In vitro biochemical reconstitution and electron microscopy of early pre-DNA replication complex intermediates using human proteins

Shenaz Allyjaun

Imperial College London
Faculty of Medicine
Medical Research Council - Laboratory of Medical Sciences

A thesis submitted for the degree of Doctor of Philosophy (PhD)

December 2023

Declaration

The following work and research findings enclosed within this thesis are my own, unless explicitly stated in the text. Where assistance has been received or the work of others reproduced, it is clearly noted in the text. Information and research carried out by others that is utilised within this thesis is appropriately cited and a full bibliography is included as part of this document.

Copyright statement

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Cell division is essential for cellular life, from single-celled organisms to higher-order metazoans. To support cellular and organism function – including proliferation, genomic information must be duplicated precisely once under highly regulated conditions to ensure genetic integrity. Dysregulation in this process can have detrimental consequences, such as carcinogenesis and developmental disorders.

Across the tree of life, DNA replication initiates from discrete sites (known as origins). These are recognised by specific protein initiators that collaborate with co-factor proteins to load the machinery used for DNA replication. The core of this machinery is the replicative helicase, essential for unwinding duplex DNA to allow for semiconservative bidirectional replication.

This research set out to examine the mechanistic aspects behind origin recognition and helicase loading using human proteins. Full-length *H. sapiens* proteins were purified and tested using a newly established *in vitro* method to assess helicase loading. Using a combination of biochemical and structural approaches, intermediates during the helicase loading process were studied in greater detail. Regulation of DNA replication is highly dependent on intrinsic ATP hydrolysis of these proteins, although the precise mechanisms behind this are not well-defined. Using hydrolysis mutants, we were able to study how coordinated ATP hydrolysis contributes to helicase loading, as well as identifying and trapping short-lived intermediates during this process.

Taken together, the results and methodologies in this thesis provide fundamental insight into how human replication machinery is assembled on a molecular level and how individual characteristics of these proteins contribute to faithful and efficient replication of the genome.

Acknowledgements

Before embarking on this PhD journey, I was convinced that long, heartfelt acknowledgement sections were cheesy – but after these four years, I can truly see that the support from so many people has been paramount in keeping me going and they all deserve an honourable mention.

Firstly, I must acknowledge Prof. Christian Speck for the space and time in your lab to carry out this research project. I appreciate the advice you've given me over the years, especially on how to keep my carnivorous plants alive.

My development as a scientist over the past few years has been supported by two fantastic postdocs in the lab. To Vera Leber, thank you for teaching me everything you could pass on about biochemistry and showing me that I don't need to be afraid of the AKTA – the AKTA needs to be afraid of me :). To Almutasem Saleh, thank you for teaching me everything you possibly could about electron microscopy, despite my uncooperative samples. Thank you for teaching me the value of truly listening to what my results are telling me – and making sure I had a fun experience along the way. I will fondly remember all the lab and office pranks and I will never look at green lab tape the same way again.

The entire DNA Replication Group cohort has been fundamental to my time here - there are so many amazing, talented, and kind scientists I have had the honour of working with. To Jess and Lucy, thank you for being the best PhD sisters, gossip buddies and partners in crime a girl could ever ask for. To Sarah and Lepakshi – the ultimate lab mums - thank you for helping to keep the lab environment so warm and for always being there whenever I needed a chat (or a distraction). To members of the lab both past and present – Chris, Indiana, Audrey, Sarah S, Yasunori, Josh, Matt, Tonia, Jen, Anna, Max, Marina – thank you for all your advice and companionship over the past four years.

There have been many people at Imperial and the LMS that have helped me during my time here. Thank you to Ricardo Aramayo for his mountains of help and patience with the electron

microscopy, and for eventually trusting me to be alone with the Talos. Thank you to everyone in the LMS second and CRB third floors for always being a friendly presence every day. To the rest of my PhD cohort – it has been a pleasure to grow and develop as a scientist around you and to watch the rest of you do the same.

There have also been so many people outside of the lab that have supported me in so many ways. A special thank you to Eathaar, for always being there to listen to me vent and giving me space to feel all my rage (and for making me salads with entire balls of mozzarella and spicy noodles). I don't know what I would have done without your friendship.

To James, who I could write an entire thesis longer than this one detailing all the patience, kindness, and support you have shown me over the years, there aren't quite enough words for me to show my gratitude. Thank you for the endless cups of tea and biscuits. Thank you for taking the time and effort to learn about this entirely new scientific concept with me, for being a constant unshakable support and shoulder to cry (and celebrate) on. It means so very much to me.

Finally, but by all means not least, I would be nowhere without the support and love from my incredible family. To my parents, there will never be a way sufficient enough to thank you for all the hard work you put in for me to have the best possible educational opportunities. To my mum, thank you for your calming, solid presence throughout everything. To my dad, thank you for your healing concoctions of ginger-honey-lemon-clove-turmeric. To Yasin and Farhana, thank you for being the parental older siblings and always stressing me out by asking what's next. To Shahid, thank you for being the chill younger brother and never stressing me out at all. To my little niblings, thank you for making life more joyful. And finally, thank you to my Nani for her gentle words of encouragement from afar and my Nana, who's efforts in his life to make sure all the women in his family had the best possible opportunity for an education has gone on to produce generations of inspiring people.

Abbreviations

A	Adenine
AAA+	ATPases Associated with diverse cellular Activities
ACS	ARS consensus sequence
APC/C	Anaphase promoting complex/cyclosome
ARS	Auonomous replicating sequence
Amp	Ampicillin
AT	Adenine-Thymine
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BAH	Bromo-adjacent homology
BeF	Beryllium fluoride
BSA	bovine serum albumin
bp	base pair
C	cytosine
CDC	Cell division cycle
Cdt1	Chromatin licensing and DNA replication factor 1
CDK	cyclin dependent kinase
CE	catalytic glutamate
ChIP-MS	Chromatin Immunoprecipitation-Mass Spectrometry
CTD	C-terminal domain
CTF	contrast transfer function
CMG	Cdc45-MCM2-7-GINS
CpG	Region of repeat CG-nucleotides
cryo-EM	cryo electron microscopy
ddH2O	double distilled H2O
DH	double hexamer
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
DTT	dithiotreitol
EB	elution buffer
EDTA	ethylenediaminetetraacetic acid
FRET	Forster resonance energy transfer
G	guanine
g	x gravity
G1	Gap phase 1
G2	Gap phase 2
G4	guanine-quadruplex
GC	guanine-cytosine
GINS	Go-Ichi-Ni-San
H4K20me2	Histone 4 lysine 20 dimethylation
H3K9me3	histone 3 lysine 9 trimethylation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HP1	heterochromatin-binding protein 1

His	histidine
IDR	intrinsically disordered region
IP	immunoprecipitation
kDa	kilodalton
Leu	leucine
M	Mitosis phase
MCM	minichromosome maintenance
MGS	Meier-Gorlin Syndrome
MP	mass photometry
MWCO	Molecular weight cut-off
ng	nanogram
nm	nanometer
nM	nanomolar
NS	negative stain
NS-EM	negative stain EM
NTD	N-terminal domain
O	ORC
OC	ORC-Cdc6
OCD	Orc-Cdc6-DNA
OD	ORC-DNA
OCCM	ORC-Cdc6-Cdt1-MCM complex
OM	ORC-MCM complex
OD600	Optical density at 600 nm
OGRE	Origin G-rich repeated elements
ORC	Origin Recognition Complex
PCR	polymerase chain reaction
pre-RC	pre-replication complex
RELION	Regularised Likelihood Optimisation
RF	arginine finger
rpm	revolutions per minute
S	Synthesis phase
S-CDK	S-phase cyclin dependent kinase
SCF	Skp1-Cullin1-F-box
SDS-PAGE	sodium-dodecyl-sulphate polyacrylamide gel electrophoresis
Strep	Streptavidin
ssDNA	single-stranded DNA
SUMO	small ubiquitin-like modifier
T	thymine
TAE	tris-acetate EDTA
TBE	tris-borate EDTA
TFIIB	Transcription Factor II B
Trp	tryptophan
Ura	uracil
v/v	volume per volume
w/v	weight per volume
WA	Walker A
WB	Walker B
WH	winged helix

WHD	winged helix domain
WT	wild-type
ZnF	zinc-finger

Table of Contents

Declaration 1
Copyright statement 2
Abstract 3
Acknowledgements 4
Abbreviations 6
Table of Figures 14
Chapter 1: Introduction 17
1.1 Replication and the Eukaryotic Cell Cycle 20
1.2 An Overview of Eukaryotic DNA Replication Licensing 22
1.3 Origin Specification in Eukaryotes 23
DNA sequence guided specification. 24
Context-dependent origin specification 25
1.4 Proteins Involved in Pre-RC formation. 28
1.4.1 The Origin Recognition Complex (ORC) 28
1.4.1.1. ORC across evolution 28
1.4.1.2. Structure of ORC 29
1.4.1.3 ATP binding and Hydrolysis of ORC 36
1.4.1.4 ORC and Chromatin Replication 39
1.4.1.4.1 ORC and euchromatin 39
1.4.1.4.2 ORC and heterochromatin 40
1.4.2 Cdc6 (Cell-division-cycle 6) 41
1.4.2.1 Structure of Cdc6 42
1.4.2 .2 ATP binding and hydrolysis of Cdc6 45
1.4.3 Cdt1 46
1.4.3.1 Structure of Cdt1 47
1.4.4. MCM2-7 48
1.4.4.1 Structure of MCM2-7 48
1.4.4.2 ATP binding and Hydrolysis by MCM 50
1.5 Intermediates involved in Loading of the Replicative Helicase 51
1.6 Activation and Downstream Activity of the Double hexamer: Replication Firing 53
1.7 Coupling DNA Replication Initiation to the Cell Cycle 54
1.7.1 ORC 55
1.7.2 Cdc6 56
1.7.3 Cdt1 and MCM 57
1.8 Pre-RC components and their relevance for human health 58
1.8.1 DNA Replication factors and Cancer 58
1.8.2 Pre-RC components and human development – Meier-Gorlin Syndrome 60

1.9 A note on *in vitro* methods to study helicase loading. 61
1.10 Transmission Electron Microscopy: The resolution revolution and its implications in structural biology 62
1.10.1 Techniques in Electron Microscopy 62
1.10.1.1 Negative Stain EM 63
1.10.1.2 Cryogenic EM 63
1.10.1.3 Single Particle Analysis (SPA) 63
1.10.2 Challenges in Structural Determination by EM 66
1.10.3 Electron Microscopy and helicase loading 66
1.11 Project aims. 67
Purification of Full-length human pre-RC proteins 67
Electron microscopy of intermediates 67
ORC and chromatin 68
Chapter 2: Materials & Methods 69
2.1 Materials 69
2.1.1 Plasmids 69
2.1.2 PCR primers/ Oligonucleotides 69
2.1.3 Bacterial Strains 69
2.2.4 Yeast Strains 69
2.2.5 Media 69
2.2 Methods 70
2.2.1 Protein Expressions and Purifications 70
2.2.1.1 Orc1-5 70
2.2.1.1.1 Transformation & Expression 70
2.2.1.1.2 Purification 72
2.2.1.2 Cdc6 75
2.2.1.2 .1 Transformation & Expression 76
2.2.1.2.2 Purification 77
2.2.1.3 Orc6, Cdt1 & MCM 79
Orc6 79
Cdt1 80
MCM2-7 80
2.2.1.4 HP1 80
2.2.1.4.1 Transformation & Expression 80
2.2.1.4.2 Purification 81
2.2.2 Selection of DNA, amplification and purification. 83
2.2.2.1 Selection of DNA sequences 83
2.2.2.2 DNA amplification and purification 85
2.2.2.2.1 Plasmid amplification 85

2.2.2.2.2 DNA oligonucleotide (100-400bp) preparation 86
2.2.2.2.3 DNA conjugation to beads 87
2.2.3 The pre-RC Assembly Assay 88
2.2.4 Pull-down Assays 89
2.2.5 Gradient Density Centrifugation to assess protein binding. 89
2.2.5.1 Gradient Preparation 90
2.2.5.2 Gradient Running and Analysis 90
2.2.6 Nucleosome Assembly 90
2.2.7 ChIP-MS Design 92
2.2.7.1 Generation of Tagged Orc2 92
2.2.7.2 Western Blotting 92
2.2.8 DNA Quality Control and Analysis 93
2.2.8.1 Agarose Gel electrophoresis 93
2.2.8.2 EMSAs – Electrophoretic Mobility Shift Assays 93
2.2.9 Protein Quality control and Analysis 93
2.2.9.1 SDS-PAGE (Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis) 93
2.2.9.2 Coomassie Staining 94
2.2.9.3 Silver Staining 94
2.2.9.4 Mass Photometry 95
2.2.10 Electron Microscopy Sample Preparation 96
2.2.10.1 Negative Stain 96
2.2.10.2 Cryogenic 96
2.2.11 Electron Microscopy Data Acquisition 97
2.2.12 Electron Microscopy Processing 97
Chapter 3: Results 99
3.1 Expression and Purification of full-length Human Pre-Replication complex proteins 99
3.1.1 ORC - Origin Recognition Complex 100
3.1.1.1 Expression - Integrated vs Non-integrated plasmids 100
3.1.1.2 Streptavidin-dependent affinity purification 103
3.1.1.3 Ion Exchange – HiTrap Heparin HP and Poros HQ 105
3.1.1.4 Size-separation methods for protein homogeneity 107
3.1.1.5 Overall 111
3.1.2 Cdc6 – Cell-dependent cycle 6 113
3.1.2.1 Removal of contaminants from Cdc6 preparations. 114
3.1.2.1.1.Size-based separation methods 114
3.1.2.1.2 Ion Exchange Chromatography 116
3.1.3 Orc6, Cdt1 and MCM 118
DNA Selection and Purification 118

3.2 Amplification and Purification of DNA Sequences 119
3.3 Reconstitution of the human pre-RC in a bulk biochemical assay 121
3.3.1 Introduction 121
3.3.2 Assay 121
3.3.3 Discussion 125
3.4. Identifying Intermediates in the pre-RC pathway 126
3.4.1 Single-particle analysis Electron Microscopy of human ORC alone 126
Negative Stain EM 127
Cryo-Electron Microscopy 129
Optimisation of cryo-grid preparation 129
Cryo-Electron Microscopy analysis of full-length ORC 133
Discussion 143
3.4.2 Binding of ORC to DNA 143
ORC is nonspecific in binding to DNA. 144
Negative Stain Microscopy of ORC binding to DNA 146
Cryo-Electron Microscopy of ORC binding to DNA 149
Reference-based template picking combined with Topaz training. 150
Ab-initio 3D modelling of Orc1-5-DNA 156
Discussion 162
3.4.3 Introduction of Cdc6 to form an ORC-Cdc6-DNA complex 166
Cdc6 enhances ORC binding to DNA 166
Cryo-Electron Microscopy of in-solution assembly of OCD 168
Preparing a homogenous prep of ORC-Cdc6-DNA 174
Cdc6 does not stably associate with ORC over time 174
Cdc6 and its ATPase activity 176
3.4.4 The Orc2-5-DNA complex 179
Over time, ORC1 can dissociate. 179
Negative Stain Microscopy of Orc2-5 on DNA 182
3.4.5 Introduction of ORC1-ATPase mutations 187
3.4.6 ATPase Mutations in the ORC-Cdc6 complex affects double hexamer formation 192
3.4.7 N-terminal Deletion of Orc1 194
Orc1-5(Orc1ΔN) on a gradient 195
..... 197
Orc1-5(Orc1ΔN) can still recruit the double hexamer. 197
Orc1-5(Orc1ΔN) in association with a Cdc6 ATPase mutant 199
Negative Stain Electron Microscopy of Orc1ΔN-Cdc6-DNA 199
Discussion 199
3.4.8 Beyond the ORC-Cdc6-DNA complex: An attempt at OCCM 201

Glycerol Gradients of Orc-Cdc6-Cdt1-MCM bound to DNA 202
Negative stain Electron Microscopy of potential OCCM 205
Discussion – The recruitment and regulation of pre-RC components to DNA to facilitate DNA replication. 208
3.5 – ORC and Chromatin 210
Introduction 210
Reconstitution of nucleosomes by salt-gradient dialysis 211
ORC does not bind to nucleosomes unless extranucleosomal DNA is present 212
Negative Stain Microscopy of ORC-nucleosomes 215
ORC-HP1 219
HP1 purification 220
Biochemical analysis of the HP1-ORC interaction. 221
..... 224
Negative Stain Microscopy of ORC-HP1 224
The ORC interactome 225
Producing a tagged ORC cell line for ChIP-MS 226
Discussion 228
4: General Discussion, Study Limitations and Future Perspectives 230
Purification of human pre-RC proteins 230
The dynamic nature of the Origin Recognition Complex 231
Limitations of using cryo-EM to study ORC and Cdc6 232
The establishment of an assay to test human pre-replication complex assembly. 234
Limitations of the pre-RC assay 235
5. Concluding Remarks 236
6: Supplementary Information 238
Supplementary Table 1: Plasmids 238
Supplementary Table 2: DNA Sequences 239
Supplementary Table 3: Primers 240
Supplementary Table 4: Yeast Strains 240
Supplementary Table 5: Bacteria Strains 240
Supplementary Table 6: Mammalian Cells 241
Supplementary Tables 7: Electron Microscopy Parameters 241
7.1 ORC1-5 NS 241
7.2 ORC1-5 Cryo 241
7.3 ORC-DNA NS 242
7.4 ORC-DNA Cryo 242
7.5 OCD Cryo 243
7.6 Orc2-5 NS 243
7.7 OCCM – F8 243

7.8 OCCM – F9 244
7.9 ORC-Nucleosome 244
7.10 ORC-HP1 245
6. References 246

Table of Figures

Figure	Title	Page
Figure 1	*An overview of the DNA replication process*	*19*
Figure 2	*Cell cycle with phases and checkpoints*	*21*
Figure 3	*An overview of helicase loading at origin sites– 'DNA licensing'*	*23*
Figure 4	Origin Specification in Eukaryotes	24
Figure 5	*Organisation of the AAA+ Origin Recognition Complex (ORC) with and without cofactor Cdc6*	*30*
Figure 6	*The inherent flexibility of human ORC.*	*32*
Figure 7	*A conserved helical Orc4 segment confers DNA specificity in some single-celled eukaryotes.*	*34*
Figure 8	*The ORC ATPase sites*	*36*
Figure 9	*Cdc6: Structure and domain organisation*	*42*
Figure 10	*Cdt1 structure and domains*	*45*
Figure 11	*MCM Structure and organisation*	*48*
Figure 12	*Pre-RC assembly with associated intermediate steps identified in S. cerevisiae and metazoans.*	*50*
Figure 13	*Cell cycle CDK activity regulates origin licensing and activation*	*53*
Figure 14	*Single-Particle Electron Microscopy: Sample preparation, Data Collection and Analysis*	*63*
Figure 15	*Selection of DNA sequences to study ORC binding and pre-RC assembly*	*82*
Figure 16	*Comparison of yields of human Orc1-5 complex expression between integrated and non-integrated plasmids*	*100*
Figure 17	Affinity purification optimisation of human Orc1-5 via the 2xStrepSumo tag on Orc1	102
Figure 18	*Ion Exchange purification of Orc1-5*	*104*
Figure 19	*Size Exclusion Chromatography for stochiometric and homogenous Orc1-5*	*106*
Figure 20	*Gradient Density Centrifugation as a matrix-free method for homogenous and stochiometric preparations.*	*108*
Figure 21	*Size-separation purification methods results in homogenous Orc1-5*	*109*
Figure 22	*An Optimised Method for human Orc1-5 purification from S. cerevisiae cells*	*110*
Figure 23	*Expression and Affinity purification of GST-Cdc6*	*111*
Figure 24	*Size-separation methodologies for a homogenous Cdc6 preparation*	*113*
Figure 25	HiTrap HP SP ion exchange chromatography produces homogenous Cdc6	115
Figure 26	*An optimised method for full-length human Cdc6 purification from E. coli*	*116*
Figure 27	*Large Scale PCR and purification methods to isolate DNA fragments*	*118*
Figure 28	*The pre-RC Assay – A method of testing double hexamer assembly*	*121*
Figure 29	*Cross-linking is crucial for ORC1-5 integrity in EM.*	*125*
Figure 30	*Purified full-length Orc1-5 shows the same architecture as other ORC structures*	126
Figure 31	*Optimisation of cryogrids – Orc1-5 alone*	*129*
Figure 32	*Relion-Integrated Picking parameters for the Orc1-5 dataset*	*131*

Figure 33	*2D classification and filtering of Reference-based picked particles*	*133*
Figure 34	*Topaz-trained autopicking*	*135*
Figure 35	*Topaz autopicking on a large scale*	*137*
Figure 36	*ab-initio electron density maps of Orc1-5*	*139*
Figure 37	*EMSAs show that purified Orc1-5 can bind and remain on DNA*	*142*
Figure 38	Negative stain microscopy of ORC bound to DNA.	145
Figure 39	*Reference-based autopicking optimisation for Orc1-5 on DNA*	*147*
Figure 40	*Topaz autopicking and extraction results in particles that can be averaged to show secondary structure elements of Orc1-5-DNA*	*149*
Figure 41	*Topaz autopicking on 11,000 micrographs and corresponding 2D classification*	*151*
Figure 42	*Ab-initio modelling and 3D electron density of Orc1-5-DNA*	*153*
Figure 43	*Filtering of 2D classes to ensure particle integrity.*	*154*
Figure 44	*Electron Density Maps of Orc1-5-DNA*	*155*
Figure 45	*Additional Filtering needed for high resolution structural information*	*157*
Figure 46	*additional processing of the Orc1-5 dataset*	
Figure 47	*Cdc6 enhances binding of ORC to DNA*	*160*
Figure 48	*Cryo-electron microscopy picking on ORC-Cdc6-DNA*	*162*
Figure 49	*ORC-Cdc6-DNA in solution results in a heterogenous population of particles*	*164*
Figure 50	*ORC-Cdc6-DNA makes up a small population of particles in solution*	*165*
Figure 51	*Cdc6 does not remain bound to ORC-DNA in the long term.*	*167*
Figure 52	*An ATPase mutant of Cdc6*	*169*
Figure 53	*Using an ATPase-dead mutant of Cdc6 traps OCD on a gradient in the long term.*	*170*
Figure 54	*Binding of Orc1-5 to DNA releases Orc1.*	*172*
Figure 55	*Orc1 is released over time*	*173*
Figure 56	*SDS-PAGE and corresponding agarose gel fraction of Orc2-5 bound to DNA*	*174*
Figure 57	*The Orc2-5-DMA complex NS processing*	*176*
Figure 58	*The Orc2-5-DNA complex*	*178*
Figure 59	*Cryogrid preparation of Orc2-5-DNA*	*179*
Figure 60	*Design and purification of an ORC1-5 ATPase mutant*	*180*
Figure 61	*Blocking ORC ATPase activity does not stop Orc1 release*	*181*
Figure 62	*ATPase dead ORC mutants can stabilise Cdc6 binding*	*182*
Figure 63	*pre-RC loading reaction with ATPase mutants stabilise intermediate complexes*	*184*
Figure 64	*Deletion of the N-terminus of Orc1*	*185*
Figure 65	*Orc1ΔN-5 can still bind to DNA and appears to stabilise OD and OCD interactions*	*187*
Figure 66	*Orc1ΔN-5 does not hinder double hexamer formation.*	*189*
Figure 67	*The intrinsically disordered regions of ORC and Cdc6 are heavily modified by phosphorylation*	*191*
Figure 68	*Pre-RC intermediates in solution are variable*	*193*
Figure 69	*ORC-Cdc6-Cdt1 and MCM on a 15-45% glycerol density gradient*	*195*
Figure 70	*ORC-Cdc6-Cdt1 and MCM on a 15-45% glycerol density gradient – NS EM analysis*	*196*
Figure 71	*MCM2-7 can self-associate into MCM2-7 asymmetric double hexamers.*	*198*
Figure 72	*Reconstitution of nucleosomes by salt gradient dialysis*	*203*
Figure 73	*ORC can only bind to mononucleosomes when flanking DNA is present*	*205*
Figure 74	*H4K20me2 mononucleosomes with 40bp flanking DNA can bind ORC.*	*206*
Figure 75	*NS-EM of ORC-H4K20me2-40bp mononucleosomes*	*208*
Figure 76	*Purification of HP1*	*211*

Figure 77	*The ORC-HP1 interaction is relatively weak.*	*213*
Figure 78	*ORC-HP1β NS EM analysis*	*215*
Figure 79	*Preparation of a Tagged Orc2 cell line for ChIP-MS studies*	*218*

Chapter 1: Introduction

Deoxyribonucleic acid (DNA) is the fundamental vehicle for information-carrying in all living organisms. A cell's DNA encodes the nucleic acid sequences that can be transcribed and translated into protein molecules, or non-protein cellular effectors including ribonucleic acid (RNA) variants, including but not limited to ribosomal RNA (rRNA), transfer RNA (tRNA) and non-coding regulatory RNA (ncRNA). This storage and processing of data is fundamental to all cell functionality and thus must be properly maintained and protected.

Across the tree of life - from single-celled bacteria to metazoans made up of trillions of cells, all organisms must undergo cellular division to grow and survive. A fundamental necessity to this is the faithful and accurate duplication of the genetic material to deliver into progenitor cells, supporting cell proliferation and maintaining normal cellular function. While precise mechanisms of these processes differ across kingdoms (Bleichert, Botchan, and Berger 2017), replication of the DNA is always tightly controlled and integrated with various cellular processes to ensure the genome is replicated once and only once per cell cycle, maintaining precise and timely replication and correct ploidy levels. Dysregulation of DNA replication is often detrimental to an organism and is associated with genomic instability leading to developmental abnormalities, cancer and cell death (Champeris-Tsaniras et al, 2014).

Eukaryotic DNA is replicated by highly functional replisomes, which are assembled onto chromatin in several discrete steps (Lewis and Costa 2020; Parker, Botchan, and Berger 2017) (Figure 1). Dedicated initiator proteins recognise defined genomic loci (deemed 'origins'), to which they recruit and deposit replicative helicases in a process termed 'origin licensing.' Post this, the loaded helicases are activated, leading to DNA unwinding and assembly of the replisome onto the active helicase core, allowing synthesis of DNA in a semi-conservative manner.

Mechanisms governing the regulation of DNA replication are conserved across eukaryotic cells. Replication proteins and mechanisms at all stages, from origin licensing to activation

have orthologs across all kingdoms of life. Over the past few decades, much progress has been made in understanding the process of genome replication using a range of model organisms, primarily S. *cerevisiae*, X. *laevis*, *D. melanogaster*, but the study of the equivalent human proteins is of increasing significance in recent years. This thesis focuses on studying the assembly of early human replication licensing proteins on DNA from an *in vitro* biochemical and structural perspective. The findings presented in this thesis set out to establish the molecular mechanisms of recognition of origin DNA and downstream recruitment of the origin licensing complex.

