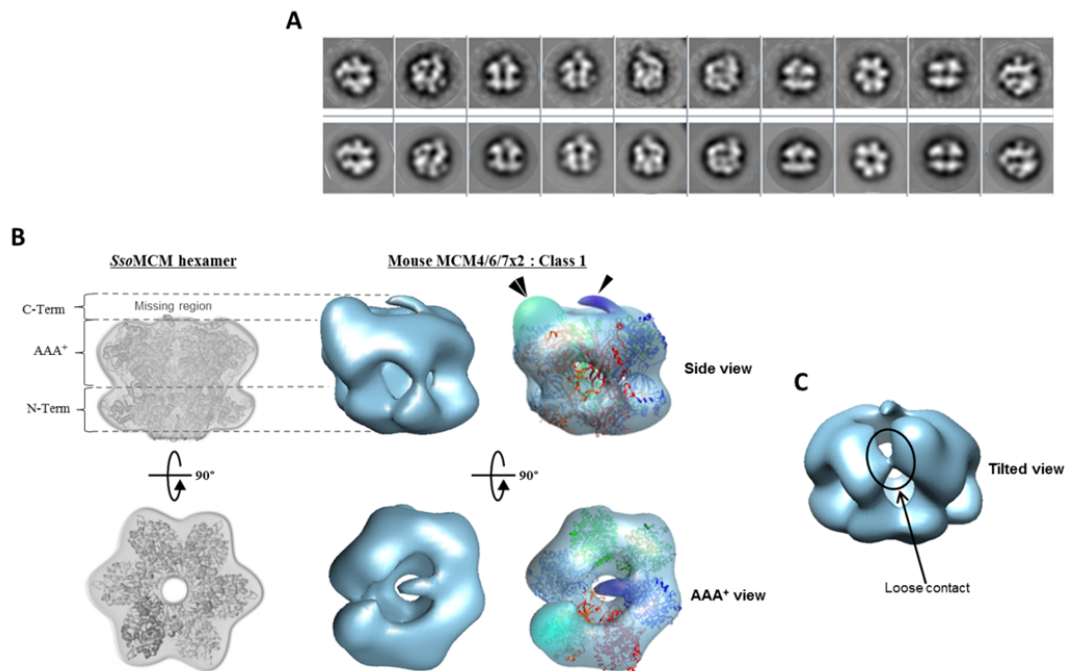
To gain further insights into MCM4/6/7x2 organization, we made a comparative analysis of the structural features between our 3D reconstructions and a number of reported MCM2-7 structures. As a group, our set of EM maps is a revealing presentation of mostly planar conformations of closed, open or nicked structures that, analyzed together, suggest a number of changes involving the movement and interaction of the putative C-terminal domains of two consecutive protomers, as well as the disruption and creation of new contacts between different subunits. In the following we will analyze each of these classes.

Map Class	Estimated resolution (FSC 0.5 cutoff)
1	27 Å
2	33 Å
3	32 Å
4	31 Å
5	30 Å
6	28 Å

**Table 2.** Estimated resolution of the 3D maps of MCM4/6/7x2 hexamers Class 1 to 6, by the Fourier Shell Correlation (FSC) criteria at a value of 0.5.

***Class 1: Closed ring structure on which the proposed architecture of the hexamer is presented***

A first MCM4/6/7x2 map, corresponding to *Class1*, presents the general shape of a two-tiered ring (Figure 43). The N-terminal opening of the central channel forms a pore with a diameter of around 25Å, and the aperture toward the AAA<sup>+</sup> region is around twice in size (diameter of around 60 Å x 45 Å). Two extra masses protruding at different points from the AAA<sup>+</sup> region of the map stand out; a small wedge and a large lump (densities displayed in dark blue and cyan, respectively, in Figure 43 B).

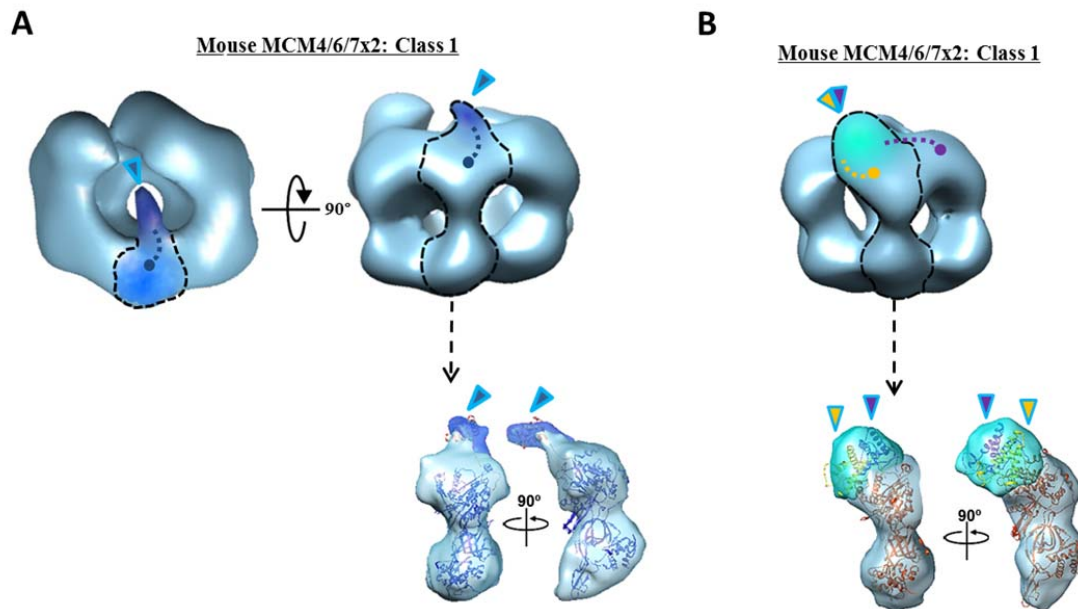


**Figure 43. Class1: MCM4/6/7x2 closed ring displaying two different extra masses, a big and a small one. (A)** Reference-free class averages (*first row*) and the corresponding forward projections (*second row*) of the 3D map. **(B)** From left to right there are shown the side views (*top panels*) and AAA<sup>+</sup> views (*bottom panels*) of: an *in silico* assembled hexamer model of *SsoMCM* lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model *Class 1*, and the fitting within the same map of six copies of *SsoMCM* monomer (PDB-3F9V), where two additional empty densities at the C-terminal region, are highlighted in cyan and dark blue, and pointed by a double or single arrowhead, respectively. **(C)** Tilted view of the map, showing a point of weak contact observed between two of the protomers.

The first mass we want to draw the attention to is the smaller density, which is clearly visible both in front and side views (density displayed in dark blue in Figure 44 A). This arched finger-like structure, which is subtly turned against the bigger extra mass, extends toward the opening of the central channel at the AAA<sup>+</sup> region, resulting in a spatial disruption that creates two slightly asymmetric invaginations of around 20 Å each, which is still big enough to accommodate ssDNA.

The length of the amino acid sequence from the HTH motif of both *MthMCM* (PDB-2MA3) and *SsoMCM* (PDB-2M45) is similar to the one of the C-terminus of mouse MCM4 (Figure 32 and Figure 33), and we observed that the atomic coordinates from PDB-2MA3 fitted reasonably well into the small extra density of our map. In addition to the small extra density (a putative HTH motif) described above, there is also a big protrusion emerging at the C-terminal region (Figure 44 B), which makes a loose contact with one of its neighbouring subunits (Figure 43 C). We observe

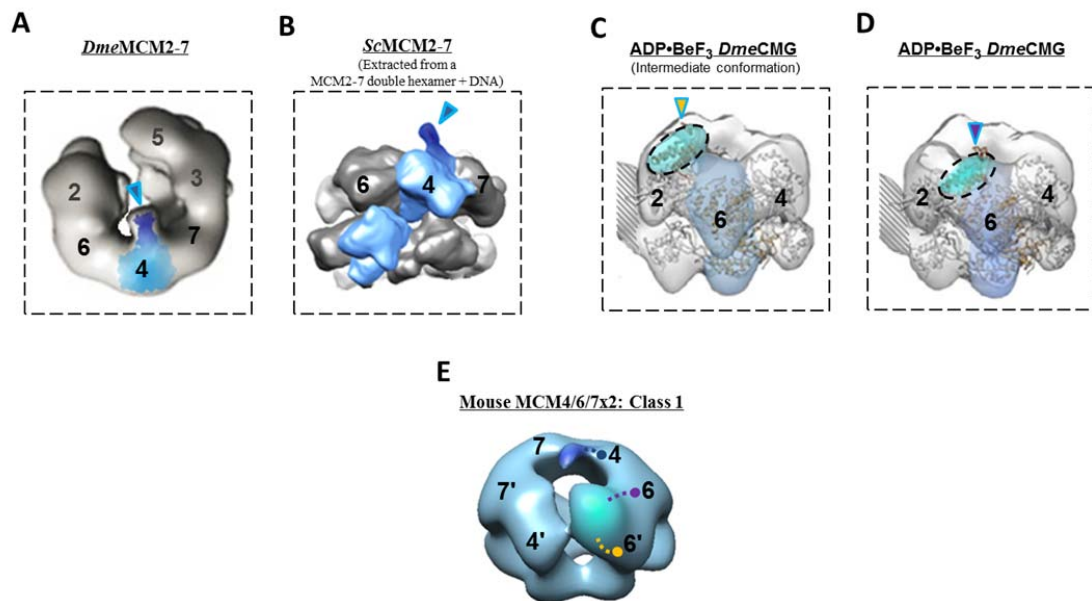
that this mass is large enough to contain two copies of the C-terminal structure of *HsMCM6* (PDB-2KLQ).



**Figure 44. Fitting of atomic structures within two extra masses protruding from MCM4/6/7x2 map Class 1. (A)** AAA<sup>+</sup> and side views of the map, where a black dashed contour highlights a protomer with a small extra density (dark blue-blushed area pointed by a cyan-contour blue arrowhead). Blue dotted-line represents a symbolic linkage between the AAA<sup>+</sup> region of a subunit and its putative HTH domain. Two different side views of the segmented subunit are shown, with the structure of *SsoMCM* (PDB-3F9V) and C-terminus of *MthMCM* (PDB-2MA3) fitted inside. **(B)** Side view of the map, where a black dashed contour highlights a protomer with a big extra density (cyan-blushed area pointed by a cyan-contour double-colored yellow/purple arrowhead). Yellow and purple dotted-lines represent a symbolic linkage between the AAA<sup>+</sup> region and the respective putative HTH domains of two consecutive subunits. Two different side views of the segmented subunit are shown, with the structure of *SsoMCM* (PDB-3F9V) and two copies of HTH *HsMCM* motif (PDB-2KLQ) fitted inside.

Interestingly, the features above-mentioned have, in some way, been partially described in previous EM maps of different MCM2-7 helicase complexes (Figure 45), a fact that has provided us with grounds to propose a tentative structural comparison between our current *Class 1* map and those other previously reported ones.





**Figure 45. Proposed architecture of MCM4/6/7x2 map *Class1*, based on structural similarities with a number of MCM2-7 EM structures featuring some comparable extra masses.** Structure of **(A)** *DmeMCM2-7* open ring (figure adapted from (Lyubimov et al. 2012)) and **(B)** *ScMCM2-7* hexamer (extracted from a MCM2-7 double hexamer + DNA complex, EMD-5857), where cyan-contour blue arrowheads point to and small density (dark blue-blushed area) protruding from MCM4. **(C)** Structure of an alternate conformation of ADP•Be<sub>3</sub> bound *DmeCMG* (Figure adapted from (Costa et al. 2011)), where cyan-contour yellow arrowhead points to the cyan-blushed area that is partially over MCM2, which have the C-terminal structure of *HsMCM6* (PDB-2KLQ) fitted. **(D)** Structure of ADP•Be<sub>3</sub> bound *DmeCMG*, where cyan-contour purple arrowhead points to the cyan-blushed area at the MCM2/6 interface with the C-terminal structure of *HsMCM6* (PDB-2KLQ) fitted (Figure adapted from (Costa et al. 2011)). **(E)** Designation of all the subunits within mouse MCM4/6/7x2 map *Class1*, where each, blue, purple and yellow dotted-lines represent a symbolic linkage between the AAA<sup>+</sup> region of a subunit and its putative HTH domain. For all cases, numbers represent the different subunits. Black-dashed boxes enclose structures resolved in studies other than the present one.

Indeed, similar to our observed small extra mass, in a report showing a EM map of *DmeMCM2-7* (structure not deposited in the Electron Microscopy Data Bank, EMD), a small wedge-like structure is observed arising from the AAA<sup>+</sup> domain of MCM4 (Figure 45 A) (Lyubimov et al. 2012). Furthermore, in a ssDNA loaded *ScMCM2-7* double-hexamer complex (EMD-5857), a small and elongated extra density extruding from MCM4 of both rings is observed, although in this case it extends upwards (Figure 45 B) (Sun et al. 2014). Perhaps this change of orientation is due to a steric exclusion produced by the presence of the ssDNA, or as a result of the general conformation acquired by the two interlocked and misaligned hexamers. It must be underlined that in the above two mentioned MCM2-7 structures this small extra mass is slightly bent in direction to MCM6. All the above observed similarities allowed us to propose that the small protrusion may correspond to the HTH domain of MCM4, so that it can be used to localize the

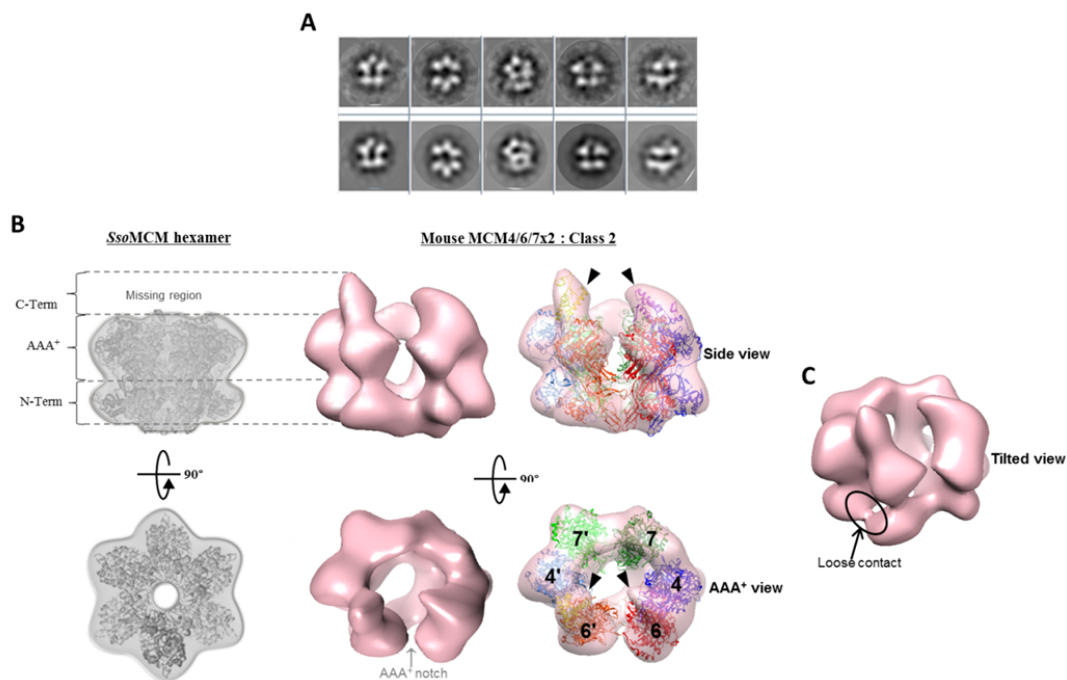
position of this subunit in our map and, at the same time, suggest the position of its neighbouring subunits, MCM6 and MCM7 (Figure 45 E).

On the other hand, following the line of the proposed fitting of two copies of the HTH MCM6 motif inside the big extra mass in our map Class1, there are some works that give support and insights about this idea. Given the putative position of MCM4 (identified by the small extra mass), the location of the lump and the model of the subunit distribution of MCM4/6/7x2 (Figure 10B), all these observations could imply that the subunit MCM6 besides the MCM4 with the finger-like protrusion extends its C-terminus all over MCM6', so the HTH parts of both, MCM6 and MCM6', can interact. A structural study on the *Drosophila* CMG complex would support similar positional shifts of the winged helix domain of MCM6 (Costa et al. 2011). Upon binding of ADP•BeF3 by DmeCMG complex (structure not deposited in EMDDB), an intermediate configuration of the DmeMCM2-7 ring has a MCM6 subunit with a putative extra mass extruding from its AAA+ domain that is partially over MCM2 (Figure 45 C). In an alternative conformation of *Drosophila* ADP•BeF3 CMG complex (EMD-1832), the putative HTH domain of MCM6 seems to be displaced over the MCM2/6 interface, apparently completely recessed against the AAA+ region (Figure 45 D). When displaying at lower contour levels the **recently** 3.8Å solved structure of the ScMCM2–7 double hexamer (EMD-6338) (Li et al. 2015), it is observed that HTH domain of MCM6 stands on the AAA+ interface of MCM2/6. Making use of all these observations, we propose that in our presented map of the closed ring the interacting HTH motifs of the MCM6-6' protomers would form the big extra electron density. In view of all the foregoing, having putatively determined the position of MCM6, MCM6' and MCM4, and following the model proposed by previous biochemical studies (Yabuta et al. 2003; Yu et al. 2004; Xu et al. 2013) (Figure 10B), the remaining subunits can also be putatively positioned (Figure 45 E).

### ***Class 2: Notched ring showing two extensions extruding from the AAA<sup>+</sup> region***

3D map *Class 2* (Figure 46) reveals the conformation of a ring with a discontinuity at the AAA<sup>+</sup> region, which is flanked by two protomers displaying a prominent mass at their respective C-terminus. Indeed, in spite of their different shapes, one molecule of the HTH motif of *Hs*MCM6 (PDB-3F9V) can be fitted inside each of these extensions; hence these protuberances are assigned to the HTH motif of each of the two MCM6 protomers in the complex. Considering that these

protuberances are flanking the notch, we propose that the two protomers at the nick interface of this map may be the copies of the protein MCM6. The protomer presenting the most extended C-terminus looks very loosely bound by its N-terminal part to the neighbouring subunit at the contrary side of the nick (Figure 46 C). This somehow apparently unstable architecture could perhaps be interpreted as an indicative of being a transitional state that moves toward a more stable conformation. Having putatively assigned the positions of the two MCM6s, and following the proposed model organization of MCM4/6/7x2 (Figure 10 B and Figure 45 E), we can complete the assignation of the remaining subunits.

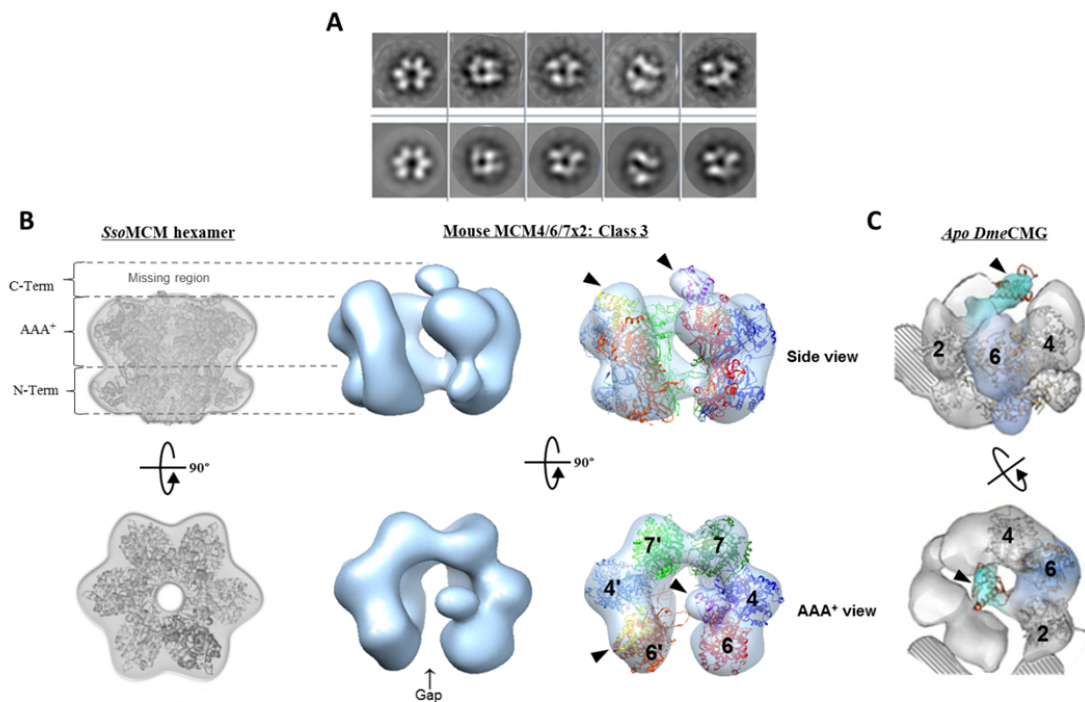


**Figure 46. Class 2: MCM4/6/7x2 complex with two densities protruding from a notched AAA<sup>+</sup> region. (A)** Reference-free class averages (first row) and the corresponding forward projections (second row) of the 3D map. **(B)** From left to right there are shown the side views (*top panels*) and AAA<sup>+</sup> views (*bottom panels*) of: an *in silico* assembled hexamer model of SsoMCM lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model *Class 2*, and the fitting within the same map of six copies of SsoMCM monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Black arrowheads point to two extra densities at the C-terminal region, each with a copy of the HTH motif of HsMCM6 (PDB-2KLQ, in light-orange and purple) fitted inside. **(C)** Tilted view of the map showing a point of weak contact between the N-terminal parts of two protomers. Numbers represent the putative organization of the different subunits.