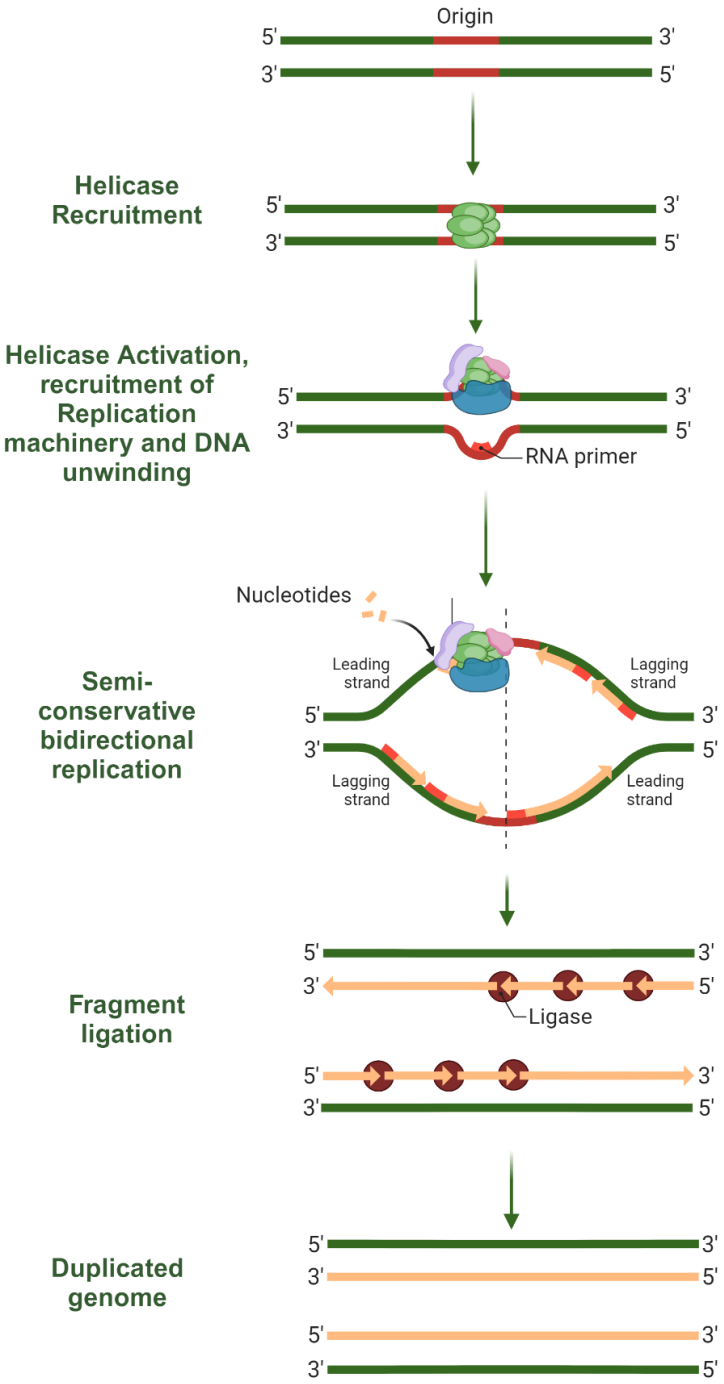


Figure 1: An overview of the DNA replication process

DNA helicases proteins are recruited to origin sites across the DNA, where they remain dormant until activation in S phase. Upon activation, the rest of the replication machinery is recruited to the DNA, facilitating unwinding of duplex DNA and initiation of semi-conservative replication. Replication proceeds in a bidirectional manner, with lagging stand ligation resulting in a fully duplicated complement of the genome. For simplicity, only one helicase is shown but this process occurs on both strands of the DNA. Figure made in BioRender.

1.1 Replication and the Eukaryotic Cell Cycle

A cell's activity is split into distinct temporal phases, each of which is heavily controlled to ensure genomic stability and cell fitness before progression into the next stage. The eukaryotic cell cycle itself is composed of four primary stages, with an ordered sequence of events that prepare the cell for division - Two 'Gap' phases (G1 and G2), the Synthesis phase (S) and the Mitotic phase (M). (Figure 2)

In G1, cells typically undergo significant growth and synthesise the proteins and organelles required for replication of the DNA. In addition to this, the G1 phase is critical to track extracellular conditions and signals from the neighbouring environment to ensure conditions are favourable for progression into the next phase, the S phase. For transition from G1 to S phase, the cell must be in a state where external signals are favourable for replication to proceed, and that all machinery required for the replication of the DNA has been synthesised and positioned in appropriate origin locations.

The S phase of the cell cycle is where the replication of the chromosomal DNA is carried out, resulting in a newly synthesised duplicate of the entire genome. During this time, the replication licensing machinery positioned across the DNA in G1 is activated, through kinase activity and the binding of a large number of helicase activation factors. Ultimately, this results in unwinding of the DNA double helix and semiconservative bidirectional replication.

Upon completion of DNA replication, the cell enters the second gap phase, G2, where material necessary of mitosis are synthesised and any DNA damage responses are activated to correct any instabilities carried over from errors in the S phase (Kousholt, Menzel, and Sørensen 2012). Progressing into M phase, - which itself is subdivided into prophase, metaphase, anaphase, telophase and cytokinesis – the cell's duplicated DNA is condensed. Duplicated chromosomes are physically segregated before the splitting of the parental cell into two daughter cells, each containing a single copy of each chromosome.

Eukaryotic cells employ a network of feedback loops to ensure progression into the next phase can only occur upon complete and accurate activity of the previous one. As a mechanism of quality control, checkpoints are present throughout the cell cycle, with the primary checkpoints being between each cell cycle phase - M/G1, G1/S and G2/M (Figure 2).

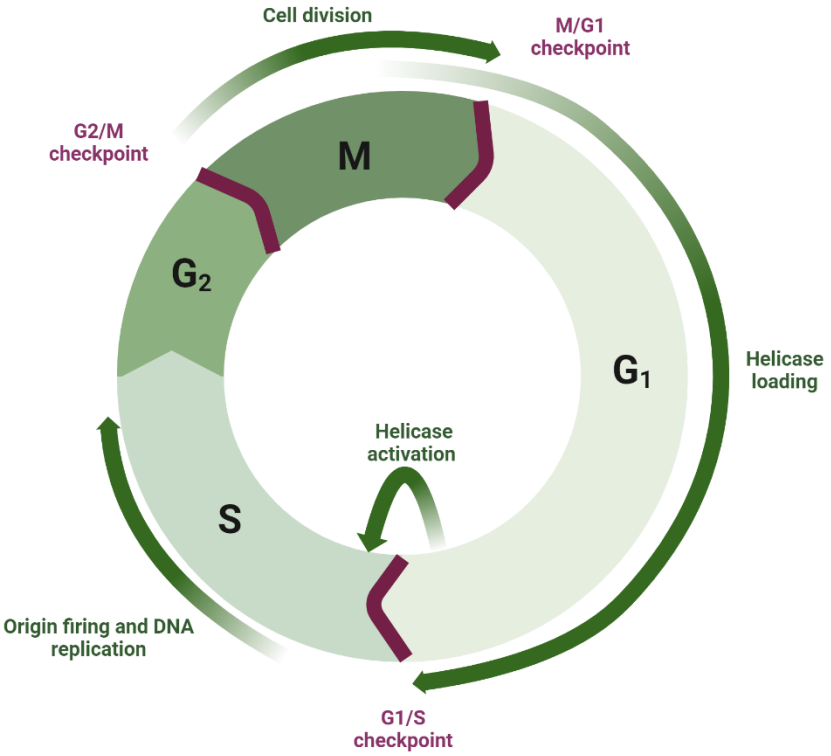


Figure 2: Cell cycle with phases and checkpoints

An overview of the cell cycle and the individual phases in relation to helicase loading and firing. In G1, the inactive replicative helicases are loaded onto origin DNA where they remain 'licensed' until activation on the onset of the S phase. Upon S-phase checkpoint mediated activation, these licensed helicases can fire and begin replication of the DNA. At the G2/M phase transition, quality control mechanisms are employed to ensure complete and accurate genome replication before the cell proceeds into the Mitotic phase where the cell divides into two daughter cells. Figure made with BioRender.

1.2 An Overview of Eukaryotic DNA Replication Licensing

Before replication can proceed, loading of the helicase occurs in a multi-step process in G1. Initial models suggest that this process begins upon an initiator protein binding to predetermined 'origin' segments of the DNA. This initiator protein would then recruit further factors that would unwind dsDNA and assemble the replisome (Jacob, Brenner, and Cuzin 1963). Further investigation identified this 'initiator' protein, dubbed the Origin Recognition Complex (ORC) as a multiprotein complex that associates with origin locations on the DNA (Bell and Stillman 1992). Further investigation showed that ORC can then act to recruit a cofactor, Cell-division-cycle 6 (Cdc6), which in turn can subsequently facilitate binding of Cdt1 (Cdc10-dependent transcription protein 1) and the heterohexameric minichromosome maintenance protein complex (MCM2-7), which serves as the replicative helicase (Labib and Diffley 2001; Shechter, Ying, and Gautier 2004). Replication licensing is incomplete until a repeated loading of the MCM2-7 helicase occurs on dsDNA and ORC, Cdc6 and Cdt1 are released, resulting in a stable head-to-head double hexamer constituting of two MCM2-7 molecules (Evrin et al. 2009; Remus, Beuron, Tolun, Griffith, Morris, and Diffley 2009; Fernández-Cid et al. 2013). Together, these proteins form the pre-replication complex (pre-RC) (Figure 3), that can remain at 'licensed' replication origins until transition into the S phase of the cell cycle, where they can undergo phosphorylation-dependent activation to catalyse the unwinding of DNA (Remus, Beuron, Tolun, Griffith, Morris, and Difflex 2009; Remus and Diffley 2009; Siddiqui, On, and Diffley 2013) and facilitate bidirectional replication.

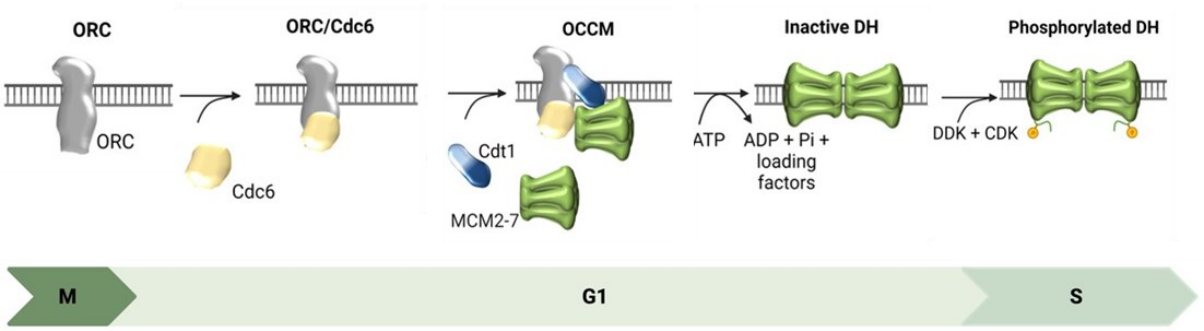


Figure 3: An overview of helicase loading at origin sites– 'DNA licensing'

Initially, ORC binds to DNA, followed by cofactor Cdc6 binding. The loaded ORC-Cdc6 complex is competent to recruit a singular complex of MCM-Cdt1 that binds to DNA via the MCM central channel. Following ATP hydrolysis, Cdc6 and Cdt1 are released and a second round of ORC-Cdc6, then Cdt1-MCM can bind, repeating the ATP hydrolysis process, resulting in a stable MCM double hexamer (DH) bound at origin sites across the DNA. Figure adapted form review in Riera et al, 2017)

Helicase loading, activation and replication fork elongation have been studied extensively over the years, with *S. cerevisiae* primarily serving as the model organism. *In vitro* reconstitutions of pre-RC formation using purified proteins have enabled biochemical, structural and single-molecule studies to identify the intermediates and dynamics involved in replication initiation. Recent studies have extended knowledge into higher-order metazoans, including *D. melanogaster* and *H. sapiens*. Characteristics of the components governing replication will be discussed in further detail below.

1.3 Origin Specification in Eukaryotes

One of the methods utilised by cells for controlling replication is the specification and distribution of sites where replication can begin. These origin sites are not random and must be tightly coordinated with transcription to avoid replication-transcription conflicts, which can result in incomplete genome duplication and genomic instability (Chen et al. 2019). Prokaryotic and archaeal genomes are primarily replicated from a single locus, with replication continuing bidirectionally around a circular genome until termination occurs – a feature suited to its comparatively smaller genome size (Jacob, Brenner, and Cuzin 1963; Barry Elizabeth and Bell Stephen 2006; Gao and Zhang 2007; Gowrishankar 2015). Eukaryotes typically have

larger genomes in comparison, often encoded on multiple chromosomes. The total genome is too large to be replicated by a single pair of forks at the speed required for complete synthesis within a single S phase cycle (Yurov and Liapunova 1977). Therefore, eukaryotes rely on significantly more origin sites, ranging in number from hundreds to thousands depending on the size and complexity of the organism (Mechali 2010). These origin sites range in their modes of specificity, efficiency and timing of activation (Mechali 2010; Mechali et al. 2013; Cvetic and Walter 2005), although can be broadly divided into two classes – one that is guided by a DNA consensus sequence and a second that is context-dependent.

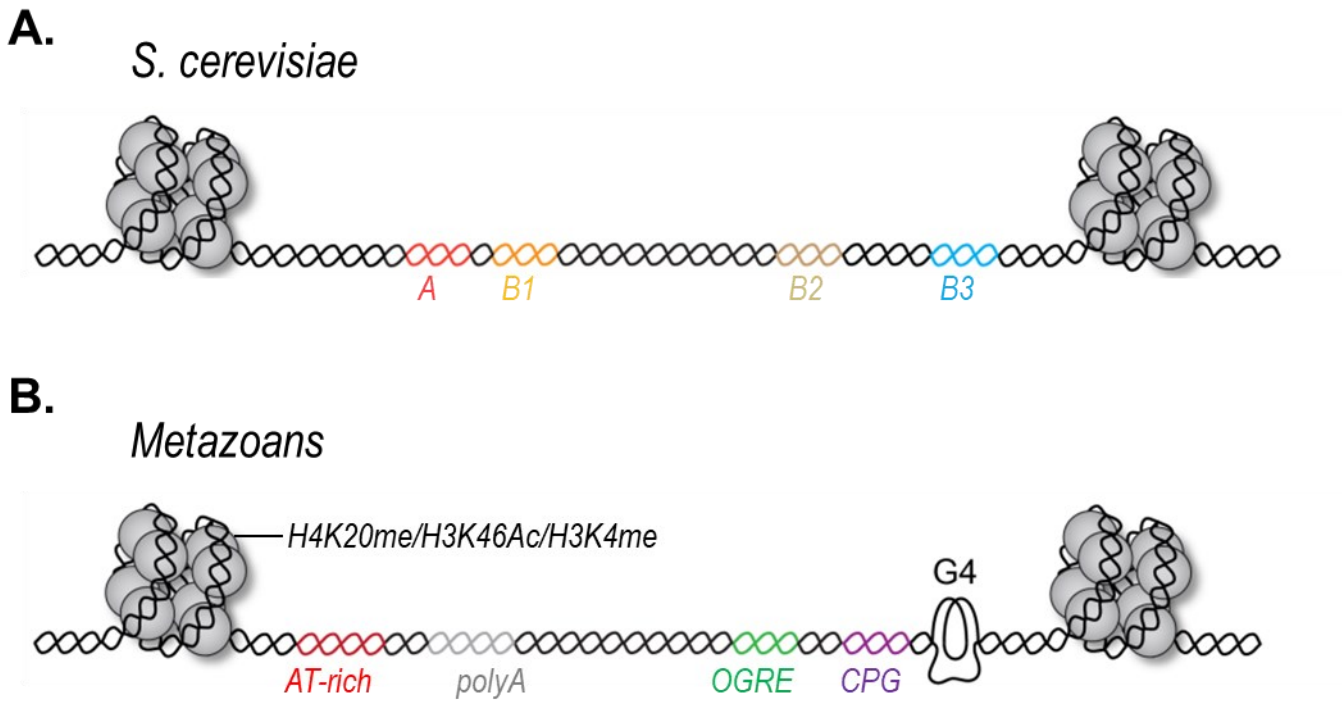


Figure 4: Origin Specification in Eukaryotes

Illustration of key characteristics of origins in S. cerevisiae and metazoans. **A:** In S. cerevisiae, origins contain an A and B1 element, the primary binding site of ORC and a downstream B2 and B3 elements, serving as secondary weaker binding sites**. B:** Metazoan origins are defined by a range of genomic features, including but not limited to: skewed AT/GC distribution, CpG islands, polyA/G stretches, OGREs, G-quadruplexes and specific histone modifications. Figure made in BioRender.

DNA sequence guided specification.

Replication origins in budding yeast *S. cerevisiae* and closely related fungi were the first eukaryotic origins to be identified, discovered by their ability to confer autonomous replication

upon fragmentation of the genome and subsequent insertion into plasmids (Stinchcomb, Struhl, and Davis 1979). These regions were termed autonomous replication sequences (ARSs) with further genomic sequence analysis uncovering sequence-specific elements dubbed autonomous consensus sequences (ACSs) (Theis and Newlon 1997; Cannon, Jenkinson, and Shepherd 1990). These sites are composed of three functional regions, the A, B1and B2 elements, while a few origins have an additional–B3 element (Marahrens and Stillman 1992) (Figure 4.A).

The budding yeast ORC recognises and binds at an 11-17bp AT-rich A element (Theis and Newlon 1997; Dhar, Sehgal, and Kaul 2012). While the presence of this sequence in the genome is abundant, only ~400 sequences are used as origin sites under normal conditions, due to the B elements, situated close to the A elements, acting in cis (Marahrens and Stillman 1992; Bell and Labib 2016). The B1, B2 and B3 elements are located downstream of the ACS. Together, the A and B1 elements represent binding sites for ORC in an ATP-dependent manner (Rao and Stillman 1995; Rowley et al. 1995). The B2 element is proposed to serve as a low-affinity ORC binding site, due to it containing an inverted ACS (Miller et al. 2019), although has also been suggested to be the binding site for the MCM2-7 helicase (Zou and Stillman 2000; Wilmes and Bell 2002). The B3 element is the least conserved element, but has been shown to recruit the transcription factor Abf1 (involved in multiple roles ranging from transcriptional activation of ribosomal proteins genes to controlling mitotic cell cycle progression) – although neither is essential for origin activity (Ganapathi et al. 2011).

While consensus sequences in budding yeast have been valuable to study the assembly of the replicative helicase onto origin sites, these sequence-guided specifications are rare in eukaryotes. This suggests an increased plasticity for origin recognition in most eukaryotic organisms that does not rely on sequence specificity.

Context-dependent origin specification

In most other eukaryotic systems, defined sequence elements are not present at replication initiation sites (ISs). Studies of larger genomic sizes in metazoans has identified between 30,000 – 50,000 distinct origins, although these sites do not follow a rigid sequence specificity (Leonard and Mechali 2013; Mesner et al. 2011). Despite this, it is still ORC binding that enables the recruitment of the rest of the pre-RC components and eventual initiation of replication. While there is no single conserved signature required for ORC binding and pre-RC assembly, distinct features have been observed at origin sites. These are flexible in terms of size, sequence, cell type and chromatin environment (Figure 4.B).

Despite the lack of causal relationship between DNA sequence and metazoan ORC binding sites, several sequence characteristics have been found to correlate with proximal ISs. Mounting evidence suggests that DNA geometry and topology is of importance for regulating origin recognition and helicase loading. Negative supercoiling and helical twisting have shown to promote *D. melanogaster* ORC binding to origins, aiding helicase loading (Remus, Beall, and Botchan 2004; Yella et al. 2020; Comoglio et al. 2015).

In vitro studies have suggested metazoan ORC binding displays a slight preference for AT-rich DNA stretches (Vashee et al. 2003; Austin, Orr-Weaver, and Bell 1999), although no defined consensus sequence has been found *in vivo* (Miotto, Ji, and Struhl 2016; Stanojcic et al. 2008). At the same time, it has also been suggested that metazoan ISs and ORC binding are associated with CpG islands and G-rich repeated elements (OGREs) that have a predisposition to form G-quadruplexes (Cayrou et al. 2012; Prorok et al. 2019; Eladl et al. 2023). It has also been proposed that asymmetric distribution of AT and GC nucleotides and flanking poly(dA:dT) tracts correlate with origin activity in mammals (Tubbs et al. 2018; Bartholdy et al. 2015). PolyA and polyG tracts are typically associated with nucleosome-free regions, due to their high intrinsic stiffness that is incompatible with nucleosome binding (Liu et al. 2018; Nelson et al. 1987).

Stretches of nucleosome-free DNA are primarily located within transcriptionally active euchromatin containing active promoters, frequently associated with early origin firing.

Nucleosomes that are present in these regions will vary according to cell type and organism, and have epigenetic features such as H4K20 methylation (Tardat et al. 2010), H3K46 acetylation (Prioleau and MacAlpine 2016; Cayrou et al. 2011; Cayrou et al. 2015) and H3K4 di- and tri- methylation (Costas et al. 2011; Eaton et al. 2011). These markers have been thought to be involved in recruiting licensing proteins to chromatic regions via an N-terminal chromatin binding domain on Orc1 (Yang and Xu 2013; Callebaut, Courvalin, and Mornon 1999; De Ioannes et al. 2019; Kuo et al. 2012), thus recruiting the rest of the core helicase components to euchromatin.

However, this does not mean that pre-RC components cannot assemble at denser regions of chromosomes. Structurally compact heterochromatin is usually replicated during mid-late S-phase (Xu et al. 2012). Studies show ORC subunits overlapping at regions of the genome containing marks typical of heterochromatin, such as H3K9me3 nucleosome modifications and HP1 proteins (Prasanth et al. 2004; Prasanth et al. 2010; Pak et al. 1997; Huang et al. 1998). Aberrant over- and under-replication occurring upon the disruption of these interactions indicate ORC is also involved in helicase recruitment at heterochromatic sites.

The presence of early and late replicating regions in metazoans indicates a level of temporal control of helicase activity and origin firing, despite the core proteins involved remaining the same. While positioning of origin sites and origin firing timing will vary according to cellular identity and environment, some of these timing changes are conserved between the same cell types of different organisms, suggesting a level of evolutionary conservation between species (Ryba et al. 2010; Hiratani et al. 2008).

Typically, the larger the genome, the higher the number of origin sites, which can result in large inter-origin distances across the genome. One mechanism to combat this is the use of dormant origins that fire only if the DNA is not already replicated by an incoming replication fork generated by the nearest origin (Santocanale et al. 1999). This adaptable activation of origins ensures replication can still progress in the event of replicative stresses, such as DNA

damage stalling incoming forks or poor growing conditions (Woodward et al. 2006; Ge et al. 2007).

Ultimately, despite the low sequence preference of origin recognition outside of budding yeast, most eukaryotes still have inherent strategies to distribute pre-RC assembly across the genome to ensure efficient S phase replication firing. Context-dependent origin licensing is not a 'one-size-fits-all,' with locations determined by a combination of genetic, epigenetic and topological features. Nonetheless, how these contextual cues are integrated to guide origin recognition and firing remains enigmatic.

1.4 Proteins Involved in Pre-RC formation.

1.4.1 The Origin Recognition Complex (ORC)

The eukaryotic Origin Recognition Complex (ORC) is a multi-subunit AAA+ (**A**TPase **A**ssociated with various cellular **A**ctivities) complex that forms a partial ring around double stranded DNA and serves as the basis of pre-RC formation. Briefly, it recognises origin DNA and can then, alongside it's cofactor Cdc6 form a closed ring around the double strand and recruit the rest of the licensing factors. It is not involved in the active replication process and must be dissociated from the DNA before helicase activation, although the precise mechanisms of this are not fully established (reviewed in Li and Stillman 2012).

1.4.1.1. ORC across evolution

Since recognition of origin sites is a critical event for chromosomal DNA replication in all living organisms, ATPase-based recognition proteins and are a common feature of replication initiation in bacteria, archaea and eukaryotes (Duderstadt and Berger 2008). In bacteria, the DnaA protein serves as the chromosomal replication initiator protein (Katayama, Kasho, and Kawakami 2017). Meanwhile, eukaryotic organisms all contain a hexameric ORC with five subunits (Orc1-5) all displaying related ATPase sites, although not all of these are functional. Orthologs of all five subunits are present in distantly related lineages, from plants to animals,

suggesting evolutionary conservation of the whole multiprotein complex. Presence of a distinct Orc1 ATPase sequence in archaea suggests a common ancestral origin (Duncker, Chesnokov, and McConkey 2009). The sub- and neo-functionalization of individual ORC subunits is thought to have developed alongside ancestral diversification (Ocaña-Pallarès et al. 2020).

1.4.1.2. Structure of ORC

Eukaryotic ORC is organised as a hetero-hexamer with subunits Orc1-6 (Figure 5.A). The largest subunits, Orc1-5 each consist of an N-terminal AAA+ or AAA+-like domain and a C-terminal winged-helix (WH) fold (Bleichert, Botchan, and Berger 2015; Speck et al. 2005; Duncker, Chesnokov, and McConkey 2009; Bell and Stillman 1992). Together, these subunits form a pentameric ring in the order of Orc1>Orc4>Orc5>Orc3>Orc2 (Figure 5.Bi), with the AAA+/AAA+-like domains and the WHDs (Liu et al. 2011) assembled into an open two-tiered ring (Bleichert, Botchan, and Berger 2015; Li et al. 2018). In this conformation, the AAA+/AAA+-like domains of adjacent subunits dock together to form composite ATPase sites (Tocilj et al. 2017; Jaremko et al. 2020). This pentameric ring can encircle a central conduit that can engage double-stranded DNA (Li et al. 2018; Tocilj et al. 2017; Bleichert, Botchan, and Berger 2015; Jaremko et al. 2020) (Figure 5.Bii).

By contrast, the Orc6 subunit consists of two cyclin-box motifs that share similarities with transcription factor IIB (TFIIB) (Liu et al. 2011). Unlike the rest of the ORC subunits, it does not participate in ring assembly, but is generally bound outside of the Orc1-5 ring, to an insertion fold on the periphery of Orc3, Orc2 and Orc5 (Li et al. 2018; Balasov, Huijbregts, and Chesnokov 2009). It is the least evolutionary and structurally conserved of all the ORC subunits and has the most variability in its role in different organisms. In S. *cerevisiae*, it appears to be an integral component for ORC recognition of origin sites, while in metazoans, it does not appear to be necessary for origin binding – many structural studies omit it completely, due to difficulties in co-purification and knockouts still being viable for replication (Bleichert, Botchan, and Berger 2015; Li et al. 2018; Tocilj et al. 2017; Jaremko et al. 2020).

While this indicates that it is not involved in origin recognition, mutations in this subunit are implicated in some cancers meaning it may still serve some function for downstream pre-RC assembly; some unrelated roles in chromosome segregation and cytokinesis have also been implied (Prasanth, Prasanth, and Stillman 2002; Chen et al. ; de Munnik et al. 2012).

In addition to the AAA+/AAA+like and WH domains, some of the ORC subunits contain ancillary domains that mediate its regulation and provide further functionalities to the complex (Figure 5.A). The Orc1 subunit alone contains an N-terminal bromo-adjacent homology (BAH) domain, attached via a flexible linker to the core of the subunit. This region has been implicated in chromatin binding (Yang and Xu 2013) and has specifically been shown to interact with specific nucleosome modifications, presumable to aid pre-RC formation at different chromatin environments (De Ioannes et al. 2019; Kuo et al. 2012; Noguchi et al.). The linker region between the core of the complex and the N-terminus is classified as an intrinsically disordered region (IDR) and is associated with a number of post-translational modifications that can mediate ORC activity (Hossain, Bhalla, and Stillman 2021).

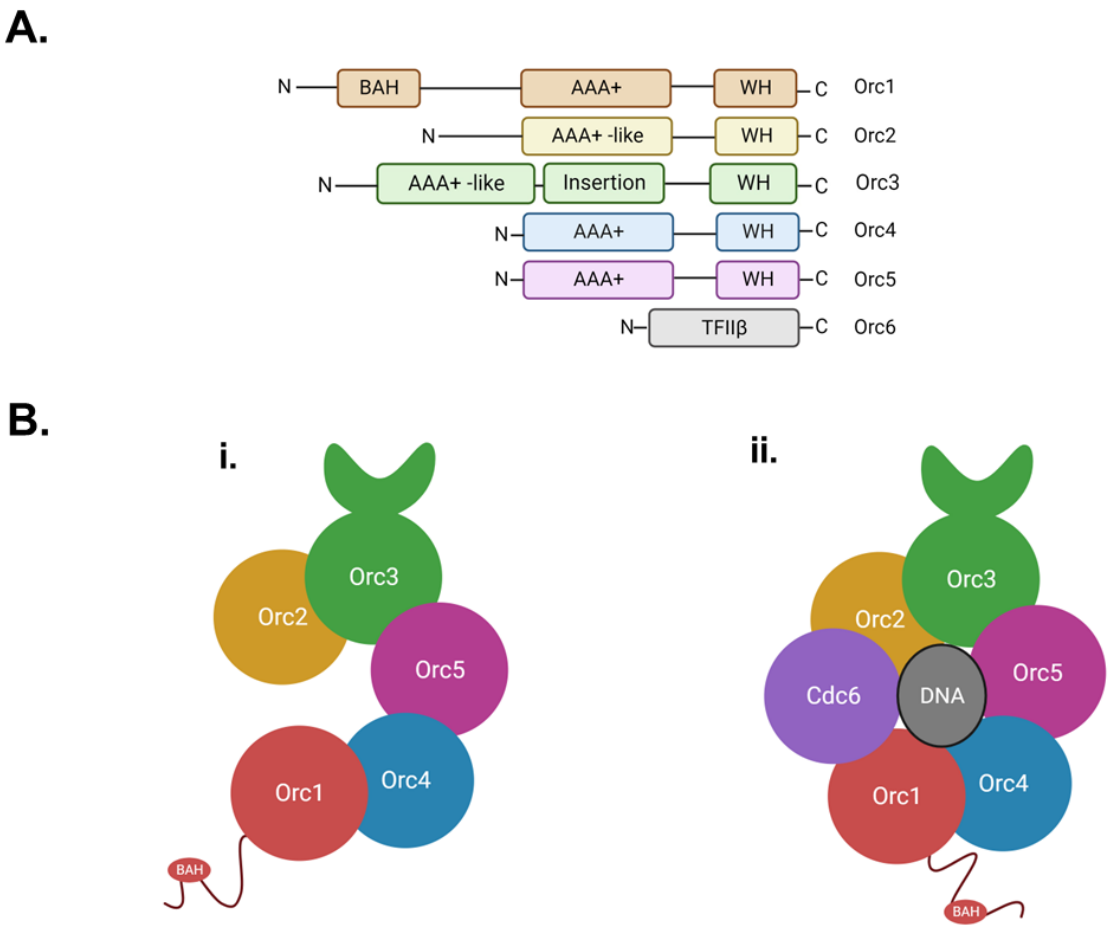


Figure 5: Organisation of the AAA+ Origin Recognition Complex (ORC) with and without cofactor Cdc6

A: *Functional domains of the six ORC subunits, adapted from Feng et al, 2021. ORC consists of five core subunits (Orc1-5) that share similar architecture in the C-terminal winged-helix domain (WH) and a central ATPase or ATPase-like domain. Orc6 follows a separate architecture of a TFIIβ-like domain and does not associate in the core Orc1-5 ring complex.* ***B: i.*** *Schematic of the assembled Orc1-5 complex, forming a partial ring that can encircle DNA.* ***ii.*** *Upon DNA binding, Cdc6 can associate with the complex, resulting in a closed ORC-Cdc6 ring around DNA (adapted from Bleichert, 2019)*

As a complex, ORC can exist in several states, both bound and free from DNA, indicating an inherent flexibility (Figure 6)(Schmidt and Bleichert 2020). Without DNA, the ring can reside in a closed or an open state. In the closed, autoinhibitory state, the ATPase modules of Orc1 and Orc4 are positioned away from each other, with the Orc2 WHD collapsed into the core of the complex, occluding the DNA cavity (Bleichert, Botchan, and Berger 2015; Bleichert et al. 2018; Schmidt and Bleichert 2020) (Figure 6.A). In the open active state, the Orc2 WHD remains away from the central cavity, closer to the Orc3 WHD, allowing the Orc1-Orc4

interaction forming the ATPase active site (Yuan et al. 2017; Tocilj et al. 2017; Li et al. 2018; Schmidt and Bleichert 2020) (Figure 6B). This open state enables access to DNA substrates and solvents, allowing DNA binding in the central core (Tocilj et al. 2017). The Orc3-5 interface is flexible, with the open and closed conformations of the complex also associated with hinge movements at this interface. The N-terminus of Orc3 contains an α-helix (aa 42 – 86) that has structurally been characterised as existing in two conformations – either an extended straightened moiety stabilising the open conformation, or bent by 32°C in the closed or DNA-bound state (Jaremko et al. 2020). These Orc3-Orc5 mediated hinge movements contribute to twisting and pinching motions, allowing the Orc2 and Orc1 subunits to move, leading to narrowing and widening of the central ring (Figure 6.A.B). It has been suggested that these movements provide the flexibility of ORC during replication events, allowing DNA binding, movement along the DNA (Sánchez et al. 2021, Sánchez et al, 2023), conformational change to allow recruitment of downstream pre-RC components, and subsequent release after MCM recruitment (Jaremko et al. 2020).

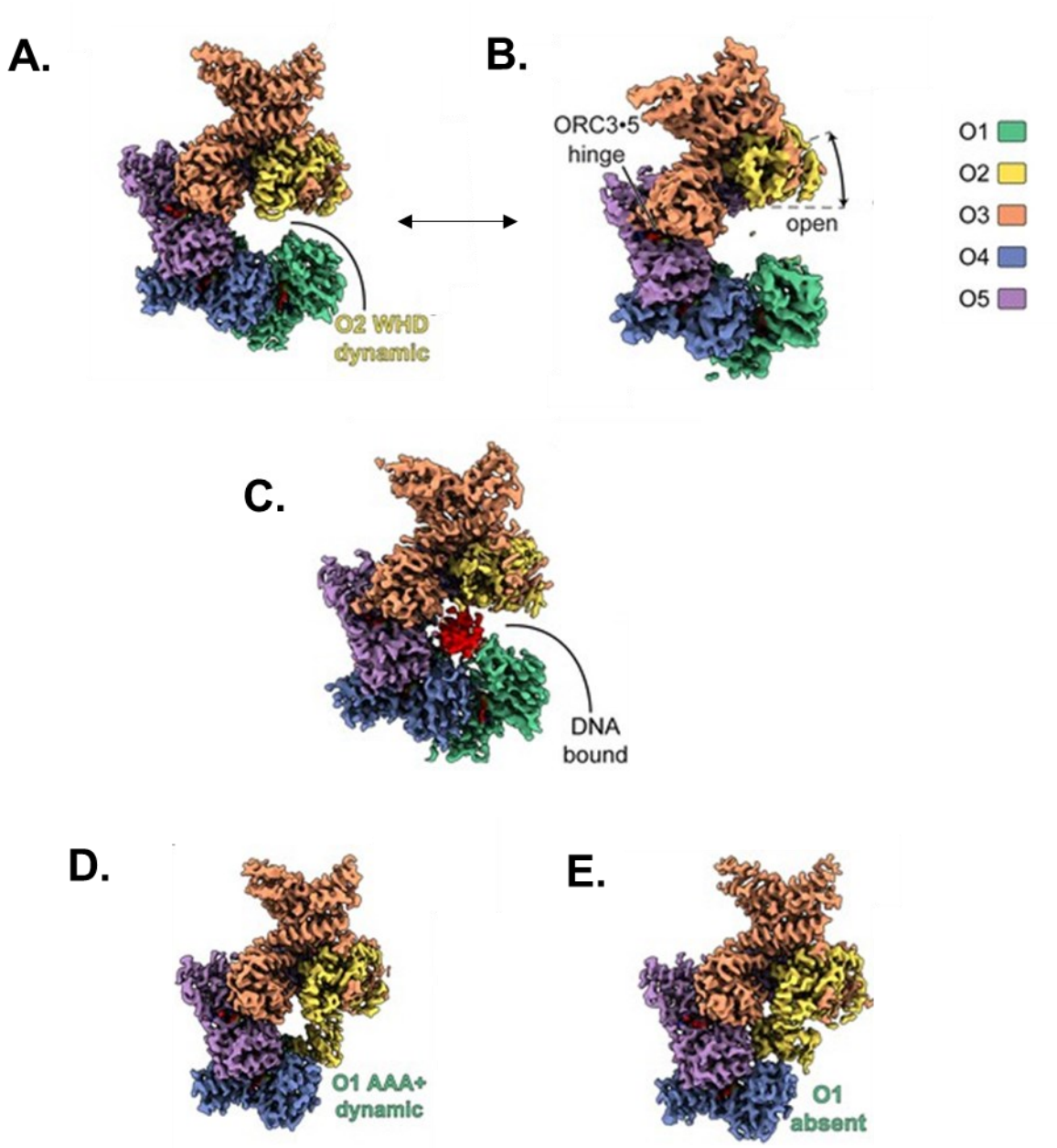


Figure 6: The inherent flexibility of human ORC.

Structural studies in the human ORC show that the complex is highly flexible, dynamic and can exist in multiple states: ***A&B:*** *Dynamic movement of the Orc1 and Orc2 WHD opens the complex to facilitate DNA binding to ORC.* ***C:*** *DNA binding to ORC, with Orc1-5 subunits forming a ring with DNA situated in the central channel. .* ***D:*** *without DNA, the Orc1 is highly flexible, large portions of the protein remain structurally unresolved.* ***E:*** *the Orc2-5 complex, where Orc1 dissociates from the core ring. EM models from Jaremko et al, 2020, PDB codes 7JPR, 7JPP, 7JPS, 7JPO, 7JPQ respectively.*

ORC bound to DNA resembles a more closed conformation of ORC, with all five main subunits contacting the DNA (Figure 6 C). The highest resolution structures of the entire ORC show

that in *S. cerevisiae*, each subunit of the budding yeast initiator contribute to DNA recognition, using different motifs that either interact with the sugar-phosphate backbone of the DNA or by making sequence-specific base contacts (Li et al. 2018). The AAA+/AAA+-like domains of Orc1, Orc2, Orc3 and Orc5 contact the DNA backbone in a nonspecific manner, while the beta-hairpin elements in the WHDs of Orc2, Orc4 and Orc5 form contacts with the DNA major grooves (Li et al. 2018). The presence of a budding-yeast specific α-helical insertion in the Orc4 WHD (aa 481 – 489) is thought to confer the ARS sequence specificity observed in budding yeast by insertion into the major groove of the ACS (Li et al. 2018; Hu et al. 2020; Yuan et al. 2017) (Figure 7 A). This helix is absent in all eukaryotes that do not display sequence-specific ORC binding and this region of Orc4 is 20 residues shorter, likely reducing its ability to confer consensus sequence specificity (Jaremko et al. 2020; Hu et al. 2020) (Figure 7 B). Despite this, presence of positively-charged residues in this region mean that it can still participate in DNA backbone interactions, albeit without a defined sequence pattern (Jaremko et al. 2020).

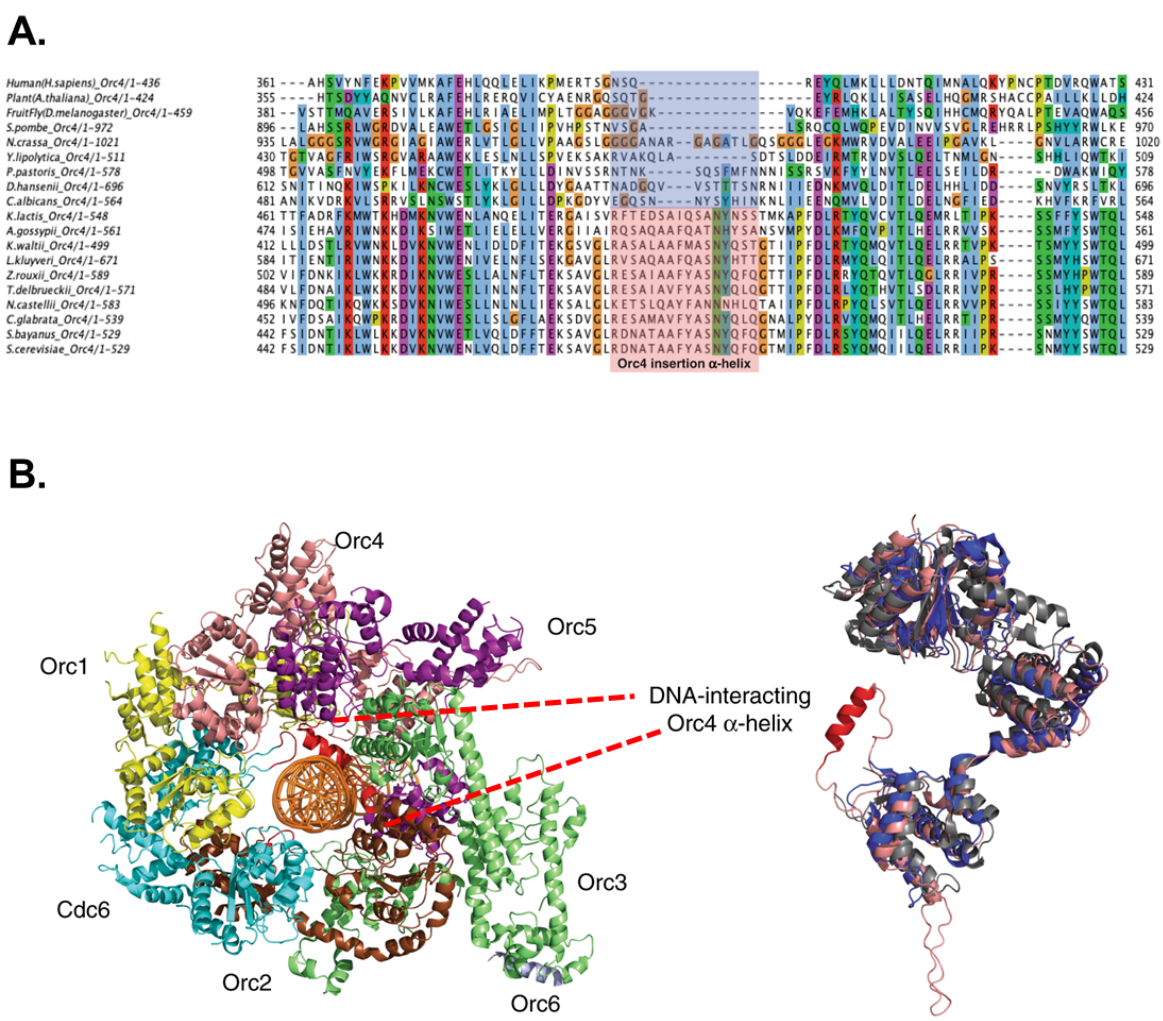


Figure 7: A conserved helical Orc4 segment confers DNA specificity in some single-celled eukaryotes.

A: *Multiple sequence alignment of Orc4 with Orc4 helix regions highlighted in pink (all single-celled related yeast or archaea) and representative higher order model organisms in blue, from Hu et al, 2020. The presence of the Orc4 helix region is likely to confer the sequence specificity of ORC to DNA in these organisms, since it is only present in those that display specific origin nucleotide sequence motifs.* ***B:*** *Top-view structure of S. cerevisiae ORC-Cdc6 bound to DNA, with DNA-interacting Orc4 helix highlighted (PDB 5UDB from Yuan et al, 2017)*

The presence of ORC subunits at all phases of the cell cycle – even outside of G1 – insinuates roles outside of replication, or a lack of degradation, presumably to reduce the transcriptional and translational toll of forming new proteins in progeny cells. Including the innate flexibility of the Orc1-5 complex, structures of a compact, stable Orc2-5 have also been identified from

cells (Jaremko et al. 2020). This has suggested a model of ORC binding in which a pre-assembled Orc2-5 binds to Orc1, promoting the opening of the core to facilitate DNA binding.

In contrast to *S. cerevisiae* ORC, low purification yields and poor behaviour of full-length *D. melanogaster* and H. *sapiens* ORC have slowed structural efforts to obtain high resolution structures of the DNA-bound metazoan complex. Nonetheless, recent publications have allowed the identification of similar architectures of metazoan ORC – *i.e*. the autoinhibitory, open, Orc2-5 and DNA bound states (Figure 6) (Jaremko et al. 2020; Schmidt and Bleichert 2020), although these cannot report on precise protein-DNA contacts. Significant structural changes, intrinsic protein flexibility and the presence of long disordered regions has meant that large domains of the complex, especially Orc1 and Orc2 remain unresolved (Figure 6 D and E). These regions have functional relevance and further studies into the structural basis of these unresolved, but functionally important sites may be key to understanding early stages of origin licensing.

1.4.1.3 ATP binding and Hydrolysis of ORC

The AAA+/AAA+-like domains of ORC can be further be divided into a RecA fold and a C-terminal lid subdomain (Wendler et al. 2012; Erzberger and Berger 2006) (Figure 8 A). The RecA fold contains a Walker A (GXXGXGKT/S) and Walker B (DD/E) motif that mediates ATP binding and catalysis (Klemm, Austin, and Bell 1997; Walker et al. 1982). It also contains a polar residue to assist the Walker B motif in nucleotide hydrolysis (Speck and Stillman 2007; Evrin et al. 2013). The composite ATPase centres also require a conserved arginine residue known as the R-finger for hydrolysis activity (Randell et al. 2006; Evrin et al. 2013; Coster et al. 2014). The C-terminal lid subdomain of the AAA+ domain also contains a conserved arginine or lysine residue involved in nucleotide sensing and transmission of conformational changes (Wendler et al. 2012) (Figure 8. B.).

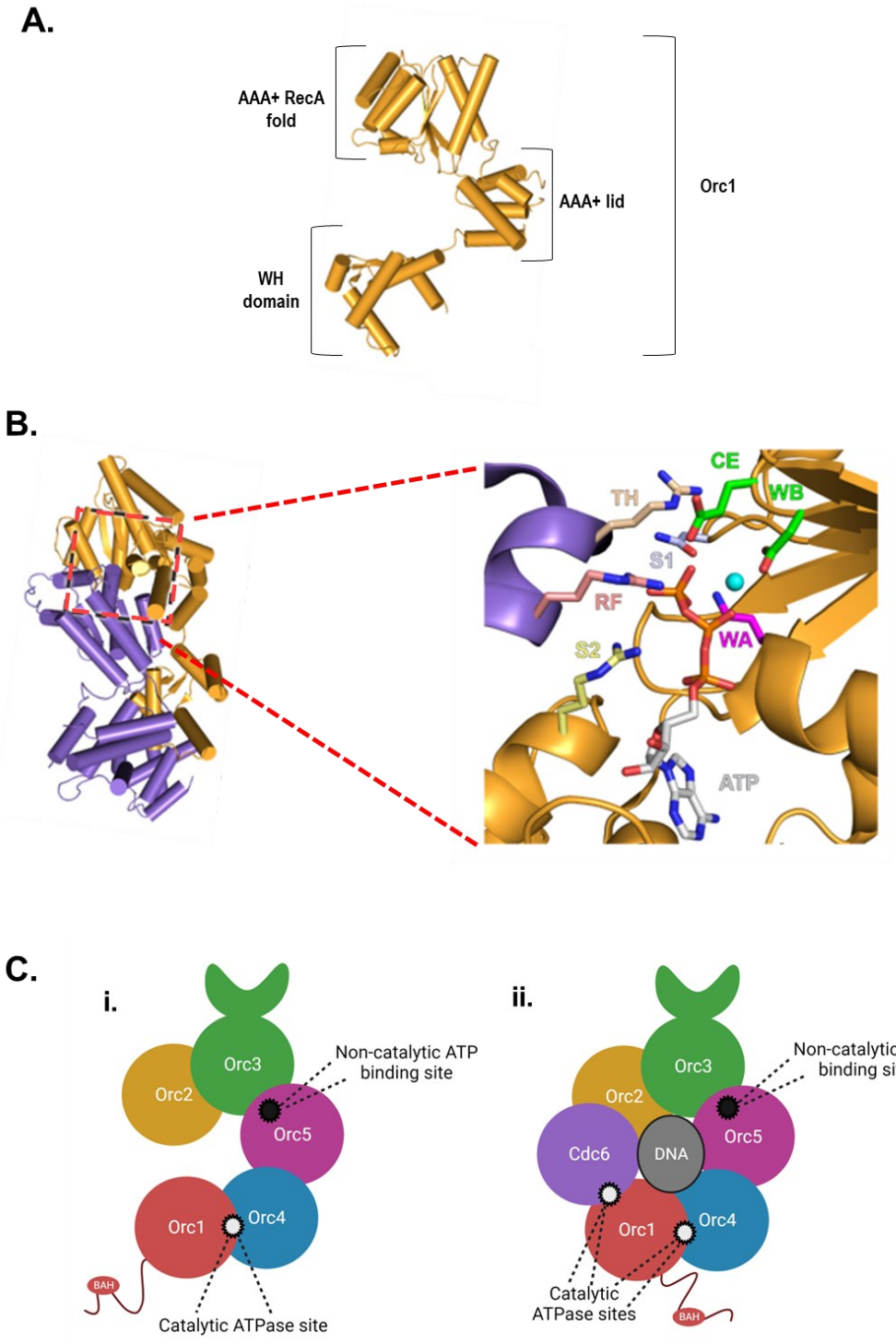


Figure 8: The ORC ATPase sites:

A*: Domain architecture of the core of human Orc1, which, like subunits Orc2-5 contains an N-terminal AAA+ domain made up of a RecA fold and lid subdomain, and a C-terminal WH domain.* ***B****: Model of the Orc1 (orange) and Orc4 (purple) composite ATPase centre (ATP – adenosine triphosphate, WA: Walker A, WB: Walker B, S1: Polar residue 1, S2: Polar residue 2, TH: tethering region, CE: catalytic glutamate, RF: arginine finger). Models of A/B from PDB 4XGC, Tocilj et al, 2017.* ***C****:* ***i*** *: Schematic of the Orc1-5 ring with its three ATP binding sites, with a canonical catalytic ATPase formed between Orc1 and Orc4.* ***ii****: an additional catalytic site is formed upon Cdc6 recruitment between Orc1 and Cdc6.*

Despite all five core subunits of the Orc1-5 complex containing AAA+ sites, functionalities of these domains have diverged through evolution (Erzberger and Berger 2006; Enemark and Joshua-Tor 2008). Only three of the subunits contain ATP binding sites (Orc1, Orc4 and Orc5). Of these sites that can bind ATP, only one bipartite catalytic centre (formed between the arginine finger of Orc4 and the ATP-binding site of Orc1) exhibits ATPase activity. It is only this catalytic centre that contains all the functional motifs (Walker A, Walker B, R-finger, Polar catalytic residues). All other AAA+/AAA+-like interfaces on the ORC ring lack one or more of these motifs, showing no hydrolase activity (Klemm, Austin, and Bell 1997; Li et al. 2018; Bleichert, Botchan, and Berger 2015; Tocilj et al. 2017; Giordano-Coltart et al. 2005) (Figure 8 C.i.).

Binding of Cdc6 to ORC on DNA results in an additional, functional composite ATPase site between Cdc6 and Orc1 (Speck et al. 2005; Liu et al. 2000; Yuan et al. 2017; Speck and Stillman 2007; Randell et al. 2006) (Figure 8 C.ii.). Ultimately, this means that only the Cdc6-Orc1 and Orc1-Orc4 ATPase centres have hydrolytic capability.

The mechanisms controlling ORC nucleotide binding and hydrolysis and its effect on downstream pre-RC assembly and initiation is poorly defined. In S. *cerevisiae*, ORC's hydrolase activity proceeds after Cdc6 activity (Randell et al. 2006) while ATP hydrolysis by ORC is likely required for iterative rounds of helicase loading by promoting its own release from DNA (Bell and Stillman 1992; Klemm, Austin, and Bell 1997). Interestingly, DNA binding to ORC's central conduit reduces the ATP activity, but it is unknown how this is coupled to substrate (DNA) binding (Klemm, Austin, and Bell 1997; Remus, Beall, and Botchan 2004; Chesnokov, Remus, and Botchan 2001). How the nucleotide hydrolysis events are coordinated and ordered by the ORC-DNA-Cdc6 complex is an open question.

While the functional roles of ATP hydrolysis in ORC is not completely defined , mutations disrupting the active catalytic sites are heavily associated with reduced cell viability and developmental disorders (de Munnik et al. 2012; Bicknell et al. 2011; Guernsey et al. 2011).

Further research is needed to decipher the precise mechanism of ORC ATPase activity during origin licensing.

1.4.1.4 ORC and Chromatin Replication

Packaging of the DNA by histone proteins into the nucleosomal core particle (NCP) affects all DNA-based reactions – including replication. The genome is packaged into highly condensed nucleosomes (heterochromatin), typically associated with repressed or silent regions of the genome; or lightly packed nucleosome particles (euchromatin), predominantly involved in active DNA processes such as transcription. This packaging is varied from cell to cell, depending on organism, cell type and cell-cycle position. Accessibility of the DNA is central to replication initiation, with heterochromatin replicating later and slower in S phase than euchromatin (Miotto, Ji, and Struhl 2016). The presence of histone modifications at different chromatin regions may be a key factor in determining the role ORC plays at each site – for example, whether it functions to recruit the rest of the pre-RC or serves additional, non-replicative functions (Chakraborty, Shen, and Prasanth 2011). However, recruitment of ORC to DNA across the varying landscape of the genome is core to ensure complete duplication of the DNA – euchromatin and heterochromatin alike.

As mentioned above, the Orc1 subunit contains an N-terminal BAH domain within its flexible linker, a nucleosome-interacting module (Yang and Xu 2013) (Figure 5.A). Mutations within this domain have shown *in vivo* effects correlating to growth delay phenotypes, particularly Meier-Gorlin syndrome (MGS) (Kuo et al. 2012; De Ioannes et al. 2019), suggesting ORC's chromatin associations having relevance to cell growth and division. However, the precise mechanism of how ORC as a complexed protein interacts with nucleosomes, steric changes in the protein structures and the downstream influence on pre-RC recruitment are yet to be revealed.

1.4.1.4.1 ORC and euchromatin

CHIP-seq studies have shown the frequency of ORC binding sites is higher in open, early replicating regions of the genome in comparison to heterochromatin, correlating replication timing to ORC density and stochastic firing from origins (Miotto, Ji, and Struhl 2016). Crystallographic studies into BAH-nucleosome interactions have shown the BAH domain can specifically recognise H4K20 dimethylation, a histone modification particularly enriched at replication origins and involved in the DNA damage response (Paquin and Howlett 2018). Direct interaction of H4K20me2 peptides with the Orc1 BAH domain *in vitro* via a dynamic aromatic dimethyl-lysine binding cage are visualised, with mutations perturbing the binding sites associated with MGS-like growth development irregularities *in vivo* (Kuo et al. 2012; De Ioannes et al. 2019) .

1.4.1.4.2 ORC and heterochromatin

ORC involvement with heterochromatin associated factors in both yeast and higher order eukaryotes has also been documented, linking it to epigenetically silenced regions of the genome (Chakraborty, Shen, and Prasanth 2011; Li and Chang 2012). This initially appears to be the converse of the open chromatin features favoured by double hexamer formation prior to replication initiation; nonetheless replication of heterochromatic regions is also crucial for complete duplication of the genome, but the facilitation and regulation of this process remains unclear. Indeed, it remains unknown whether ORC proteins can also serve auxiliary roles outside of replication, particularly those associated with genome silencing.

In *S. cerevisiae*, Orc1 can recruit heterochromatin protein Sir1 (Silent Information Regulator 1) to the mating loci of haploid cells, promoting the formation of a silencing complex (Triolo and Sternglanz 1996; Kueng, Oppikofer, and Gasser 2013). In *S. pombe*, ORC in association with Cdc6 can associate with Swi6, responsible for gene silencing at telomeres both *in vitro* and *in vivo* (Li et al. 2011). Despite this, replication in the pericentromeric chromatin region is reduced on silencing of Swi6, showing that a heterochromatic protein can positively regulate replication initiation (Hayashi et al. 2009).

A metazoan homologue of Swi6, Heterochromatin Protein 1 (HP1) is a fundamental protein responsible for condensing chromatin into its silenced state (Zeng, Ball, and Yokomori 2010; Kumar and Kono 2020). It forms a symmetrical dimer and binds to repressive H3K9me3 marks, inducing compaction of nucleosomes to form heterochromatin (Hiragami-Hamada et al. 2016; Machida et al. 2018). In *D. melanogaster*, ORC can co-immunoprecipitate with HP1 isoforms, suggesting a co-localisation of the two (Pak et al. 1997). Furthermore, siRNA-mediated knockdowns or mutations in some ORC subunits have resulted in abnormal chromosome positioning and disrupted HP1 localisation within pericentromeric chromatin and satellite repeats in mammalian cells (Prasanth et al. 2010; Prasanth et al. 2004).

The Orc2 subunit harbours an ability to bind to TRF2 (telomeric repeat binding factor 2), a protein involved in telomere maintenance and protection against end-to-end fusion of chromosomes. ORC can also specifically recognise ORCA (ORC-associated protein A), a factor that preferentially binds to H3K9 methylated marks enriched at heterochromatin, with ORCA-deficient cells displaying aberrant heterochromatin replication (Giri et al. 2015; Giri and Prasanth 2015).

Since ORC has a high affinity for open chromatin, its ability to associate with heterochromatic factors may be core to facilitate pre-RC recruitment and replication at heterochromatic regions. However, the precise mechanisms of recruitment at these regions remains unclear, as does any potential auxiliary role ORC may have in organising chromatin structure. The presence of ORC subunits at all stages of the cell cycle indicates auxiliary, non-replicative functions and understanding this may be key to linking its roles in replication to chromatin architecture formation.

1.4.2 Cdc6 (Cell-division-cycle 6)

First identified in budding yeast, Cdc6 was found as an essential factor in G1, where mutants displayed cells arrested in the G1-S phase boundary. It is a key initiation factor involved in the assembly of the pre-RC, considered to be a co-factor of ORC, involved in locating origin sites

on the DNA and recruiting downstream factors. It is a single subunit AAA+ protein, sharing homology with Orc1-5 and is responsible for the stabilisation of the pentamer onto DNA.