### ***Class 3: Open ring with a C-terminal protrusion that partially occludes the central pore***

The structure of an almost planar, open ring, is observed in *Class 3* map (Figure 47 A and B). We draw the attention to both protomers at the gap, which are notably longer than the rest at the C-terminal region, letting enough space for the HTH structure of *HsMCM6* (PDB-2KLQ) to be fitted in addition to the respective copy of the almost full-length *SsoMCM* monomer (PDB-3F9V). Following a similar reasoning used in the analysis of map *Class 2*, we propose that the opening of this volume occurs at the MCM6'-6 interface, allowing us to design the position of the remaining subunits. The putative C-terminus of MCM6' looks almost fused with the AAA<sup>+</sup> domain, suggesting that the HTH motif could be in a recessed position. On the other hand, the putative MCM6 monomer is somehow displaced upward in relation to the horizontal axis of the bottom wheel, and its C-terminal part looks like a lobe projected toward the central channel of the ring, so that it is partially obstructing this pore. Similar C-terminal structures can be recognized in some MCM helicase density maps from other organism. For example, an extended density at the C-terminus of MCM6 is observed, at low contour levels, in the structure of an apo *DmeCMG* complex (Figure 47 C) (Costa et al. 2011). HTH domains positioned against the central channel had been also visualized in one of the rings of the *MthMCM* double hexamer, in presence of dsDNA (Costa et al. 2006a). Furthermore, the **recently** solved structure of a *HsMCM2-7* + DNA complex shows one prominent C-terminal extension (Hesketh et al. 2015).

In a study of the C-terminal winged helix fold domain of *hsMCM6*, this region, by itself, failed to bind DNA (Wei et al. 2010). An enhanced affinity for ssDNA and dsDNA was observed in a *MthMCM* C-terminal deletion mutant, suggesting that the HTH motif might be implicated in controlling DNA accessibility to the helicase pore (Jenkinson & Chong 2006). The aforementioned observations let us think that perhaps, the putative HTH MCM6 that extend toward the central channel, in the *Class 3* map, could be a flexible structure acting as a transient steric impediment for DNA to go into the central channel of the helicase under certain conformation of the hexamer.

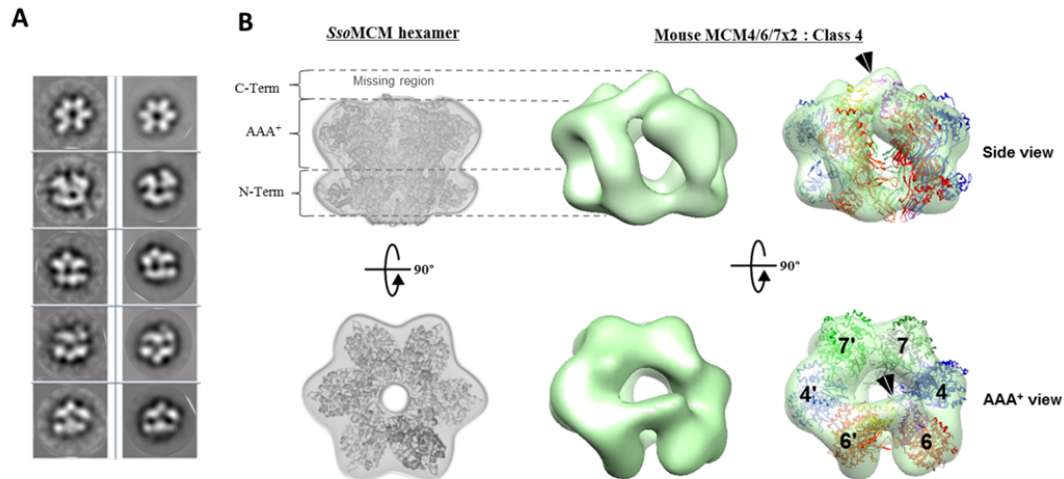


**Figure 47. Class 3: MCM4/6/7x2 open ring with two longer subunits at the gap interface.** (A) Reference-free class averages (first row) and the corresponding forward projections (second row) of the 3D map. (B) From left to right there are shown the side views (*top* panels) and AAA<sup>+</sup> views (*bottom* panels) of: an *in silico* assembled hexamer model of *SsoMCM* lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model *Class 3*, and the fitting within the same map of six copies of *SsoMCM* monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Black arrowhead points to an extra density at the putative C-terminal region of the complex, where the HTH motif from *HsMCM6* (PDB-2KLQ, in purple) is fitted. (C) Tilted and AAA<sup>+</sup> views of apo *DmeCMG* (EMD-1833), with models of MCM subunits 2, 4 and 6, and the *HsMCM6* HTH structure (PDB-2KLQ) fitted inside. The density corresponding to MCM6 is highlighted with blue blush and the one containing the atomic coordinates of human HTH MCM6 motif (PDB-2KLQ), pointed by a black arrowhead, is highlighted with cyan blush. (Figure adapted from (Costa et al. 2011)). For all cases, numbers represent the different subunits.

#### ***Class 4: Loosely closed ring with two interacting C-terminal protrusions***

The toroidal structure of map *Class 4* (Figure 48) presents a highlighting feature at the connecting interface between two of the subunits, which is looser than in the rest, both at the bottom and at the top. At this interface, the link of the upper part seems to be formed between the C-terminus of the two flanking protomers, contrasting with other pair interaction between proteins, which make contact at the level of their AAA<sup>+</sup> domains. The fact that two copies of the HTH motif of *HsMCM6* (PDB-3F9V) can be fitted along this C-terminal connection, suggest that the loose interface is formed by the putative MCM6'-6 proteins. This putative HTH-HTH connection goes in the same line of our idea that the big lump presents in *Class 1* map could be formed by the interacting HTH domains of both MCM6s.

protomers, the rest of the subunits were organized according with the model presented in Figure 10 B.

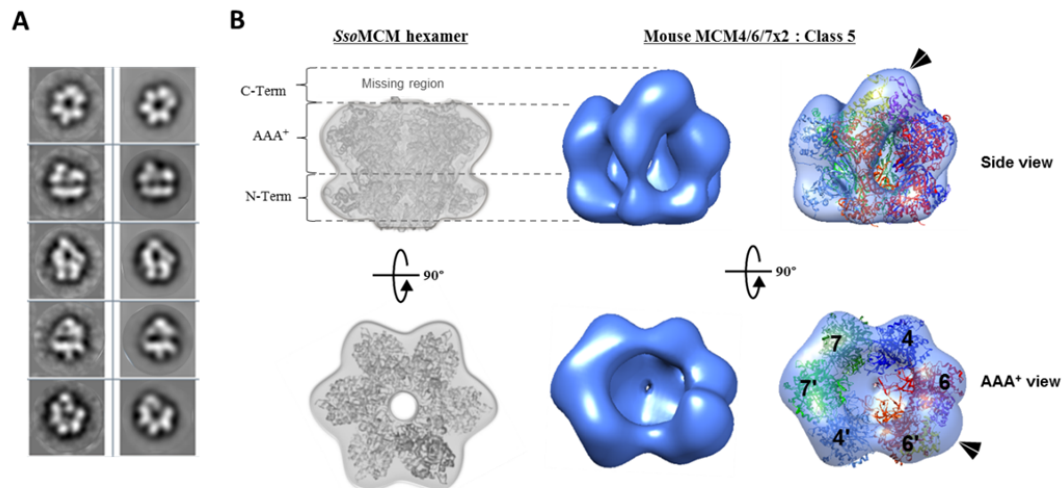


**Figure 48. Class 4: MCM4/6/7x2 loosely closed ring with a C-terminal connecting bridge.** (A) Class average of experimentally acquired images (*left column*) and matching forward projections (*right column*) of the reconstructed model. (B) From left to right there are shown the side views (*top panels*) and AAA<sup>+</sup> views (*bottom panels*) of: an *in silico* assembled hexamer model of *SsoMCM* lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model *Class 4*, and the fitting within the same map of six copies of *SsoMCM* monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Black double-arrowhead points to an extra density that creates a connecting bridge between the putative AAA<sup>+</sup> regions from two adjacent subunits, where two copies of the HTH motif from *HsMCM6* (PDB-2KLO, in light orange and purple) can be fitted. Numbers represent the putative organization of the different subunits.

### ***Class 5: Closed ring with a N-terminal sealed pore and a big lump at its C-terminal region***

In *Class 5* map (Figure 49), the opening of the central channel at the AAA<sup>+</sup> region of the ring is quite wide (diameter of around 50Å), contrasting with its N-terminal region, which is constricted in such a way that is practically sealing the pore at this side, a feature that will avoid DNA to cross along the helicase central pore. Another remarkable feature is a prominent mass placed at the C-terminal region, extruding between two subunits. Given that two copies of the HTH motif of *HsMCM6* (PDB-3F9V) can be accommodated within the lump, and following the previously exposed reasoning for map *Class 1*, we propose that this extra mass could be designated to be formed by the interacting C-terminal motifs of the putative MCM6<sup>1-6</sup> protomers, positioned at

both sides of the mass. The model presented in Figure 10 B was followed to complete the arrangement of the remaining subunit.

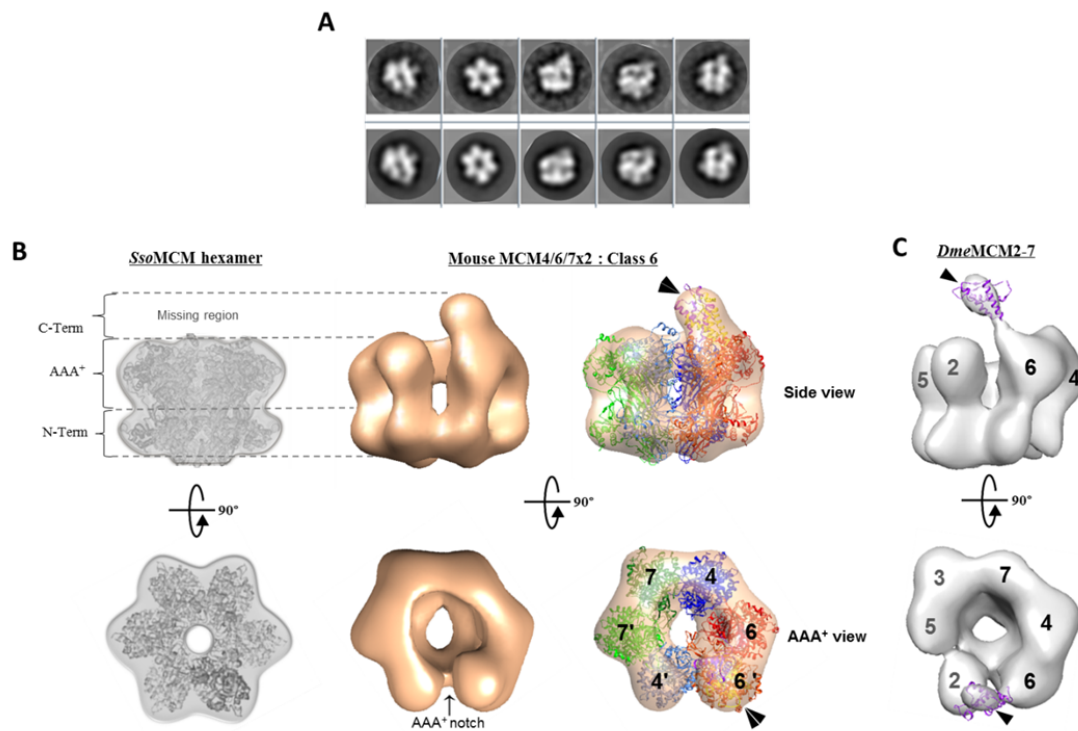


**Figure 49. Class 5: N-terminal sealed MCM4/6/7x2 ring with a big C-terminal electron density. (A)** Reference-free class averages (left column) and matching forward projections (right column) of the density map. **(B)** From left to right there are shown the side views (top panels) and AAA<sup>+</sup> views (bottom panels) of: an *in silico* assembled hexamer model of *SsoMCM* lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model Class 5, and the fitting within the same map of six copies of *SsoMCM* monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Black double-arrowhead points to a big extra density at the putative C-terminal region of the complex, where two copies of the HTH motif from *HsMCM6* (PDB-2KLQ, in light orange and purple) can be fitted. Numbers represent the putative organization of the different subunits.

### ***Class 6: Notched ring with a big lump at its C-terminal region***

*Class 6* map (Figure 50) shows the structure of a ring with a discontinuity at the AAA<sup>+</sup> region. One of the protomers that resides at the nick has an extra mass protruding at the C-terminus, in which two copies of the HTH structure of MCM6 (PDB-2KLQ) can be fitted, similarly to what was previously observed in the maps of *Class 1* and *Class 5*. *Class 6* EM map resembles the shape of the planar, notched conformation of a *Drosophila* MCM2-7 map (EMD-1834) when displayed at a lower contour level (Figure 50 C), where it is observed an extra mass departing from the MCM6 AAA<sup>+</sup> domain, although in this case it extends across the nick reaching a position over the next subunit and, as expected, allows to fit only one copy of the HTH structure (PDB-3F9V) (Costa et al. 2011). Considering all above observations and the model of the subunit distribution of MCM4/6/7x2 (Figure 7), we propose, again, that the extra mass corresponds to the interacting C-

terminals of both MCM6<sup>1-6</sup> subunits and that the two subunits at the AAA<sup>+</sup> notch interface should be the putative MCM4<sup>1-6</sup>. A nick or weak association between MCM4 and MCM6 has also been observed in EM models of *Dme*MCM2-7 (Costa et al. 2011) and yeast OCCM complex (Sun et al. 2013), which reflects the trend of these subunits to dissociate.



**Figure 50. Class 6: MCM4/6/7x2 complex with a big additional density protruding from one side of a notched AAA<sup>+</sup> region.** (A) Reference-free class averages (first row) and the corresponding forward projections (second row) of the 3D map. (B) From left to right there are shown the side views (*top* panels) and AAA<sup>+</sup> views (*bottom* panels) of: an *in silico* assembled hexamer model of *Sso*MCM lacking the C-terminal region (modified from (Brewster et al. 2008)); a solid colored 3D map of MCM4/6/7x2 model Class 6, and the fitting within the same map of six copies of *Sso*MCM monomer (PDB-3F9V, in orange, light-blue, light-green, red, blue and green). Black double-arrowhead points to a big extra density at the putative C-terminal region of the complex, where two copies of the HTH motif from *Hs*MCM6 (PDB-2KLQ, in light orange and purple) can be fitted. (C) Structure of apo *Dme*MCM2-7 (EMD-1834). Black arrowhead points to a C-terminal extra density protruding from MCM6, where it is fitted the structure of the *Hs*MCM6 HTH motif (PDB-2KLQ, Purple). In all cases, numbers represent the different subunits.

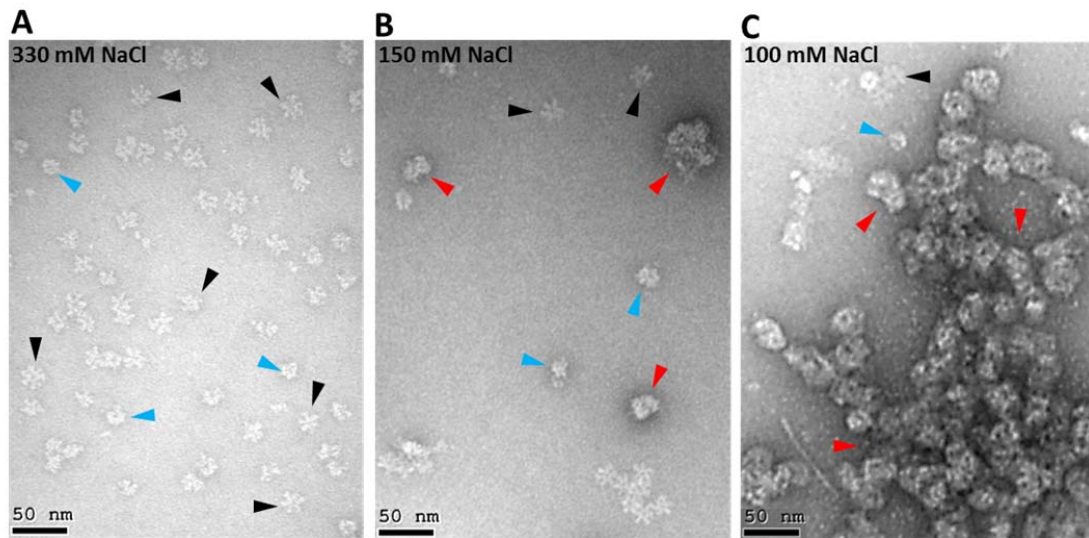


## **4.3. Human Twinkle**

### **4.3.1. Aggregation of helicase particles occurring at low salt concentration is reversible by increasing the ionic strength**

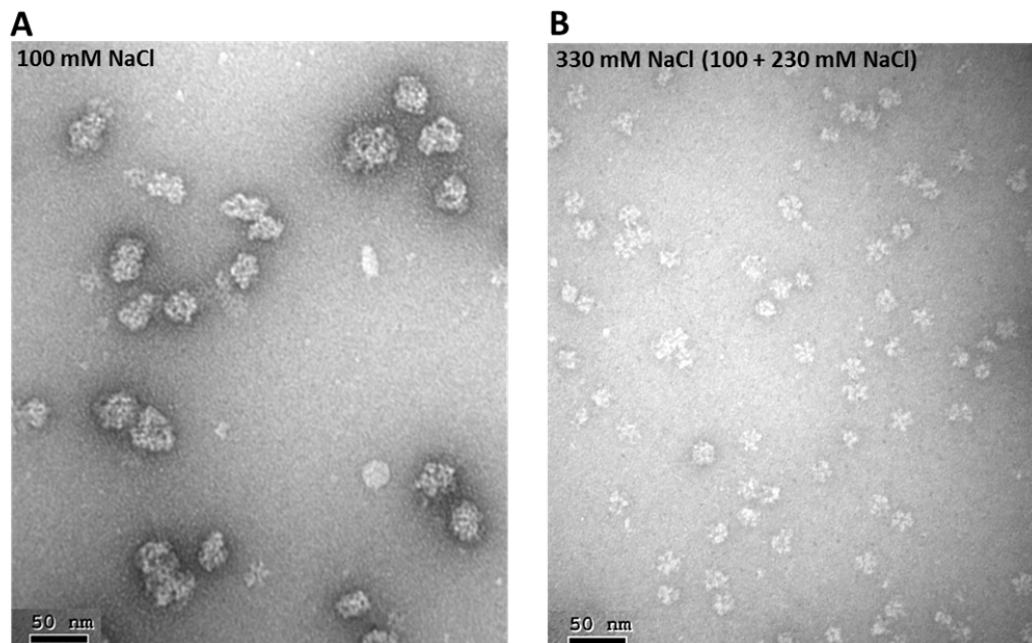
The solubility of Twinkle is greatly affected by the ionic strength of the media (Ziebarth et al. 2010). Even in presence of cofactors (MgCl<sub>2</sub> and ATP $\gamma$ -S) that help stabilize Twinkle, low concentrations of salt results in an important loss of soluble protein. On the other hand, increasing the temperature along a range of 4-37°C have shown to enhance the recovery of soluble protein when using concentrations above 150 mM NaCl (Ziebarth et al. 2010).

In order to have a direct and detailed visualization of the effect of different salt concentrations on the protein state, we monitored by EM a soluble sample at 330 mM NaCl (without adding cofactors), and then we diluted it until reaching 150 mM NaCl and 100 mM NaCl concentrations, adding MgCl<sub>2</sub> and ATP $\gamma$ -S as stabilizing cofactors (Figure 51 A, B and C, respectively). NS-EM images of the sample at 330 mM NaCl (Figure 51 A) allowed a clear visualization of a heterogeneous group of free particles, most of which showed a star-like shape with a variable number of flexible arms (each monomer). The protomers are bound to each other by one end creating a central channel. A mixture of extended and more compacted particles was observed with no signs of sample aggregation. Conversely, NS-EM evaluation of the samples at 150 mM NaCl (Figure 51 B) and 100 mM NaCl (Figure 51 C) showed clear signs of an important degree of protein aggregation, which was more pronounced in the lower salt condition, corroborating that the level of aggregation is inversely proportional to the salt concentration of the sample (Ziebarth et al. 2010).



**Figure 51. Representative NS-EM images showing the aggregation effect on a wild type Twinkle sample under three different concentrations of salt. (A) 330 mM NaCl, (B) 150 mM NaCl + Cofactors and (C) 100 mM NaCl + cofactors. Black and blue arrowheads point to either extended or more compacted helicase particles, respectively, while red arrowheads point to protein aggregates.**

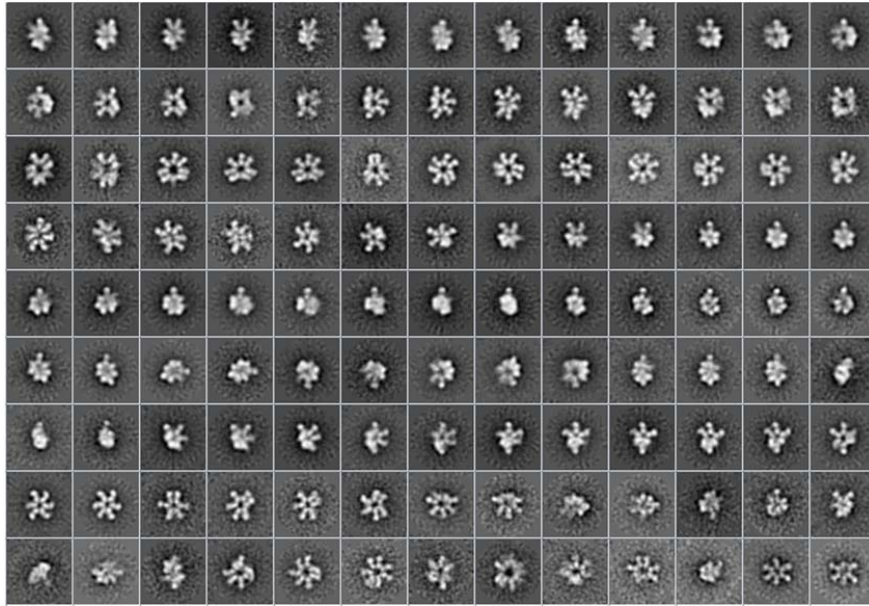
We were interested to know whether the protein aggregates observed at the lowest salt conditions could be solubilized back upon increasing salt concentration. For that purpose, we started from a partially aggregated sample at 100 mM NaCl plus cofactors (Figure 52 A), and we increased the salt concentration to 330 mM NaCl (Figure 52 B). Remarkably, it was observed not only that the aggregation disappeared, but also that particles showed their initial appearance of ring-shaped complexes with flexible arms.



**Figure 52. Recovery of helicase particles from an aggregated sample.** Representative NS-EM micrographs showing (A) a protein sample almost completely aggregated at 100 mM NaCl + cofactors, and (B) the same sample after increasing the salt concentration to 330 mM NaCl, where it can be seen the complete recovery of soluble helicase particles.