Knockouts or aberrations in ATPase activity resulted in accumulation of cells in G1. Cells that continued into the cell cycle resulted in a 1C complement of DNA, suggesting an absence of DNA replication (Piatti, Lengauer, and Nasmyth 1995).

1.4.2.1 Structure of Cdc6

Cdc6 shares much sequence and structural homology with individual ORC subunits. It contains a homologous AAA-type nucleotide binding fold involved in ATP hydrolysis and a C-terminal winged helix domain that interacts with DNA (Speck et al. 2005; Speck and Stillman 2007). It shares the most sequence and structural conservation with Orc1 in comparison to all other ORC subunits (Liu et al. 2000). Like Orc1, it contains an N-terminal disordered extension that contains multiple phosphorylation sites that regulate its and function across the cell cycle (Perkins, Drury, and Diffley 2001; Liu et al. 2000) (Figure 9.A)

The only available high-resolution structure of Cdc6 alone is a crystal structure of an archaeal orthologue from *P. aerophilium* (Liu et al. 2000). Here, it was shown that two motifs, analogous to the Walker A/B motifs, form an AAA+ module capable of forming a nucleotide binding pocket (Figure 9.B.). This study also characterised the winged helix fold, a region that was critical for cell cycle checkpoint control and progression into S phase.

Like Orc1, because of its inherent regions of flexibility and disorder, most structural studies can only visualise Cdc6 when bound and stabilised by associated proteins such as CDK2/Cyclin A (Cheng et al. 2006) or, most relevantly, in combination with other pre-RC components, primarily ORC (Schmidt and Bleichert 2020; Feng et al. 2021).

In *S. cerevisiae*, the WHD of Cdc6 contributes to ORC-based origin DNA recognition via an initiator specific motif. When compared to ORC-DNA, the addition of Cdc6 expands the footprint of nuclease-protected (and therefore protein-bound) DNA around origin sites (Diffley et al. 1994). Most ORC-DNA contact points are retained upon the addition of Cdc6 (Feng et

al. 2021). Exceptionally, the Orc2 WHD is displaced from the centre of the ring structure and is replaced by the Cdc6 AAA+ domain, to form a closed ring of six ATPase domains around the ACS element of the DNA. The WHD of Cdc6 remains flexible, with transient binding possible – overall likely to provide freedom to engage with downstream MCM association (Feng et al. 2021) (Figure 9.C.).

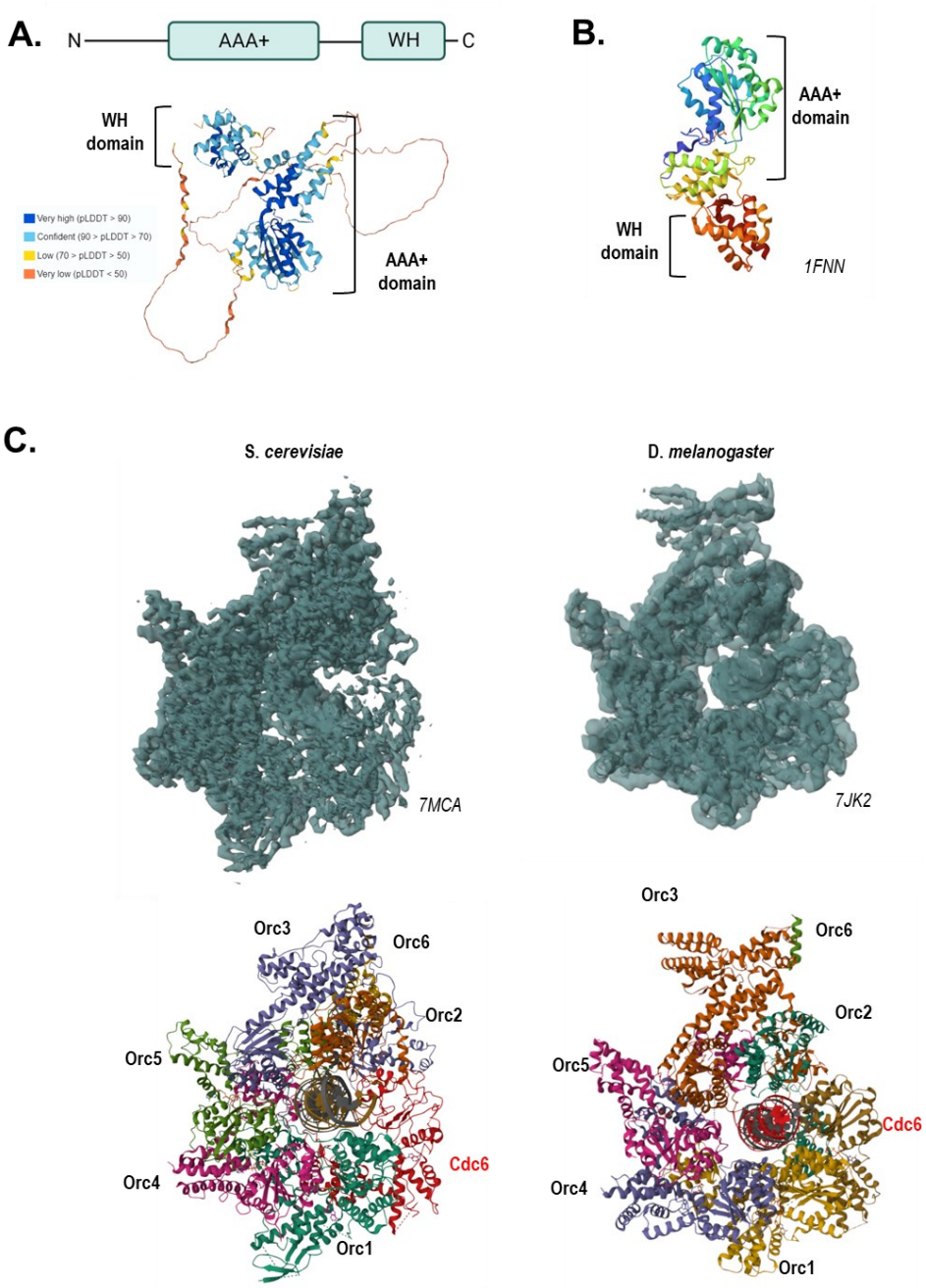


Figure 9: Cdc6: Structure and domain organisation

***A**: Domain organisation of Cdc6, and corresponding Alphafold prediction, containing an N-terminal AAA+ domain and a C-terminal WH domain, connected and flanking by long flexible linkers. **B**: Crystal structure of P. aerophilium (PDB 1FNN from Liu et al, 2000) showcasing the core domains. **C**: S. cerevisiae (Feng et al, 2021) and D. melanogaster (Schmidt and Bleichert, 2020) EM density and corresponding models of Cdc6 bound to ORC-DNA, displaying a similar closed ring architecture, with Cdc6 binding between Orc1 and Orc2 on the ring.*

Studies into the structure of *D. melanogaster* of ORC-Cdc6-DNA show that it is arranged similarly around DNA as *S. cerevisiae*, suggesting structural conservation between the two, despite the differences in DNA sequence requirements. *D. melanogaster* ORC-Cdc6 encircles

DNA as a hexameric, two-tiered almost planar ring (Bleichert et al. 2018). Like in S. *cerevisiae*, the presence of Cdc6 does not show any significant remodelling of the ORC-DNA configuration, although it does stabilise the Orc2 WHD (Schmidt and Bleichert 2020).

Upon binding to the ORC ring, it specifically adopts a similar 3D structure to Orc1 and is positioned between Orc1 and Orc2, closing the ring around DNA. It engages in canonical and non-canonical AAA+ interactions with Orc1 and 2. It forms a composite ATPase site with Orc1 (Speck et al. 2005; Yuan et al. 2017; Randell et al. 2006; Wang et al. 1999; Mizushima, Takahashi, and Stillman 2000) but also has two conserved arginine residues that are involved in mediating protein-protein interactions at the non-catalytic Orc2-Cdc6 ATPase interface, resulting in increased complex stability (Schmidt and Bleichert 2020).

1.4.2 .2 ATP binding and hydrolysis of Cdc6

Cdc6 alone is hydrolytically inactive, without displaying any ATPase activity. Upon binding to ORC-DNA, formation of the Orc1-Cdc6 catalytic core occurs. Coordinated ATP hydrolysis at this site is required for efficient loading of downstream pre-RC components onto DNA and encourages the release of nonproductive intermediates. However, the mechanisms controlling ATP binding and hydrolysis are poorly understood, especially in higher-order metazoans.

In studies done in *S. cerevisiae*, Cdc6 contributes to the unique case of DNA origin specificity. In the presence of non-origin DNA, Cdc6 rapidly hydrolyses ATP, destabilising the Orc1-Cdc6 interface and resulting in the disassociation of both itself and ORC from DNA (Coster et al, 2014, Frigola et al, 2013, Fernández-Cid et al, 2013, Samel et al, 2014, Speck et al, 2017) (Figure 8.C.). Conversely, in the presence of a canonical origin site, ATP hydrolysis is delayed, stabilising the ORC-Cdc6 association with DNA until the MCM proteins are loaded (Speck and Stillman 2007; Mizushima, Takahashi, and Stillman 2000). Secondly, the physical loading of the MCM double hexamer is dependent on Cdc6-based ATP hydrolysis. ATP hydrolysis by the Orc1-Cdc6 catalytic site can promote Cdt1-MCM loading but can also initiate release of pre-RC components in the event of missing pre-RC proteins or phosphorylation-induced ORC

inactivation (Frigola et al. 2013; Coster et al. 2014). Cdc6 ATPase activity disengages Cdc6 from the pre-RC, allowing double hexamer formation (Chang et al. 2015).

An interesting observation of note is that some *H. sapiens* cell lines have shown proliferation ability even in the disruption of Orc2 or Orc1 subunits, provided that Cdc6 is still fully functional. Considering the evolutionary conservation between Orc1 and Cdc6, it could be that this provides an additional element of protection – that even in the absence of the ATPase subunit, Cdc6 provides a level of redundancy, potentially providing the same purpose as Orc1 to load MCM2-7 onto origins (Shibata et al. 2016). However, this remains controversial, and more studies are needed for further insight into this potential role.

As of yet, Cdc6 has not been shown to influence the selection of origin sites in higher-order eukaryotes as it does in *S. cerevisiae*. Despite structural advances shedding light on *D. melanogaster* ORC-Cdc6 structure, the precise role of ORC and Cdc6 ATPase activity during origin licensing for *H. sapiens* proteins remains unstudied. While likely it serves homologous functions, further research into this will enable a deeper understanding of origin licensing regulation by ATP hydrolysis.

1.4.3 Cdt1

Cdc10-dependent transcript protein 1 (Cdt1) is another key component involved in loading MCM2-7 onto chromatin during replication licensing. First discovered in fission yeast, it coordinates with Cdc18 (a Cdc6 homologue in fission yeast) to load MCM proteins into DNA (Nishitani et al. 2000). This was further confirmed to act similarly in budding yeast, with Cdt1 interacting directly with MCM2-7 hexamers in solution and loading cooperatively onto the DNA to result in stable, head-to-head double hexamers (Remus et al. 2009; Tanaka and Diffley 2002; You and Masai 2008; Fernández-Cid et al. 2013; Liu et al. 2012). Corresponding studies using *X. laevis*, *M. musculus* and *H. sapiens* cells show Cdt1 functioning in an analogous role (Nishitani et al. 2001; Maiorano, Moreau, and Méchali 2000; Pozo et al. 2018).

Unlike the rest of the essential proteins involved in the pre-RC loading and licensing reaction, Cdt1 does not harbour any intrinsic enzymatic activity. Instead, it functions as a loading cofactor, to stabilise the MCM2-7 hexamer in an open conformation that renders it compatible to DNA binding and double hexamer formation (Fernández-Cid et al. 2013; Zhai et al. 2017; Frigola et al. 2017). The protein itself dissociates from origins upon double hexamer formation and is not part of the active replisome (Ticau et al. 2017). Despite its lack of enzymatic activity, it remains a crucial component of replication licensing – inactivation or deletion of Cdt1 in cells results in the failure of MCM loading (Fujita 2006).

1.4.3.1 Structure of Cdt1

Compared to other pre-RC components, Cdt1 shares the lowest level of sequence homology across eukaryotes (Tanaka and Diffley 2002; Devault et al. 2002; Fujita 2006). Despite this, the general domain architecture is still preserved across species.

Cdt1 is largely flexible, but can be divided into three globular domains; the N-terminal domain (NTD), the central helical domain (CH) and the C-terminal domain (CTD), with a flexible linker between them (Zhai et al. 2017; Pozo and Cook 2017) (Figure 10.A.). These domains form multiple contacts with the MCM2-7 hexamer (Figure 10.B.), modulating it in an open ring conformation to permit DNA loading (Fernández-Cid et al. 2013; Zhai et al. 2017)

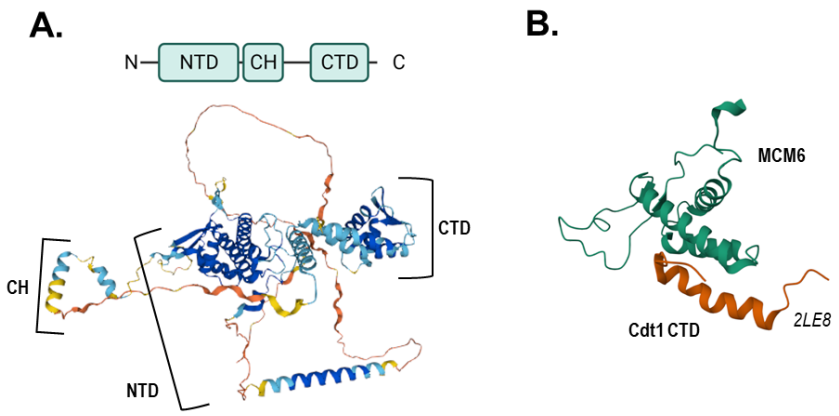


Figure 10: Cdt1 structure and domains

A: *Domain organisation and corresponding AlphaFold prediction of human Cdt1. It consists of a C-terminal and N-terminal and central domain (NTD, CTD and CH) that all form separate helical structures joined by flexible linkers. Due to the large regions of flexibility and overall small size of the protein, structural biology techniques have struggled to isolate Cdt1 alone.* ***B:*** *Crystal structure of the Cdt1 CTD helical segment lined against the with the WHD tail of Mcm6 (Liu et al, 2011)*

1.4.4. MCM2-7

The core of the replicative helicase is made up of a heterohexameric assembly of proteins, termed the minichromosome maintenance proteins subunits 2-7 (MCM2-7). Together, they form the molecular motor that unwinds duplex DNA and is central to fork progression during DNA replication elongation (Bochman and Schwacha 2008). In replication licensing, the end goal of ORC, Cdc6 and Cdt1 is to recruit two MCM2-7 hexamers to origin sites where they can remain until activated in the G1-S phase transition (Kelman and Kelman 2013; Forsburg. 2004).

Members of the MCM family have been found conserved in all eukaryotes, as well as related proteins existing in archaea. While individual MCMs within a single species are not overly homologous to each other, conservation of each MCM subunit across species is what has allowed the identification is distinct MCM classes (Forsburg. 2004).

1.4.4.1 Structure of MCM2-7

MCM proteins are comprised of three primary domains – the N-terminal extension that contains an oligonucleotide binding domain for capturing single-stranded DNA, and the C terminal domain consisting of a core AAA+ domain (or MCM 'box') and a winged helix domain. (Li et al. 2015; Yuan et al. 2017) (Figure 11.A.). The AAA+ domains form an active ATPase between each pair of neighbouring subunits that play roles in helicase loading, activation and unwinding (Bochman, Bell, and Schwacha 2008; Coster et al. 2014; Kang, Warner, and Bell 2014; Schwacha and Bell 2001; Bochman and Schwacha 2008). It consists of the Walker A and Walker B elements and a short motif downstream of the Walker B element, defining the arginine finger (Forsburg, 2004). ATP binding enables the subunits to arrange into a left-handed toroidal ring in the order Mcm2 – Mcm5 – Mcm3 – Mcm7 – Mcm4 – Mcm6 (Figure 11.B.). This ring is not closed, but possesses an opening 'gate' between Mcm2 and Mcm5 that can open and close, depending on nucleotide occupancy to allow substrate (DNA) binding in the central channel of the MCM ring (Samel et al. 2014; Zhai et al. 2017; Xu et al. 2022). The C-terminal WHD of Mcm3 is responsible in S. *cerevisiae* for binding and stabilising the MCM hexamer against Cdc6 and Orc2 (Frigola et al, 2012, Yuan et al, 2017). The C-terminal WHD of Mcm6 interacts with Cdt1 during helicase loading (Fernandez-Cid et al, 2013; Figure 10.B.) and with Polε at the active replication fork (Jones et al. 2023). The N-terminal extensions – as well as containing the oligonucleotide binding fold that binds single-stranded DNA once DNA unwinding occurs - are involved in regulatory roles, controlling helicase activation via phosphorylation (Saleh et al. 2022; Greiwe et al. 2022; Cheng et al. 2022).

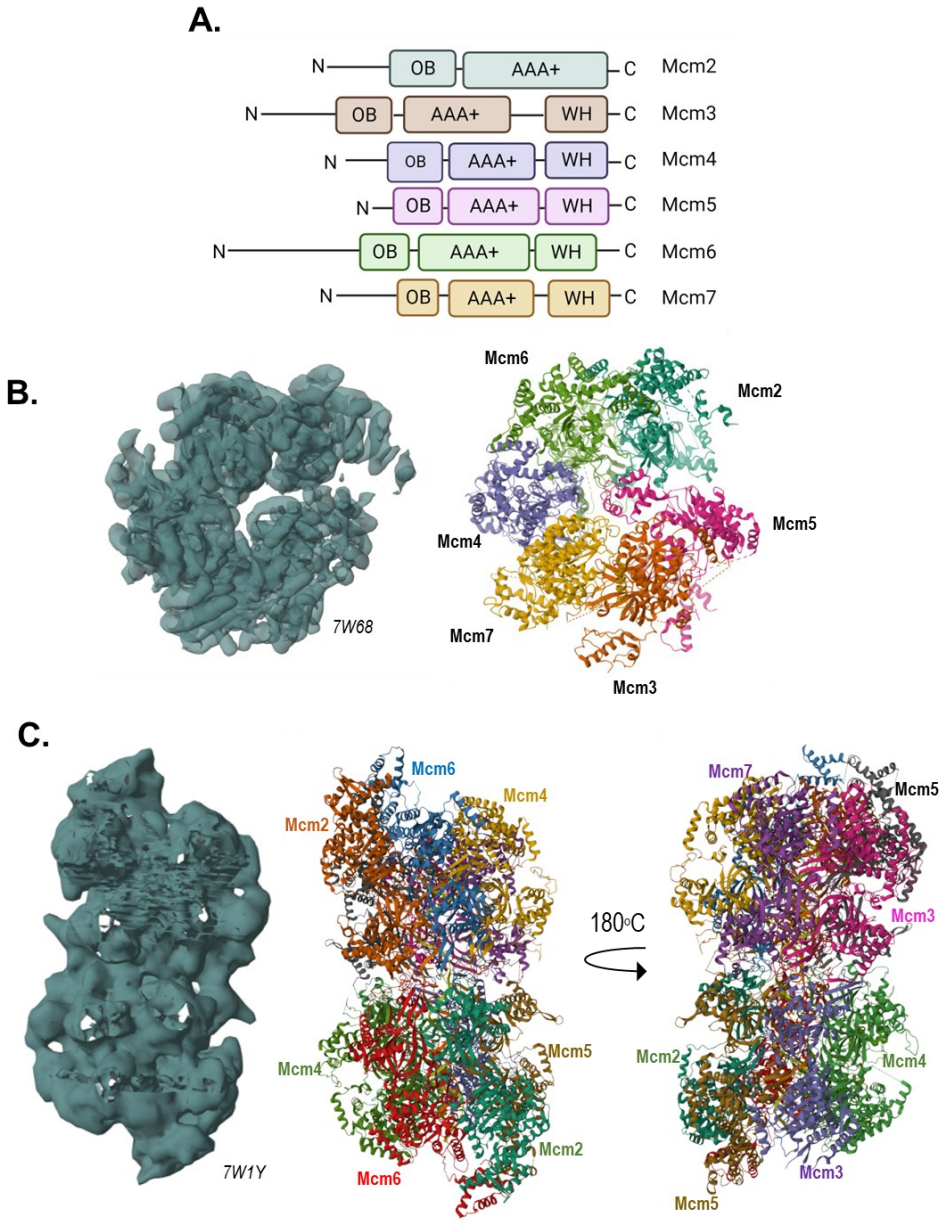


Figure 11: MCM Structure and organisation

A: *Schematic representation of Mcm2-7 subunits, with oligonucleotide binding domains (OB), AAA+ domains and WH domains highlighted. All six subunits present similar architecture, with only Mcm2 lacking in a C-terminal WH domain.* ***B:*** *EM density map and model of a single human hexamer of Mcm2-7, forming a ring with a central DNA binding channel. PDB 7W68 from Xu et al, 2023* ***C:*** *EM density map of the assembled human double hexamer, bound in a head-to-head fashion around DNA. The N-terminal regions of each subunit form the central contact points between hexamers. PDB 7W1Y from Li et al, 2023*

1.4.4.2 ATP binding and Hydrolysis by MCM

Like ORC, the ATPases in MCM assemble into oligomers with the active site sitting at the interface of adjoining subunits. Here, one subunit contributes the Walker A and B motifs while

the arginine finger required for catalysis is contributed by the other subunit (Davey, Indiani, and O'Donnell 2003). ATP binding and hydrolysis by MCM subunits are required for release of Cdt1 and full loading of the double hexamer (Coster et al. 2014).

Closing of the Mcm2/5 gate is hydrolysis dependent, as is activation of the double hexamer helicase at the G1 – S phase transition and interactions with Cdc45 and GINS complex stimulates further hydrolysis activity to separate the double strand and regulate replisome progression (Douglas et al. 2018; Perera and Trakselis 2019; Ilves et al. 2010; Ying and Gautier 2005).

1.5 Intermediates involved in Loading of the Replicative Helicase

Unravelling the molecular mechanisms of helicase loading has been a major research focus for several decades. Most of the work in determining the mechanisms and pathways that load the eukaryotic replicative helicase have been carried out in *S. cerevisiae*, although more recent studies in metazoans such as *D. melanogaster* and *X. laevis* have identified paralogous steps (Ekundayo and Bleichert 2019; Bleichert, Botchan, and Berger 2017). Loading of the two hexamers is an orchestrated sequence of initiator-cofactor intermediates that undergo ATP hydrolysis-based conformational changes to achieve coordinated recruitment of the core MCM helicase to the DNA.

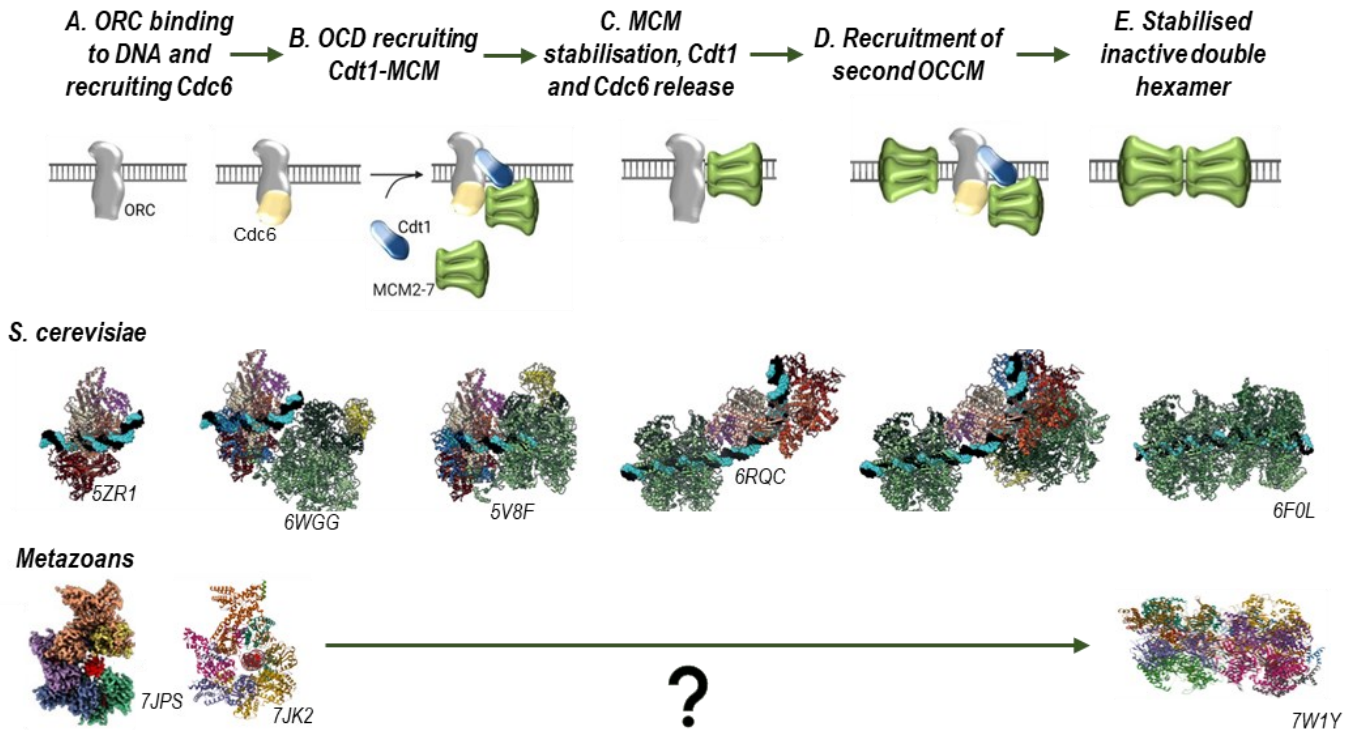


Figure 12: Pre-RC assembly with associated intermediate steps identified in S. cerevisiae and metazoans.

The associated intermediate steps of loading of the replicative helicase – from ORC binding to DNA to stable double hexamer - identified through structural studies of S. cerevisiae and D. melanogaster/H. sapiens proteins. High resolution structures of the intermediates between ORC binding to DNA to inactive double hexamer fully assembled onto origin sites have been heavily studied in S. cerevisiae, enabling the ordering of intermediates along the pathway. So far, for metazoans, structure of only D. melanogaster Orc1-5 and Cdc6 bound to DNA and the fully formed double hexamer have been published. Intermediates along the pathway have not yet been studied on a structural scale – the differences between higher order pre-RC assembly and the established S. cerevisiae mechanism is still an open question. Adapted from reviews Riera et al, 2017 and Costa and Diffley, 2022, with corresponding PDB models for individual intermediates.

Once ORC has bound to DNA in its central channel, Cdc6 is recruited, forming an ORC-Cdc6-DNA (OCD) complex. This closes the initiator ring and clamps it onto DNA, bending the double strand (Bleichert et al. 2018; Yuan et al. 2017; Sun et al. 2014) (Figure 12.A.). This closed OCD is then competent to recruit a single Cdt1-bound MCM hexamer by interactions with Mcm3 and Mcm7 (Yuan et al, 2020) that can latch onto DNA via it's Mcm2/5 gate (Samel et al. 2014). This forms a short lived intermediary ORC-Cdc6-Cdt1-MCM2-7 (OCCM) complex, where the bent DNA is aligned with the gap on the MCM ring, then released from ORC to enter

the central channel of the MCM2-7 ring. (Fernández-Cid et al. 2013; Sun et al. 2014; Miller et al. 2019; Ticau et al. 2017). (Figure 12.B.).

The mechanism of recruitment of the second hexamer is still a topic of debate. In *S. cerevisiae*, engagement of the high-affinity ACS element by ORC is what leads to the recruitment of the first MCM2-7 hexamer and formation of the OCCM intermediate. Subsequent ATP hydrolysis releases Cdc6 and Cdt1, closing the Mcm2/5 gate and forming another intermediary ORC-MCM (OM) complex bound to the DNA (Kang, Warner, and Bell 2014; Coster et al. 2014; Ticau et al. 2017) (Figure 12.C-D.). Following this, ORC detaches from the loaded single hexamer and binds to the inverted B2 element, a weaker ORC binding site, where it can associate with the N-terminus of the loaded MCM hexamer. The repeated recruitment of a second Cdc6 and a subsequent Cdt1-MCM can then result in a head-to-head loaded double hexamer, with the N-terminal domains of each MCM2-7 facing each other. (Figure 12.E.).