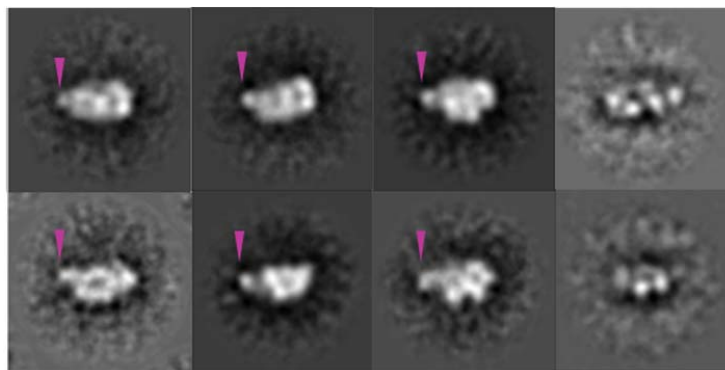
### 4.3.2. Twinkle forms multiple homo-oligomeric complexes showing long and flexible arms

The 2D image analysis and classification of 14,500 single particles selected from the full-length Twinkle sample under 330mM NaCl conditions (Figure 53), revealed a heterogeneous collection of particles composed of different oligomers with star-like shapes showing both variable extended and more compacted conformations (Figure 55). The compact particles constitute around 35% of the total, while only around 15% of the particles have all the protomers completely extended; the remaining 50% was formed by particles with a mix of extended and flexed monomers.



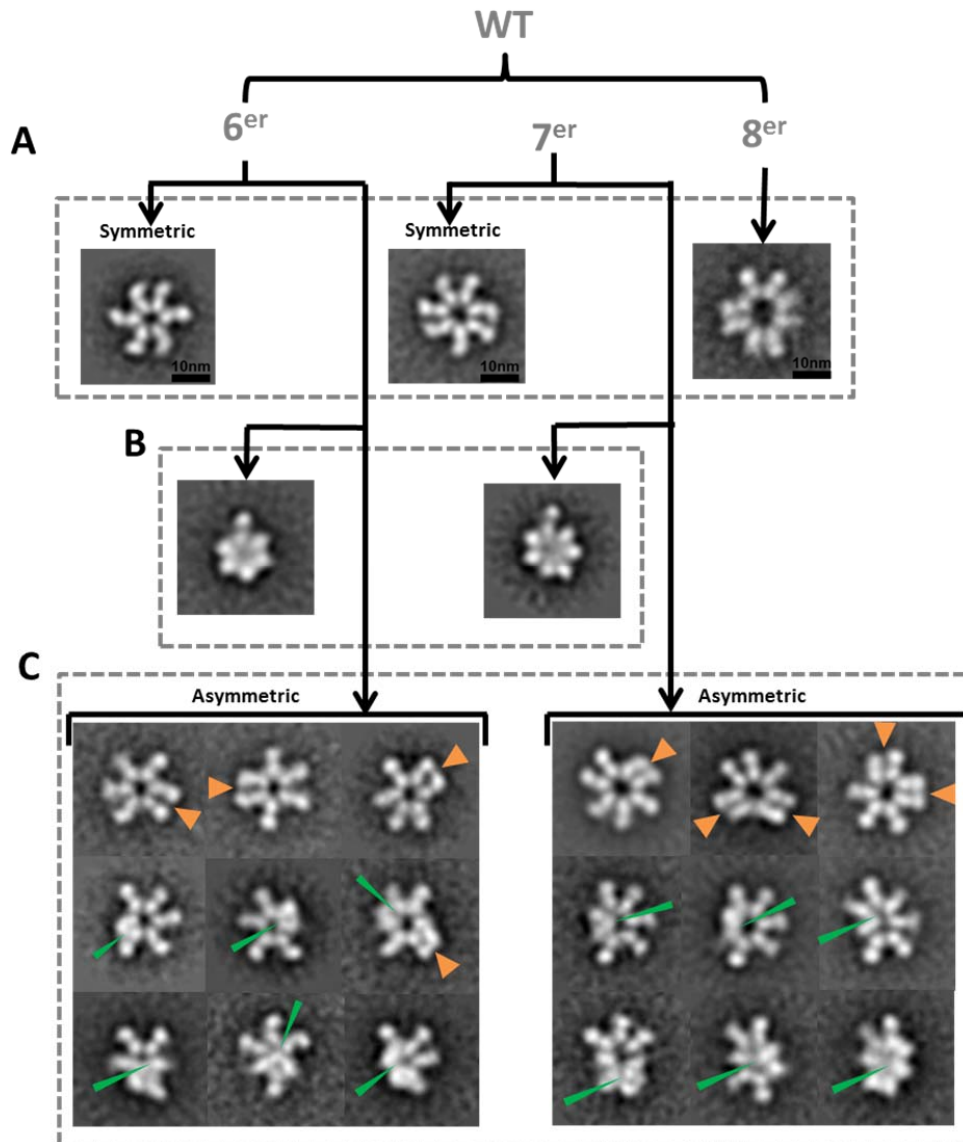
**Figure 53.** Reference-free class average classification (CL2D) of the complete set of 14500 particle images of the full-length Twinkle protein sample at 330 mM NaCl.

Most of the observed images were frontal views of the particles (Figure 53) and, based on longitudinal measurements and the classical two-tier shape of most ring-like helicases, we concluded that only a limited number of particles (about 5%) corresponded to side-views of compact conformations (Figure 54). Possible side-views of the extended particles were only observed in 55° tilted images, which was not surprising considering that the extended conformations seems to have a relatively flat disk-like shape.



**Figure 54.** Lateral views of the wild type (WT) Twinkle. NS-EM 2D class averages showing possible lateral views of the particles, where a density that could correspond to the NTD of an extended protomer is pointed with a fuchsia arrowhead. The two images in the last column at the right side of the panel came from 55° tilted images.

In addition to the already reported hexameric and heptameric complexes formed by Twinkle (Ziebarth et al. 2010; Fernández-Millán et al. 2015), here we identified a new octameric species form of the protein (Figure 55 A). Similarly to the other two lower order oligomers, the octamers presented a radial organization of its monomers, but it was a minority class with only 5% of the images.




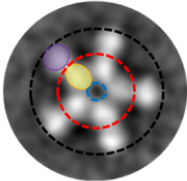
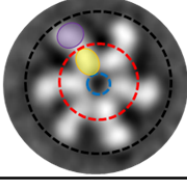
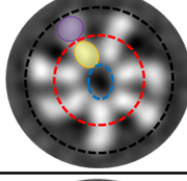
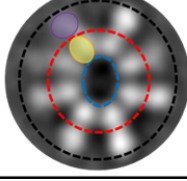
**Figure 55. Diagram showing representative NS-EM 2D classes of the wild type (WT) Twinkle sample at 330 mM NaCl. (A)** Different type of oligomers (hexamer,  $6^{er}$ ; heptamer,  $7^{er}$  and octamer,  $8^{er}$ ) formed by the protein. **(B)** Compacted particles with only one arm extended. **(C)** Variable conformational states where extended protomers that tend to form "close pairs" are pointed by orange arrowheads. Some of the protomers seem to be bent toward the center of the complex (pointed by a green arrowhead).

The observed sample population, mostly formed by similar proportions of hexamers and heptamers (Table 3), was in agreement with the expected oligomeric forms of the selected purified fractions coming from the glycerol gradients.

Oligomer state	WT (%)
Hexamer	45
Heptamer	50
Octamer	5

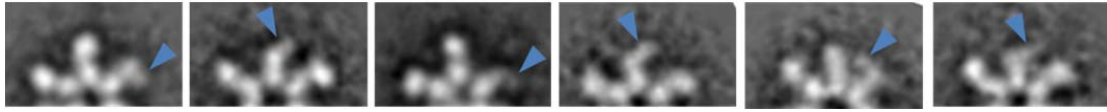
**Table 3. Approximate percentage (%) of the different oligomeric state particles for the full-length wild type (WT) Twinkle sample.**

Both the crystal structure of a T7 gp4 heptamer (Toth et al. 2003) and the 3D map of a chemically fixed hexameric form of Twinkle (Fernández-Millán et al. 2015) show a compact, modular arrangement where the CTDs form a closed ring topped by the NTDs loosely arrayed (Figure 14). Contrasting with the above mentioned compact conformations, dynamic studies of Twinkle in solution have shown that this helicase tends to acquire an extended conformation where the flexible NTDs are primarily orientated on the sides of the CTD ring (Fernández-Millán et al. 2015), which is in agreement with our NS-EM images showing mostly extended particles of Twinkle. We observed that each monomer has an articulated arm-like structure with two main globular electron densities separated by a soft density in between, which may correspond to the flexible linker connecting the CTD and the NTD. Within a Twinkle helicase particle, each of its protomers binds to the next one by an edge lobe putatively corresponding to the oligomerization CTD, creating a central channel. The lobes at the opposite side of the central channel are putatively designate as the NTDs, and showed a wide number of different orientations. A simplified, general description of the domains arrangement within the complexes can be made based on the more symmetric, arms-extended conformations (Table 4). These extended particles have a central channel composed by the putatively closely interacting CTDs, followed by a final ring area occupied by the more flexible putative NTDs. The diameter of the central channel increased accordingly with the oligomeric state of the particles, and for the three cases, hexamer, heptamer and octamers, their pore is big enough to easily accommodate ssDNA or dsDNA.

Oligomeric state		Average external diameter (nm)		
Scale bar: 		Central channel	CTD ring	NTD ring
Pentamer		~2	~11	~20
Hexamer		~3	~12.5	~21
Heptamer		~4	~13.5	~22
Octamer		~6	~14.5	~24

**Table 4.** Colored dashed circles demark the approximate external diameter of the particles' central channel (blue), CTD (red) and NTD (black) rings, for each of the different oligomeric states. A yellow and a purple oval, respectively highlight the putative CTD and NTD of one monomer.

Though very subtle, some of the subunits show a small density positioned at one side of the NTD lobe. It is noteworthy that the NTD lobes presenting this extra mass, look less dense than most of the ones where only the larger globule is observed, which suggest that this part may be formed by two modules that tend to be very close but sometimes can be more separated (Figure 56). Considering that the NTD is structurally divided into a bulky RPD and a much smaller and flexible ZBD at the terminal edge (Fernández-Millán et al. 2015), the small density observed in our images could be the ZBD. The above is in agreement with the recently solved EM structure of a compact conformation of Twinkle (Fernández-Millán et al. 2015), where the detected ZBDs are positioned at one side of its respective RPD.



**Figure 56. Putative NTD.** Blue arrowheads point to a small extra density protruding at one side of the NTD lobe which, in turn, seems to be less electron dense than the NTDs that do not present the extra mass. By its position and size, the small extra mass could be the ZBD.

A closer analysis of the conformational variability in the full set of 2D class images suggests two main types of movement of the arms (its monomers) that can be combined in multiple ways. The first class of movement is an axial displacement of the NTDs, where there is an observed tendency of the extended arms to arrange in close pairs (Figure 55 C). The second form of movement is the bending of the arms, where a variable number of NTDs bent toward the ring area of the CTDs. Several combinations of bent NTDs occur, showing different number of both neighbouring and distant subunits contracted against the center of the complex (Figure 55 C). Whether each NTD is placed on its own RPD or the one of a neighbor subunit, is a question that cannot be unambiguously answered only with this 2D image analysis. However, the solved structure of both compact conformations of a Twinkle hexamer (Fernández-Millán et al. 2015) and a T7 gp4 heptamer (PDB-1Q57) (Toth et al. 2003), showed NTDs placed on top of the CTDs from a neighbouring protomer. A second question that arises is which of the interdomain interactions, ZBD/RPD or RPD/CTD, could be responsible for a determined type of observed movements, for which it could be useful a comparative analysis between the EM images of the wild type Twinkle and those coming from a protein lacking the ZBD.

It specially caught our attention the apparent absence of conformations with all the arms retracted. Instead, close to 20% of the particles showed a conformation where all but one (n-1) of the protomers bent toward the central pore letting only one arm extended with its NTD fully exposed to the solvent. The above described one-arm-extended conformation is in agreement with a conformation previously reported (Fernández-Millán et al. 2015), where a chemically fixed Twinkle sample show a complex with only one NTD isolated, while the remaining RPDs are interconnected through contacts with neighbouring ZBDs. The apparently structural predisposition of Twinkle to have one loosed monomer could be supported by the proposed ring-opening mechanism of T7 gp4 (Ahnert et al. 2000), where the initial binding of the DNA by the N-terminal



primase domain produces a conformational change that triggers the ring opening for its subsequent loading onto the DNA molecule, followed by the closing of the ring.

### 4.3.3. The ZBD has a role on the structural flexibility

It has been reported that deletion of ZBD of the human Twinkle have a negligible effect on the stability of the complex and its helicase activity (Farge et al. 2008), so the enzyme is still able to support nearly normal levels of DNA synthesis during *in vitro* conditions. On the other hand, the lack of ZBD reduces the ssDNA binding ability of the helicase, and consequently, its ssDNA-dependent stimulation for ATP hydrolysis (Farge et al. 2008).

We were interested in evaluating possible structural effects of ZBD beyond oligomerization. To this aim, we analyzed by NS-EM a Twinkle peptide lacking ZBD ( $\Delta$ ZBD, our protein construct Del.E147K684). The initial inspection of the sample showed similar images to those coming from the wild type protein (Figure 57), which is in agreement with the fact that ZBD is not required for the formation of higher order oligomers. No sign of sample aggregation was observed at 250 mM NaCl conditions.

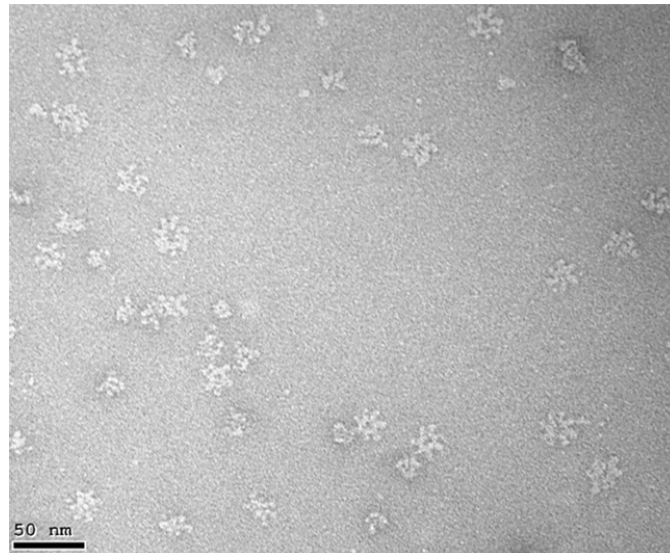
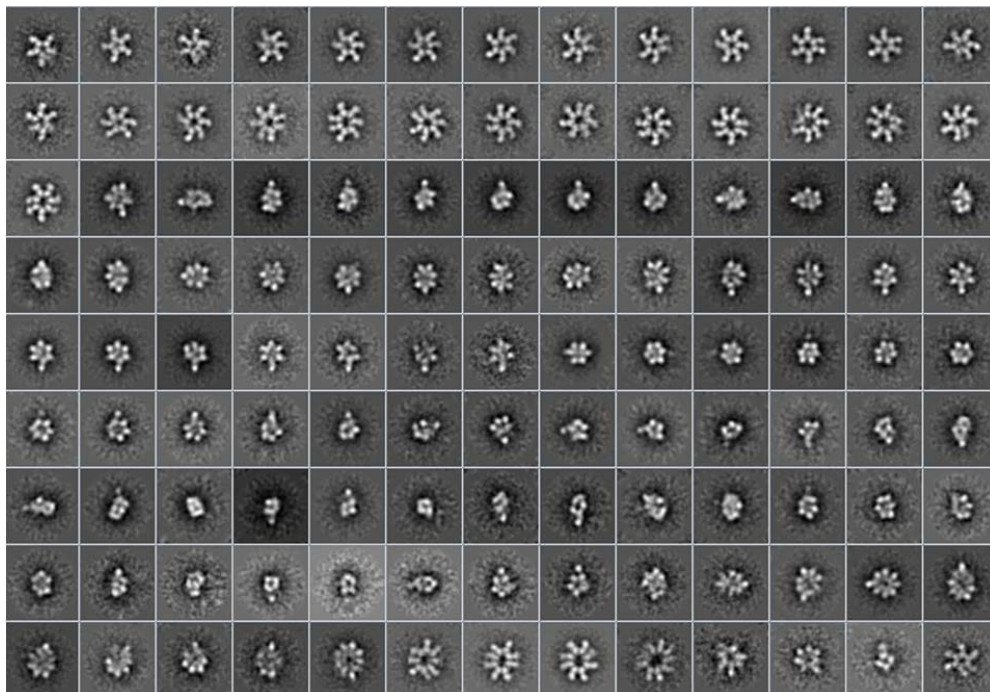


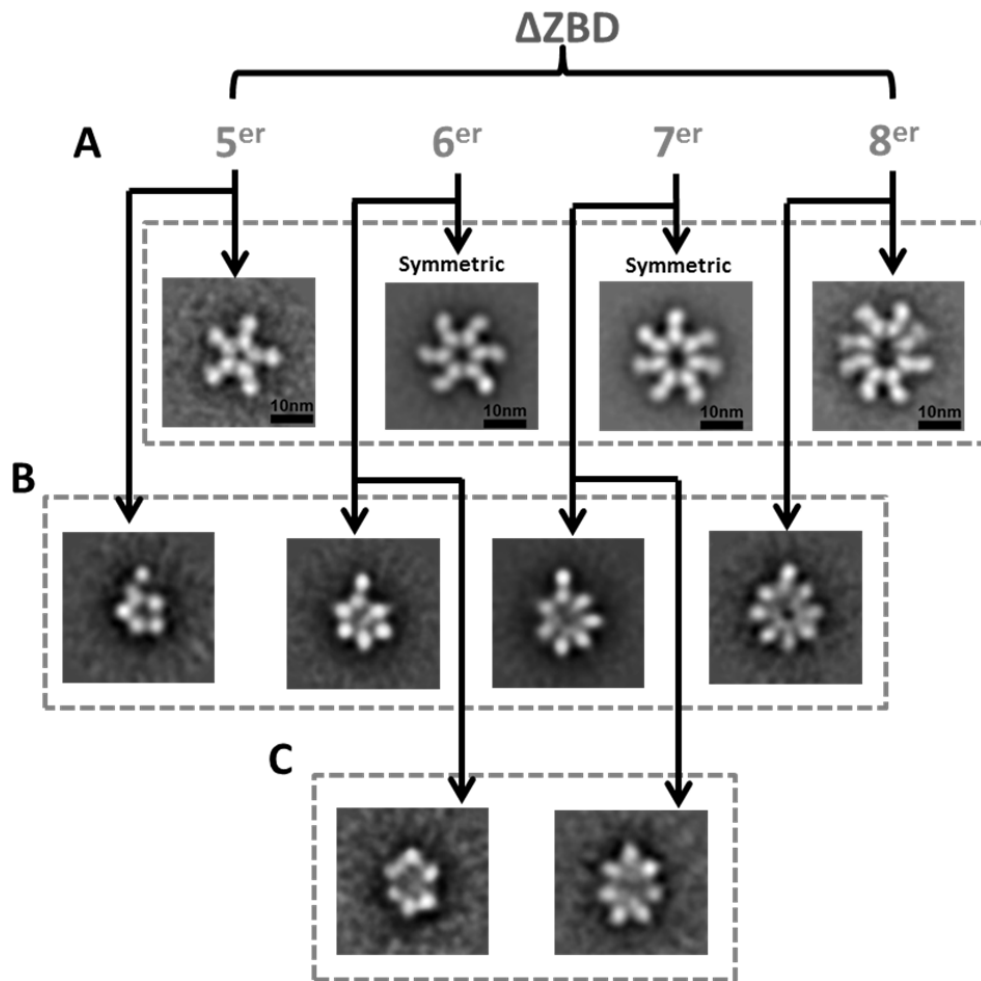
Figure 57. Representative NS-EM micrograph of the truncate  $\Delta$ ZBD protein at 250 mM NaCl.

9,991 single particles images of the negatively staining  $\Delta$ ZBD sample were subjected to reference-free 2D image classification (Figure 58).



**Figure 58.** Reference-free 2D class average classification (CL2D) of the complete set of 9991 particle images of the Twinkle deletion construct lacking the ZBD ( $\Delta$ ZBD) of protein.

The analysis of the 2D classified images revealed important differences between the  $\Delta$ ZBD protein and the full-length Twinkle. Both completely extended and compacted particles are formed by the  $\Delta$ ZBD protein (Figure 59A), but aside from the complexes with only one arm extended, there were no other intermediate conformations with variable number of extended/flexed arms which, moreover, are very commonly formed by the full-length protein. Indeed, compared with the full-length protein, the extended conformations of the  $\Delta$ ZBD protein showed a relatively symmetric distribution of the arms where the NTDs are well separated one from each other. This contrasts with the mostly asymmetric extended particles observed with the full-length protein, where pairs of protomers showing NTDs in close proximity occur frequently.



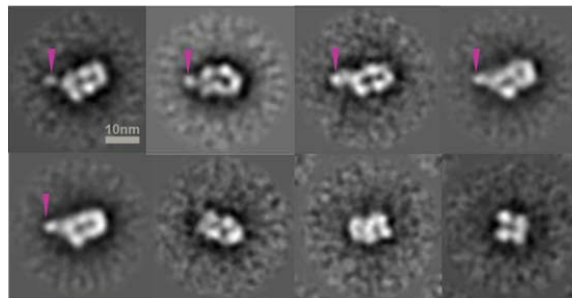
**Figure 59. Diagram showing representative NS-EM 2D classes of the  $\Delta$ ZBD construct at 250 mM NaCl. (A)** Different type of oligomers (pentamer, 5<sup>er</sup>; hexamer, 6<sup>er</sup>; heptamer, 7<sup>er</sup> and octamer, 8<sup>er</sup>) formed by the protein. **(B)** Compacted particles with only one arm extended. **(C)** Completely compacted particles. With all the arms flexed.

Additionally to the n-1 arms retracted conformation (Figure 59 B), the  $\Delta$ ZBD protein acquires conformations where all the monomers bend towards the center of the ring (Figure 59 C), which was not observed in the sample of the wild type Twinkle. While the number of hexamers of  $\Delta$ ZBD (Table 5) remained similar to that observed in the wild-type sample (Table 3), the amount of heptamers decreased considerably. Besides detecting around 10% of octameric particles, the  $\Delta$ ZBD protein also showed a similar percentage of pentamers (Figure 59 A and Table 5). Despite pentamers not being detected in our wild type sample under the used conditions, Twinkle pentamers have been previously reported (Ziebarth et al. 2010) by SDS-PAGE analysis under conditions of 100 mM NaCl plus cofactors (MgCl<sub>2</sub> and ATP $\gamma$ -S).

Oligomer state	$\Delta$ ZBD (%)
Pentamer	10
Hexamer	50
Heptamer	30
Octamer	10

**Table 5. Approximate percentage (%) of the different oligomeric state particles for the  $\Delta$ ZBD construct sample.**

We also observed a marked change in the proportion of extended vs compact conformations, being the compact ones more than 70% of the total of particles in the  $\Delta$ ZBD sample, while only 35% in the sample of the full-length protein. Concomitantly with an increase in the number of compact particles, it was observed a higher number of side-views (Figure 60) when compared with the wild type Twinkle, although it was still quite low to make a 3D reconstruction.

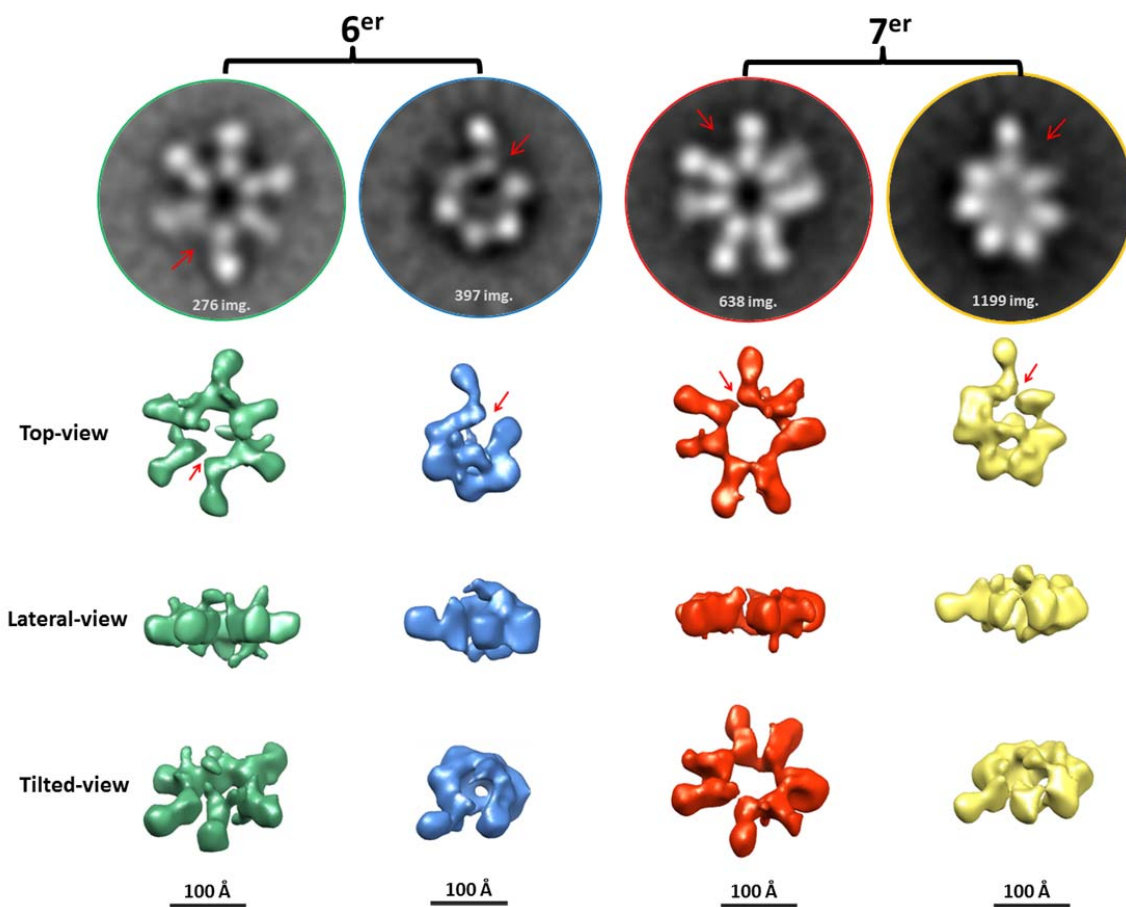


**Figure 60. Lateral views of the  $\Delta$ ZBD construct.** NS-EM 2D class averages showing possible lateral views of the particles, where a density that could correspond to the NTD of an extended protomer is pointed with a fuchsia arrowhead.

The observed conformational differences between the full-length and the  $\Delta$ ZBD proteins suggest that ZBD is implicated in a broad number of different conformations that do not occur in its absence.

#### **4.3.4. Initial 3D maps: One and two floor structures**

In agreement with previous reports (Ziebarth et al. 2010; Fernández-Millán et al. 2015), our results show that Twinkle is a highly heterogeneous sample. The presence of multiple oligomeric and conformational states together with the quite low number of side-views of the particles, made a 3D reconstruction by NS-EM of the complexes unfeasible. Although the principal goal of the present work was making a 2D image comparative analysis between full-length and  $\Delta$ ZBD proteins, we were still interested in obtaining preliminary NS-EM 3D maps of the most representative conformations of Twinkle, in order to make further refinements using cryo-EM data. To overcome the problem of the reduced number of side views, tilted pair images of the negatively stained wild-type specimen were taken. 9,721 tilted pair particles were selected and subjected to 2D average classification. Representative 2D classes from hexameric and heptameric complexes from both fully extended conformations and compacted forms with only one extended arm were selected and, for each case, an initial 3D map was reconstructed by the RCT method (Radermacher 1988) (Figure 61). Consistently with the mass of hexameric and heptameric oligomers of Twinkle, the maps of the extended conformations showed a flat disc-like shape, while the contracted conformations presented bulky lateral views suggesting a two-floor structure with one thinner density projected outside, which may correspond to the extended protomer.



**Figure 61. RCT reconstructions of both extended and compact conformations of hexameric ( $6^{er}$ ) and heptameric ( $7^{er}$ ) complexes.** The first row showed the 2D classes whose corresponding  $55^\circ$  tilted pair images were used for obtaining the respective volumes (In white is showed the number of images on each class). A red arrow points to an area of lower density in the 3D maps (and the corresponding position in their 2D classes) which could suggests the presence of a thin gap.

All four 3D maps, especially those from hexamers, showed, apparently, one gap at the interface between two subunits. Although it is acknowledged that at the maps resolution level, this kind of detail should not be blindly trusted, in support of a possible ring opening some of our 2D average classes showed a thin gap (Figure 62) and, it has also been previously reported an open conformation of a Twinkle heptamer (Fernández-Millán et al. 2015). Indeed, there is a natural ability in other ring-shaped helicases to acquire open-ring conformations (Costa et al. 2011; Arias-Palomo et al. 2013) as a mechanism for DNA loading, providing some level of functional support to the existence of a real opening or, at least, a weaker point of contact between two protomers in our Twinkle maps.

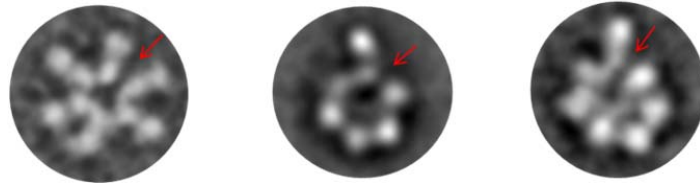


Figure 62. 2D classes of particles with an apparent thin gap (pointed by a red arrow) on the ring.

### 4.3.5. Establishing sample conditions for cryo-EM studies

Compared with NS-EM, cryo-EM is a higher resolution technique that avoids preferred orientations of the particles on the grid, so we started working to establish the cryo-EM conditions for future structural studies of Twinkle. Besides being highly sensitive to low salt concentrations, we also observed that in the absence of glycerol Twinkle forms a kind of inclusion bodies (Figure 63 A) that, interestingly, present some form of internal organization (Figure 63B), suggesting that the helicase complexes could still be assembled.

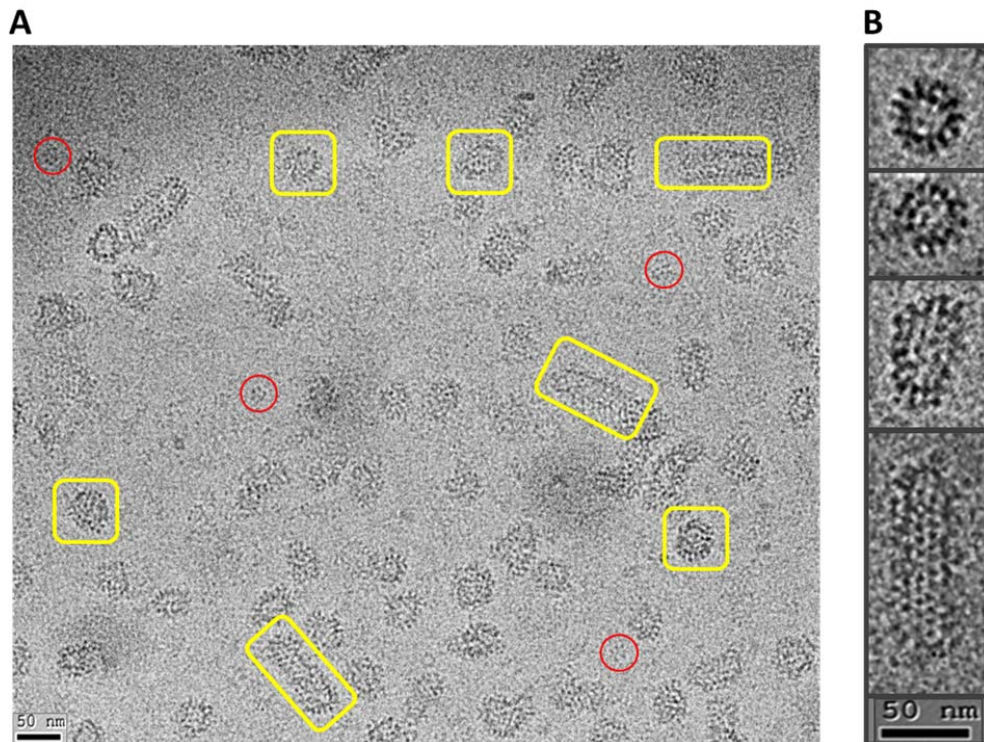
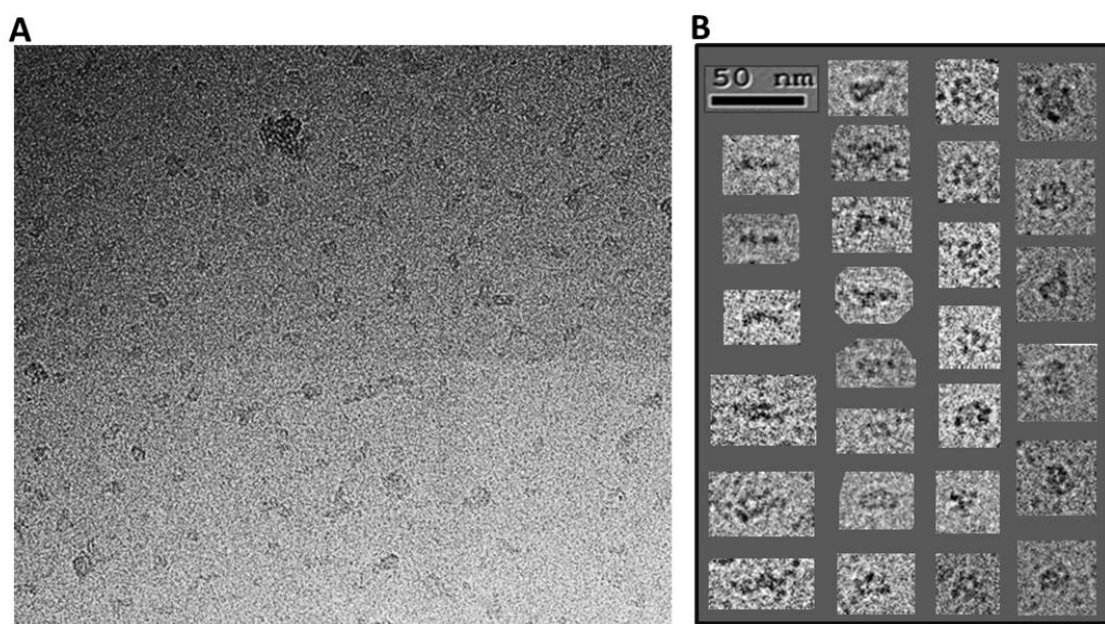


Figure 63. Representative cryo-EM images of Twinkle at 330mM NaCl and in absence of glycerol. (A) Micrograph showing protein aggregates (highlighted in yellow) of relatively regular rounded or elongated shapes. Some few free helicase particles can be identified (highlighted in red). (B) 2 x zoom of some aggregates where there can be inferred some form of internal organization.

Use of sample additives such as glycerol could be very problematic for vitrification due to its cryoprotectant effect. An additional problem of glycerol is the negative effect for cryo-EM imaging due to the contrast reduction it causes (Cong & Ludtke 2010). Because our initial sample came from a glycerol gradient purification, we had to work on finding the optimal equilibrium condition to maintain both the stability of the helicase complexes and an acceptable amount of glycerol for making cryo-EM grids. The final buffer conditions for cryo-EM were: 35 mM Tris-HCl, pH 7.5, 2 mM  $\beta$ -mercaptoethanol, ~5% glycerol and 330 mM NaCl, with a protein concentration of 50ng/ $\mu$ l. At this moment, we are in the process of cryo-EM image acquisition (Figure 64), as well as planning new efforts to improve the homogeneity of the samples. Higher resolution maps for different oligomeric and conformational states of Twinkle are expected to be obtained in the future.



**Figure 64. Cryo-EM of the wild type Twinkle sample at 330 mM NaCl. (A)** Representative micrograph image. **(B)** Individual particles (zoomed images), where some relatively flat particles (first, left column) likely correspond to lateral-views of extended conformations.



## Chapter 5

# 5. Discussion

## 5.1. Intrinsic protein flexibility facilitates quaternary structure assembly and conformational/oligomeric fluctuations

The assembly of proteins into higher order complexes is ubiquitous within the cell (Havugimana et al. 2012; Robinson et al. 2007) and is required to perform many of the diverse biochemical activities essential to cell homeostasis, growth and proliferation. The intrinsic flexibility of proteins is intimately related to their functions, which often go hand in hand with their ability to interact with other molecules of the same (homomeric) or different (heteromeric) class to form either transient or stable complexes (Marsh & Teichmann 2014; Marsh & Teichmann 2013). Furthermore, the flexibility of the unbound state of a protein generally correlates with the magnitude of binding-induced conformational changes (Marsh & Teichmann 2011; Dobbins et al. 2008). In many cases, the subunits (that is, the individual polypeptide) constituents of a protein complex can be assembled into a wide variety of symmetric and/or asymmetric quaternary structure topologies, where each of them often carried out a different function. Although it is expected that proteins tend to acquire a more stable and defined structure upon the interactions occurring within a macromolecular assembly, this does not imply the absolute loss of flexibility. Indeed, it is a common phenomenon that specific regions of a protein maintain a high degree of flexibility in the new formed complex, and even in some cases, there are parts that remain unstructured, waiting to establish additional interaction with other molecules.

The three cases of study presented in this thesis; CPAP, MCM4/6/7 and Twinkle, constitute excellent examples of the natural structural versatility of proteins, showing the ability of establishing both flexible and multiple oligomeric complexes, which can be intuitively associated with the functional complexity of the biological processes where each of these macromolecular machines has been related with.

## 5.2. Structural insights into CPAP conformational and oligomeric behavior: The possible hierarchical building of a scaffold

The centrosome, a conserved organelle found in nearly every animal cell that acts as the primary microtubule organizing center (MTOC), is composed by a pair of centrioles immersed in a proteinaceous network of pericentriolar material (PCM). Centrosomes are involved in the organization of the mitotic spindle and in cytokinesis, as well as in the nucleation of cilia and flagella; they also serve as anchorage site for a wide number of regulatory processes.

CPAP (Centrosomal P4.1 associated protein, also known as Centromere Protein J –CENPJ-), a cell cycle regulated protein fundamental for centriole assembly, is a centrosomal protein present across the eukaryotic taxon. This protein is formed by one conserved globular domain at its C-terminus while the rest of the sequence is predicted to have disordered and coiled-coil regions. In humans, the carboxyl region of CPAP between the residues 897-1338 (CPAP<sup>897-1338</sup>) is responsible for its interaction with several other important centrosomal proteins, and it includes a homodimerization domain. Mutations of CPAP have been associated with primary autosomal recessive microcephaly (MCPH). Despite of the widely demonstrated biological and clinical relevance of CPAP, its mechanistic behavior remains blurry.

This work constitutes the first report of different oligomers of CPAP<sup>897-1338</sup> and their structural analysis. Here we have shown and analyzed the 2D images of such structures and resolved the 3D-volume of both a putative CPAP<sup>897-1338</sup> dimer and a tetramer. Based in our observations, we propose that CPAP presents a discrete but dynamic oligomeric behavior, and that a tetrameric complex of the protein could be the structural brick of higher order supramolecular structures which could be working as a scaffold that tethers the PCM to centrioles.

### **5.2.1. CPAP<sup>897-1338</sup> forms different homo-holigomeric complexes *in vitro***

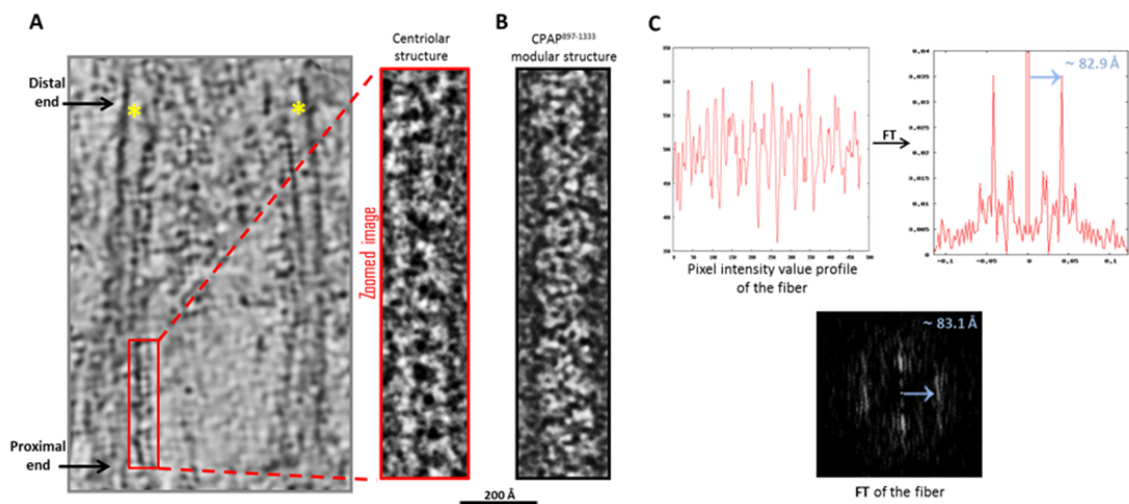
In this study we report, for the first time, that the human CPAP<sup>897-1338</sup> construct forms different types of homo-oligomeric complexes. Taking together the analysis of the data given by the SEC purifications; the 2D images and the 3D volume obtained by the EM work; the solved crystallographic structure of the G-Box domain; and, finally, the *in silico* structure modeling of the CC4/CC5 coiled-coils and a gross estimation of the dimensions of the CCGb-linker of CPAP, we propose that CPAP presents a discrete but dynamic oligomeric behavior that includes the formation of dimers, tetramers and larger structures formed by stacks of tetramers.

Taking into account the multiple binding partners of CPAP, it is not surprising that this protein is predicted to be largely unstructured, being this a characteristic that can confer the structural flexibility necessary to interact with different proteins/complexes. Indeed, it is reasonable to consider that each of the oligomeric states of CPAP would be more prone to establish an interaction with a specific protein/complex depending on its own initial configuration; the building up of different binding sites at the intersubunit interfaces of the different CPAP complexes could be a possible mechanism contributing to this.

It is well known that concentration can be a driving factor affecting the oligomerization status of many proteins, which in turn, modifies its basic structure and by definition also its function (Giese & Vierling 2002; Kley et al. 2011; Chen et al. 2003; Kutter et al. 2007). There exist several reports of proteins forming filaments in response to factors such as the increment in concentration (a phenomenon known as molecular crowding) (Noree et al. 2010), in some cases, showing a modular behavior (Petrovska et al. 2014) resembling the one we observe in CPAP<sup>897-1338</sup>. Indeed, CPAP concentration is regulated along the cell cycle and its levels increase gradually from the beginning of S phase until mitosis, being this a lapse of time that coincides with the procentriole formation (in early S phase), elongation (in late S phase) and with centrosome maturation (along the G2 phase). Remarkably, centrosomal CPAP maintains a continuous exchange with a cytoplasmic CPAP pool (Kitagawa et al. 2011), reaching its highest level in G2, when there is a maximal recruitment of proteins to the PCM. CPAP shows a significant decrease at the end of mitosis/early G1, when the dynamic of formation and maturation of the centrosome is already

finished (Tang et al. 2009; Azimzadeh & Bornens 2007; Kim et al. 2012). Furthermore, the CPAP protein level is also regulated during centriole amplification in multiciliated cells (Zhao et al. 2013). Protein concentration-dependent conformational changes could be a mechanism contributing to direct the network of diverse interactions that CPAP must carry out. Thus, it is tempting to speculate that cell cycle regulation of CPAP concentration could be a complex and highly synchronized strategy, which together with other determinant factors (e.g., protein phosphorylation (Chang et al. 2010; Chen et al. 2006; Zhao et al. 2010)) controls the oligomeric state of CPAP in order to direct some of its multiple functions.