The question of whether the recruitment of a second ORC molecule binding event is required for double hexamer formation is still a topic of investigation. Structural studies have suggested that the first bound single hexamer can recruit a secondary ORC molecule (Miller et al. 2019). Meanwhile, single-molecule FRET studies have indicated that the initial ORC can carry out a positional 'flip', to the opposite side of the loaded single hexamer to recruit the second Cdc6, then Cdt1-MCM2-7 hexamer (Gupta et al. 2021). Both mechanisms have been suggested, although it is still unclear whether loading of the double hexamer involves one or two ORC molecules, and whether both or only one mechanism is relevant *in vivo*.

1.6 Activation and Downstream Activity of the Double hexamer: Replication Firing

After loading, the MCM2-7 double hexamer is activated by the combined action of cell-cycle regulated kinases and firing factors at the G1 – S phase transition. The double hexamers are activated by phosphorylation of the Dbf4-dependent Cdc7 kinase (DDK) (Saleh et al. 2022; Greiwe et al. 2022; Cheng et al. 2022). Subsequent activation by Sld2/3/7 (synthetic lethality

with dpb11) and Dpb11 (Tanaka et al. 2011) can then facilitate the loading of the Cell division protein 45 (Cdc45) and Go-Ichi-Ni-San (GINS) to form two Cdc45-MCM-GINS (CMG) complexes. This forms the core of the motor module of the moving replisome, each with DNA unwinding activity. Additional cofactors including Mcm10, Polymerases alpha and epsilon (Polα/ε), Ctf4, RPA and Top11 form the minimal DNA replication machinery in vitro (Yeeles et al. 2015; Baris et al. 2022), although additional cofactors are needed for the efficiency of the same process *in vivo*. Activation of the replisome initiates leading and lagging strand synthesis, ultimately resulting in a cell with a duplicated complement of its genome (Lewis and Costa 2020) (Figure 1).

Activation and beyond of the MCM2-7 double hexamer is beyond the scope of this thesis, although it is still of value to consider these events to appreciate the importance of precise and regulated helicase loading in G1.

1.7 Coupling DNA Replication Initiation to the Cell Cycle

Cells go through iterative phases of genome duplication and segregation in the cell cycle. Transition through phases is tightly controlled, primarily by cyclin-dependent kinases (CDKs) that transfer phosphate groups to target proteins to modulate their functionalities (Poon 2016). The activity of CDKs oscillates throughout the cell cycle, determined by cycle-dependent cyclin expression and degradation. Cyclins are marked for proteasomal degradation by ubiquitination, either by the E3 ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C) (Peters 2006) or the Skp, Cullin F-box containing complex (SCF) (Ang and Wade Harper 2005). The APC/C is activated during mitosis by M-CDK-mediated phosphorylation and operates until the G1-S phase transition, maintaining the degradation of S- and M-phase cyclins (Kraft et al. 2003). This is inhibited in late G1/early S phase, allowing the expression and accumulation of S-phase cyclins that activate licensed origins and promote further repression of APC/C. This interplay between CDKs and APC/C results in discrete periods of activity, which is reinforced by additional and interconnected regulatory factors such as cyclin inhibitors, phosphatases, ubiquitinases and transcription factors.

Multiple mechanisms of CDKs restrict helicase loading to G1 and to help safeguard against re-replication of the DNA. CDK activity is generally low in G1 to allow helicase loading, but increases upon entry into S-phase, simultaneously activating the loaded helicases while preventing further helicase loading. S-phase CDKs primarily target ORC, Cdc6 and MCM2-7 (Nguyen, Co, and Li 2001). (Figure 13)

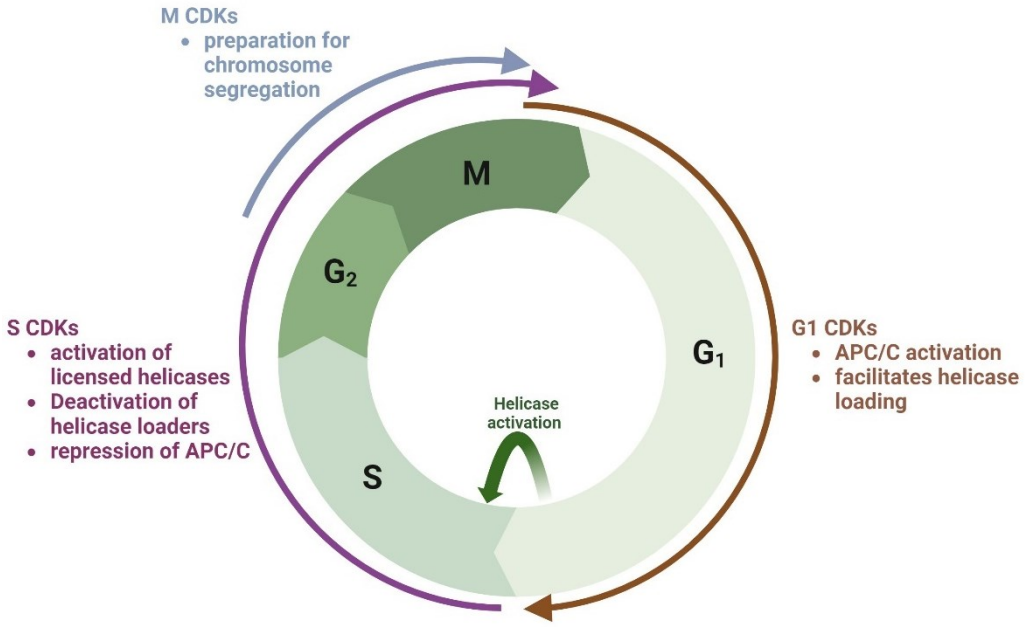


Figure 13: Cell cycle CDK activity regulates origin licensing and activation.

Schematic of relative activity over the cell cycle, in M phase, S-CDK and M-CDK levels are high, preventing origin licensing from occurring. At mitotic exit and the onset of G1, APC/C is activated to facilitate helicase loading by inhibiting S/M-CDKs. Upon S-phase activation, S-CDK activity rises again to prevent further helicase loading that can result in over-replication of the DNA. Figure made in BioRender.

1.7.1 ORC

Multiple ORC subunits are subject to CDK-dependent regulation – functionally, this process is conserved from budding yeast to metazoans, although this is done through different mechanisms (Siddiqui, On, and Diffley 2013; Diffley 2011). In budding yeast, S-phase CDKs promote phosphorylation of Orc2 and Orc6 – while this doesn't specifically affect MCM recruitment to DNA-bound ORC, it does appear to hinder helicase loading and prevents re-

replication *in vivo* (Nguyen, Co, and Li 2001; Chen and Bell 2011). In higher-order metazoans, the N-terminal regions on Orc1 and Orc2 are major targets for phosphorylation-dependent regulation. CDK-mediated phosphorylation of *D. melanogaster* Orc1 and Orc2 reduces ORC's ATPase activity, suggesting a regulatory role in ORC's ATP-dependent functions (Remus et al. 2005). Specifically, S-phase CDK/Cyclins E and A regulate DNA licensing at this level, with low amounts activating DNA licensing via modifications of the helicase loading proteins, and vice-versa (Furstenthal at al, 2001, Ekholm-Reed et al, 2004, Coverly et al, 2002). Similarly, S-phase CDK-dependent phosphorylation in the Orc1 disordered linker region reduces Orc1's affinity to chromatin in S phase and prevents the formation of a stable ORC-chromatin complex until mitosis is completed (Li, Vassilev, and DePamphilis 2004). Interestingly, the presence of Orc1 oscillates throughout the cell cycle, but is the only subunit of ORC to do so. It accumulates in mid-G1 phase, reaching a peak at the G1-S boundary, then reducing via CDK or ubiquitination activity to a basal – but not completely absent – level in S phase (Méndez et al. 2002; Li, Vassilev, and DePamphilis 2004). While Orc2-Orc5 and Orc6 also contain regions of phosphorylation potential, they remain at a constant level throughout the cell cycle, suggesting that the Orc1 subunit is key to temporal control of helicase loading.

1.7.2 Cdc6

Post pre-RC formation and upon checkpoint activation for progression of G1 into S phase, CDKs phosphorylate Cdc6, targeting it for ubiquitination and degradation by the proteasome in *S. cerevisiae* (Drury, Perkins, and Diffley 1997; Perkins, Drury, and Diffley 2001). In budding yeast, this regulation avoids reactivation of helicase assembly within the same cell cycle. While in some proliferating human cells, the APC/C can target Cdc6 for ubiquitin-mediated degradation (Petersen et al. 2000; Méndez and Stillman 2000), often Cdc6 is not eliminated in S or G2, with alternative mechanisms in place to prevent aberrant reformation of the pre-RC per cell cycle. The S/G2 phase Cyclin A-CDK2 complex can phosphorylate residues on the N-terminus of Cdc6 proteins not associated with chromatin other initiator counterparts,

rendering it inactive and preventing reinitiation in human cells post the G1-S checkpoint (Méndez and Stillman 2000; Petersen et al. 2000; Coverley et al. 2000).

1.7.3 Cdt1 and MCM

Cdt1 is present strictly in G1 alone – upon the onset of S phase, levels decrease and only become detectable again upon a new entry into G1 (Nishitani et al. 2001). Recombinant Cdt1 introduced into cells outside of G1 results in high levels of illegitimate origin firing and unregulated reinitiation of DNA synthesis in Xenopus egg extracts (Maiorano et al. 2005).

Despite its lack of ATP activity, it is perhaps the most extensively regulated of all the pre-RC components. In metazoans, Cdt1 function is inhibited post S-phase onset by Geminin, a small protein that functions by sequestering Cdt1 and rendering it incapable of binding to MCM (Scott and Gheorghe 2022); thus making replicative hexamer recruitment to origin-bound ORC-Cdc6 impossible (Tada et al. 2001; Wohlschlegel et al. 2000). At the end of mitosis, Geminin undergoes proteasome-mediated degradation, allowing the licensing system to reactivated on the onset of a new cell cycle (Li and Blow 2004; McGarry and Kirschner 1998). Cdt1 itself can also be abolished by proteasome-mediated degradation through the APC/C pathway (Arias and Walter 2006). Since pathological dysregulation of Cdt1 results in strong genotoxic stress and oncogenic transformation (Vaziri et al. 2003; Fujita 2006), both mechanisms act as redundant pathways to prevent genome over-replication.

In theory, a double hexamer of MCM2-7 loaded at each replication origin is sufficient to replicate the entire genome. However, MCM2-7 hexamers exist in the cell in a vast stochiometric excess in comparison to the levels of other replication factors, or indeed replication origins (Hyrien, Marheineke, and Goldar 2003; Takahashi, Wigley, and Walter 2005). This excess offers protection against potential fluctuations of MCM proteins, allowing replication to occur even in the event of an MCM level reduction (Lei, Kawasaki, and Tye 1996), or can activate dormant origins under conditions of replication stress (Woodward et al. 2006; Ibarra, Schwob, and Méndez 2008).

In a dividing cell, MCM proteins are present in all stages of the cell cycle and are strictly regulated at the steps of licensing, elongation and termination (Deegan and Diffley 2016; Hyrien 2016). In *S. cerevisiae*, free MCM2-7 proteins concentrate in the nucleus in G1, but are exported to the cytoplasm via activation of a nuclear export signal on Mcm3 outside of replication licensing, leaving only the assembled complexes already involved in the replication process and preventing re-licensing outside of G1 (Liku et al. 2005). As described above, metazoan MCM proteins are primarily mediated by Cdt1 regulation, with Geminin-facilitated inactivation of Cdt1 preventing over-replication of DNA.

In summary, eukaryotic DNA replication is controlled by an intricate network of cell-cycle dependent systems. While specific pathways may differ, phosphorylation-mediated regulation of the pre-RC components is conserved in eukaryotes, allowing segmentation of activity throughout the cell cycle. Bypassing these mechanisms has the potential to lead to over- or under-replication of the genome, resulting in genotoxic stress and potential cell death.

1.8 Pre-RC components and their relevance for human health

DNA replication needs to occur accurately to prevent the accumulation of mutations that can eventually lead to genomic instability, growth defects and uncontrolled cell cycle regulation. Due to the highly vital and conserved nature of the proteins involved in replication assembly, mutations affecting function are often lethal. Those that are not lethal, but still display aberrations in DNA licensing can have catastrophic consequences for the cell – including genomic rearrangements, altered genome dosage and promotion of tumorigenesis (Neelsen et al. 2013).

1.8.1 DNA Replication factors and Cancer

Early detection of cancers is paramount for efficient treatment, with changes in replication being one of the fundamental biomarkers of cancerous cells. It is a significant hallmark of multiple cancer types and pre-RC components have been demonstrated to be relevant markers for early detection of many common cancers.

In recent years, there has been evidence to show that ORC genes can be differentially expressed in several tumours, including lung, colon, liver, breast and brain cancers (Chen et al. 2021; Xiong et al. 2020; Liu et al. 2021; Gavin et al. 2008; Wang et al. 2020; Zhang et al. 2020; Han and Li. 2022). Some subunits of the complex – namely Orc1, Orc5 and Orc6 have been identified as having prognostic value in some cancers because of their differential expression in these diseases (Mao et al. 2020; Hu et al. 2019; Tang et al. 2023).

As well as this, MCM proteins are another key component in cancer diagnostics (Yu et al. 2020). While MCMs are expressed at high levels in proliferating cells, they are significantly downregulated in quiescent or senescent cells (Madine et al. 2000), making them a potential biomarker for clinical use in cancer detection (Gonzalez et al. 2005). Mutations or overexpression of one or more MCM subunits has been identified in various oncogenic tissues (Valverde et al. 2018; Rusiniak et al. 2012; Lau et al. 2010; Choy et al. 2016; Andreas et al. 2011). Upregulated MCM are often correlated with increased metastasis, malignancy and tumour sizes and decreased life expectancy (Yu et al. 2020; Wu et al. 2018; Guan et al. 2015).

Targeting helicase loading specifically in cancer cells has been proposed as a mechanism to curb uncontrolled replicative activity (Simon and Schwacha 2014; Seo and Kang 2018). To date, a small number of molecules have been identified as inhibitors of MCM activity, although these primarily focus on MCM expression (Seo and Kang 2018). Geminin, a small helical protein that prevents MCM loading by binding and inhibiting Cdt1 association with MCM has been identified as a potential druggable target to cease helicase loading (Saxena et al. 2004).

While pre-RC components have undoubted use as oncogenic biomarkers, studies are consistently aiming to identify ways they can be repurposed as therapeutic targets themselves. Development of assays to assess each step of assembly and test protein-protein interactions is paramount to pave the way to develop high-throughput screens of small molecules that can inhibit helicase loading and replication. For this, understanding the mechanism of assembly, activation, action and structural conformation and dynamics will be vital for improving drug design while limiting off-target applications.

1.8.2 Pre-RC components and human development – Meier-Gorlin Syndrome

Correct function of initiation and DNA replication factors is crucial for cell division and human development, with mutations significantly disrupting this causing impaired growth and tissue development. Developmental disorders such as dwarfism are often associated with deregulated cell cycle progression resulting in an overall reduced cell number (Klingseisen and Jackson 2011). One such disorder is Meier-Gorlin Syndrome (MGS), a rare autosomal recessive disorder considered a form of primordial dwarfism (Gorlin and Cervenka, 1975) Patients with MGS show a broad spectrum of phenotypic features, although the most prevalent characteristics are microtia, short stature and hypoplastic patellae (de Munnik et al. 2012), suggesting specific compromise in bone and cartilage development. Other skeletal abnormalities, difficulty in breathing and feeding problems are also associated with MGS, although patients do not display a reduction in life expectancy (de Munnik et al. 2012; Bicknell et al. 2011; Cohen et al. 2002; Bongers et al. 2005).

Genetic mutations observed in patients with MGS are strongly associated with helicase loading factors. Approximately 75% of patients with MGS have mutations in Orc1, Orc4, Orc6, Cdc6 or Cdt1, with most cases containing missense mutations in Orc1 (Bicknell et al. 2011; Guernsey et al. 2011; Bongers et al. 2001). Hypomorphic mutations affecting the Orc1-Orc4 or Orc1-Cdc6 ATPase sites as well as the Orc1-BAH domain have been found to have the most significant effect on the observed disease phenotype (Kuo et al. 2012; De Ioannes et al. 2019). In zebrafish mutants, Orc1-BAH domain perturbations result in an MGS-like phenotype that can be rescued by introduction of wild-type Orc1, but not Orc1-BAH mutants (Kuo et al. 2012). However, there is still no detailed characterisation of MGS associated pre-RC protein mutations and how they affect licensing and firing both in vitro and in vivo. Further scientific efforts are needed to understand how and where in the licensing mechanism these mutations have an effect and why cartilage and tissue development are specifically affected.

1.9 A note on *in vitro* methods to study helicase loading.

Loading helicases onto DNA involves rapid transitions between multiprotein intermediates with distinct protein-protein and protein-DNA interactions.

Bulk biochemical assays reflect the average of multiple asynchronous events, requiring a multi-step and time-heavy process to remove proteins not involved in associating with DNA. While this removes the temporal resolution of protein binding, release and rearrangements that typically occur within seconds or fractions of a second, it does allow the identification of stable, long-lived intermediates. This is particularly valuable for gaining high-resolution structural data, identifying incredible detail on inter- and intra- molecular interactions.

Many of these biochemical assays have been developed to study pre-RC formation in *S. cerevisiae* (Evrin et al. 2009; Remus et al. 2009) and more recently, *D. melanogaster* (Schmidt and Bleichert 2020). This has facilitated the insights gained on helicase loading intermediates in these organisms, supported by high-resolution structural data of these intermediates (Miller et al. 2019; Riera et al. 2017; Fernández-Cid et al. 2013) (Figure 12). More recently, parallel single-molecule studies of *S. cerevisiae* helicase loading have elucidated the temporal dynamics between trapped intermediates, allowing for a deeper mechanistic understanding of helicase loading on both a structural and time-resolved scale (Sánchez et al. 2021; Sánchez et al. 2023).

While these studies have been invaluable for setting the foundation to understand the helicase loading pathway, scientific efforts in studying the same process in humans has only recently become viable. Complexities in expression and sample preparation using wild-type *H. sapiens* proteins has resulted in the field still lacking much information on the similarities and differences involved in the human pre-RC loading pathway. There have been recent structural studies elucidating the structural characteristics of human ORC and MCM proteins (Tocilj et al. 2017; Jaremko et al. 2020; Xu et al. 2022) (Figure 10, 12), but a biochemical assay to study

the assembly of the human pre-RC is yet to be published, as are paralogous structural intermediates identified in other organisms.

1.10 Transmission Electron Microscopy: The resolution revolution and its implications in structural biology

The ability to visualise the 3D structure of proteins has become fundamental to understanding how they function. With ongoing technological advancements, structural biology has been able to tackle complex inter- and intra-molecular interactions to support the mechanisms of action of macromolecules at near-atomic resolution.

For decades, the determination of macromolecular structures was dominated by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. While allowing near-atomic resolution, crystallisation of samples can prove challenging – in many cases, impossible (Callaway 2015). Both techniques are only suited to molecules with a relatively low molecular weight, typically below 100 kDa, making them a poor choice for large macromolecular complexes (Quentin and Raunser 2018). Electron microscopy (EM) has provided an alternative for solving the structure of large multi-protein assemblies, with advancements in sample preparation, detection and analysis allowing a relatively quick turnover time, without the need for well-diffracting crystals or limitations on molecule size. The development of EM and its value in visualisation of biological molecules earned Jacques Dubochet, Joachim Frank and Richard Henderson the 2017 Nobel Prize in Chemistry (Shen 2018)

1.10.1 Techniques in Electron Microscopy

Briefly, in transmission EM, monodispersed macromolecules are imaged by bombardment with an electron beam. Prepared samples on a supporting grid are inserted into a transmission electron microscope, where an electron gun (either a thermoionic electron source or a field emission gun) fires electrons that interact with the atoms of the specimen, resulting in electron

scattering. Scattered electrons are recorded using a detector, producing 2D images (micrographs) of density (Figure. 14). Picking and averaging of electron densities allows the production of 2D images of different projections of the molecule, which can be used to construct a 3D model.

1.10.1.1 Negative Stain EM

Staining electron microscopy allows for the relatively quick and simple visualisation of a sample, achieved using a contrast enhancing stain reagent. Usually, this is achieved by coating the specimen with heavy metal salts commonly derived from molybdenum, uranium or tungsten. These heavy ions react with the electron beam, allowing greater phase contract and a higher signal-to-noise ratio (SNR). In most cases, the stain is excluded from the molecule itself, giving rise to the term 'negative staining'. While this method provides relatively quick and easy visualisation, major downfalls include harsh conditions of the stain resulting in protein denaturation and a total resolution limited by the grain size of the stain (typically a maximum resolution of 15 – 20 Å (Scarff et al. 2018; Ohi et al. 2004) (Figure 14.A.)

1.10.1.2 Cryogenic EM

Unlike NS EM, cryogenic EM (Cryo-EM), specimen immobilisation involves vitrification, achieved by rapidly plunge-freezing of the sample into a cryogenic fluid, typically liquid ethane. The low melting point (-188°C) and high heat capacity properties of liquid ethane promotes the formation of vitreous ice, preventing crystal formation and maintaining an aqueous environment. This rapid freezing process leaves the sample in an as-near-as-possible native state, fixed at different orientation. The lack of stain in cryo-EM reduces the SNR, making particles more difficult to identify with the naked eye but ultimately removing the resolution limits imposed by staining EM (Dubochet and McDowall 1981) (Figure 14.A.).

1.10.1.3 Single Particle Analysis (SPA)

SPA involves the extraction of individual particles in various orientations from micrographs, either manually or automatically. Aligned particles can be sorted into 2-dimensional classes where an average projection representing all particles in each class is derived. Averaged 2D projections of different angular orientations can then be used to reconstruct a 3D model by calculating the angular relationships between classes (Lyumkis 2019; Thompson et al. 2016). Over the years, data processing software has been developed using different statistical models for particle reconstruction, requiring different levels of user input. These include, but are not limited to EMAN (Ludtke, Baldwin, and Chiu 1999), RELION (Scheres 2012), CryoSPARC (Punjani et al. 2017) and cisTEM (Grant, Rohou, and Grigorieff 2018). An overview of the single particle data collection and analysis workflow is presented in Figure 14.B-C.)

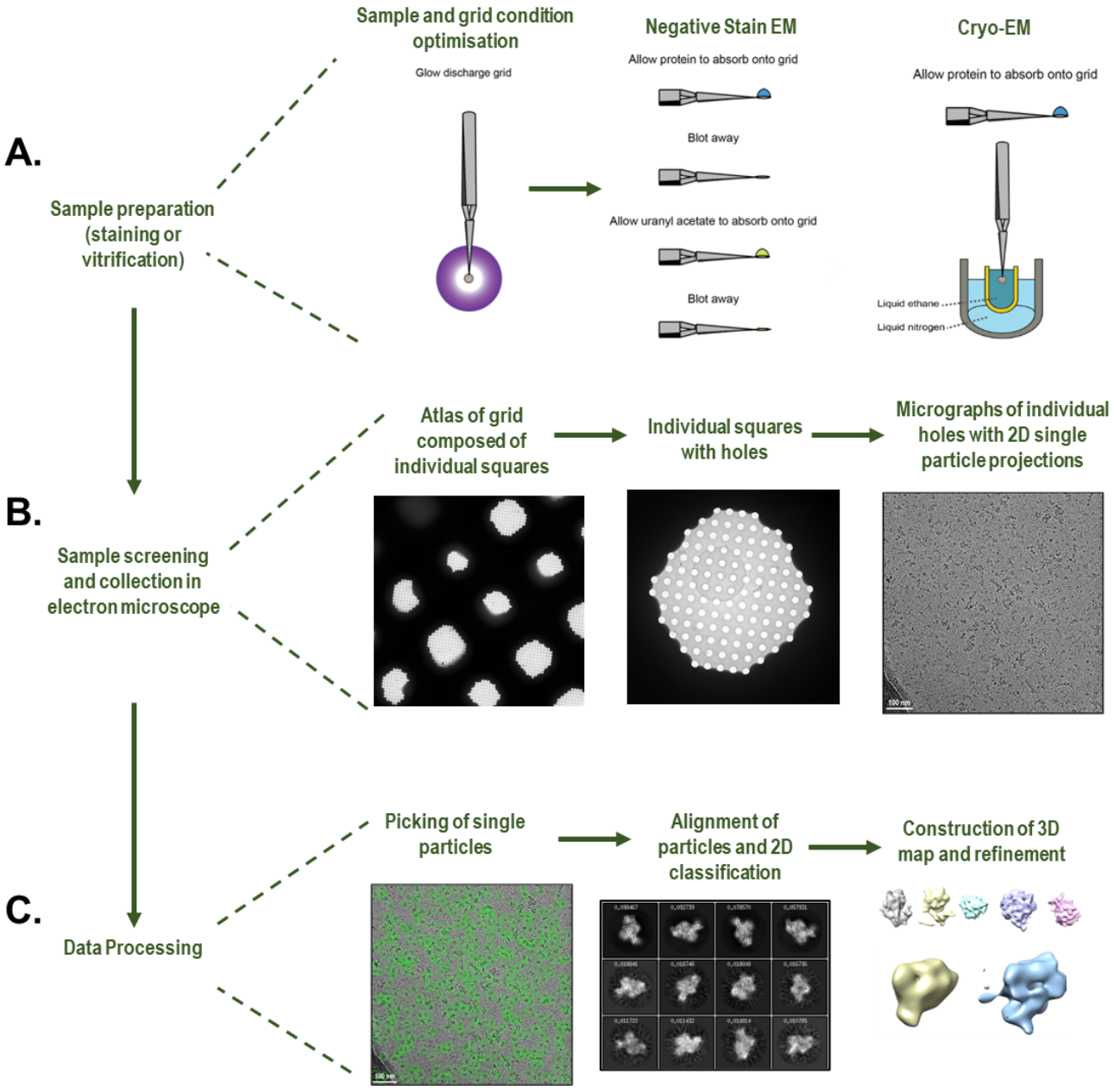


Figure 14: Single-Particle Electron Microscopy: Sample preparation, Data Collection and Analysis

*Overview of the **A**: sample preparation of protein-DNA complexes being applied to grids for staining or vitrification, **B:** screening and collection of grids on an electron microscope, with representative micrographs generated and **C**: data processing pipelines for single-particle cryo-EM experiments as carried out in this work. The sample screening and data processing figures are from the results presented in this thesis.*

1.10.2 Challenges in Structural Determination by EM

The choice of EM holds multiple advantages in structural biology in comparison to NMR and crystallography. Molecules can be studied in more native conditions, are not limited by their size and the technique typically requires comparatively lower quantities of sample (usually in the nanomolar range). To date, resolutions up to 1.2 Å have been reached (Nakane et al. 2020), with further improvements expected as new sample preparation, detection and image processing methods are developed.

However, electron microscopy still comes with some pitfalls. It is still near impossible to achieve high resolution snapshots of dynamic or highly disordered protein regions, many of which are critical for protein functionality. As with all structural biology techniques, the sample preparation process requires extensive optimisation, often unique to each specimen and its inherent physical properties. Variations in solubility, flexibility and stability of proteins will affect sample preparation and data processing. Specimens may have preferred orientations when applied to the grids and either stained or frozen, resulting in interactions and potential molecular disruptions at the air-water interface (D'Imprima et al. 2019). Grids themselves may have disparities in ice or stain thickness, compromising the quality of collected images (Lyumkis 2019). Sample-specific challenges encountered during this project will be discussed in more detail in following chapters.

1.10.3 Electron Microscopy and helicase loading

As mentioned in previous chapters, EM has been an invaluable tool for studying the intermediates involved in the assembly of the eukaryotic helicase. It has allowed the identification and high-resolution identification of key stages of the process that have the potential to be targeted to modulate replication initiation. Most structural studies have used *S. cerevisiae* or *D. melanogaster* as model systems – the focus now is on producing an equivalent platform using human proteins that will provide a benchmark mechanism that can

be exploited downstream for its therapeutic potential in targeting aberrant DNA replication processes.

1.11 Project aims.

DNA replication initiation and helicase loading has long been a focus of research, mostly focussing on the *S. cerevisiae* system. While these studies have undoubtedly been valuable, equivalent studies for the metazoan system of replication of initiation have not advanced as quicky. Crucially, there has yet to be a model assay to study helicase loading using human proteins – mostly down to the difficulties in producing high quality, functional preparations of human pre-RC proteins. However, establishment of reproducible, high-quality protein purification methods has the potential to advance studies to assess helicase loading, identify key intermediates in the pathway that could serve as potential drug targets and understand how aberrant replication machinery loading influences genome integrity.

Purification of Full-length human pre-RC proteins

My first aim is to establish a reproducible method of purifying full-length Orc1-5 and Cdc6. While studies have managed to purify and structurally characterise human ORC and Cdc6, often these are highly modified or truncated to allow for protein stability. The production of full-length proteins will allow for greater insight into protein behaviour *in vivo*. The DNA Replication Group has previously established methods of purifying other pre-RC components, as well as developing a helicase loading assay, which will also be used to study human pre-RC assembly.

Electron microscopy of intermediates

Electron microscopy has been key in identifying individual intermediate in the helicase loading pathway. Using purified human proteins, I aim to study these intermediates on a structural level, starting from human ORC binding to DNA, then recruiting Cdc6. Merging biochemical

studies with structural information has the potential to identify the key steps in the process of origin recognition and pre-RC recruitment in humans.

ORC and chromatin

The final aim is to increase the understanding of how the entire ORC structure interacts with nucleosomes, as this can explain how origins are specified. ORC must be distributed across the genome – including nucleosome-dense regions, to ensure complete replication during S-phase. Although current data suggests that ORC does not possess an intrinsic DNA sequence specificity, multiple protein-protein interactions between ORC and chromatin/chromatin associated factors are likely utilised to stabilise ORC binding across the genome to regulate origin licensing. Biochemical and structural approaches will be used to study ORC's interaction with nucleosomes with specific epigenetic modifications and chromatin-associated factors, such as HP1.

Chapter 2: Materials & Methods

2.1 Materials

2.1.1 Plasmids

- See Supplementary Table 1

2.1.2 PCR primers/ Oligonucleotides

- See Supplementary Table 3

2.1.3 Bacterial Strains

See Supplementary Table 5

2.2.4 Yeast Strains

- See Supplementary Table 4

2.2.5 Media

Media recipes provided in this section were produced by the MRC-LMS media kitchen. Media specific to an experiment will be provided in the appropriate sections of the methodologies.

LB (1L)	
5 g	Yeast Extract
10 g	Tryptone
10 g	NaCl

LB + agar (for plates - 1L)	
5 g	Yeast Extract
10 g	Tryptone
10 g	NaCl
20 g	Agar

2.2 Methods

2.2.1 Protein Expressions and Purifications

Note: These protocols represent optimised methods – the optimisation process will be discussed in greater detail in the Results section. Any optimisation/experimental credit will be explicitly stated. Wild-type and mutant strains of proteins were expressed and purified in the same manner, unless otherwise stated.

2.2.1.1 Orc1-5

2.2.1.1.1 Transformation & Expression

Competent *S. cerevisiae* YC658 cells (MATa, lys2::pGAL1 GAL4::LYS2, pep4::HIS3, bar1::hisG derived from W30) (Matheson, Parsons, and Gammie 2017) were used to co-express codon-optimised human ORC subunits 1-5 under control of a galactose inducible promoter.

300 ng of each plasmid (Supplementary Table 1 – PLAS1, PLAS2, PLAS3) were added to 30 µl of chemically competent YC658 cells and mixed gently. 180 µl of PEG solution was added, then solution was left to incubate for 30 minutes at room temperature. Following this, DMSO was added to a final concentration of 10 % (v/v) and mixed well, then cells were heat shocked at 42 °C for 15 minutes. Cells were centrifuged at 2000 rpm in a benchtop centrifuge for 3 minutes, then cell pellets were resuspended in distilled water and spread on selective agar plates. Plates were grown for three days at 30 °C. Selected colonies were then grown at 30 °C with shaking at 220 rpm overnight in 50 mL of Selective Growth media, until an OD600 of 2-4 was reached. For storage, cells were aliquoted and frozen at -80 °C in a 1:1 volumetric ratio of culture:glycerol.

Selective Plates for *S. cerevisiae* (1L)	
6.75 g	Yeast Nitrogen Base (without amino acids) (ForMedium)
20 g	Glucose
0.6 g	Amino Acid mix (-His, -Leu, -Trp, -Ura) (ForMedium)
20 g	Agar
to p.H. 5.5	HCl
to 1L	ddH2O

Selective Growth Media (Galactose Induction) (1L)	
6.9 g	Yeast Nitrogen Base (without amino acids) (ForMedium CYN0405)
20 mg	Histidine
0.6 g	Amino Acid mix (-His, -Leu, -Trp, -Ura) (COMMERCIAL INFORMATION)
5 g	Glucose
30 g	Glycerol
2 % (w/v)	Lactic Acid (85%)
To p.H. 6.8	NaOH
to 1L	ddH2O

Rich Media (diluting cells) (1L)	
10 g	Yeast extract (Bacto™ ThermoScientific)
20 g	Peptone (Bacto™ ThermoScientific)
0.1 g	Glucose
30 g	Glycerol
2 % (w/v)	Lactic Acid (85%)
To p.H. 5.5	NaOH
to 1L	ddH2O

Lysis Buffer (for harvesting cells)	
50 mM	HEPES-NaOH (p.H. 7.6)
200 mM	NaCl
2 mM	DTT
20 mM	NaF

2 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.02 % (w/v)	NP-40
1 mM	ATP

For expression, cells were diluted and grown overnight as described in a larger volume of selective media (6 – 12 L) to an OD600 between 2-4, then diluted with equal volumes of rich media. Alpha-factor (α-factor) was added to a final concentration of 5 µg/mL to arrest cells in G1. Cells were left to adjust for 1.5 – 2 hours before protein expression was induced by addition of 2% galactose (w/v). Cells were left to express for 4 hours, then harvested by centrifugation at 3,000 g at 4 °C for 10 minutes in a Sorvall Lynx 6000 centrifuge (ThermoScientific) fitted with a precooled F9-6x1000 LEX rotor (ThermoScientific. Pellets were washed and resuspended in 10 mL of Lysis Buffer for harvesting and re-centrifuged in benchtop centrifuge (Eppendorf). Pellets were scraped and pooled into 20 mL syringes and syringed directly into a clean plastic bowl of liquid nitrogen to form snap-frozen 'noodles,' then stored at -80 °C until use.

2.2.1.1.2 Purification

2.2.1.1.2.1 Lysis

Frozen noodles were ground into a powder using a SPEX™ freezer-mill cooled by liquid nitrogen. ~40 g of powder equivalent to ~6 L of culture was used per purification. Powder was resuspended and homogenised in Lysis Buffer supplemented with 12 µL benzonase (Merck), cOmplete EDTA-free protease inhibitor tablets (Roche) and PhosSTOP phosphatase inhibitor cocktail tablets (Roche) for 15 – 30 minutes. Lysate was then centrifuged for 50 minutes at 18,000 rpm on an acceleration gradient (Sorvall Lynx 6000 centrifuge (ThermoScientific) fitted with a precooled F21-8x50Y rotor (ThermoScientific); resulting supernatant was filtered twice using 5 µm and 0.45 µm filters, respectively.

hORC1-5 Lysis Buffer	
50 mM	HEPES-NaOH (p.H. 7.6)
250 mM	KCl
50 mM	NaCl
2 mM	DTT
5 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.2 % (w/v)	NP-40
10 mM	B-glyceroPO4

2.2.1.1.2 .2 StrepXT – Affinity

Filtered lysate was loaded onto a 5 mL StrepXT column (Cytiva) pre-equilibrated with Lysis Buffer using a peristaltic pump (Pharmacia LKB P1). Lysate was loaded into the column, then the column was washed with 5x column volumes of lysis buffer. Bound protein was eluted by the addition of 75 mM biotin, suspended in Lysis buffer. Eluate was collected in fractions. Samples at each stage were run on SDS-PAGE gels and Coomassie stained to identify fractions containing the desired protein. Appropriate fractions were pooled, and an additional 2 mM $MgCl_2$ and 1 mM ATP was added, alongside 15 units SUMOStar protease (LifeSensors) to cleave the Orc1 N-terminal 2x Sumo-Strep tag. The sample was left to cleave overnight (12 – 16 hours) on a rotating wheel at 4 ^{o}C.

2.2.1.1.2 .3 HiTrap Heparin – Ion Exchange

A 1 mL HiTrap Heparin column was equilibrated with Low Salt Buffer on an AKTA Pure (Cytiva). Sample was diluted to 200 mM NaCl using No-salt lysis buffer and injected onto the column using a 50 mL superloop (Cytiva). The column was then washed with 10x column volumes of Low Salt Buffer, then protein was eluted with a salt gradient up to 1 M NaCl in 15 minutes at a flow rate of 1 mL/min. Fractions were collected in 0.5 mL volumes. UV absorbance at 230, 260 and 280 nm was tracked throughout the process. 10 µl samples of column flow-through, washes and predicted protein fractions were run on SDS-PAGE gels, then Coomassie stained to identify fractions containing the desired protein. These were

pooled, then concentrated using a 4 mL Amicon® Ultra Centrifugal Filter (50 kDa MWCO) to a maximum concentration of ~30 µM (~10 mg/mL). Pooled and concentrated samples were aliquoted and either subject to further purification by density gradient centrifugation or snap-frozen in liquid nitrogen and stored at –80°C.

Low Salt Buffer (hORC1-5 purification)	
50 mM	HEPES-NaOH (p.H. 7.6)
200 mM	NaCl
1.5 mM	DTT
25 mM	NaF
5 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.02 % (w/v)	NP-40

High Salt Buffer (hORC1-5 purification)	
50 mM	HEPES-NaOH (p.H. 7.6)
1000 mM	NaCl
1 mM	DTT
20 mM	NaF
5 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.02 % (w/v)	NP-40

2.2.1.1.2 .4 Density Gradient Centrifugation

Glycerol density gradients were prepared manually, (see Chapter 2..2.5 for more details) by injection of a high-density buffer underneath an equal volume of low-density buffer. 400 µl of low-density buffer (10% glycerol) was added to the bottom of a (INFORMATION) 0.8 mL tube. Using a syringe and blunt-tip needle, 400 µl of high-density buffer (30% glycerol) was applied

underneath the low-density buffer. Gradients were left to form and settle by upright incubation at 4 ºC for >2 hours.

Low/High Glycerol Density Gradient Buffer (hORC1-5 purification)	
20 mM	HEPES-NaOH (p.H. 7.6)
300 mM	NaCl
1 mM	DTT
2 mM	$MgCl_2$
1 mM	ATP
10/30 % (w/v)	Glycerol

40 µl was removed from the top of the formed gradients and replaced with 40 µl of protein sample dropwise, taking care not to allow mixing into the gradient. Centrifuge tubes were placed into pre-cooled swing buckets fitted with appropriate adaptors. Gradients were run at 29,000 rpm at 4 ºC for 16 hours in a precooled Optima™ L-80 XP ultracentrifuge (Beckman Coulter) using a SW55 Ti rotor (Beckman).). Acceleration and deceleration values were set to 3.

Gradients were fractionated manually - 32 µl fractions were collected and any pellets collected at the bottom of the tubes were resuspended in 10 µl of high-density gradient buffer. Fractions were run on an SDS-PAGE and Coomassie stained, with fractions containing stochiometric protein pooled, concentrated and buffer -exchanged into low-density buffer using a 0.5 mL Amicon® Ultra Centrifugal Filter (50 kDa MWCO). Protein was concentrated to between 3 – 9 µM, then aliquoted into appropriate volumes. Aliquots were snap-frozen in liquid nitrogen and stored at –80 ºC.

2.2.1.2 Cdc6

Codon-optimisation and initial protein expression/purification methods were designed by Dr. Vera Leber (DNA Replication Group). Purification methods were optimised by myself and Lucy Edwardes (DNA Replication Group).

2.2.1.2 .1 Transformation & Expression

50 ng of codon-optimised plasmid (Supplementary Table 1 – PLAS10) were added to ice-thawed competent BL21 Codon+ E. *coli* cells (New England Biolabs) and gently mixed. The mixture was incubated on ice for 30 minutes, then heat shocked at 42 ºC for 45 seconds. Sample was then placed back on ice for 2 minutes, then mixed with 600 µl of pre-warmed S.O.C media (ThermoFisher Scientific) and incubated at 37 ºC for 45 minutes – 1 hour with shaking at 300 rpm in a thermomixer (Eppendorf). 50 – 100 µl was spread onto an LB-agar selection plate containing 300 µM Ampicillin (ThermoFisher Scientific). Plates were incubated overnight at 37 ºC.

Individual colonies from fresh transformations were picked using a sterile pipette tip and grown overnight in 5 mL of LB containing 300 µM Amp at 37 ºC with shaking at 200 rpm in an LTX incubator shaker (Kuhner). The following morning, cells were diluted to an OD600 of 0.15 with LB + Amp and further grown with shaking at 200 rpm. Once the culture reached an OD600 of 2 (~3.5 hours), it was used to inoculate 5 L of LB + Amp to a final OD600 of 0.05. Large cultures (6 – 12L) were grown in a shaking incubator (INFORS) at 37 ºC with shaking at 200 rpm until an OD600 of 0.2-0.4 was reached. Cultures were then relocated to a 16 ºC shaker and allowed to cool for ~1 hour. Cells were then induced with 0.5 mM IPTG and incubated overnight at 16 ºC with shaking at 200 rpm. The following morning, cells were spun down at 4000 rpm, 4 ºC for 10 minutes in a Sorvall LYNX 6000 centrifuge fitted with a precooled F9-6x1000 LEX rotor (ThermoScientific). Cell pellets were resuspended in Lysis buffer, transferred into falcon tubes and re-centrifuged as above. Clean pellets were snap frozen in liquid nitrogen and stored at -80 ºC until further use.

Lysis Buffer (Cdc6 purification)	
50 mM	HEPES-NaOH (p.H. 7.6)
250 mM	KCl
50 mM	NaCl
2 mM	DTT
5 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.2 % (w/v)	NP-40

2.2.1.2.2 Purification

2.2.1.2.2.1 Lysis

Cell pellets correlating to 5 L of culture were resuspended in 60 mL lysis buffer supplemented with cOmplete EDTA-free protease inhibitor tablets (Roche) and 15 units benzonase (Merck). Lysate was sonicated 7 x 30 seconds at an amplitude of 20% using a Vibra-Cell™ sonicator then centrifuged for 50 minutes at 18,000 rpm, 4 °C in a Sorvall LYNX 6000 centrifuge fitted with a precooled F21-8x50y LEX rotor (ThermoScientific). Supernatant was filtered using a 5 µm syringe filter.

2.2.1.2.2.2 GST purification - Affinity

Per 5 L of culture, 1 g of glutathione agarose beads (10 mL of 100% resin) (Pierce™) was washed in 50 mL ddH2O for at least 2 hours. Beads were then washed with 2x 50 mL lysis buffer rinses using a gravity flow column (BioRad). Once washed, the lysate was incubated with the beads for 1 hour on a rotating wheel. Following this, the beads were allowed to settle and the lysate flow-through was collected. Beads were then washed twice with 50 mL of lysis buffer, then 50 mL of high-salt buffer and Cleavage Buffer supplemented with 0.5 mM ATP respectively, and finally transferred to a smaller gravity flow column. The sample was resuspended in Cleavage Buffer, supplemented with PreScission protease (homemade) and left to cleave overnight (12 – 16 hours) on a rotating wheel at 4 °C. Cleaved eluate was

collected, and the beads were washed twice with Cleavage Buffer. Samples at all stages were run on SDS-PAGE gels and Coomassie stained to identify fractions containing desired protein. Fractions of interest were pooled and filtered using a 0.22 µm syringe filter for further purification.

High Salt Wash (for GST beads)	
50 mM	HEPES-NaOH (p.H. 7.6)
1000 mM	KCl
50 mM	NaCl
1 mM	DTT
5 % (w/v)	Glycerol
2 mM	$MgCl_2$
0.05 % (w/v)	NP-40

Cleavage Buffer (Cdc6 Purification)	
30 mM	HEPES-NaOH (p.H. 7.6)
167 mM	KCl
33.2 mM	NaCl
1 mM	DTT
5 % (w/v)	Glycerol
1 mM	$MgCl_2$
0.05 % (w/v)	NP-40

2.2.1.2.2.1.3 HiTrap SP HP – Ion exchange

A 1 mL HiTrap SP HP column (Cytiva) was washed and equilibrated with Cleavage Buffer on an AKTA Pure (Cytiva). Post-filtration sample was injected onto the column using a 50 mL superloop (Cytiva). The column was then washed with 10x column volumes of Cleavage Buffer, then protein was eluted on a salt gradient into Elution Buffer in 15 mL at a flow rate of 1 mL/min. Fractions were collected in 0.3 mL volumes. UV absorbance at 230, 260 and 280 nm was tracked throughout the process. 10 µl samples of column flow-through, washes and predicted protein fractions were run on SDS-PAGE gels, then Coomassie stained to identify

fractions containing the desired protein. These were pooled, then concentrated and buffer exchanged to a maximum concentration of ~10 µM (~0.6 mg/mL) in Cleavage buffer supplemented with 1 µM ATP using a 4 mL Amicon® Ultra Centrifugal Filter (50 kDa MWCO). Pooled and concentrated samples were aliquoted and snap-frozen in liquid nitrogen and stored at –80°C.

Elution Buffer (Cdc6 Purification)	
30 mM	HEPES-NaOH (p.H. 7.6)
825 mM	KCl
165 mM	NaCl
1 mM	DTT
5 % (w/v)	Glycerol
1 mM	$MgCl_2$
0.05 % (w/v)	NP-40

2.2.1.3 Orc6, Cdt1 & MCM

The expression and purification of Cdt1 and MCM2-7 will not be discussed in detail, as these were primarily done by other members in the lab. Protocols for transformation, expression and purification were optimised by Dr. Vera Leber; communal lab stocks of these proteins were produced by Vera Leber, Jennifer Wells, Joshua Tomkins and Matthew Peach (DNA Replication Group) respectively.

Orc6

Orc6 was expressed as an individual subunit in E. coli with an N-terminal His-tag, inducible with IPTG. The protein was isolated by lysing cells, a HisTrap Excel column (Cytiva) overnight cleavage of the His-tag and finally ion-exchange purification with a Poros HQ column (ThermoFisher Scientific)

Cdt1

Cdt1 was expressed in E. coli with an N-terminal GST tag, inducible with IPTG. Lysed cells were loaded onto GST beads (Pierce™), washed, then bound protein was cleaved with Precision Protease overnight, then underwent ion exchange purification with a Poros HS column.

MCM2-7

Expression of MCM2-7 as a hexamer in HEK293F cells with an N-term His-tag on Mcm4 and N-term 2xStrep-tag on Mcm2 was performed by Oxford Expression Technologies (OET). Purifications were performed in-house by lysing pellets and performing two rounds of affinity chromatography – first with NiNTA beads (ThermoFisher Scientific), then with a StrepXT column (Cytiva).

2.2.1.4 HP1

All isoforms of HP1 (α, β and γ) were expressed and purified in the same way. Plasmid constructs and purification methods were designed by Noelia Fernandez-Rivero.

2.2.1.4.1 Transformation & Expression

50 ng of codon-optimised plasmid (Supplementary Table 1 – PLAS17/18/19) were added to ice-thawed BL21 E. coli cells (New England Biolabs) and gently mixed. The mixture was incubated on ice for 30 minutes, then heat shocked at 42 °C for 45 seconds. Sample was then placed back on ice for 2 minutes, then mixed with 600 ul of pre-warmed S.O.C media (ThermoFisher Scientific) and incubated at 37 °C for 45 minutes – 1 hour with shaking at 300 rpm in a thermomixer (Eppendorf). 50 – 100 µl was spread onto an LB-agar selection plate containing 300 µM Ampicillin. Plates were incubated overnight at 37 °C. Single colonies were then selected and grown overnight at 37 °C with 200 rpm shaking (INFORS) in 50 mL LB

supplemented with Ampicillin. The following morning, cells were diluted into 1 L of LB supplemented with Ampicillin and grown at 37 ºC with 200 rpm shaking until an OD600 of 0.3. Temperature of the shaking incubator was decreased to 30 ºC until culture reached an OD600 of 0.6, then protein expression was induced by addition of 0.5 mM IPTG for 3 hours.

Cells were spun down at 4000 rpm, 4 ºC for 10 minutes in a Sorvall LYNX 6000 centrifuge fitted with a precooled F21-8x50y LEX rotor. Cell pellets were resuspended in Lysis buffer supplemented with cOmplete protease inhibitor tablets (Roche), transferred into falcon tubes and re-centrifuged as above. Clean pellets were snap frozen in liquid nitrogen and stored at -80 ºC until further use.

Lysis Buffer (HP1 purification)	
	PBS (p.H. 7.4) (137 mM NaCl, 2.7 mM KCl, 8 mM Na2HPO4, and 2 mM KH2PO4)
150 mM	NaCl
2 mM	DTT

2.2.1.4.2 Purification

2.2.1.4.2.1 Lysis

Frozen cell pellets equivalent to 50 mL of culture were thawed on ice, then resuspended in 20 mL of Lysis buffer supplemented with cOmplete protease inhibitor tablets (Roche) and lysozyme to a final concentration of 0.1 mg/ml. Resuspended cells were incubated for 30 minutes on ice, then sonicated at a 20% amplitude in 5 second bursts for a total of 2 minutes. Supernatant was then centrifuged at 38,800 g at 4 ºC for 30 minutes (centrifuge/rotor info).

2.2.1.4.2.2 GST-agarose – affinity

GST-agarose beads were washed three times with 10x volumes of lysis buffer. Sample supernatant was then applied to the beads and incubated at 4 ºC for 2 hours on a rotating wheel. Sample was then transferred to a gravity flow column (BioRad) and the flow-through

collected. Beads were washed twice with Wash Buffer, then rinsed three times with Cleavage buffer. Beads were then transferred to a clean, sealed gravity flow column and resuspended in 20 mL of Cleavage buffer supplemented with 150 µl PreScission protease (homemade), then incubated briefly at 4 °C on a rotating wheel. Elutions containing the soluble protein were collected in fractions, then were run on an SDS-PAGE gel and Coomassie stained. Desired fractions were pooled and concentrated using a 4 mL Amicon® Ultra Centrifugal Filter (10 kDa MWCO).

Wash Buffer (HP1 purification)	
	PBS (p.H. 7.4) (137 mM NaCl, 2.7 mM KCl, 8 mM Na2HPO4, and 2 mM KH2PO4)
150 mM	NaCl
1 mM	DTT
5 % (w/v)	Glycerol
0.2 % (w/v)	Triton X-100

Cleavage Buffer (HP1 purification)	
	PBS (p.H. 7.4) (137 mM NaCl, 2.7 mM KCl, 8 mM Na2HPO4, and 2 mM KH2PO4)
1 mM	DTT
5 % (w/v)	Glycerol

2.2.1.4.2.3 Size Exclusion Chromatography

A Superose 6 Increase 10/30 column (Cytiva) was washed and equilibrated with Gel Filtration Buffer on an AKTA Pure (Cytiva). Concentrated protein was loaded via a suitably sized injection loop and run through the column at 0.5 mL/min, collecting 1 mL fractions. UV absorbance at 230, 260 and 280 nm was tracked throughout the process. Protein peaks were run on SDS-PAGE gels and Coomassie stained, then fractions containing pure protein were concentrated using a 0.5 mL Amicon® Ultra Centrifugal Filter (10 kDa MWCO) to a final

concentration of ~25 μM. Concentrated protein was aliquoted, snap frozen in liquid nitrogen and stored at –80 °C.

Gel Filtration Buffer	
	PBS (p.H. 7.4) (137 mM NaCl, 2.7 mM KCl, 8 mM Na2HPO4, and 2 mM KH2PO4)
5 % (w/v)	Glycerol

2.2.2 Selection of DNA, amplification and purification.

2.2.2.1 Selection of DNA sequences

DNA sequences to be used to study pre-RC component binding to DNA were selected and designed according to published ChIP-Seq data, focussing on sites showing strong Orc1 (Dellino et al. 2013) or Mcm7 (Sugimoto et al. 2018) peaks across the human genome. 164-bp long segments were picked, with varying AT/GC content; sites were also picked that were in proximity to nucleosomes (Schwartz et al. 2019), for downstream pre-RC nucleosome studies (Figure 15). Sequences selected, their corresponding AT/GC content and genomic location can be found in Supplementary Table 2. These sequences were then fused to the 147-bp Widom-601 sequence, a published high-affinity nucleosome binding sequence (Chua et al. 2012) for nucleosome interactor studies. Relevant primers relating to either the chosen sequence alone or the sequence + Widom-601 were designed for amplification of the appropriate sections of the DNA.

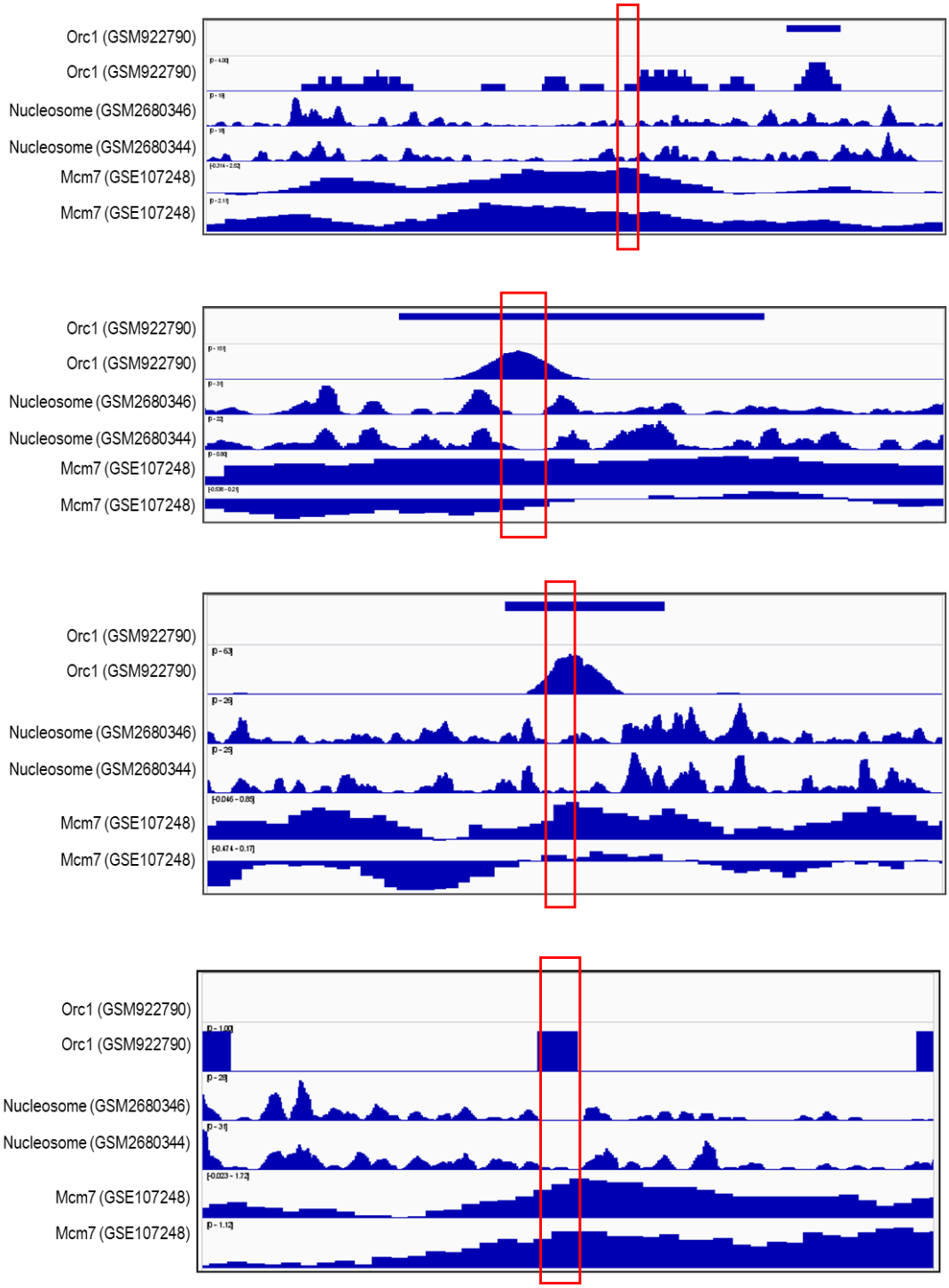


Figure 15: Selection of DNA sequences to study ORC binding and pre-RC assembly

IGV (integrated Genome Viewer) peaks correlating to picked sequences showing strong Orc1 (GSM922790, Dellino et al, 2013) and Mcm7 (GSE107248, Sugimoto et al, 2018) binding, situated between nucleosomes (GSM2680346, GSM2680344, Schwartz et al, 2019). Precise sequences and genome locations are detailed in Supplementary Table 2.