Column-like structures with a repeating unit around 8nm have been observed at the luminal side in a tomographic reconstruction of the basal body triplet in *C. reinhardtii* (Li et al. 2012), and it has been suggested that these elongated structures could correspond to CPAP (Hatzopoulos et al. 2013). Furthermore, we note that tomographic images of centrioles from calf thymus (Figure 65 A) purified in our group, show modular, column-like structures with an axial periodicity of approximately 83 Å (Figure 65 C) running along the inner wall. These long centriolar structures look very similar to our purified stacks of CPAP<sup>897-1338</sup> (Figure 65 B) and, indeed, their repeated unit of close to 80 Å, match the dimension of the narrow side of the putative CPAP<sup>897-1338</sup> tetramers.



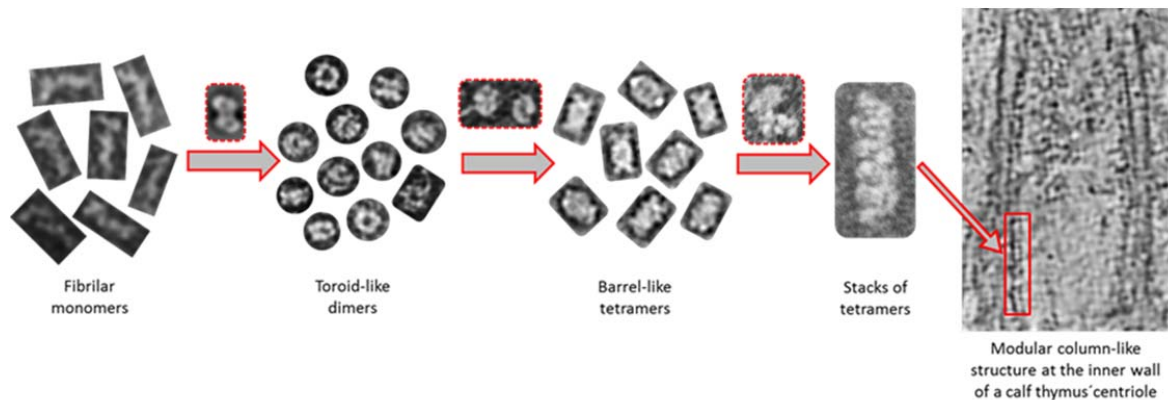
**Figure 65. Centriolar localization of long, modular structures, proposed to be putative stacks of CPAP tetramers. (A)** Highly filtered image of a cross-section of a calf thymus' centriole tomogram (*left panel*), where part of a modular column-like structure localized at the inner wall of the centriole is highlighted in red. The proximal end of the centriole is at the bottom and the distal end is at the top. The position of the centriole walls are marked with a yellow asterisk (\*). Zoomed image of the column-like structure (red panel), which would be putatively formed by stacks of CPAP tetramers.

(B) NS-EM image of long, modular structures formed by the purified CPAP<sup>897-1338</sup> protein. This image is displayed at a similar scale to the one of the centriolar zoomed image showed in subparagraph A. (C) The centriolar modular structure has a ~ 83 Å axial repetition pattern (*left pannel*), as measured by both the Fast Fourier Transform (FFT) of the pixel intensity profile of the image (*top panel*) and by the diffraction pattern of the structure in the Fourier space (Fourier transform, FT) (*bottom panel*). The magnitude of the pattern is represented with a blue arrow and the value is written in the same color.

CPAP interacts with the 8 nm length  $\alpha/\beta$ -tubulin heterodimers (Hung et al. 2004) that form the MT that constitute the walls of the centriole. It could be expected that proteins/complexes that interact with MT have patterns related to the one of the MT themselves. We propose that the interaction of putative CPAP tetramer stacks with the  $\alpha/\beta$ -tubulin heterodimer repeats occurring along centriolar MTs every ~8 nm, would fix the structure of the modular CPAP filaments producing the 83 Å axial periodicity pattern observed in the putative CPAP column-like structure localized at the inner walls of the calf thymus' centriole (Figure 65 A). Studies showing that CPAP concentrate within the proximal lumen of both parental and nascent centrioles (Kleylein-sohn et al. 2007; Chang et al. 2010) support our proposed localization of the CPAP modular fibers, although the identification of CPAP within the purified calf thymus' centrioles by direct labeling of the protein must still be done.

An overview of our results and its analysis allowed us to propose a putative model for the formation of higher order modular filaments of CPAP (

Figure 66), where first, the highly flexible monomer strings of CPAP<sup>897-1338</sup> dimerize acquiring a globular and more stable toroidal structure. Then, dimers of CPAP<sup>897-1338</sup> would dimerize into a tetrameric structure, which, in turn, could acts as a somewhat flexible building block for larger, modular and elongated rope-like supramolecular structures with an axial periodicity around 8 nm. Finally, we propose that these modular rope-like structures formed *in vitro* could correspond to the column-like structures observed in the inner-walls of the centriole.



**Figure 66. Tentative model for the progressive self-assembly of CPAP into gradually higher oligomeric subcomplexes until formation of a modular rope-like structure.** The highly flexible monomers may dimerize by meshing into a globular toroid-like complex, which in turn could dimerize again resulting in the formation of a hollow, barrel-like tetrameric structure. Finally, stacks of tetramers may form linear arrangements of variable length with a periodicity close to 80 Å, which could also correspond to those structures known to exist at the inner walls of the centriole. Red dashed lines highlight putative intermediate states between one oligomeric state and the next one.

The proposed tentative organization of the G-Box and CC4/CC5 domains on opposite sides of the putative tetramer structure of CPAP<sup>897-1338</sup> would confer to it a structural (and functional) polarity, which, in turn, would also apply to the modular, higher order rope-like complexes. However, further work labeling the different domains of CPAP and the respective structural studies must be carried out to unambiguously determine the organization of the monomers within the different complexes.

The presented work reinforces the idea that CPAP forms organized higher order structures allowing it to act as a scaffold that connects PCM proteins and complexes with the nascent centriole (Gopalakrishnan et al. 2011; Hatzopoulos et al. 2013; X. Zheng et al. 2014), contributing to the progression of procentriole assembly and elongation (Tang et al. 2009). We propose that molecular crowding (Kuznetsova et al. 2014; Ellis & Ellis 2001) could be an important driving force affecting both the structure and oligomeric state of CPAP. The precise mechanism by which CPAP goes from one oligomeric state to a different one, as well as the exact biological role of each of these complexes, is a matter of future studies that would include works with CPAP mutants as well as interaction assays with some of its already known partners. Our results showing that CPAP<sup>897-1338</sup> displays a diverse and dynamic oligomeric behavior lay the foundation not only for future structural works, but also for new mechanistic/functional studies of this critical factor in

centrosome formation and, consequently, in cell biogenesis. Mutations of CPAP, which could potentially affect its structure and ability to associate with other proteins, are associated with primary autosomal recessive microcephaly disease, thus the present study may lead to important future biomedical implications.



### **5.3. Insights into the architecture and flexibility of the hexameric MCM4/6/7 ring and other lower order sub complexes**

The minichromosome maintenance (MCM) family is a conserved group of proteins proposed to work as replicative helicases in archaea and eukaryotes. They are usually assembled as ring-shaped molecular motors that are essential in the initiation and progression steps of DNA replication. In eukaryotes, each of the six homologous proteins MCM2-7 present different catalytic and regulatory roles and all of them are necessary to maintain the integrity of the genome. Eukaryote MCMs can form different types of oligomers, though it is not clear the biological role of most of them. A growing number of studies suggest that beyond replication, MCMs proteins must also be involved in DNA transcription, chromatin remodeling, activation of dormant replication origins and checkpoint responses to maintain genome stability. The hexameric MCM4/6/7 is a subassembly formed by two MCM4/6/7 heterotrimers that has an intrinsic helicase activity. There are many studies focused on the biochemical properties of the MCM4/6/7 hexameric helicase; however, structural works are practically inexistent. Here we present a set of electron microscopy (EM) maps that include different conformational states of the mouse hexameric MCM4/6/7 helicase, as well as some putative trimeric structures. In agreement with previous biochemical studies of MCM4/6/7, and based on feature comparisons between our volumes and equivalent proteins in several of the density maps already resolved for the MCM2-7 heteroexameric helicase, we propose a putative architecture of MCM4/6/7 hexamer as well as the possible identification of the protomers flanking discontinuities present in different conformations of the MCM4/6/7 helicase. Finally, we present a tentative time sequence model with the structures reported in this work, which provides a potential mechanistic explanation of a reconfiguration process beginning with the dimerization of trimers and ending with the formation of a planar, hexameric ring (*Class 1* map). This ring complex presents a weak contact localized between the putative intersubunit AAA<sup>+</sup> MCM4'/HTH MCM6' interface, a feature that could create a dynamic gate that facilitates the helicase opening during its loading on DNA.

### 5.3.1. Structural polymorphism

A number of EM maps have shown the helicase MCM2-7 as a single or double hexamer displaying conformational variations, which can be also found in complex with other proteins of DNA molecules. On the other hand, there is very scant structural information on the helicase MCM4/6/7x2, which is restricted to some studies presenting some EM 2D images of this complex (Bochman & Schwacha 2007; Sato et al. 2000; Yabuta et al. 2003). The fact that the trimer MCM7→4→6 forms part of the MCM2-7 helicase (Figure 10 A) lets us think that the already reported 3D maps of this complex can provide important clues about the structure and subunit distribution within MCM4/6/7x2. Furthermore, proteins MCM4, 6 and 7 can associate into hexamers and trimers; even some of them have been purified as free monomers (Musahl et al. 1995; Sherman et al. 1998; Ma et al. 2010; Ichinose 1996; Bochman et al. 2008; Davey et al. 2003; Xu et al. 2013), but currently there are no solved structures for any of these states.

Up to now the structural observation of the C- and N-terminal extension of full-length eukaryotic MCM proteins has been elusive due to their high flexibility. The identification of these extensions within the context of the whole MCM4/6/7x2 helicase complex can provide unique important clues about its functional role, hence the relevance of working on this front.

It has been observed in different helicases that binding of nucleotide can induce the formation of active toroidal oligomers which, along with the hydrolysis process, bring about additional conformational changes required to carry out the enzyme functions (Hydrolysis et al. 1984; Yong & Romano 1995; Ziebarth et al. 2010). The mouse MCM4/6/7 sample analyzed during this work, which was initially purified as an hexamer in the presence of ATP (which by the moment of our EM analysis might be partially hydrolyzed), allowed us to visualize different structures of the hexameric ring as well as smaller oligomeric complexes that include putative trimers and monomers. In agreement with the abovementioned, in the presence of ADP, MCM4/6/7 likely presents an equilibrium between different forms of MCM4/6/7 that exchange rapidly (Ma et al. 2010).

In this work, aside from presenting mouse MCM4, 6 and 7 low resolution 3D maps of a number of trimeric and hexameric complexes, we have carefully made an exercise of putting them in the context of the available structural and functional works on MCM proteins, which allow us to putatively locate the relative position among subunits and speculate about some functional

consequences of the observed structural features. The presented volumes are EM snapshots that suggest transition states going from the isolated trimers  $MCM7 \rightarrow 4 \rightarrow 6$  and  $MCM6' \rightarrow 4' \rightarrow 7'$  to the formation of a planar ring with a MCM4 C-terminal domain extended toward the central channel, and the interacting HTH motifs of the subunits MCM6 and 6' protruding, maintaining weak contact with the AAA<sup>+</sup> region of MCM4'. We speculate that this structure may be able to isomerize to an open ring with a MCM4'/6'gap, which could be a suitable conformation for DNA-loading. Due to the putative architecture similarities of this MCM4/6/7x2 volume with MCM2-7, we propose that the subunit rearrangement within this ring conformation could be close to the active state required by the helicase to bind and, subsequently, unwind dsDNA, although it will be expected that upon DNA binding additional conformational changes take place. For example, in *Dme*CMG it is observed that the AAA<sup>+</sup> region of the MCM2-7 ring changes from a flat conformation to a right-handed spiral after DNA binding, maintaining in both states a practically flat N-terminal region (Costa et al. 2014). A recent study showed that in a recombinant human MCM2-7 complex, the binding of DNA induces conformational changes including the protrusion of a mass emerging from the AAA<sup>+</sup> region of the helicase (Hesketh et al. 2015); the structure of this active *Hs*MCM2-7 + DNA complex shows a closed, N-terminal planar-ring configuration, where the prominent C-terminal extension is suspected to be the C-terminus of some of the subunits (Hesketh et al. 2015).

Summarizing, the following are some of the new contributions of this work to MCMs' helicase research field:

1. The structure of the trimeric MCM oligomers resembles that of half MCM hexameric rings, suggesting that the spatial organization of the monomers within the trimer already have the general architecture observed in the hexameric assembly.
2. There is a conformational state where there is a connecting bridge between two MCM trimers, which localize at the putative N-terminal region of the complex, arising near the middle protomer from each side, both putatively corresponding to MCM4 subunits. We propose that this connection could be responsible for the incipient dimerization of the trimers. Nevertheless, we cannot discard the possibility of additional functional roles that could be related with this conformation.
3. Putative HTH MCM6 domains show drastic positional changes, going from recessed localizations to different extended positions, some of which project toward the central channel (*Class 3* map).

We propose that HTH motif from MCM6 could work as a structural regulator for DNA for entry through the helicase pore by making a steric impediment, for example, when the conformational state of the ring is not yet a “mature” active one.

4. The sealing of the N-terminal pore of the central channel in the more compacted ring closed conformation (*Class 5* map) could be a specific mechanism for avoiding linear DNA to access the inside of the helicase. This could be indicative that the conformation of this ring is not yet a “mature” active one.

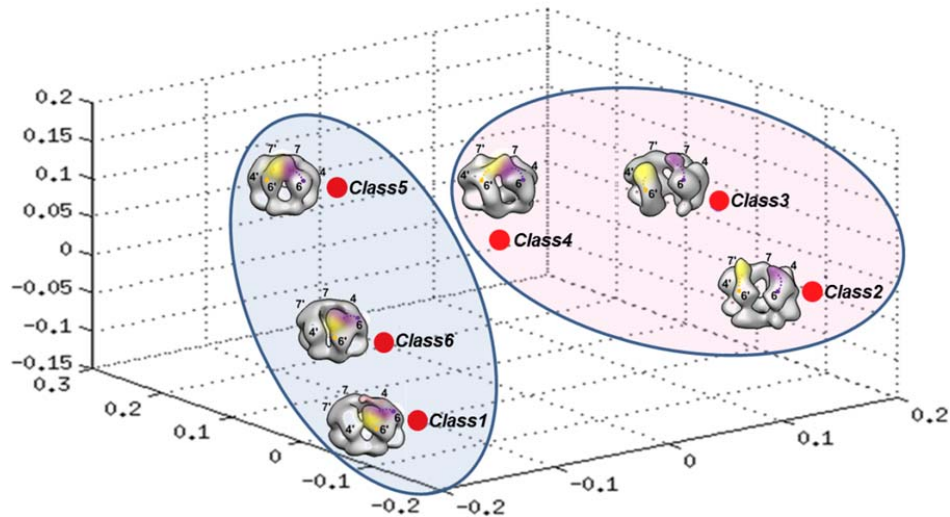
5. We report a putative interaction between the HTH motifs from the two copies of MCM6 within the helicase (maps from *Class 1*, *Class 4*, *Class 5* and *Class 6*). We propose that this interaction plays an important role on the establishment of a potentially active architecture of MCM4/6/7x2. Additionally, it is plausible that the big mass resulting from this HTH-HTH interaction, which localized at an interface of the ring that seems to be more prone to get open in the *Class 1* map, could work as a bulky clasp. We speculate that this clasp could be transiently displaced, promoting the formation of a gapped ring for allowing the helicase to load into DNA, and would be closed again to capture the DNA molecule during helicase translocation.

6. In a proposed possibly active (“mature”) conformation of the MCM4/6/7x2 ring (*Class 1* map), the C-terminal domain of one of the two putative MCM4 protomers forms a finger-like structure extended toward the central channel, reaching a position that would allow it to potentially interact with any DNA molecule inside the ring, suggesting a possible important role for this structure.

7. We propose that the mostly planar state observed in all the different ring conformations, and the ability to get closed by itself, could be an explanation for the intrinsic helicase activity of MCM4/6/7x2. This contrast with the tendency of MCM2-7 to form inactive opened non-planar rings, and its requirement for additional factors to acquire a planar configuration and sealing its gap, which seems to promote the activation of its helicase activity.

A further computational analysis using normal mode analysis (NMA) (Jonic et al. 2015) of the six different conformational states (*Class 1* to *Class 6* 3D maps) of the ring-like MCM4/6/7x2 complex (Figure 67), showed two main groups containing the more structurally similar maps. In one group there were included maps *Class 2*, *3* and *4*, where the putative HTH domains of the MCM6

subunits are separated or loosely interacting. In the second group there were maps *Class 1, 5* and *6*, all of them showing a big extra density putatively assigned to the closely interacting HTH domains of the MCM6 subunits.

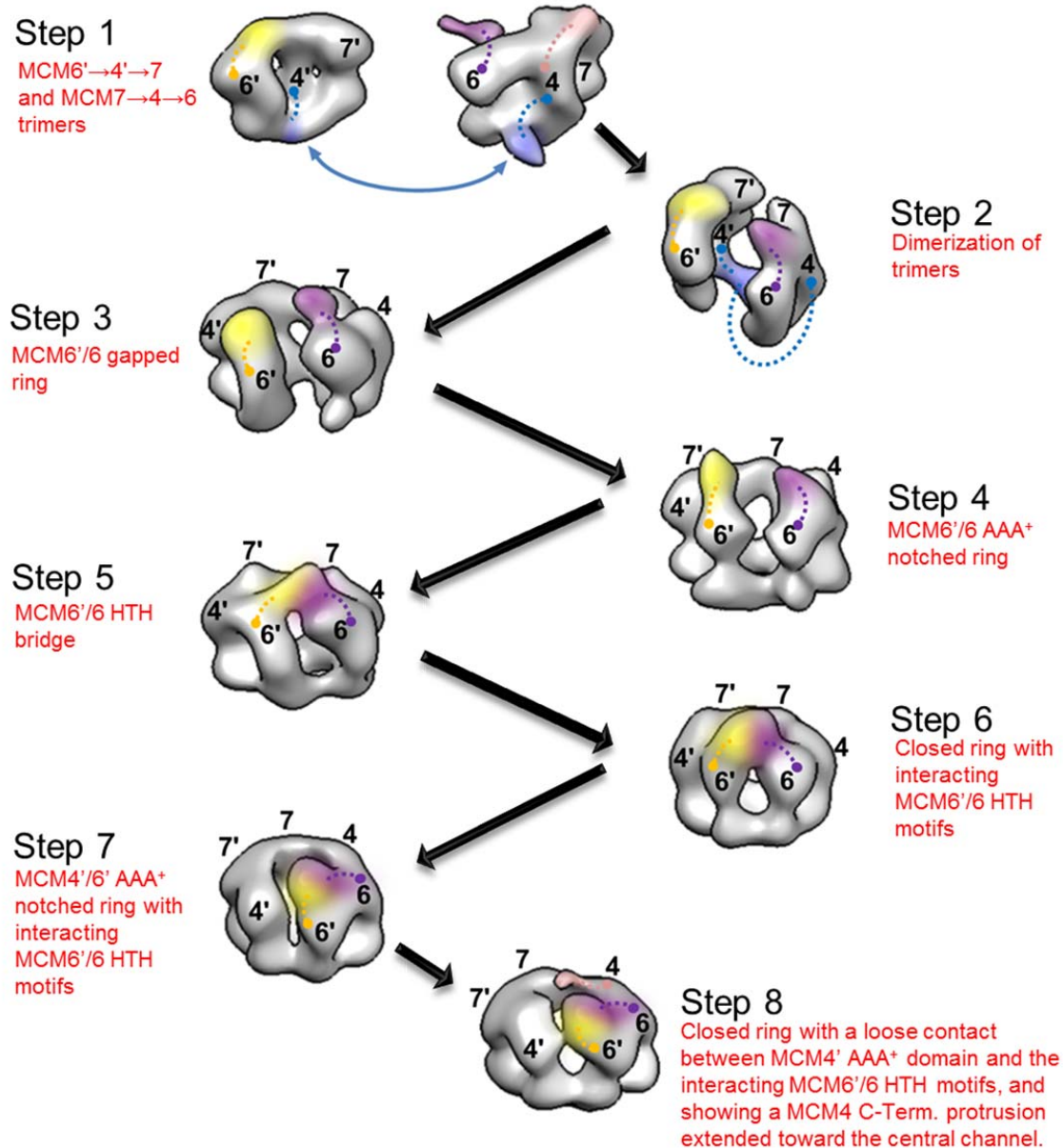


**Figure 67. Normal mode analysis (NMA) of the MCM4/6/7x2 ring-like structures *Class 1* to *Class 6*.** Graphical representation showing the relative distance correlation (Red dots) among all the analyzed maps, as projected in a 3D space. Blue and red shadowed figures highlight two groups, each containing the more closely related structures. Numbers at the three different axes of the graphic, represent relative values, instead of absolute ones.

In the following we are going to propose a tentative time sequence of the different EM maps reported in this work (Figure 68):

1. MCM4/6/7 trimers are formed. We speculate that two different configurations, MCM7→4→6 and MCM6'→4'→7', could coexist.
2. The two putative trimers, MCM7→4→6 and MCM6'→4'→7', dimerize by making an initial contact through a bridge between the N-terminal domains of the MCM4-4' subunits.
3. The connected trimers approximate by one edge, so that MCM7-7' protomers bind side by side. In this way an open ring with a gap is created at the MCM6-6' interface. The HTH motif of MCM6 is extended toward the central channel making a partial obstruction of the pore, which could work as a steric impediment for DNA passing through at this stage.
4. The N-terminals of both MCM6-6' make contact, closing the lower part of the ring, while their HTH motifs protrude, in different degrees, from the top of the ring.
5. The HTH parts of the MCM6-6' protomers make contact, creating a loose junction that partially obstructs the opening at the top of the ring.