2.2.2.2 DNA amplification and purification

Selected sequences as specified in Supplementary Table 2 were synthesised and cloned into pUC57 plasmid vectors via an EcoRI restriction site by Genscript Biotech Corp.

2.2.2.2.1 Plasmid amplification

Genscript-synthesised plasmids resuspended to a volume equalling roughly 1-3 µg/µl. Chemically competent E. coli DH5α cells (Invitrogen) were used for standard plasmid amplifications. Cells were transformed as described in Section 2.2.1.2.1, plated in a sterile hood on LB-agar plates containing ampicillin then grown overnight at 37 ºC. Colonies were individually selected the following morning and added to 5 mL of liquid LB containing appropriate selective antibiotics. Cells were grown either overnight or grown for ~8 hours, then the culture was transferred to a larger volume (100 – 250 mL), depending on the desired yield of plasmid DNA. Cells were incubated at 37oC with shaking at 200 rpm in an Excella (Eppendorf) shaker.

Once grown to a density of approximately OD600 2, cells were harvested at 4 ºC by centrifugation at 4000 rpm in a benchtop centrifuge. Supernatant was discarded and the plasmid DNA was extracted from pelleted cells using the QIAprep™ Miniprep or Midiprep kits (QIAGEN) depending on culture volume, according to manufacturer's instructions. Elution of plasmid DNA from kit columns was carried out in commercial EB buffer (10 mM Tris-HCl pH 8).

DNA concentration was determined using a NanoDrop Spectrophotometer (ThermoFisher Scientific), using 1.5 - 2 µl of eluate for each reading. Readings for each sample were taken three times and averaged for a final stock concentration. Between 300 – 500 ng of DNA was run on a 0.75 % agarose gel to determine the integrity and purity of the final sample.

2.2.2.2.2 DNA oligonucleotide (100-400bp) preparation

All small DNA fragments with a length between 100 – 400 bp were amplified by PCR using Pfu polymerase (Agilent) or Q5 polymerase (New England Biolabs) for oligonucleotides with a GC content of >50 %. Extension times and annealing temperatures varied depending on the length and nucleotide content of oligonucleotides. An example PCR reaction and cycle is shown below. PCR products were separated on a 0.75 % agarose gel to verify the desired product, then purified either using the QIAquick PCR purification kit (QIAGEN) or a Gel purification kit (QIAGEN), according to manufacturer's instructions.

PCR Reaction (10 mL)	
5x Q5 polymerase buffer	1960ul
dNTPs	200ul
Forward primer 10uM	392ul
Reverse primer 10uM	392ul
Template (~10ng/ul)	80ul
Q5 DNA polymerase	35ul
Q5 High GC enhancer buffer	500ul
Nuclease-free H2O	To 9.8mL

Thermocycler conditions		
Temp (°C)	Time	
98	1 min	
98	30 sec	x35
63	30 sec	
72	20 sec	
72	1 min	
4	∞	

For large-scale oligonucleotide amplification, 2-4 96-well plate reactions were set up (9.8 mL total per plate) and PCR was performed as usual. Reactions were checked on a 0.75 % agarose gel, then pooled and aggregates were removed using a 0.2 µm syringe filter. DNA was precipitated by adding NaAc to a final concentration of 0.3 M, then 2.5x vol of ice-cold

100% EtOH was added and mixed well. DNA was left to precipitate for at least 1 hour at -80 ºC, or overnight at – 20 ºC. The sample was then centrifuged for 30 minutes at 4 ºC, 4000 rpm. Supernatant was carefully decanted, and pellet was rinsed with 70% EtOH for 10 minutes at 4 ºC, 4000 rpm. Once again, the supernatant was decanted, and the pellet allowed to air dry at room temperature.

Dried DNA pellets were resuspended in a buffer containing 10 mM Tris-HCl (pH 8) and 1 mM EDTA, then loaded onto a cleaned and equilibrated Poros HQ anion exchange column (ThermoFisher Scientific). Once all the sample had been loaded, the column was washed with the same buffer until UV readings at 280, 260 and 230 nm had stabilised. Free nucleotides and polymerase proteins were removed by elution on a salt gradient to 500 mM NaCl. Oligonucleotides were eluted on a salt gradient up to 1 M NaCl. Typically, the desired PCR-fragments were eluted from the column at approximately 700 mM NaCl and collected in 0.5 mL fractions. Aggregates and plasmid DNA were released from the column using a 2 M NaCl wash. Following agarose gel electrophoresis of fractions, desired fractions were pooled and desalted by ethanol precipitation, as described above. Resulting DNA pellets were resuspended in EB (10 mM Tris-HCl pH 8.5) to the desired concentration.

2.2.2.2.3 DNA conjugation to beads

To produce biotinylated DNA for pull-down experiments, DNA fragments were PCR amplified as described above, using 5’ biotinylated primers. Following PCR, purification of the oligonucleotide and quality control, NaCl was added to the DNA sample to a final concentration of 0.83 M. For every 10 µg of DNA, 40 µl of magnetic MyOne Streptavidin T1 dynabeads (Invitrogen) was washed in Bind/Wash buffer, then added to the DNA in a 3:1 v/v ratio respectively. The reaction mixture was incubated overnight at 24 oC at 15050 rpm in a thermomixer (Eppendorf). Following overnight incubation, supernatant was removed, and unbound DNA was measured using a NanoDrop spectrophotometer to allow quantification of DNA bound to the beads. DNA-bound beads were resuspended in an appropriate volume of Bind/Wash buffer.

Bind/Wash Buffer (DNA conjugation to magnetic beads)	
5 mM	Tris-HCl, p.H. 7.5
0.5 mM	EDTA
1 mM	NaCl

2.2.3 The pre-RC Assembly Assay

The in vitro helicase loading assay was adapted from previously established protocols for the *S. cerevisiae* and *D. melanogaster* system (Coster et al. 2014; Coster and Diffley 2017; Schmidt and Bleichert 2020) and optimised with the help of Dr. Vera Leber (DNA Replication Group).

Orc1-5, Orc6, Cdc6, Cdt1 and MCM2-7 were mixed in a 1:2:2:1:1.5 molar ratio (typically 47:94:94:47:47:70 nM) in Low Salt Buffer and incubated for 10 minutes at 30 °C in a thermomixer shaking at 300 rpm. Biotinylated-conjugated DNA beads were washed three times in Low Salt Buffer, added to the protein mixture in a 2-fold molar excess and incubated for 20 minutes as above. To remove unbound protein, beads were washed twice with either Low Salt Buffer or High Salt Buffer for 30 seconds at a time using a DynaMag magnetic stand (Invitrogen) to identify loading intermediates of double hexamer loading, respectively. Following washes, DNA was eluted from beads in Low Salt Buffer supplemented with 5 mM $CaCl_2$ and 1 unit DNAseI (ThermoFisher Scientific). Elutions were performed for 15 minutes, in the same conditions as above, digesting any DNA unbound by protein and releasing protein-bound DNA from the magnetic beads. Eluates were either prepared for SDS-PAGE analysis or used for grid preparation for electron microscopy. All assays were performed at least twice as independent experiments.

Low Salt Buffer (pre-RC assay)	
25 mM	HEPES-KOH p.H. 7.5
250 mM	KGlu
3 mM	ATP
3 mM	$MgCl_2$
1 mM	DTT
0.1 % (w/v)	Triton X-100

High Salt Buffer (pre-RC assay)	
25 mM	HEPES-KOH p.H. 7.5
500 mM	KGlu
3 mM	ATP
3 mM	$MgCl_2$
1 mM	DTT
0.1 % (w/v)	Triton X-100

2.2.4 Pull-down Assays

Individual DNA-protein pulldown assays were performed similarly as above in the Pre-RC assembly Assay (Section 2.2.3).

ORC-protein pulldowns were performed by incubating Strep-ORC alongside other proteins of interest in the same buffer as the pre-RC assay in appropriate ratios, as discussed further in results. Proteins were pulled down via magnetic Strep T3 XT beads (IBA); beads were washed twice, and bound proteins were subsequently eluted with the same buffer supplemented with 8 mM D-biotin. Eluates were either prepared for SDS-PAGE analysis or used for grid preparation for electron microscopy. All assays were performed at least twice as independent experiments.

2.2.5 Gradient Density Centrifugation to assess protein binding.

2.2.5.1 Gradient Preparation

Glycerol density gradients were manually prepared by injection of a high-density buffer underneath an equal volume of low-density buffer. Individual components of buffers and density gradient depended on individual experiments and will be specifically discussed per experimental result. 400 µl of the low-density buffer was added to the bottom of an 0.8 mL Open-Top Thinwall Ultra-clear tube (Beckman Coulter). Using a syringe and blunt-tip needle, 400 µl of high-density buffer was applied underneath the low-density buffer. Gradients were left to form and settle by upright incubation at 4 °C for >2 hours.

2.2.5.2 Gradient Running and Analysis

40 µl was removed from the top of the formed gradients and replaced with 40 µl of sample dropwise, taking care not to allow mixing into the gradient. Centrifuge tubes were placed into pre-cooled swing buckets fitted with appropriate adaptors. Gradients were run at 29,000 rpm at 4 °C for 16 hours in a precooled Optima™ L-80 XP ultracentrifuge (Beckman Coulter) using a SW55 Ti rotor (Beckman). Acceleration and deceleration values were set to 3.

Gradients were fractionated manually - 32 µl fractions were collected and any pellets collected at the bottom of the tubes were resuspended in 10 µl of high-density gradient buffer.

Fractions were separated on an SDS-PAGE and either Coomassie or Silver stained for analysis. Fractions intended for structural analysis were fixed with 0.1% glutaraldehyde and buffer exchanged to remove glycerol using Zeba Spin Desalting Columns (ThermoScientific). Optimisation of this process will be discussed further in the Results section.

2.2.6 Nucleosome Assembly

DNA for nucleosome assembly was designed by using sequences from Supplementary Table 2, coupled to the 149-bp Widom 601 nucleosome binding sequence. Sequences were PCR amplified and purified as described in Section 2.2.2.

Wild-type histone octamers were purchased from The Histone Source (Colorado State University); modified nucleosomes were a gift from the Bartke Lab (Helmholtz Institute Munich).

Nucleosomes were assembled using the salt gradient dialysis method described in (Luger, Rechsteiner, and Richmond 1999). DNA and octamers were combined in various molar ratios in High Salt Buffer in flat-bottom membrane dialysis tubes (Slide-A-Lyzer MINI Dialysis Units – 3,500 MWCO (ThermoScientific)). Samples were dialysed into Low Salt Buffer over 24 hours at 4 °C on a slow magnetic stirrer (150 rpm) using a dual-channel peristaltic pump (Watson Marlow). Dialysed samples were centrifuged at 14,000 g in a tabletop centrifuge at 4 °C for 5 minutes to remove aggregates. Supernatant was retained and the DNA:protein concentration was determined using a NanoDrop Spectrophotometer (ThermoFisher Scientific), using 1.5 - 2 µl of eluate for each reading. Readings for each sample were taken three times and averaged for a final stock concentration. Between 300 – 500 ng of the nucleosome was run on a 1.75 % EMSA as a quality control to check the occupancy of DNA with the octamer.

High Salt Buffer (nucleosome assembly)	
20 mM	HEPES-KOH p.H. 7.5
2000 mM	KCl
1 mM	EDTA
0.5 mM	DTT

High Salt Buffer (nucleosome assembly)	
20 mM	HEPES-KOH p.H. 7.5
100 mM	KCl
1 mM	EDTA
0.5 mM	DTT

2.2.7 ChIP-MS Design

2.2.7.1 Generation of Tagged Orc2

The full length Orc2 subunit was cloned by ligation independent cloning into a pCAG-derived vector with the help of Tianyi Zhang (Lymphocyte Development Group) to generate an N-terminal Flag-2XStrepII tagged Orc2 subunit. HEK-293T cells were cultured in DMEM supplemented with 10% foetal bovine serum, glutamate and antibiotics. The plasmid containing Flag-2XStrep-Orc2 was transfected using JetPEI reagents (Polyplus) according to manufacturer's protocols Cells were selected in 1.5 ug/ml puromycin two days after transfection. Monoclonal cell populations were isolated through limited dilution cloning and expanded.

2.2.7.2 Western Blotting

Expression of endogenous and tagged Orc2 were evaluated using western blotting. One 10cm plate at 80% confluency (~7 x 106 cells) was lysed and cell debris was removed by centrifugation. Lysate was run on a 4-12% SDS-PAGE at 200V, then transferred to nitrocellulose membranes in a wet transfer at 400 mA for 75 minutes. Membranes were blocked in 1X TBS-T with 4% milk and then incubated with either Orc2-HRP (Santa Cruz Biotech) or Flag-HRP (Sigma) overnight at 4oC. Membranes were washed in 18 TBS-T, incubated with SuperSignal West Dura (Thermofisher) and imaged using Amersham Imager 600 (GE Healthcare).

Lysis Buffer (HEK293T cells)	
50 mM	Tris-HCl p.H. 7.4
150 mM	NaCl
1 mM	EDTA
1% v/v	Triton X-100

2.2.8 DNA Quality Control and Analysis

2.2.8.1 Agarose Gel electrophoresis

Agarose gel electrophoresis was used for quality control to check size distribution of the DNA molecules post amplification or PCR. Agarose gels were prepared at suitable percentages (0.5-2 %, depending on the length of DNA) by dissolving the appropriate weight of agarose powder in 1x Tris-acetate-EDTA buffer (TAE) and allowing to set for >45 minutes. For visualisation, SYBR safe intercalating agent (ThermoScientific) was added in a 1:10,000 concentration from commercial stock. DNA samples were mixed with 6x DNA loading dye (New England Biolabs) and loaded onto set gels, then run at 100-120 V in 1x TAE buffer for 30 – 90 minutes, depending on the predicted size of DNA fragments.

2.2.8.2 EMSAs – Electrophoretic Mobility Shift Assays

EMSAs were used as bulk assays to determine protein binding to DNA. Agarose gels were prepared at suitable percentages (0.5-1.5%) by dissolving the appropriate weight of agarose powder in 1x Tris-borate (TB) buffer and allowing to set for >45 minutes. Protein-DNA interactions were left to incubate, then mixed with 50% glycerol loading buffer with no dye. Samples were loaded onto gel and run at 80 V for 120 minutes at 4 ^{o}C.

2.2.9 Protein Quality control and Analysis

2.2.9.1 SDS-PAGE (Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis)

Protein samples were prepared for SDS-PAGE analysis by mixing with appropriate ratios of either 6x or 5x Laemmli buffer, then incubating at 95 ^{o}C for 5 minutes.

SDS-PAGE was carried out using either homemade 5/10% polyacrylamide gels or commercially available gradient gels (ThermoFisher NuPAGE^{TM} 4-12% Bis-Tris). Gels were

run in MOPS running buffer (Formedium). All gels were run on a BioRad powerpack between 200-220 V for 45-60 minutes. Once gels were run, bands were visualised by either Coomassie (Abcam) or silver staining, depending on sample concentration.

5X Laemmli Buffer	
250 mM	Tris-HCl p.H. 6.8
0.25%	Bromophenol Blue
50% (w/v)	Glycerol
500 mM	DTT
10% (v/v)	SDS

2.2.9.2 Coomassie Staining

Post running, gels were briefly rinsed in ultrapure water (VWR AnalaR NORMAPUR), then incubated with Coomassie (Abcam) for 15 minutes with gentle shaking. Once stained, excess Coomassie was rinsed off with ultrapure water.

2.2.9.3 Silver Staining

After gel electrophoresis, polyacrylamide gels were rinsed briefly in water, then incubated in a fixing solution for a minimum of 1 hour, up to overnight with gentle shaking. After fixation, gels were rehydrated in ultrapure water for two 10-minute washes. Following wash and rehydration, gels were incubated in chilled sensitivity solution for 60 – 90 seconds with gentle shaking, then rinsed twice for 20 seconds each with ultrapure water. Chilled staining solution was added to gels and incubated for 15 – 30 minutes with gentle shaking. Following this, gels were rinsed briefly in ultrapure water, then incubated in developer solution until desired intensity of bands had been reached. To halt further development, 15 mL of 48% (w/v) citric acid was added. Gels were stored in distilled water and scanned for documentation.

Fixing solution	
40% v/v	MeOH
13.5% v/v	Formaldehyde

Sensitivity solution	
0.02% w/v	$Na_2S_2O_3$

Staining solution	
0.1% w/v	$AgNO_3$

Fixing solution	
3% w/v	$CaCO_3$
0.05% v/v	Formaldehyde
0.0004% w/v	$Na_2S_2O_3$

2.2.9.4 Mass Photometry

Glass slides (VWR) were cleaned with water and isopropanol before air drying. 7 µl of poly-L-lysine (Sigma-Aldrich) was added to the centre of the slide for 30 seconds before washing with water again and air drying. A 6-well gasket (Grace Bio Labs) was applied to the slide and positioned on the Refeyn TwoMP mass photometer accordingly. 19 µl of sample buffer was applied to the selected well and the focusing procedure was performed, followed by 1 µl of sample addition and data acquisition using the AcquireMP software. Data was analysed using the DiscoverMP software. For samples with a particularly low or high concentration, ratios of buffer and sample were adjusted accordingly, to allow for counts between approximately 1000 and 4000.

2.2.10 Electron Microscopy Sample Preparation

Techniques described in this section are general protocols. Specific conditions and optimisations used to prepare individual samples will be specified in the Results section where appropriate.

2.2.10.1 Negative Stain

Carbon coated on 300 Mesh Cu grids (Agar Scientific) were glow-discharged on the carbon coated surface using a PELCO easiGlow Discharge Cleaning system at 0.39 mBar, 15 mA with a glow-hold cycle for 25 and 10 seconds respectively. 3.5 µl of sample was applied to the carbon coated surface of the grid and incubated for 60 seconds at room temperature. Excess liquid was blotted away using filter paper (Whatman), then the grid was rinsed with two blot washes using the same buffer as the sample, using two application-blot cycles without letting the grid fully dry in between. 3.5 µl of 0.22 µm filtered 2% (w/v) uranyl acetate (pH 4.5) was applied onto the same face of the grid and left to stain for 60 seconds at room temperature before being blotted away as above. The stained grids were left to air-dry for >5 minutes prior to storage or visualisation.

2.2.10.2 Cryogenic

R2/2 2 nm carbon-coated grids with a 300 Mesh Cu support, both commercial and home-made (Quantifoil) were glow-discharged on the carbon coated surface using a PELCO easiGlow Discharge Cleaning system, as described above.

A Vitrobot Mark IV (FEI) was set up for 100% humidity at 4 ^{o}C in the sample application chamber. Glow discharged grids were clipped into the application chamber and 3.5 µl of sample was applied and incubated for 30 seconds. Sample was blotted away using filter paper (Whatman) for 1 second, using a blot force of 2. Sample was then plunge-frozen into a liquid ethane cooled by liquid nitrogen to promote the formation of vitreous ice. Grids were stored in liquid nitrogen prior to screening and data collection.

2.2.11 Electron Microscopy Data Acquisition

In-house data collections on negative stain and cryo- grids were performed on a Talos F200i TEM microscope (ThermoFisher Scientific) fitted with a Falcon 3EC Direct Electron Detector (ThermoFisher Scientific). Collection parameters will be specified for each individual dataset in the supplementary section. Data acquisition was carried out using EPU software (FEI), and micrographs were exported in the .mrc format for data processing.

Microscope calibrations and data collections were facilitated by Dr. Ricardo Aramayo (Laboratory of Medical Sciences/Imperial College London)

The large-scale cryo-EM data collections for ORC and ORC-DNA samples were collected on a Titan Krios 300 kV FEG microscope (ThermoFisher Scientific) equipped with a K3 detector (Gatan) at the London Consortium for Electron Microscopy (LonCEM) at the Francis Crick Institute in London. Collection parameters will be specified for each individual dataset in the supplementary section. Movies were exported in the .mrcs format for data processing. Microscope preparation, grid handling and data acquisition with EPU (FEI) was carried out by Dr. Nora Cronin (LonCEM, Francis Crick Institute).

2.2.12 Electron Microscopy Processing

Both negative stain and cryo-EM data were processed using RELION 3.1 or RELION 4.0 beta (Scheres 2012) on the LMS cluster or individual workstation. Recorded movie frames were motion-corrected using MotionCor2 (Zheng et al. 2017) and contrast transfer parameters (CTF) were determined using CTFFIND4.0 (Rohou and Grigorieff, 2015). Averaged .mrc files and power spectra were manually inspected and micrographs containing severe stain/ice contamination or containing aberrant CTF estimation values worse than 5 A resolution were discarded. Particle picking on a subset of micrographs was performed either manually or automatically using RELION's integrated log-function blob picker, Topaz (Bepler et al. 2019) in a reference-free manner and subject to 2D classification. Suitable class averages

representing different sample views were selected and served as templates for reference-based particle autopicking. Particles were extracted and subject to further 2D classification and filtering before construction of an ab-initio 3D classification and 3D modelling. Specific parameters for each dataset will be discussed individually in the Results section.

Chapter 3: Results

3.1 Expression and Purification of full-length Human Pre-Replication complex proteins

To understand and develop *in-vitro* mechanisms of pre-replication assembly, it is imperative to have high-quality, homogenous and functional preparations of all proteins involved in the process. To date, the lack of publications studying the biochemistry and structural biology of the human pre-RC have been limited compared to the *S. cerevisiae* system, primarily because of the difficulties in generating stable protein complexes. This is likely due to a combination of protein flexibility and post-translational modifications – many of these proteins contain long, flexible linkers or extensions, subject to an array of post-translational modifications depending on the cell cycle stage or cellular environment. This innate flexibility results in proteins that may behave in unexpected ways upon purification due to heterogeneity in modification landscapes, or do not average particularly well in structural studies. One way around this is to purify truncated mutants of these proteins, to remove the variability of flexible domains. Indeed, this has been done for structural studies using human replication proteins – so far, several available 3D structures of the human ORC have truncated forms of Orc1 and Orc2, with the N-terminal regions removed. A study published in 2020 at the same time as these describe experiments managed to purify full-length ORC (Cheng et al. 2020), although this made EM analysis difficult and large portions of these subunits remain unmapped in available electron density data. While truncating the proteins may improve sample homogeneity and ease purification methods for these studies, the deletion of these regions may affect the functionality of these proteins, generate pseudo functions, and may not provide a thorough depiction of their *in vivo* activities.

One of the goals of this project was to develop a reproducible method of purification of full-length human ORC and Cdc6 proteins. This will allow the exploration of the functionalities of the proteins in a state more comparable to an *in vivo* system. With this, as well as using these for pre-RC assembly assays and potentially uncovering structural data of previously

unmapped regions, we can study chromatin interactions with the BAH region of the Orc1 NTD and alternative behaviour with differing NTD modification states.

For the timescale and scope of this project, only some of these have been addressed. However, the method to purify the proteins of interest outlined in this study now opens a multitude of future experimental possibilities that will allow a deeper investigation into pre-RC component behaviour.

3.1.1 ORC - Origin Recognition Complex

The Origin Recognition Complex was expressed and purified as an entire complex containing subunits ORC1-5. While previous studies have managed to isolate individual subunits (Thome et al. 2000), it was unknown whether this method would introduce additional complexities in assembling the subunits together. It was imperative to express these proteins in a eukaryotic expression system, to allow for G1 arrest, post-translational modifications necessary for protein stability, proper protein folding and disulfide bond formation and reduced risk of bacterial endotoxin contamination. Both insect cell and yeast cell expressions were explored for their up-scalability, although early success with yeast cells resulted in Sf9 expression being abandoned. Using *S. cerevisiae* meant that rapid growth and ease of expression volume up-scaling up to 20 L was possible.

3.1.1.1 Expression - Integrated vs Non-integrated plasmids

Both integratable and non-integratable plasmids containing the ORC subunits were tested to determine the best option for expression. Both sets of plasmids contained an ampicillin resistance marker for bacterial plasmid amplification, auxotrophic selection markers (Orc1/5 – uracil, Orc2 – tryptophan, Orc3/4 – leucine) for selective growth and a galactose-inducible promoter upstream of the gene of interest open reading frame to control expression of the genes upon G1 arrest. Non-integratable, episomal plasmids also had an *S. cerevisiae* origin sequence to ensure replication without integration, while integrated plasmids lacked this to avoid division of cells without plasmid integration. (Figure 16.A.B.). These bands were

confirmed by Western blot once to confirm the presence and molecular weight of all subunits (not shown), but Coomassie staining was used thereafter since antibody concentrations were too high to use consistently. While both methods successfully produced each of the human ORC subunits as shown in Figure 16, non-integrated plasmids produced a much higher concentration of protein upon lysis of cells and initial affinity purification – and average yield of ~4 mg of protein in comparison to ~1 mg of protein from 6L of cells, identified via Coomassie staining of elution and Bradford absorption measurements of protein (Figure 16.C.). It must be noted that the total protein yields at this stage of the purification process include various contaminants and some degradation of subunits, which would have provided an inaccurate measurement of total protein of interest. Nonetheless, from band intensities on Coomassie stained SDS-PAGE gels, it was assumed that the integrated plasmids allowed for higher amounts of protein, especially noted when considering the Orc4 subunit, barely observed in the equivalent elution samples for the integrated plasmids. For this reason, non-integrated plasmids were chosen to transform yeast cells for expression

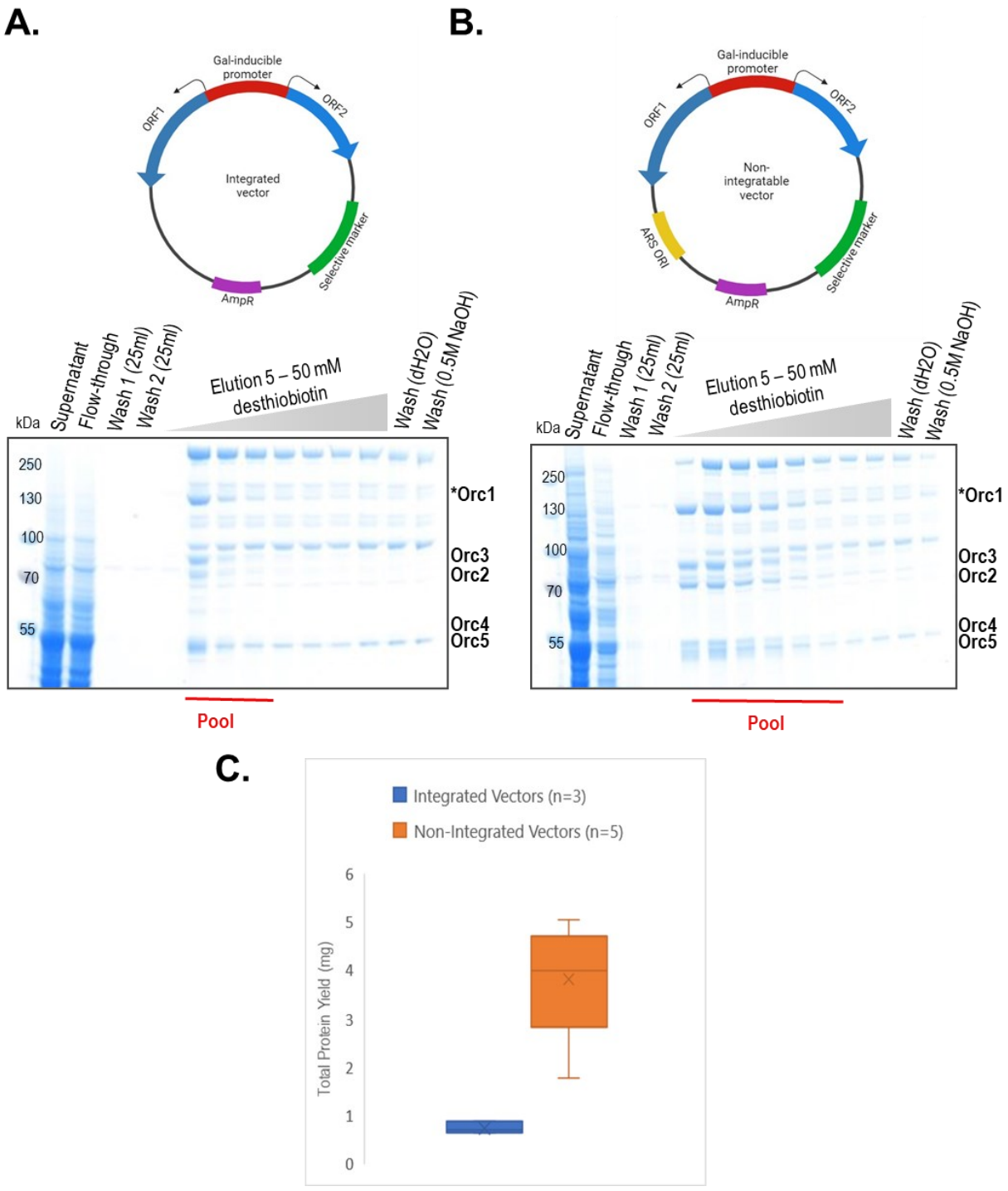


Figure 16: Comparison of yields of human Orc1-5 complex expression between integrated and non-integrated plasmids

Plasmid design (above) and corresponding Coomassie-stained SDS-PAGE gels for Streptavidin affinity purification step (below) following growth and lysis of S. cerevisiae expression culture of ***A:*** *Integrated plasmids and* ***B:*** *Non-Integrated plasmids. For both protocols, 6L of cells (corresponding to 25-30g of milled cell powder) were lysed and filtered supernatant was loaded onto a pre-loaded StrepTrap HP column. 1mL fractions were eluted with a gradient of desthiobiotin, with 10ul run on an SDS-PAGE gel. Fractions containing all ORC subunits (labelled) were pooled and protein yield was quantified via Bradford assay. Note: *Orc1 denounces Strep-tagged Orc1.* ***C:*** *Graph quantifying expressed and purified protein pools isolated after first step from integrated versus non-integrated plasmids, showing non-integrated plasmids have an increased total protein yield (including impurities) of ~4x in comparison to integrated plasmids.*

3.1.1.2 Streptavidin-dependent affinity purification

Following lysis and centrifugation, cell lysate was passed through a packed Streptavidin affinity column, intending to bind and elute via the 2x-StrepSumo tag on the NTD of Orc1. Initially, this was done using a StrepTrap HP column (Cytiva) to decent success, with an average yield of about 4 mg total protein from ~6L of expression culture (Figure 16.C). However, elution from these columns required a high concentration of desthiobiotin (~25 mM as opposed to the advertised 5 mM). Discontinuation of this column meant optimisation of the same process with the updated Strep-Tactin XT (Cytiva) Sepharose matrix, which resulted in an improved yield of protein following the above procedure. Recommended elution concentrations of 25 – 75 mM biotin were tested, with 50 mM being appropriate as the most effective (and more cost-effective) to elute bound protein (Figure 17). The Strep-Tactin XT columns, due to their higher affinity than the StrepTrap HP columns, also bound and eluted fewer impurities and degradation products. Overall, this resulted in a cleaner, higher yield preparation of Orc1-5 (~7 mg of protein from 6L of culture) (Figure 17.C.). Additionally, the reduction in levels of Orc4 across the elution profile were not observed when using the StrepTactinXT resin in comparison to the StrepTrapHP resin (Figure 17.A. and 17.B.), suggesting a state in which the entire complex is stable.