6. Two HTH motifs coming from the MCM6-6' subunits merge together, creating a lump at the top of the ring, in the middle point between the respective two subunits. It would be expected that the interaction of the flexible C-terminal domains stabilizes their HTH fold in a more compact architecture, strengthening the whole union between the MCM6-6' protomers. At this stage the ring is completely closed and the pore at the N-terminal region is practically sealed, preventing any potential DNA molecule to pass through.
7. The AAA<sup>+</sup> domains from MCM4' and MCM6' separate, giving rise to a notched ring, where the HTH MCM6-6' lump had moved to the top of MCM6, near to the nick.
8. Now that the new interactions between the MCM6-6' subunits had been correctly established within the ring (MCM4'→7'→7→4→6→6'), the lump of the two interacting HTH motifs shifts its position by moving over the notch, making slight contact with the AAA<sup>+</sup> domain of MCM4' such that the nick gets closed again. The central channel is open by both sides but the big aperture at the AAA<sup>+</sup> region is partially divided into two smaller chambers by the putative MCM4 C-terminal finger-like structure projected towards the central pore. In this position the extended C-terminal structure could potentially get in close contact with any DNA crossing through the helicase channel. The pores placed parallel to the central channel, are all big enough for allowing DNA to pass through. The new structural rearrangement of intersubunit contacts in this ring could potentially give rise to a “mature” conformation able to load itself onto DNA and support the helicase activity.



**Figure 68. Proposed sequence of a putative multiple-step assembly process of MCM4/6/7x2.** The sequence begins with the dimerization of the trimers followed by a number of intermediate conformational changes of the hexamer in a way that protruding HTH domains of both MCM6 and MCM6' interact very closely, and the C-terminal domain of MCM4 protrudes toward the central channel of the helicase. Numbers represent the putative organization of the subunits. Some putatively designated structural features had been highlighted as follows: MCM6' HTH domain (Yellow); MCM6 HTH domain (Purple); N-terminal extension of both MCM4 and MCM4' (either separated or interacting) (Blue), and C-terminal domain of MCM4 (pale-pink). Colored dotted-lines represent a symbolic linkage between the respective colored domain and the putative subunit to which it belongs.

Our tentatively proposed sequence of conformational changes provides a structural-based model for the formation of a potentially active MCM4/6/7x2 subunit arrangement (MCM4'→7'→7→4→6→6'), which could somehow emulate the general organization of MCM7→4→6 within the helicase MCM2-7 (MCM5→3→7→4→6→2). We have shown that in the presence of ATP, which might be partially hydrolyzed, coexist different conformations of hexamers, trimers and monomers of the proteins MCM4, 6 and 7. Which would be the nucleotide binding state (ATP, ADP or apo) of each of the presented conformations is a question that still must be answered. Future work on EM analysis of MCM4/6/7 samples in presence of non-hydrolyzable ATP and ADP analogs would help to solve the question. Meanwhile, possible clues can be found on previous studies of MCM. The observation that under apo condition mostly trimeric and smaller conformation of MCM appear, while the presence of nucleotide promotes the hexamerization (Ma et al. 2010), strongly points to trimers as being apo states while hexamers represent different nucleotide-binding states. When compared with the apo state, ATPγS bound *EcuMCM2-7* and *DmeMCM2-7* open rings presents a more compact shape (Lyubimov et al. 2012). These observations would suggest that the closed and, perhaps, even the more compact configurations of MCM4/6/7x2, could correspond to ATP bound states, while the more distended structures could be a mixture of ATP/ADP-binding states.

Studies on *DmeMCM2-7* (Costa et al. 2011) and *EcuMCM2-7* (Lyubimov et al. 2012) have shown that under both conditions, absence or presence of nucleotide, the most prominent conformational state is a non-planar open-ring configuration. It has been proposed that MCM2-7 needs additional factors for closing (through the use of specific anions, as Glutamate or Acetate) (Bochman & Schwacha 2008; Hesketh et al. 2015) or sealing (making a complex with GINS and Cdc45) (Costa et al. 2011) its MCM2/5 gate, in order to function as an helicase (O'Shea, V.L. and Berger 2014). Interestingly, planar states of MCM2-7 had only been reported in a minority population of partially closed ring particles (apo notched *DmeMCM2-7* (Costa et al. 2011)); in an helicase-DNA complex in presence of specific ions (*HsmMCM2-7* with helicase activity (Hesketh et al. 2015), or when the helicase is in complex with proteins implicated in its DNA loading (*ScOCCM* (Sun et al. 2013; Sun et al. 2014)) or dsDNA unwinding activation (*DmeCMG* (Costa et al. 2011) and). This led us to tentatively propose that a planar state could be required for a correct structural/functional alignment of the subunits, in order to carry out the activation of the helicase activity. Our study demonstrates that mouse helicase MCM4/6/7x2 forms, by itself, mostly N-



terminal planar configurations of the open, notched or completely closed ring, which would give a structural support to its already known intrinsic DNA-unwinding activity (You et al. 1999).

The extensive variety of MCMs complexes that have been identified throughout this work and several others put in evidence a dynamic and highly flexible group of proteins that can be associated with the functional complexity of the eukaryotic replication and transcription regulatory processes, in which they had been found to play important roles. The ensemble of maps presented in this study is, perhaps, just a small but representative sample of what should be a whole range of continuous apo and ATP-binding/hydrolyzed conformational states. These structures represent an important progress toward the understanding of the mechanistic behavior of MCM4/6/7x2 helicase and, to some extent, of MCM2-7 and other ring-shaped helicase complexes.

## **5.4. Structural Characterization of TWINKLE, the human mitochondrial DNA helicase: flexibility and heterogeneity**

The ring-shaped mitochondrial replicative DNA helicase Twinkle is an essential component of the minimal replisome required for the correct maintenance of mtDNA. Mutations in the Twinkle gene and mtDNA deletions are both associated with adPEO syndrome. In addition to its DNA unwinding activity, Twinkle presents an antagonistic annealing activity, so it is thought that this enzyme could be implicated in different aspects of DNA processing aside from replication.

Both hexameric and heptameric complexes of Twinkle have been previously reported. Here, we have provided further conformational insights using EM techniques that allowed us to identify, aside from the more commonly found hexamers and heptamers, novel octameric and pentameric states of the protein. We have described a wide number of conformations of Twinkle where protomers behave as flexible arms, showing different axial displacements and bending movements of their NTD. Flexibility of NTD give rise to multiple forms of both extended and more compacted complexes, agreeing with previous dynamic studies of Twinkle protein in solution. Additionally, our structural comparative analysis between the full-length Twinkle and a truncated construct lacking the ZBD of the protein puts in evidence a notable role of ZBD on the ability of the enzyme to acquire a wide number of conformational states that in absence of this domain do not occur. Our structural characterization of Twinkle showing its highly dynamic and flexible behaviour can be associated with the structural/functional complexity required for both DNA unwinding and annealing activities.

### **5.4.1. Effect on the helicase sample to changes in ionic environment**

Function and structure of mitochondria are altered by changes in osmotic and ionic conditions of both its internal and external environment, so variations of these factors within the mitochondria

matrix, being importantly influenced by the metabolic state, are likely related to regulation of mitochondria functions (Hackenbrock 1968; Bradshaw & Pfeiffer 2006; Murphy & A. Eisner 2010). The precise mechanisms that allow synchronization of cell proliferation with mitochondrial genome replication are still far from being completely understood, but ionic conditions seem to play an essential role. To date, important effects on the state and function of proteins in solution caused by changes in factors such as ionic strength, temperature and molecular crowding (Schroer et al. 2012; Kuznetsova et al. 2014) have been widely described. Specifically, studies of the dynamic effect of salt concentration on the state of Twinkle protein showed its strong susceptibility to aggregate at concentrations below 250 mM NaCl, especially in the absence of additional stabilizing cofactors (Ziebarth et al. 2010). In agreement with the aforementioned study, we observed that Twinkle is completely soluble at concentrations of 250 mM NaCl or higher, while an important aggregation occurred at 150 mM NaCl, which was much more pronounced at 100 mM NaCl, even in presence of those cofactors that had been described to increase the solubility at these lower salt conditions (Ziebarth et al. 2010). Complete recovery of soluble helicase particles by increasing salt concentration to 330 mM NaCl from partially insoluble samples at 100 mM NaCl plus cofactors, suggests that at the analyzed low salt conditions the oligomers are not dissociated, instead just a tight conglomeration of still assembled complexes could be occurring in response to the buffer conditions. If the latter is true, perhaps this could be used as a mechanism of mitochondria to regulate the availability of Twinkle by using a “sequestration” strategy regulated by changes in the ionic strength of the environment, which in turn would affect mtDNA replication.

### **5.4.2. Multiple oligomer states of the enzyme**

Twinkle is able to form a number of oligomeric complexes which are predominately hexameric and heptameric forms, although we also detected new populations of octamers and pentamers. The diameter of central channel in all the different complexes can easily accommodate ssDNA and, with exception of the pentameric form, the rest could also accommodate a dsDNA molecule. We observed that the population of hexamers was around half of the particles in both full-length and  $\Delta$ ZBD samples, while the proportion of other oligomeric states was smaller and vary significantly between the two proteins. The above suggests that the hexameric configuration could be the most

stable form of Twinkle, which would not be surprising considering that many other different ring-like helicases tend to be organized as hexamers (Patel & Picha 2000).

### **5.4.3. Structural flexibility, possible opening of the ring and effect of the ZBD**

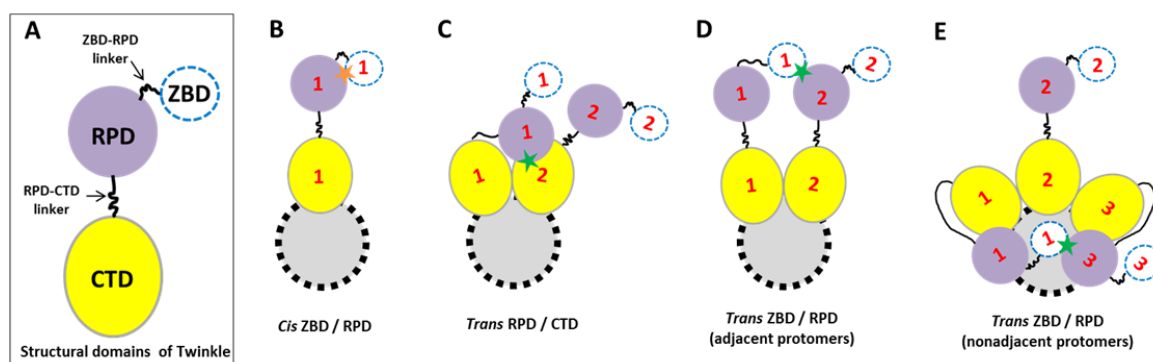
Our NS-EM image analysis revealed that hexameric and heptameric complexes formed by the full-length Twinkle present a wide range of mostly asymmetric conformations produced by multiple combinations of axial displacement and bending of the protomers. The full-length protein sample showed an important number of the n-1 bent arms form, but no particles with all arms contracted were detected. Interestingly, the  $\Delta$ ZBD protein presented both n-1 bent arms extended and fully contracted particles, demonstrating that ZBD is not required to acquire these conformations and suggesting that, instead, RPD/CTD and/or RPD-CTD linker/CTD interactions could likely be involved. On the other hand, the absence of completely contracted conformations in the full-length Twinkle suggests that the presence of ZBD somehow precludes the simultaneous bending of all the protomers, probably by a steric impediment. Considering the absence of fully contracted particle states in the wild type protein and, instead, the presence of a significant proportion of n-1 arms bent particles, it seems important to consider whether there is any functional advantage for the enzyme to acquire a conformation where only one NTD is extended, fully exposed to the solvent. It has been proposed a ring-opening mechanism for ssDNA binding in the central channel of the T7 gp4 (Ahnert et al. 2000) where a transient binding of NTD to a DNA molecule produces conformational changes that open the ring. One possibility would be that in Twinkle, multiple simultaneously NTD-DNA interactions could interfere with the opening of the ring, so only one NTD should initially bind to DNA to correctly activate the helicase opening for its subsequent loading onto the DNA. If only one DNA-binding site is available (the one of the extended monomer) that would restrict the binding of multiple DNA molecules to the helicase. Additionally, a full exposition of only one NTD would help to avoid an initial competition for binding the same DNA molecule by multiple protomers. In the same line of the putative ability of Twinkle to acquire an open conformation, our EM data of the wild type protein showed a small group of particles with a thin gap, which is also suggested in all our 3D initial map reconstructions and in a previous structural report of Twinkle (Fernández-Millán et al. 2015). The low frequency of open-ring

particles observed could be reflecting the stability of closed conformations in absence of DNA. It would be interesting to know whether the presence of DNA promotes an increase in the number of particles in open conformation; unfortunately, our attempts to obtain a soluble sample at the low salt concentration that the enzyme require to bind DNA, have failed.

The particles of the wild type protein showed a tendency to arrange some of their extended protomers into closer pairs, which could be in agreement with the report of highly stable dimeric forms of Twinkle resistant to the denaturant conditions of SDS-PAGE (Ziebarth et al. 2010). This observation suggests that multimerization of stable, preformed dimers, could contribute to the formation of more stable ring-like higher order species of Twinkle. In support of this idea, it was proposed in T7 gp4 and T4 gp41, an oligomerization mechanism (Notarnicola et al. 1995; Dong et al. 1995) where, first, protein dimers are formed through N-terminal “head-to-head” interactions and, subsequently, the new formed dimers oligomerize through C-terminal “tail-to-tail” interactions to form hexameric complexes. All the above-mentioned points would imply the presence of different protein-protein interfaces within the helicase ring. In contrast to the broad number of mostly asymmetric particles formed by the full-length Twinkle protein, the  $\Delta$ ZBD construct showed, aside from the more compacted states, particles with a predominantly symmetric distribution of extended protomers, where the NTDs are isolated from each other.

Contacts of Twinkle *in trans* between ZBDs and RPDs, and between RPDs and CTDs of adjacent protomers are suggested by the fitting of a Twinkle homology model (based on both T7 gp4 and DnaG) into the EM map of a chemically fixed, compact hexameric state of the enzyme (Fernández-Millán et al. 2015). Some of the ZBDs within the EM map were not detected, presumably due to their higher flexibility, which could explain why the putative ZBDs in our EM images of Twinkle are so difficult to detect (Figure 56). Additionally, both *cis* and *trans* interactions between ZBD and RPD are reported in T7 gp4 (Lee & Richardson 2002) and DnaG (Corn et al. 2005); a putative tight interaction *in cis* between ZBD and RPD of Twinkle (Figure 69 B) could also make difficult to identify the ZBD by NS-EM. The aforementioned domain interactions are all compatible with our EM results, which allowed us to propose their possible role on the different conformations observed, as follows: both fully compact (Figure 59 C) and n-1 bent arms (Figure 55 B and Figure 59 B) particle conformations could be produced by *trans* RPD/CTD (Figure 69 C) interactions, while *trans* ZBD/RPD interactions between either two neighbouring (Figure 69 D) or nonconsecutive (Figure 69 E) subunits seem to be responsible for the asymmetric conformations exhibited by the

full-length Twinkle (Figure 55 C), which are not displayed by the  $\Delta$ ZBD protein. Our EM analysis shows that ZBD of Twinkle is required for such additional conformations observed in the full-length protein, putting in evidence an essential role of this domain for the naturally highly dynamic behavior of the enzyme.



**Figure 69. Putative domain interactions of Twinkle that could explain the conformations observed by NS-EM. (A)** Representation of structural domains of Twinkle. CTD is shown in yellow and RPD in purple. ZBD is shown as a blue dashed circle, meaning that is not always detected by EM. The RPD-CTD and ZBD-RPD linkers are shown as wavy black lines. **(B)** *Cis* ZBD/ RPD interaction. **(C)** *Trans* RPD/ CTD and **(D)** ZBD/ CTD interactions between two consecutive protomers. **(E)** *Trans* ZBD/ RPD interactions between two nonconsecutive subunits. *Cis* and *trans* contacts between domain are represented by an orange and green star, respectively. In figures **B** to **E**, the gray circle delineated by a black dashed line represents the central channel of a ring-shaped complex; for simplifying the scheme, there is shown the minimal number of subunits that is required to explain each of the different types of domain interactions. Domains belonging to one subunit are labeled with the same number.

Taking together with the current knowledge about Twinkle, which include the identification of several highly flexible oligomeric states, the different intra and intersubunit interactions, the ability to carry out the antagonist helicase and unwinding activities (in a NTPase-dependent or NTP-independent manner, respectively), the presence of three different DNA binding sites and its affinity for binding different single-stranded and double-stranded DNA substrates, it seems likely that Twinkle plays a role in different aspects of DNA processing beyond replication. It has been reported that although truncation of the ZBD decreases the ability of the enzyme to bind ssDNA which, in consequence, also affects its ssDNA-stimulation of the ATPase activity, it does not significantly alters the activity of the replisome *in vitro* (Farge et al. 2008). Our results showing a notable structural role of the ZBD on the ability of the enzyme to acquire a wide number of conformational changes that, in absence of this domain do not occur, suggest that there must be an important, although not yet described functional role for these interdomain interactions in the activity of Twinkle. Since deletion of the ZBD does not seem to importantly affect the helicase activity of Twinkle *in vitro*, it could be playing a more essential role in other activities of the

enzyme, such as DNA annealing ability (Sen et al. 2012; Wu 2012), or in other regulatory activities under *in vivo* conditions. Our observations raise important new questions that, we expect, will lead the foundation for future research exploring further structural features and functional roles of the flexibility of Twinkle and its ZBD.

## Chapter 6

# 6. Conclusions

Concerning human CPAP subproject:

1. CPAP is a highly versatile protein essential for centriole biogenesis, reported to interact with many different proteins along the cell cycle. Together, the biochemical, biophysical and structural studies carried out allowed, for the first time, the isolation and identification of multiple homomeric complexes formed by CPAP<sup>897-1338</sup>. Structural characterization of the different forms of the protein showed that, in its monomeric state, CPAP<sup>897-1338</sup> behaves as a flexible filament, while the three dimensional reconstructions of both a dimeric and a tetrameric assemblies of the protein revealed globular and well-structured conformations.
2. The analysis of purified modular column-like structures with an axial repetition pattern around 8 nm suggested that they are formed by stacks of the CPAP<sup>897-1338</sup> tetramer, and, furthermore, that they could correspond to similar structures observed at the inner walls of the centriole.

Concerning mouse MCM4/6/7 subproject:

3. MCM4/6/7 helicase has a dynamic structural behavior in presence of ATP. The three dimensional analysis of the heteromeric MCM4/6/7 sample revealed, for the first time, the structure of different complexes that include: trimers, what seems to be the incipient dimerization of trimers and several ring-like hexameric conformations.
4. Fitting of homologue atomic structures of MCM proteins into the obtained three dimensional maps allowed the putative localization of structural domains, showing especially flexible regions, putatively designated to either separate or interacting C-



terminus of the two MCM6 copies in the helicase ring assemblies. Additionally, a small finger-like structure in one of the maps could correspond to the C-terminal domain of one of the two MCM4 protomers in the complex.

5. Supported by previous biochemical and structural studies of MCM helicases, structural features of our MCM4/6/7 density maps suggest that the subunit organization within the helicase hexamer is: MCM4'→MCM7'→MCM7→MCM4→MCM6→MCM6'

Concerning human Twinkle subproject:

6. The homomeric helicase Twinkle forms highly flexible assemblies, primarily hexamers and heptamers, although pentamers and octamers are found in a minority proportion.
7. Differently extended or compact conformations of Twinkle result as consequence of multiple combinations of the axial movements and bending of the monomers.
8. Comparison between the full-length protein and a N-terminal deletion mutant showed a significative structural role of the zinc binding domain on the ability of the enzyme to acquire a wide number of conformational changes.
9. Initial three dimensional reconstructions showed that extended conformations of the helicase are relatively flat, while the compact ones present a more bulky structure suggesting the presence of a second floor.

# Conclusiones

En referencia al subproyecto de CPAP:

1. CPAP es una proteína versátil, esencial en la biogénesis centriolar, la cual se ha reportado que interacciona con distintas proteínas a lo largo del ciclo celular. En su conjunto, los estudios bioquímicos, biofísicos y estructurales que se realizaron, han permitido, por primera vez, la identificación y el aislamiento de múltiples complejos de oligómeros formados por CPAP<sup>897-1338</sup>. La caracterización estructural de las distintas formas en que se encuentra la proteína mostraron que en su estado monomérico CPAP<sup>897-1338</sup> se comporta como hebras flexibles, mientras que las reconstrucciones en tres dimensiones de la forma dimerica y tetramérica revelaron conformaciones globulares bien estructuradas.
2. El análisis de unas estructuras modulares en forma de columna y con un patrón de repetición axial cercano a los 8 nm sugieren que dichas estructuras están formadas por pilas de tetrámeros de CPAP<sup>897-1338</sup>, y además, que podrían corresponder a estructuras similares a las que se encuentran localizadas en las paredes internas del centriolo.

En referencia al subproyecto de MCM4/6/7:

3. La helicasa MCM/6/7 muestra un comportamiento estructural muy dinámico en presencia de ATP. El análisis en tres dimensiones de la muestra heteromérica MCM4/6/7 permitió mostrar, por primera vez, la estructura de diferentes complejos, entre los cuales se incluyen: trímeros, lo que pareciera ser un estadio de dimerización de los trímeros, y varias conformaciones de hexámeros en forma de anillo.
4. El ajuste de modelos atómicos por homología dentro de los mapas en tres dimensiones obtenidos permitió localizar dominios estructurales de la proteína de manera putativa, mostrando zonas especialmente flexibles que fueron asignadas, también en forma putativa, a la región C-terminal de las dos copias de MCM6 que forman parte del complejo helicasa. Adicionalmente, se identificó una estructura en forma de dedo que podría corresponder a la región del C-terminal de una de las dos copias de MCM4.

5. Apoyándose en previos estudios bioquímicos y estructurales de MCM, varios de los elementos estructurales en nuestros mapas de densidad sugieren la siguiente organización de los protómeros dentro del complejo MCM4/6/7:  
MCM4'→MCM7'→MCM7→MCM4→MCM6→MCM6'

En referencia al subproyecto de Twinkle:

6. Twinkle forma complejos altamente flexibles que, es su mayoría, corresponden a formas hexaméricas y heptaméricas, aunque también se encuentra un número minoritario de pentámeros y octámeros.
7. El movimiento de los protómeros de la helicasa en distintas direcciones da lugar a múltiples conformaciones extendidas o compactas.
8. La comparación entre la proteína completa y un mutante con una delección en la región N-terminal de Twinkle muestra que el dominio de unión a zinc juega un papel estructural importante para que la enzima adquiera un amplio número de cambios conformacionales.
9. Las diferentes reconstrucciones iniciales en tres dimensiones que se obtuvieron muestran que las conformaciones extendidas de Twinkle son relativamente planas, mientras que las compactas presentan una estructura más abultada que sugiere la presencia de un segundo piso.

# References

- Adachi, Y., Usukura, J. & Yanagida, M., 1997. A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast. *Genes to Cells*, 2(7), pp.467–479.
- Ahnert, P., Picha, K.M. & Patel, S.S., 2000. A ring-opening mechanism for DNA binding in the central channel of the T7 helicase - primase protein. *EMBO*, 19(13), pp.3418–27.
- Alves-Cruzeiro, J.M.D.C., Nogales-Cadenas, R. & Pascual-Montano, A.D., 2014. CentrosomeDB: a new generation of the centrosomal proteins database for Human and Drosophila melanogaster. *Nucleic acids research*, 42(Database issue), pp.D430–6.
- Amos, L.A., Henderson, R. & Unwin, P.N.T., 1982. Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Progress in Biophysics and Molecular Biology*, 39, pp.183–231.
- Aravind, L. & Koonin, E. V., 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic acids research*, 27(23), pp.4658–4670.
- Aravind, L., Leipe, D.D. & Koonin, E. V., 1998. Toprim - a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*, 26(18), pp.4205–4213.
- Arias-Palomo, E. et al., 2013. The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell*, 153(2), pp.438–48.
- Azimzadeh, J. & Bornens, M., 2007. Structure and duplication of the centrosome. *Journal of cell science*, 120(Pt 13), pp.2139–42.
- Bae, B. et al., 2009. Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure*, 17(2), pp.211–222.
- Barry, E.R. et al., 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic acids research*, 35(3), pp.988–98.
- Barry, E.R. et al., 2009. Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), pp.1051–6.
- Bell, S.D. & Botchan, M.R., 2013. The minichromosome maintenance replicative helicase. *Cold Spring Harbor perspectives in biology*, 5(11), p.a012807.
- Bettencourt-Dias, M. & Glover, D.M., 2007. Centrosome biogenesis and function: centrosomes brings new understanding. *Nature reviews. Molecular cell biology*, 8(6), pp.451–63.
- Biswas-Fiss, E., Khopde, S. & Biswas, S., 2005. The Mcm467 complex of *Saccharomyces cerevisiae* is preferentially activated by autonomously replicating DNA sequences. *Biochemistry*, pp.2916–2925.
- Bochman, M.L., Bell, S.P. & Schwacha, A., 2008. Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Molecular and cellular biology*, 28(19), pp.5865–73.

- Bochman, M.L. & Schwacha, A., 2007. Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step. *The Journal of biological chemistry*, 282(46), pp.33795–804.
- Bochman, M.L. & Schwacha, A., 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiology and molecular biology reviews*, 73(4), pp.652–83.
- Bochman, M.L. & Schwacha, A., 2008. The Mcm2-7 complex has in vitro helicase activity. *Molecular cell*, 31(2), pp.287–93.
- Bradshaw, P.C. & Pfeiffer, D.R., 2006. Release of Ca<sup>2+</sup> and Mg<sup>2+</sup> from yeast mitochondria is stimulated by increased ionic strength. *BMC biochemistry*, 12(February), pp.1–12.
- Brewster, A.S. et al., 2008. Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA hexameric helicase. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51), pp.20191–20196.
- Brewster, A.S. & Chen, X.S., 2010. Insights into MCM functional mechanism: lessons learned from the archaeal MCM complex. *Critical reviews in biochemistry and molecular biology*, 45(3), pp.243–256.
- Burger, G., Gray, M.W. & Lang, B.F., 2003. Mitochondrial genomes: anything goes. *Trends in cell biology*, 19(12), pp.709–716.
- Burgess, S. a et al., 2004. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *Journal of structural biology*, 147(3), pp.247–58.
- Burkhardt, R., Schulte, D. & Hu, B., 1995. Interactions of human nuclear proteins P1Mcm3 and P1Cdc46. *European Journal of Biochemistry*, 438(2), pp.431–438.
- Carvalho-Santos, Z. et al., 2010. Stepwise evolution of the centriole-assembly pathway. *Journal of cell science*, 123(Pt 9), pp.1414–26.
- Chang, J. et al., 2010. PLK2 phosphorylation is critical for CPAP function in procentriole formation during the centrosome cycle. *The EMBO journal*, 29(14), pp.2395–406.
- Chen, C.-Y. et al., 2006. CPAP interacts with 14-3-3 in a cell cycle-dependent manner. *Biochemical and Biophysical Research Communications*, 342(4), pp.1203–1210.
- Chen, Y., Wei, L. & Mu, J.D., 2003. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp.15492–15497.
- Chenna, R., 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31(13), pp.3497–3500.
- Cho, J.-H. et al., 2006. Depletion of CPAP by RNAi disrupts centrosome integrity and induces multipolar spindles. *Biochemical and Biophysical Research Communications*, 339(3), pp.742–747.
- Chong, J.P.J., Thmmes, P. & Blow, J.J., 1996. The role of MCM / PI proteins in the licensing of DNA replication. *Trends in Biochemical Sciences*, 21(3), pp.102–106.
- Cizmecioglu, O. et al., 2010. Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *The Journal of cell biology*, 191(4), pp.731–9.

- Claycomb, J.M. et al., 2002. Visualization of replication initiation and elongation in *Drosophila*. *The Journal of cell biology*, 159(2), pp.225–36.
- Cole, C., Barber, J.D. & Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. *Nucleic acids research*, 36(Web Server issue), pp.W197–201.
- Comartin, D. et al., 2013. CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Current biology : CB*, 23(14), pp.1360–6.
- Cong, Y. & Ludtke, S.J., 2010. Single particle analysis at high resolution. *Methods in Enzymology*, 482, pp.211–235.
- Cormier, A. et al., 2009. The PN2-3 domain of centrosomal P4.1-associated protein implements a novel mechanism for tubulin sequestration. *The Journal of biological chemistry*, 284(11), pp.6909–17.
- Corn, J.E. et al., 2005. Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Molecular Cell*, 20(3), pp.391–401.
- Costa, A. et al., 2014. DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *eLife*, 3, p.e03273.
- Costa, A. et al., 2006a. Structural basis of the *Methanothermobacter thermautotrophicus* MCM helicase activity. *Nucleic acids research*, 34(20), pp.5829–38.
- Costa, A. et al., 2006b. Structural studies of the archaeal MCM complex in different functional states. *Journal of structural biology*, 156(1), pp.210–9.
- Costa, A. et al., 2011. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nature structural & molecular biology*, 18(4), pp.471–7.
- Costa, A. & Onesti, S., 2009. Structural biology of MCM helicases. *Critical reviews in biochemistry and molecular biology*, 44(5), pp.326–42.
- Cottee, M. a et al., 2013. Crystal structures of the CPAP/STIL complex reveal its role in centriole assembly and human microcephaly. *eLife*, 2, p.e01071.
- Coué, M., Amariglio, F. & Maiorano, D., 1998. Evidence for different MCM subcomplexes with differential binding to chromatin in *Xenopus*. *Experimental Cell Research*, 289(2), pp.282–289.
- Crampton, D.J. et al., 2006. Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of Molecular Biology*, 360(3), pp.667–677.
- Crevel, G. et al., 2007. Differential requirements for MCM proteins in DNA replication in *Drosophila* S2 cells. *PLoS one*, 2(9), p.e833.
- DaFonseca, C.J., Shu, F. & Zhang, J.J., 2001. Identification of two residues in MCM5 critical for the assembly of MCM complexes and Stat1-mediated transcription activation in response to IFN-gamma. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6), pp.3034–9.
- Dalton, S. & Whitbread, L., 1995. Cell cycle-regulated nuclear import and export of Cdc47, a protein essential for initiation of DNA replication in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7), pp.2514–2518.

- Daniel, D.C., Dagdanova, A. V & Johnson, E.M., 2013. The MCM and RecQ Helicase Families : Ancient Roles in DNA Replication and Genomic Stability Lead to Distinct Roles in Human Disease. In D. D. Stuart, ed. *The Mechanisms of DNA Replication*. pp. 59–89.
- Das, M. et al., 2014. MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. *Molecular Biology International*, pp.1–11.
- Davey, M.J., Indiani, C. & O'Donnell, M., 2003. Reconstitution of the Mcm2-7p heterohexameric, subunit arrangement, and ATP site architecture. *The Journal of biological chemistry*, 278(7), pp.4491–9.
- Debec, A. & Sullivan, W., 2010. Centrioles : active players or passengers during mitosis ? *Cell Molecular Life Science*, 67, pp.2173–2194.
- Dimitrova, D. & Todorov, I., 1999. Mcm2, but not RPA, is a component of the mammalian early G1-phase prereplication complex. *The Journal of cell biology*, 146(4), pp.709–722.
- Dobbins, S.E., Lesk, V.I. & Sternberg, M.J.E., 2008. Insights into protein flexibility : The relationship between normal modes and conformational change upon protein – protein docking. *Proceedings of the National Academy of Sciences of the United States of America*, 105(30), pp.10390–5.
- Dong, F., Gogol, E.P. & Von Hippel, P.H., 1995. The phage T4-coded DNA replication helicase (gp41) forms a hexamer upon activation by nucleoside triphosphate. *Journal of Biological Chemistry*, 270(13), pp.7462–7473.
- Doxsey, S., 2001. RE-EVALUATING CENTROSOME FUNCTION. *Nature Reviews Molecular Cell Biology*, 2, pp.688–698.
- Doxsey, S., Zimmerman, W. & Mikule, K., 2005. Centrosome control of the cell cycle. *Trends in Cell Biology*, 15(6), pp.303–311.
- Dünder, H. et al., 2012. Identification of a novel Twinkle mutation in a family with infantile onset spinocerebellar ataxia by whole exome sequencing. *Pediatric Neurology*, 46(3), p.2012.
- Dziak, R. et al., 2003. Evidence for a role of MCM (mini-chromosome maintenance)5 in transcriptional repression of sub-telomeric and Ty-proximal genes in *Saccharomyces cerevisiae*. *The Journal of biological chemistry*, 278(30), pp.27372–81.
- Egli, M., 2010. Diffraction techniques in structural biology. In *Current Protocols in Nucleic Acid Chemistry*.
- Ellis, R.J. & Ellis, R.J., 2001. Macromolecular crowding : obvious but underappreciated. *tre*, 26(10), pp.597–604.
- Erzberger, J.P. & Berger, J.M., 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure*, 35, pp.93–114.
- Evrin, C. et al., 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), pp.20240–5.
- Fan, L. et al., 2006. A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase. *Journal of molecular biology*, 358(5), pp.1229–1243.

- Farge, G. et al., 2008. The N-terminal domain of Twinkle contributes to single-stranded DNA binding and DNA helicase activities. *Nucleic acids research*, 36(2), pp.393–403.
- Fernández-Millán, P. et al., 2015. The hexameric structure of the human mitochondrial replicative helicase Twinkle. *Nucleic acids research*, March, pp.1–12.
- Fitch, M.J., Donato, J.J. & Tye, B.K., 2003. Mcm7, a subunit of the presumptive MCM helicase, modulates its own expression in conjunction with Mcm1. *The Journal of biological chemistry*, 278(28), pp.25408–16.
- Fletcher, R.J. et al., 2003. The structure and function of MCM from archaeal *M. Thermoautotrophicum*. *Nature structural biology*, 10(3), pp.160–7.
- Forsburg, S.L. et al., 1997. Mutational analysis of Cdc19p, a *Schizosaccharomyces pombe* MCM protein. *Genetics*, 1041(3), pp.1025–1041.
- Frank, J., 1996. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Oxford University Press.
- Fratter, C. et al., 2010. The clinical, histochemical, and molecular spectrum of PEO1 (Twinkle)-linked adPEO. *Neurology*, 74(20), pp.1619–1626.
- Frueh, D.P. et al., 2013. NMR methods for structural studies of large monomeric and multimeric proteins. *Current opinion in structural biology*, 23(5), pp.734–9.
- Giaginis, C. & Vgenopoulou, S., 2010. MCM proteins as diagnostic and prognostic tumor markers in the clinical setting. *Histology and Histopathology*, 25(3), pp.351–370.
- Giese, K.C. & Vierling, E., 2002. Changes in oligomerization are essential for the chaperone activity of a small heat shock protein in vivo and in vitro. *The Journal of biological chemistry*, 277(48), pp.46310–8.
- Gineau, L., Cognet, C. & Kara, N., 2012. Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. *The Journal of Clinical Investigation*, 122(3), pp.821–832.
- Glaeser, R.M. & Hall, R.J., 2011. Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects. *Biophysical journal*, 100(10), pp.2331–7.
- Gómez, E.B., Catlett, M.G. & Forsburg, S.L., 2002. Different phenotypes in vivo are associated with ATPase motif mutations in *Schizosaccharomyces pombe* minichromosome maintenance proteins. *Genetics*, 160(April), pp.1305–1318.
- Gopalakrishnan, J. et al., 2011. Sas-4 provides a scaffold for cytoplasmic complexes and tethers them in a centrosome. *Nature Communications*, 2(359), pp.1–11.
- Graham, B.W. et al., 2011. Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Research*, 39(15), pp.6585–6595.
- Gray, M.W., Burger, G. & Lang, B.F., 1999. Mitochondrial Evolution. *Science*, 283(5407), pp.1476–1482.
- Green, M. & Sambrook, J., 2012. *Molecular cloning: a laboratory manual* 4th ed., New York: Cold Spring Harbor Laboratory.



- Grigorieff, N., 2007. FREALIGN: high-resolution refinement of single particle structures. *Journal of structural biology*, 157(1), pp.117–25.
- Gudi, R. et al., 2014. Centrobin-centrosomal protein 4.1-associated protein (CPAP) interaction promotes CPAP localization to the centrioles during centriole duplication. *The Journal of biological chemistry*, 289(22), pp.15166–15178.
- Guo, S., Tabor, S. & Richardson, C., 1999. The linker region between the helicase and primase domains of the bacteriophage T7 gene 4 protein is critical for hexamer formation. *Journal of Biological Chemistry*, 274(42), pp.30303–30309.
- Hackenbrock, B.Y.C.R., 1968. Chemical and physical fixation of isolated mitochondria in low-energy and high-energy states. *Progress in Biophysics and Molecular Biology*, 61(2), pp.598–605.
- Hatzopoulos, G.N. et al., 2013. Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture. *Structure*, 21(11), pp.2069–77.
- Haven, N., 1994. Domains of Escherichia coli primase: functional activity of a 47-kDa N-terminal proteolytic fragment. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp.11462–11466.
- Havugimana, P.C. et al., 2012. A census of human soluble protein complexes. *Cell*, 150(5), pp.1068–1081.
- Heel, M. Van & Schatz, M., 2005. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3), pp.250–262.
- Henderson, R. et al., 2012. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20(2), pp.205–214.
- Hennessy, K.M. et al., 1991. A group of interacting yeast DNA replication genes. *Genes & Development*, 5(6), pp.958–969.
- Hesketh, E.L. et al., 2015. DNA induces conformational changes in a recombinant human minichromosome maintenance complex. *Journal of Biological Chemistry*, 290(12), pp.1–15.
- Hingorani, M.M. & Patel, S.S., 1993. Interactions of bacteriophage T7 DNA primase/helicase protein with single-stranded and double-stranded DNAs. *Biochemistry*, 32(46), pp.12478–12487.
- Holthoff, H.P., 1998. Human protein MCM6 on HeLa cell chromatin. *Journal of Biological Chemistry*, 273(13), pp.7320–7325.
- Hsu, W.-B. et al., 2008. Functional characterization of the microtubule-binding and -destabilizing domains of CPAP and d-SAS-4. *Experimental Cell Research*, 314(14), pp.2591–2602.
- Hu, B. et al., 1993. The P1 family: a new class of nuclear mammalian proteins related to the yeast Mcm replication proteins. *nuc*, 21(23), pp.5289–5293.
- Hughes, C.R. et al., 2012. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. *The Journal of clinical investigation*, 122(3), pp.814–820.
- Hung, L. et al., 2004. Identification of a novel microtubule-destabilizing motif in CPAP that binds to tubulin heterodimers and inhibits microtubule assembly. *Molecular Biology of the Cell*, 15(June), pp.2697–2706.

- Hung, L.Y., Tang, C.J. & Tang, T.K., 2000. Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the gamma-tubulin complex. *Molecular and Cellular Biology*, 20(20), pp.7813–7825.
- Hydrolysis, D.A.T.P. et al., 1984. Structural and functional studies of the dnaB protein using limited proteolysis. Characterization of domains for DNA-dependent ATP hydrolysis and for protein association in the primosome. *Journal of Biological Chemistry*, 259(1), pp.88–96.
- Hyrien, O., Marheineke, K. & Goldar, A., 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 25(2), pp.116–25.
- Ibarra, A., Schwob, E. & Méndez, J., 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proceedings of the National Academy of Sciences of the United States of America*, 105(26), pp.8956–61.
- Ichinose, S., 1996. Binding of human minichromosome maintenance proteins with histone H3. *Journal of Biological Chemistry*, 271(39), pp.24115–24122.
- Ilves, I. et al., 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Molecular cell*, 37(2), pp.247–58.
- Ilyina, T. V., Gorbalenya, A.E. & Koonin, E. V., 1992. Organization and evolution of bacterial and bacteriophage primase-helicase systems. *Journal of molecular evolution*, 34(4), pp.351–357.
- Ishimi, Y., 1997a. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *The Journal of biological chemistry*, 272(39), pp.24508–13.
- Ishimi, Y., 1997b. A DNA Helicase Activity Is Associated with an MCM4, -6, and -7 Protein Complex. *Journal of Biological Chemistry*, 272(39), pp.24508–24513.
- Ishimi, Y. et al., 2001. Biochemical activities associated with mouse Mcm2 protein. *The Journal of biological chemistry*, 276(46), pp.42744–52.
- Iyer, L.M. et al., 2004. Evolutionary history and higher order classification of AAA+ ATPases. *Journal of structural biology*, 146(1-2), pp.11–31.
- Jemt, E. et al., 2011. The mitochondrial DNA helicase TWINKLE can assemble on a closed circular template and support initiation of DNA synthesis. *Nucleic acids research*, 39(21), pp.9238–49.
- Jenkinson, E.R. & Chong, J.P.J., 2006. Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), pp.7613–7618.
- Ji, K. et al., 2014. Twinkle mutations in two Chinese families with autosomal dominant progressive external ophthalmoplegia. *Neurological Sciences*, 35(3), pp.443–448.
- Jin, Q. et al., 2014. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure*, 22(3), pp.496–506.
- Jonic, S. et al., 2015. StructMap: Multivariate distance analysis of elastically aligned electron microscopy structures for exploring pathways of conformational changes. *Structure*, p.Under revision.

- Kanter, D.M., Bruck, I. & Kaplan, D.L., 2008. Mcm subunits can assemble into two different active unwinding complexes. *The Journal of biological chemistry*, 283(45), pp.31172–82.
- Kasiviswanathan, R. et al., 2004. Biochemical characterization of the Methanothermobacter thermoautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. *The Journal of biological chemistry*, 279(27), pp.28358–66.
- Kato, M. et al., 2003. Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. *Molecular cell*, 11(5), pp.1349–1360.
- Kearsey, S.E. & Labib, K., 1998. MCM proteins: evolution, properties, and role in DNA replication. *Biochimica et biophysica acta*, 1398(2), pp.113–36.
- Kearsey, S.E., Labib, K. & Maiorano, D., 1996. Cell cycle control of eukaryotic DNA replication. *Current opinion in Genetics and Development*, 6(2), pp.208–214.
- Keck, J.L. et al., 2000. Structure of the RNA polymerase domain of E. coli primase. *Science*, 287(5462), pp.2482–2487.
- Kelman, Z., Lee, J.-K. & Hurwitz, J., 1999. The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum Delta H contains DNA helicase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26), pp.14783–14788.
- Kim, M.K., Dudognon, C. & Smith, S., 2012. Tankyrase 1 regulates centrosome function by controlling CPAP stability. *EMBO reports*, 13(8), pp.724–32.
- Kimura, H. et al., 1995. Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and characterization of the products: physical interaction between P1 (MCM3) and CDC46 proteins. *Nucleic Acids Research*, 23(12), pp.2097–2104.
- Kinoshita, Y. & Johnson, E.M., 2004. Site-specific loading of an MCM protein complex in a DNA replication initiation zone upstream of the c-MYC gene in the HeLa cell cycle. *The Journal of biological chemistry*, 279(34), pp.35879–89.
- Kirkham, M. et al., 2003. SAS-4 is a C. elegans centriolar protein that controls centrosome size. *Cell*, 112(4), pp.575–87.
- Kitagawa, D. et al., 2011. Spindle positioning in human cells relies on proper centriole formation and on the microcephaly proteins CPAP and STIL. *Journal of cell science*, 124(Pt 22), pp.3884–93.
- Kley, J. et al., 2011. Structural adaptation of the plant protease Deg1 to repair photosystem II during light exposure. *Nature structural & molecular biology*, 18(6), pp.728–31.
- Kleylein-sohn, J. et al., 2007. Plk4-induced centriole biogenesis in human cells. *Developmental Cell*, 13(2), pp.190–202.
- Kohlmaier, G. et al., 2009. Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. *Current Biology*, 19(12), pp.1012–1018.
- Koonin, E. V., 1993. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic acids research*, 21(11), pp.2541–7.

- Korhonen, J. a et al., 2004. Reconstitution of a minimal mtDNA replisome in vitro. *The EMBO journal*, 23(12), pp.2423–9.
- Korhonen, J. a, Gaspari, M. & Falkenberg, M., 2003. TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *The Journal of biological chemistry*, 278(49), pp.48627–32.
- Krude, T. & Musahl, C., 1996. Human replication proteins hCdc21, hCdc46 and P1Mcm3 bind chromatin uniformly before S-phase and are displaced locally during DNA replication. *Journal of cell science*, 318(Pt 2), pp.309–318.
- Krueger, S. et al., 2014. The solution structure of full-length dodecameric MCM by SANS and molecular modeling. *Proteins*, (April), pp.1–11.
- Kuchta, R.D. & Stengel, G., 2010. Mechanism and evolution of DNA primases. *Biochimica et Biophysica Acta*, 1804(5), pp.1180–1189.
- Kusakabe, T. & Richardson, C.C., 1996. The role of the zinc motif in sequence recognition by DNA primases. *Journal of Biological Chemistry*, 271(32), pp.19563–19570.
- Kutter, S. et al., 2007. The influence of protein concentration on oligomer structure and catalytic function of two pyruvate decarboxylases. *The Protein Journal*, 26, pp.585–591.
- Kuznetsova, I.M., Turoverov, K.K. & Uversky, V.N., 2014. What macromolecular crowding can do to a protein. *International Journal of Molecular Science*, 15(12), pp.23090–23140.
- Kyriakouli, D.S. et al., 2008. Progress and prospects: gene therapy for mitochondrial DNA disease. *Gene Therapy*, 15(14), pp.1017–1023.
- De la Rosa-Trevín, J.M. et al., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *Journal of structural biology*, 184(2), pp.321–8.
- Larkin, M. a et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8.
- Leal, G.F. et al., 2003. A novel locus for autosomal recessive primary microcephaly (MCPH6) maps to 13q12.2. *Journal of medical genetics*, 40(7), pp.540–2.
- Lee, J.K. & Hurwitz, J., 2000. Isolation and characterization of various complexes of the minichromosome maintenance proteins of *Schizosaccharomyces pombe*. *The Journal of biological chemistry*, 275(25), pp.18871–8.
- Lee, J.K. & Hurwitz, J., 2001. Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. *Proceedings of the National Academy of Sciences of the United States of America*, 98(1), pp.54–9.
- Lee, S. et al., 2012. Zinc-binding domain of the bacteriophage T7 DNA primase modulates binding to the DNA template. *Journal of biochemistry*, 287(46), pp.39030–39040.
- Lee, S. & Richardson, C.C., 2002. Interaction of adjacent primase domains within the hexameric gene 4 helicase-primase of bacteriophage T7. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp.12703–12708.

- Li, N. et al., 2015. Structure of the eukaryotic MCM complex at 3.8 Å. *Nature*, (July), pp.1–6.
- Li, S. et al., 2012. Three-dimensional structure of basal body triplet revealed by electron cryo-tomography. *The EMBO journal*, 31(3), pp.552–62.
- Lin, Y.-C. et al., 2013. Human microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly. *The EMBO journal*, 32(8), pp.1141–54.
- Lin, Y.-N. et al., 2013. CEP120 interacts with CPAP and positively regulates centriole elongation. *The Journal of cell biology*, 202(2), pp.211–9.
- Linding, R., 2003. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research*, 31(13), pp.3701–3708.
- Liu, C. et al., 2012. Structural insights into the Cdt1-mediated MCM2-7 chromatin loading. *Nucleic acids research*, 40(7), pp.3208–17.
- Liu, W. et al., 2008. Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic acids research*, 36(10), pp.3235–43.
- Longley, M.J. et al., 2010. Disease variants of the human mitochondrial DNA helicase encoded by C10orf2 differentially alter protein stability, nucleotide hydrolysis, and helicase activity. *The Journal of biological chemistry*, 285(39), pp.29690–702.
- Ludtke, S.J., Baldwin, P.R. & Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of structural biology*, 128(1), pp.82–97.
- Lupas, A., M, V.D. & Stock, J., 1991. Predicting coiled coils from protein sequences. *Science*, 252(5009), pp.1162–4.
- Lyubimov, A.Y. et al., 2012. ATP-dependent conformational dynamics underlie the functional asymmetry of the replicative helicase from a minimalist eukaryote. *Proceedings of the National Academy of Sciences of the United States of America*, 109(30), pp.11999–2004.
- Ma, X. et al., 2010. The effects of oligomerization on *Saccharomyces cerevisiae* Mcm4/6/7 function. *BMC biochemistry*, 11(37), pp.1–15.
- Madine, M. et al., 1995. The nuclear envelope prevents reinitiation of replication by regulating the binding of MCM3 to chromatin in *Xenopus* egg extracts. *Current Biology*, 5(11), pp.1270–1279.
- Mahrenholz, C.C. et al., 2011. Complex networks govern coiled-coil oligomerization--predicting and profiling by means of a machine learning approach. *Molecular & cellular proteomics : MCP*, 10(5), p.M110.004994.
- Maine, G., Sinha, P. & Tye, B., 1984. Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. *Genetics*, (Hand 1978).
- Manuscript, A., 2011. Definition and estimation of resolution in single-particle reconstructions. *Structure*, 18(7), pp.768–775.
- Mao, C. & Holt, I.J., 2009. Clinical and molecular aspects of diseases of mitochondrial DNA instability. *Chang Gung Med.*, 32, pp.354–369.

- Marsh, J. a & Teichmann, S. a, 2014. Protein flexibility facilitates quaternary structure assembly and evolution. *PLoS biology*, 12(5), p.e1001870.
- Marsh, J.A. & Teichmann, S.A., 2013. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *Bioessays*, 36(2), pp.209–218.
- Marsh, J.A. & Teichmann, S.A., 2011. Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure/Folding and Design*, 19(6), pp.859–867.
- Matson, S.W., Richardson, C.C. & A, A.S.U.S., 1985. Nucleotide-dependent Binding of the Gene 4 Protein of Bacteriophage T7 to Single-stranded DNA. *Journal of Biological Chemistry*, 260(4), pp.2281–7.
- Matsushima, Y. & Kaguni S., L., 2009. Functional importance of the conserved N-terminal domain of the mitochondrial replicative helicase. *Biochimica et biophysica acta*, 1787(5), pp.290–295.
- Mcbride, H.M. & Neuspiel, M., 2006. Mitochondria: More than just a powerhouse. *Current Biology*, 16(14), pp.551–560.
- McGeoch, A.T. et al., 2005. Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nature structural & molecular biology*, 12(9), pp.756–62.
- Mcm, T., 1994. The MCM2-3-5 proteins : are they replication licensing factors ? *Trends in Cell Biology*, 4(5), pp.160–166.
- Milazzo, A. et al., 2011. Characterization of a direct detection device imaging camera for transmission electron microscopy. *Ultramicroscopy*, 110(7), pp.744–747.
- Miller, J.M. et al., 2014. Analysis of the crystal structure of an active MCM hexamer. *eLife*, 3, pp.1–13.
- Ming, D. et al., 2002. How to describe protein motion without amino acid sequence and atomic coordinates. *Proceedings of the National Academy of Sciences of the United States of America*, 99(13), pp.8620–5.
- Moir, D.O.N. & Botstein, D., 1982. Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics*, 100(4), pp.547–563.
- Murphy, E. & A. Eisner, D., 2010. Regulation of intracellular and mitochondrial sodium in health and disease. *Circulation Research*, 104(3), pp.292–303.
- Musahl, C. et al., 1995. A human homologue of the yeast replication protein Cdc21. Interactions with other Mcm proteins. *European Journal of Biochemistry*, 230(3), pp.1096–1101.
- Nguyen, T. et al., 2012. Interactions of the human MCM-BP protein with MCM complex components and Dbf4. *PloS one*, 7(4), p.e35931.
- Nitani, N. et al., 2008. Mcm4 C-terminal domain of MCM helicase prevents excessive formation of single-stranded DNA at stalled replication forks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), pp.12973–8.
- Nogales-Cadenas, R. et al., 2009. CentrosomeDB: a human centrosomal proteins database. *Nucleic acids research*, 37(Database issue), pp.D175–80.

- Nogueira, C. et al., 2014. Syndromes associated with mitochondrial DNA depletion. *Italian Journal of Pediatrics*, 40(34), pp.1–10.
- Noree, C. et al., 2010. Identification of novel filament-forming proteins in. *Journal of Cell Biology*, 190(4), pp.541–551.
- Notarnicola, S. et al., 1995. A domain of the gene 4 helicase/primase of bacteriophage T7 required for the formation of an active hexamer. *Journal of Biological Chemistry*, 270(34), pp.20215–24.
- O’Shea, V.L. and Berger, J.M., 2014. Loading strategies of ring-shaped nucleic acid translocases and helicases. *Current opinion in structural biology*, 25, pp.16–24.
- Pascual-Montano, A. et al., 2006. Optimization problems in electron microscopy of single particles. *Ann. Oper. Res.*, 148, pp.133–165.
- Patel, S.S. & Picha, K.M., 2000. Structure and function of hexameric helicases. *Annu. Rev. Biochem.*, 3(69), pp.651–97.
- Patelt, S.S., 1995. Bacteriophage T7 helicase / primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9), pp.3869–3873.
- Petrovska, I. et al., 2014. Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *eLife*, (April), pp.1–19.
- Petterson, E.F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–12.
- Pham, X.H. et al., 2006. Conserved sequence box II directs transcription termination and primer formation in mitochondria. *Journal of Biological Chemistry*, 281(34), pp.24647–24652.
- Podobnik, M. et al., 2000. A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomerases. *Journal of molecular biology*, 300(2), pp.353–362.
- Prokhorova, T.A. & Blow, J.J., 2000. Sequential MCM/P1 subcomplex assembly is required to form a heterohexamer with replication licensing activity. *Journal of Biological Chemistry*, 275(4), pp.2491–2498.
- Pucci, B. et al., 2007. Modular organization of the Sulfolobus solfataricus mini-chromosome maintenance protein. *The Journal of biological chemistry*, 282(17), pp.12574–82.
- Radermacher, M., 1988. Three dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of Electron Microscopy Technique*, 9(4), p.1988.
- Remus, D. et al., 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell*, 139(4), pp.719–30.
- Remus, D. & Diffley, J.F.X., 2009. Eukaryotic DNA replication control: lock and load, then fire. *Current opinion in cell biology*, 21(6), pp.771–7.
- Reyes, A. et al., 2013. Mitochondrial DNA replication proceeds via a “ bootlace ” mechanism involving the incorporation of processed transcripts. *Nucleic acids research*, 41(11), pp.5837–5850.

- Richter, A. & Knippers, R., 1997. High-molecular-mass complexes of human minichromosome-maintenance proteins in mitotic cells. *European Journal of Biochemistry*, 141, pp.136–141.
- Robinson, C. V., Sali, A. & Baumeister, W., 2007. The molecular sociology of the cell. *Nature*, 450(7172), pp.973–982.
- Romero, P. & Obradovic, Z., 1998. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.*, pp.437–48.
- 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), pp.721–745.
- Rossmann, M.G. et al., 2005. Combining X-ray crystallography and electron microscopy. *Structure (London, England : 1993)*, 13(3), pp.355–62.
- Rothenberg, E. et al., 2007. MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *Journal of Biological Chemistry*, 282(47), pp.34229–34234.
- Roy, A., Kucukural, A. & Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, 5(4), pp.725–38.
- Sakakibara, N., Kelman, L.M. & Kelman, Z., 2009. Unwinding the structure and function of the archaeal MCM helicase. *Molecular Microbiology*, 72(2), pp.286–296.
- Dos Santos, H.G. et al., 2013. Structure and non-structure of centrosomal proteins. *PloS one*, 8(5), p.e62633.
- Sato, M. et al., 2000. Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *Journal of molecular biology*, 300(3), pp.421–31.
- Scheffler, I.E., 2007. *Mitochondria* 2nd ed. John Wiley & Sons, ed., Hoboken.
- Scheres, S.H.W., 2015. Semi-automated selection of cryo-EM particles in RELION-1 . 3. *Journal of Structural Biology*, 189(2), pp.114–122.
- Scheres, S.H.W. & Chen, S., 2012. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9(9), pp.853–854.
- Schroer, M.A. et al., 2012. The effect of ionic strength, temperature, and pressure on the interaction potential of dense protein solutions: from nonlinear pressure response to protein crystallization. , 102(June), pp.2641–2648.
- Schulte, D., Burkhart, R. & Musahl, C., 1995. Expression, phosphorylation and nuclear localization of the human P1 protein, a homologue of the yeast Mcm 3 replication protein. *Journal of cell science*, 108(Pt 4), pp.1381–1389.
- Schwacha, A. & Bell, S.P., 2001. Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. *Molecular Cell*, 8(5), pp.1093–1104.
- Sen, D. et al., 2012. Human mitochondrial DNA helicase TWINKLE is both an unwinding and annealing helicase. *The Journal of biological chemistry*, 287(18), pp.14545–56.



- Sherman, D. & Forsburg, S., 1998. Schizosaccharomyces pombe Mcm3p, an essential nuclear protein, associates tightly with Nda4p (Mcm5p). *Nucleic acids research*, 26(17), pp.3955–3960.
- Sherman, D.A., Pasion, S.G. & Forsburg, S.L., 1998. Multiple domains of fission yeast Cdc19p (MCM2) are required for its association with the core MCM complex. *Molecular Biology of the Cell*, 9(7), pp.1833–1845.
- Sheu, Y.-J. et al., 2014. Domain within the helicase subunit Mcm4 integrates multiple kinase signals to control DNA replication initiation and fork progression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), pp.E1899–908.
- Shi, Y. et al., 2008. Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), pp.11122–7.
- Shutt, T.E. & Gray, M.W., 2006a. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in genetics*, 22(2), pp.90–95.
- Shutt, T.E. & Gray, M.W., 2006b. Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *Journal of molecular evolution*, 62(5), pp.588–99.
- Singleton, M.R. et al., 2000. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, 101, pp.589–600.
- Singleton, M.R., Dillingham, M.S. & Wigley, D.B., 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry*, 76, pp.23–50.
- Sorzano, C.O.S. et al., 2010. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of structural biology*, 171(2), pp.197–206.
- Spelbrink, J.N. et al., 2001. Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nature genetics*, 28(3), pp.223–31.
- Sterner, J.M. et al., 1998. Negative regulation of DNA replication by the retinoblastoma protein is mediated by its association with MCM7. *Molecular and cellular biology*, 18(5), pp.2748–2757.
- Su, T.T., Feger, G. & Farrell, P.H.O., 1996. Drosophila MCM protein complexes. *Molecular Biology of the Cell*, 7(2), pp.319–329.
- Sun, J. et al., 2013. Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nature structural & molecular biology*, 20(8), pp.944–51.
- Sun, J. et al., 2014. Structural and mechanistic insights into Mcm2 – 7 double-hexamers assembly and function. *Genes & development*, 28(20), pp.2291–2303.
- Suomalainen, A. et al., 1997. Autosomal dominant progressive external ophthalmoplegia with multiple deletions of mtDNA: clinical, biochemical, and molecular genetic features of the 10q-linked disease. *Neurology*, 48(5), pp.1244–53.
- Takahashi, T.S., Wigley, D.B. & Walter, J.C., 2005. Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends in biochemical sciences*, 30(8), pp.437–44.

- Tama, F., Wriggers, W. & Chaco, P., 2003. Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *Journal of molecular biology*, 2836(02), pp.485–492.
- Tama, F., Wriggers, W. & Iii, C.L.B., 2002. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *Journal of molecular biology*, 2836(02), pp.297–305.
- Tang, C.-J.C. et al., 2009. CPAP is a cell-cycle regulated protein that controls centriole length. *Nature Cell Biology*, 11(7), pp.825–831.
- Tang, C.-J.C. et al., 2011. The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. *The EMBO Journal*, 30(23), pp.4790–804.
- Tompa, P. & Fuxreiter, M., 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in biochemical sciences*, 33(1), pp.2–8.
- Toth, E. a. et al., 2003. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Molecular Cell*, 12(5), pp.1113–1123.
- Tougu, K., Peng, H. & Marians, J., 1994. Identification of a domain of Escherichia coli primase required for functional interaction with the DnaB helicase at the replication fork. *Journal of Biological Chemistry*, 269(6), pp.4675–4682.
- Tower, J., 2015. Programmed cell death in aging. *Ageing Research Reviews*, (pii: S1568-1637(15)00036-7), pp.1–11.
- Tran, N.Q. et al., 2010. A single subunit MCM6 from pea forms homohexamer and functions as DNA helicase. *Plant molecular biology*, 74(4-5), pp.327–36.
- Treviño, M. a et al., 2014. Emergence of structure through protein-protein interactions and pH changes in dually predicted coiled-coil and disordered regions of centrosomal proteins. *Biochimica et biophysica acta*.
- Tsuruga, H. et al., 1997. Expression, nuclear localization and interactions of human MCM/P1 proteins. *Biochemical and biophysical research communications*, 236(1), pp.118–25.
- Tye, B.I.K.K., 1996. Physical interactions among Mcm proteins and effects of Mcm dosage on DNA replication in Saccharomyces cerevisiae. *Molecular and cellular biology*, 16(9), pp.5081–5090.
- Tye, B.K. & Sawyer, S., 2000. The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts. *The Journal of biological chemistry*, 275(45), pp.34833–6.
- Tyynismaa, H. et al., 2004. Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number. *Human molecular genetics*, 13(24), pp.3219–3227.
- Uhn, L.E.A.K., 1999. The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: Implications for a dual role as a primer recognition factor and processivity clamp. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17), pp.9527–9532.
- Vargas, J. et al., 2014. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics (Oxford, England)*, pp.1–8.

- Vincent, T.L., Green, P.J. & Woolfson, D.N., 2013. LOGICOIL--multi-state prediction of coiled-coil oligomeric state. *Bioinformatics (Oxford, England)*, 29(1), pp.69–76.
- Vulprecht, J. et al., 2012. STIL is required for centriole duplication in human cells. *Journal of cell science*, 125(Pt 5), pp.1353–62.
- Wanrooij, S. et al., 2010. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Molecular Cell*, 37(1), pp.67–78.
- Wanrooij, S. & Falkenberg, M., 2010. The human mitochondrial replication fork in health and disease. *BBA - Bioenergetics*, 1797(8), pp.1378–1388.
- Watanabe, E., Ohara, R. & Ishimi, Y., 2012. Effect of an MCM4 mutation that causes tumours in mouse on human MCM4/6/7 complex formation. *Journal of biochemistry*, 152(2), pp.191–8.
- Wei, Z. et al., 2010. Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. *The Journal of biological chemistry*, 285(17), pp.12469–73.
- Wiedemann, C. et al., 2015. Structure and regulatory role of the C-terminal winged helix domain of the archaeal minichromosome maintenance complex. *Nucleic Acids Research*, 43(5), pp.2958–2967.
- Woodward, A.M. et al., 2006. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *The Journal of cell biology*, 173(5), pp.673–83.
- Wu, Y., 2012. Unwinding and rewinding: double faces of helicase? *Journal of nucleic acids*, 2012, p.140601.
- Xu, M., Chang, Y.P. & Chen, X.S., 2013. Expression, purification and biochemical characterization of *Schizosaccharomyces pombe* Mcm4, 6 and 7. *BMC biochemistry*, 14, p.5.
- Yabuta, N. et al., 2003. Mammalian Mcm2/4/6/7 complex forms a toroidal structure. *Genes to Cells*, 8(5), pp.413–421.
- Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M.D., 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research*, 34(2), pp.564–74.
- Yankulov, K. et al., 1999. MCM proteins are associated with RNA polymerase II holoenzyme. *Molecular and cellular biology*, 19(9).
- Ying, C.Y. & Gautier, J., 2005. The ATPase activity of MCM2-7 is dispensable for pre-RC assembly but is required for DNA unwinding. *The EMBO journal*, 24(24), pp.4334–44.
- Yong, Y. & Romano, L.J., 1995. Nucleotide and DNA-induced conformational changes in the bacteriophage T7 gene 4 protein. *Journal of Biological Chemistry*, 270(41), pp.24509–24517.
- You, Z. et al., 2002. Roles of Mcm7 and Mcm4 subunits in the DNA helicase activity of the mouse Mcm4/6/7 complex. *The Journal of biological chemistry*, 277(45), pp.42471–9.
- You, Z., Komamura, Y. & Ishimi, Y., 1999. Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. *Molecular and cellular biology*, 19(12), pp.8003–15.
- You, Z. & Masai, H., 2008. Cdt1 forms a complex with the minichromosome maintenance protein (MCM) and activates its helicase activity. *The Journal of biological chemistry*, 283(36), pp.24469–77.

- You, Z. & Masai, H., 2005. DNA binding and helicase actions of mouse MCM4/6/7 helicase. *Nucleic acids research*, 33(9), pp.3033–47.
- Yu, X. et al., 2002. The Methanobacterium thermoautotrophicum MCM protein can form heptameric rings. *EMBO reports*, 3(8), pp.792–797.
- Yu, Z., Feng, D. & Liang, C., 2004. Pairwise interactions of the six human MCM protein subunits. *Journal of molecular biology*, 340(5), pp.1197–206.
- Zhang, L. et al., 2013. Structural basis of transfer between lipoproteins by cholesteryl ester transfer protein. *Nature Chemical Biology*, 8(4), pp.342–349.
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9, p.40.
- Zhao, H. et al., 2013. The Cep63 paralogue Deup1 enables massive de novo centriole biogenesis for vertebrate multiciliogenesis. *Nature cell biology*, 15(12), pp.1434–44.
- Zhao, L. et al., 2010. Dimerization of CPAP Orchestrates Centrosome Cohesion Plasticity. *The Journal of Biological Chemistry*, 285(4), pp.2488–2497.
- Zheng, T. et al., 2014. Plasma minichromosome maintenance complex component 6 is a novel biomarker for hepatocellular carcinoma patients. *Hepatology research*, pp.1–10.
- Zheng, X. et al., 2014. Conserved TCP domain of Sas-4/CPAP is essential for pericentriolar material tethering during centrosome biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), pp.E354–63.
- Zhu, J. & Frank, J., 1997. Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *Journal of structural biology*, 118(3), pp.197–219.
- Ziebarth, T.D. et al., 2010. Dynamic effects of cofactors and DNA on the oligomeric state of human mitochondrial DNA helicase. *The Journal of biological chemistry*, 285(19), pp.14639–47.
- Ziebarth, T.D., Farr, C.L. & Kaguni, L.S., 2007. Modular architecture of the hexameric human mitochondrial DNA helicase. *Journal of molecular biology*, 367(5), pp.1382–91.

# Apendix

Publications resulting from projects apart from the ones included in the present thesis:

Sorzano CO, Vargas J, de la Rosa-Trevin JM, Oton J, Alvarez-Cabrera AL, Abrishami V, Sesmero E, Marabini R, Carazo JM. 2015. A statistical approach to the initial volume problem in Single Particle Analysis by Electron Microscopy. *Journal of Structural Biology*, 189 (3), pp 213-219.

Vargas J, Álvarez-Cabrera AL, Marabini R, Carazo JM, Sorzano CO. 2014. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics*, 30(20), pp 2891-2898.

Jonic S, Sorzano CO, Álvarez-Cabrera AL, Carazo JM. StructMap: Multivariate distance analysis of elastically aligned electron microscopy structures for exploring pathways of conformational changes. 2015. *Paper under revision*.