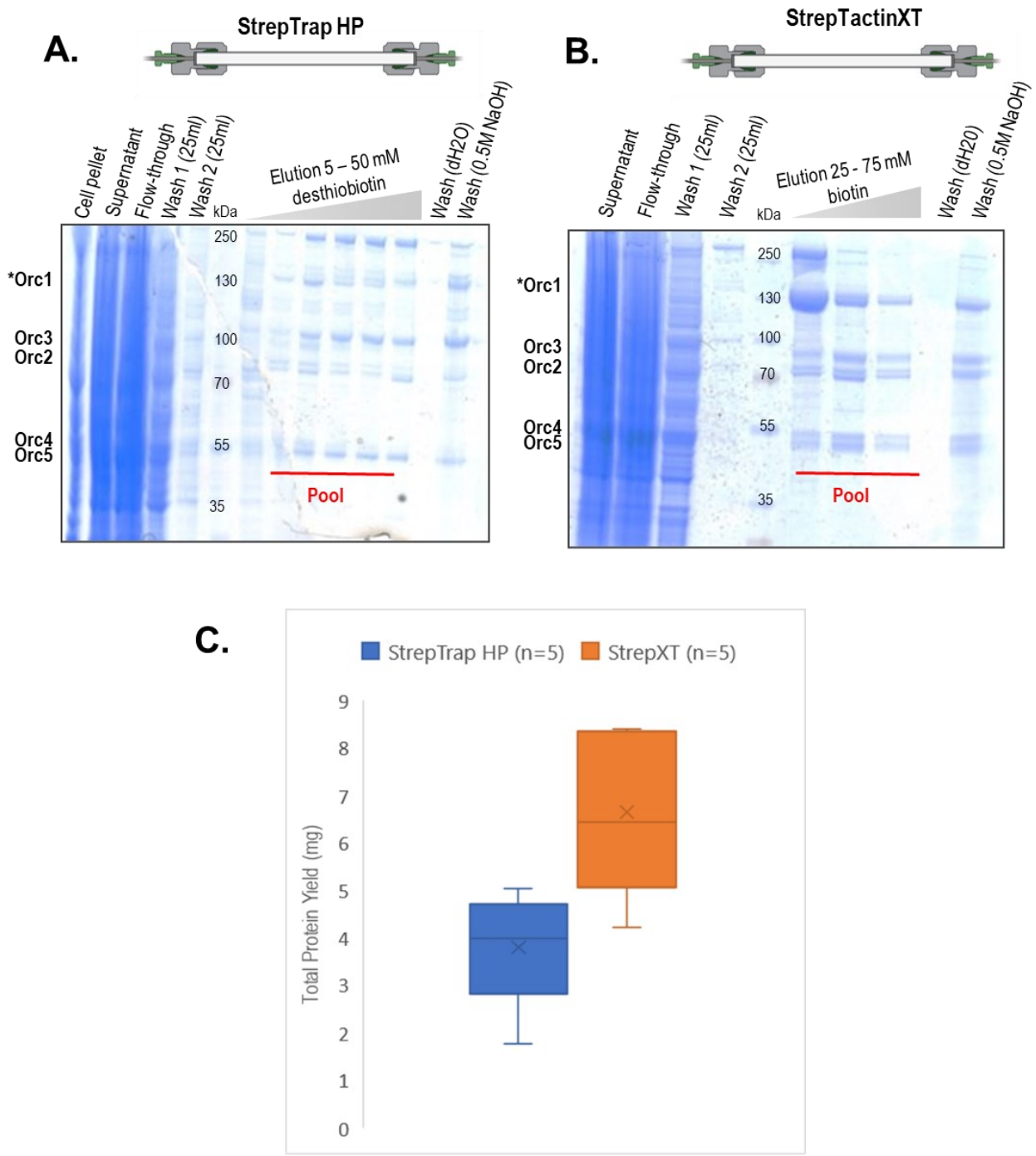


Figure 17: Affinity purification optimisation of human Orc1-5 via the 2xStrepSumo tag on Orc1: StrepTrap HP versus StrepTactin XT.

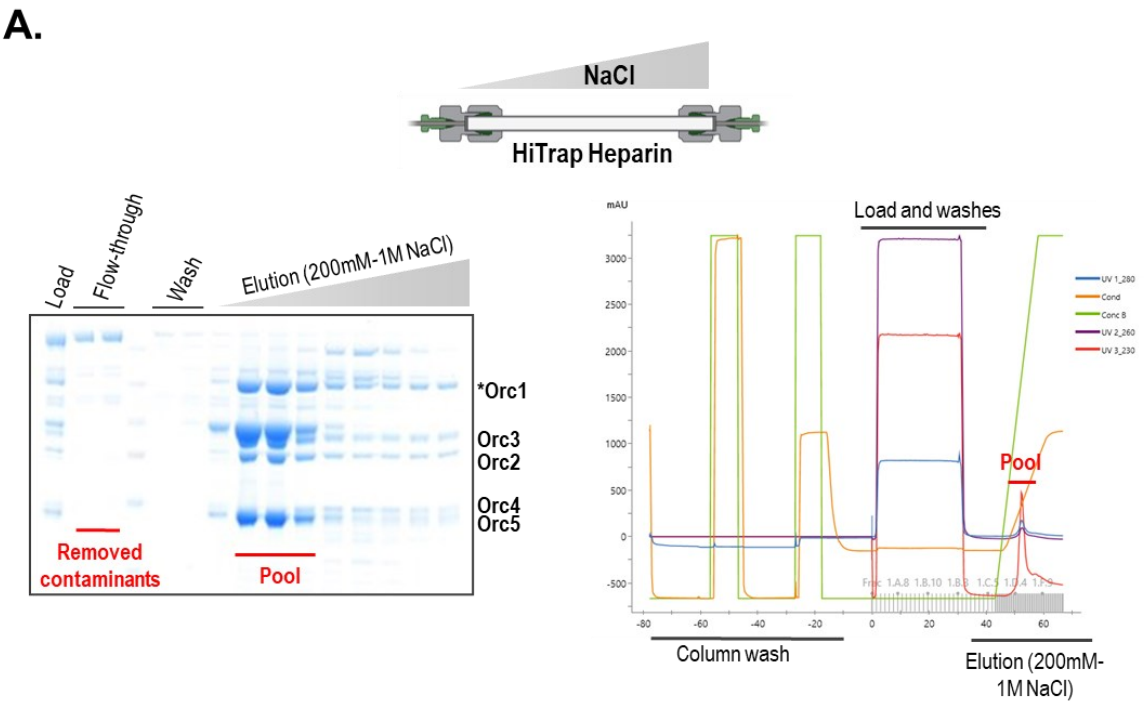
A: *Coomassie-stained SDS-PAGE gel with Orc1-5 labelled subunits following affinity purification steps post growth and lysis of cells with* ***A:*** *a StrepTrapHP prepacked column and* ***B****: a StrepTactinXT column. Elutions were done with a gradient of desthiobiotin and D-biotin respectively, according to manufacturer's recommendations and 10ul of elution was loaded per lane.* ***C:*** *Graph quantifying expressed and purified protein pools isolated after elution from StrepTrapHP versus StrepTactinXT columns. Note that this pool is post the first step of cleanup post cell lysis, thus protein quantification also includes impurities. Quantifications are from repeated experiments (n=5 of each condition). Gels shown are representative of all experiments. Note: *Orc1 denotes Strep-tagged Orc1.*

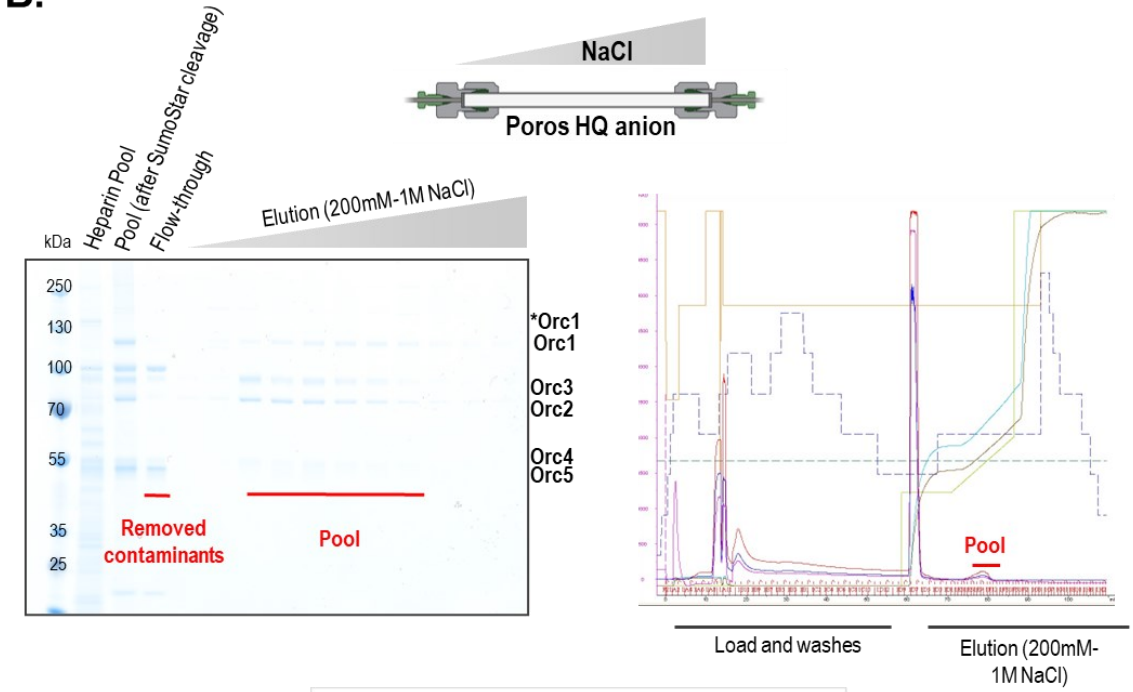
3.1.1.3 Ion Exchange – HiTrap Heparin HP and Poros HQ

Orc1-5's DNA binding ability was exploited during the next stage of the purification, by using a heparin column and a salt gradient ion-exclusion based elution method. Pooled protein from the Streptavidin-based affinity purification was diluted to 200 mM NaCl, then injected into a preequilibrated HiTrap Heparin HP (Cytiva) column using an AKTA Pure system (Cytiva). Gradual increase of NaCl concentration on a gradient eluted bound protein. Orc1-5 eluted as a whole complex between 400-550 mM NaCl (Figure 18.A.), with aggregates and impurities separated out at either lower or higher concentrations of salt.

Sample was pooled stringently to contain few impurities and observed subunit stoichiometry, then, if required, incubated overnight with SumoStar protease to cleave the N-terminal Orc1 2x StrepSumo tag, as shown in Figure 18.B.).

If the sample still contained impurities – usually the case if the StrepTrap HP column was used in the first purification step (Figure 17), the sample was re-diluted into 200 mM NaCl buffer, then loaded and eluted onto a Poros HQ anion exchange column in the same manner as described for the Heparin Column purification. While this method successfully removed the mid-weight molecular contaminants (Figure 18.B, in the flow-through), there was a significant loss in the yield of protein (Figure.18.C.), with a significant reduction in subunits Orc4 and Orc5 in particular. Increasing the molarity of the elution buffer up to 1 M NaCl did not improve the elution of the complex. In addition, pressure build-up of the Poros HQ column suggested protein aggregation in the column or column filter, making this an unideal method for removing impurities while retaining high yield. This problem was rectified upon using the Strep-Tactin XT column in the initial stages of the purification, where these mid-molecular weight impurities were not purified alongside the Orc1-5 complex (Figure 17.B-C.).


A.
NaCl
HiTrap Heparin
Load
Flow-through
Wash
Elution (200mM-1M NaCl)
*Orc1
Orc3
Orc2
Orc4
Orc5
Removed contaminants
Pool
Load and washes
Pool
Column wash
Elution (200mM-1M NaCl)


B.
NaCl
Poros HQ anion
Heparin Pool
Pool (after SumoStar cleavage)
Flow-through
Elution (200mM-1M NaCl)
kDa
250
130
100
70
55
35
25
*Orc1
Orc1
Orc3
Orc2
Orc4
Orc5
Removed contaminants
Pool
Pool
Load and washes
Elution (200mM-1M NaCl)

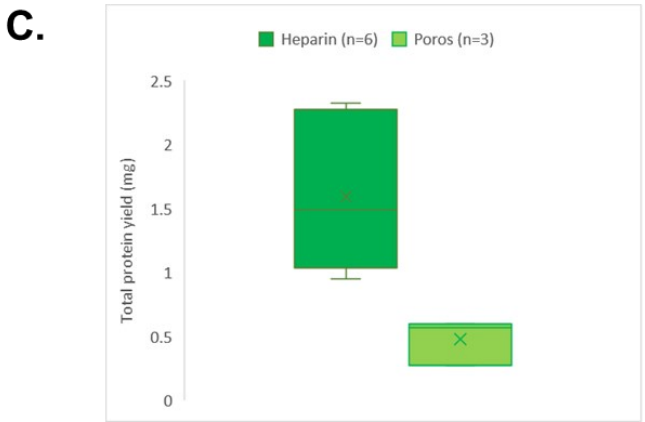

C.
Heparin (n=6)
Poros (n=3)
Total protein yield (mg)
2.5
2
1.5
1
0.5
0

Figure 18: Ion Exchange purification of Orc1-5

A: *Ion exchange purification SDS-PAGE gel and corresponding AKTA trace following salt-gradient exchange on a Heparin column.* ***B:*** *: Ion exchange purification SDS-PAGE gel and corresponding AKTA trace following salt-gradient exchange on a Poros HQ column. Elutions were measured by Bradford quantification and fractions containing only the ORC subunits were pooled. Note: the * denotes the 2x Strep-SUMO tag on Orc1, that was cleaved by SumoStar protease into Orc1* ***C:*** *Graph quantifying expressed and purified protein isolated after Heparin and then Poros, showing a significant reduction in protein yield following Poros column purification. Gels and traces shown are representative of repeats of each described step of the purification.*

3.1.1.4 Size-separation methods for protein homogeneity

At this point, the purified sample contained only subunits Orc1-5. However, upon Mass Photometry measurements, it was evident that while some of these subunits were assembled into the full complex (342 kDa), there were also sub-populations present of semi-assembled or single subunits (Figure 21). To obtain homogenous, fully assembled Orc1-5 populations, size-exclusion methods were explored to remove partially assembled complexes.

Initially, pooled samples were concentrated to ~ 500 µl at approx. 1.5 µM before being loaded onto a preequilibrated Superose 6 10/300 Increase column (Cytiva) in a buffer containing 50 mM HEPES-NaOH, 200 mM NaCl, 2 mM $MgCl_2$, 1 mM ATP and 10 % glycerol. Samples typically eluted at the ~10-15 mL mark and were pooled and concentrated using 0.5 mL Amicon concentrator 30kDa MWCO (Figure 19.A.) However, there was a significant loss of protein (>90%) after loading onto the column (Figure 19.B). It was predicted that much of the protein was incompatible with the matrix of the column and became stuck or aggregated, since much protein debris was present on washing the column with 0.5 M NaOH and 6 M guanidinium hydrochloride. Because of this, alternate methods of matrix/bead-free purification based on molecule size were explored.

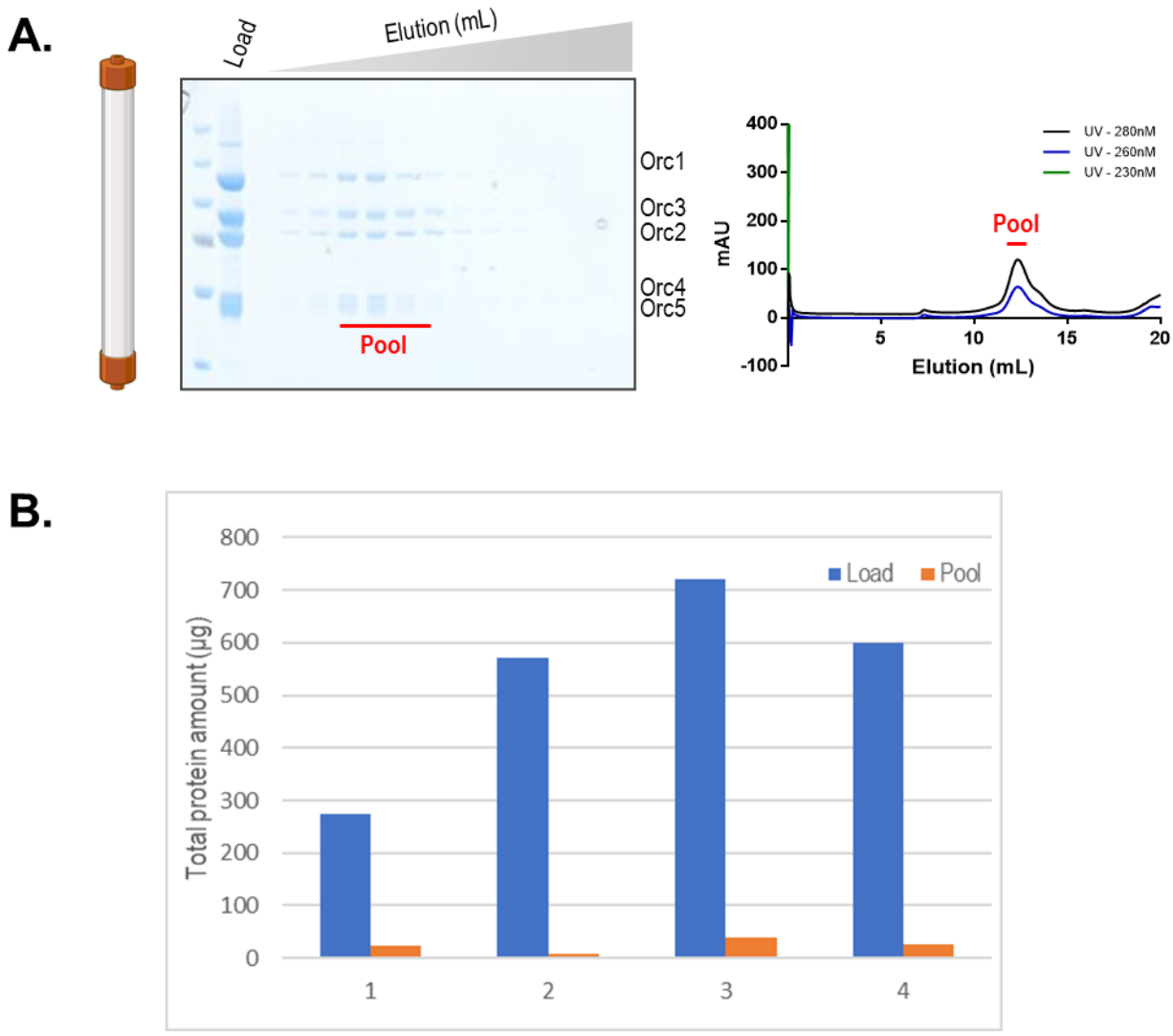


Figure19: Size Exclusion Chromatography for stochiometric and homogenous Orc1-5

A: *SDS-PAGE gel and corresponding AKTA trace of ORC purification sample (hitherto purified via Strep and Heparin affinity purification steps) following gel-filtration with Superrose 10/30 Increase column.* ***B:*** *Graph quantifying expressed and purified protein isolated after four separate gel filtration experiments – showing a repeated >90% loss in yield post gel-filtration.*

Glycerol density gradients were preferred, due to their previous use in the lab for separating out protein complexes of differing composition (Saleh et al, 2022). 800 µl mini-density gradients were prepared and run as described in Chapter 2.2.5, with a glycerol distribution of 10 – 30%. Upon fractionation and SDS-PAGE analysis (Figure 20), fractions containing all subunits were pooled and concentrated with a 0.5 ml 30 kDa cut-off protein concentrator (Amicon), and then buffer exchanged into a final 10% glycerol buffer. Initially, buffer conditions

were prepared at 150 mM NaCl, to remain as close as possible to physiological salt conditions. However, the protein pools would routinely become aggregated and stick to the concentrator membrane upon reaching a concentration value of 3 mg/mL, resulting in the loss of usable protein and lower overall yield (Figure 20.A.C.).

It was hypothesised that lowering the glycerol and salt concentration simultaneously resulted in destabilisation and aggregation of the protein pool. Since retaining high glycerol concentrations was unideal for downstream density gradient~~s~~ and cryo-electron microscopy experiments, increasing the NaCl concentration of the storage buffer was explored.

The same gradient density centrifugation experiments were conducted with 300 mM and 500 mM NaCl (Figure 20.A.), as the complex eluted between these values during the ion exchange chromatography purification steps. With equivalent pooling and concentration steps following gradient fractionation and SDS-PAGE analysis, there was a decrease in the amount of protein lost during the concentration step (Figure 20.C.). This suggests a higher salt concentration can serve to stabilise the protein and prevent aggregation in lower glycerol concentrations. The protein also eluted earlier in the gradient, suggesting less aggregation and stabilisation of the complex. Since there was little difference between the yields of protein concentrated in 300 versus 500 mM salt, 300 mM NaCl was chosen as a gradient and storage buffer salt concentration to allow for easier dilution into downstream experimental conditions.

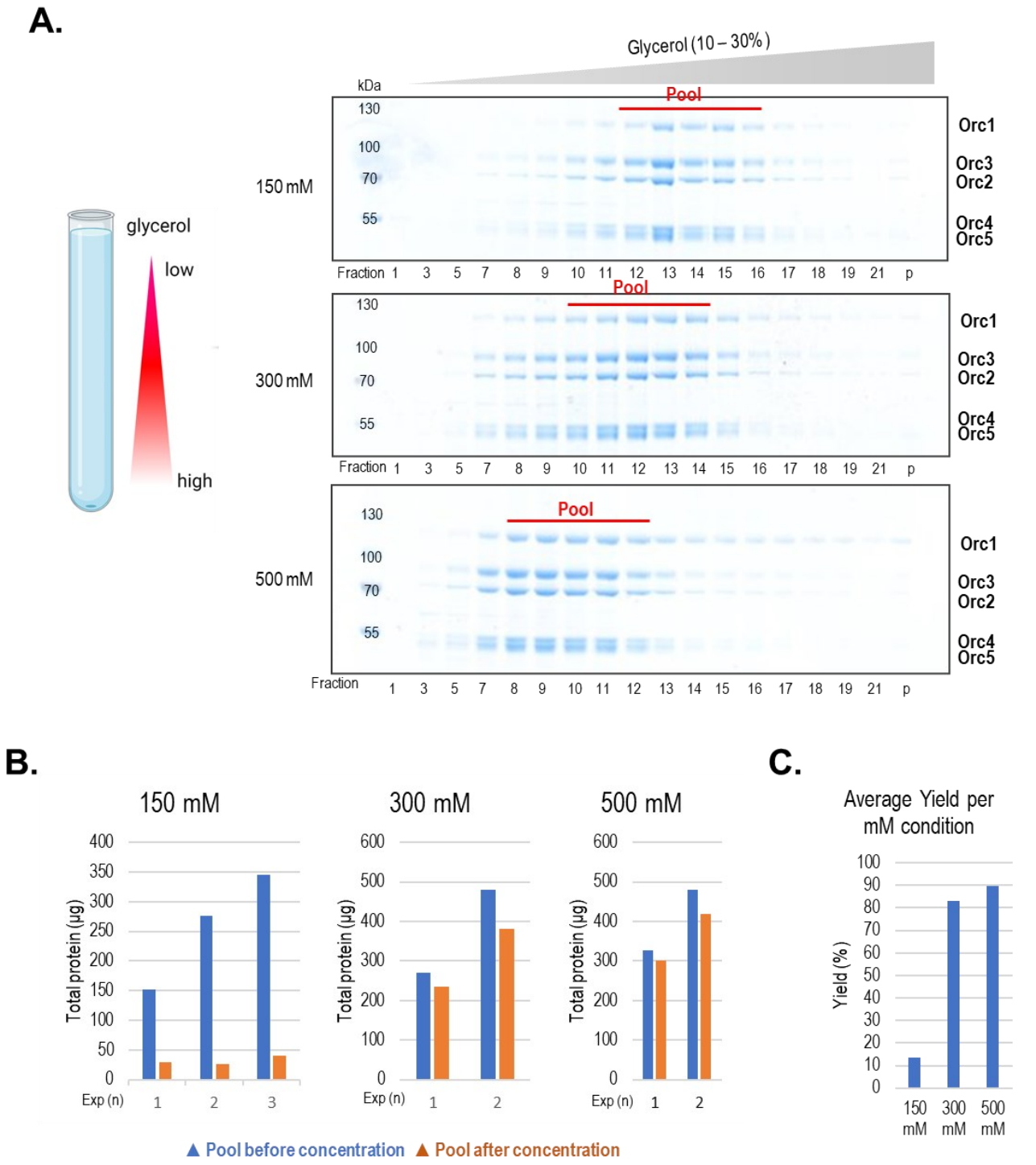


Figure 20: Gradient Density Centrifugation as a matrix-free method for homogenous and stochiometric preparations.

A*: Representative Coomassie-stained SDS-PAGE gels per salt conditions of density-gradient centrifugation of purified Orc1-5.* ***B:*** *Graph quantifying purified protein yields isolated after density gradient centrifugation and spin-column concentration per salt condition (150 mM n=3, 300 mM n=2, 500 mM n=2), showing that lower salt concentrations (150 mM NaCl) results in a larger loss of protein recovery post concentration in comparison to 300 nM and 500 mM.* ***C:*** *Average yield of protein post-concentration per salt condition. Using a higher salt condition (300 or 500 mM) restores yield to almost pre-concentration levels. To maintain a 'closer to native' salt concentration, 300 mM NaCl was chosen for concentration and storage buffer conditions of the ORC complex.*

This additional step of glycerol density-based separation of full versus partial complex assemblies provided a way of ensuring homogenous protein with all five subunits present. This was confirmed by mass photometry analysis of the gradient-purified sample, which showed a single peak of protein at ~350 kDa, the expected molecular weight of assembled Orc1-5 (~350 kDa) (Figure 21).

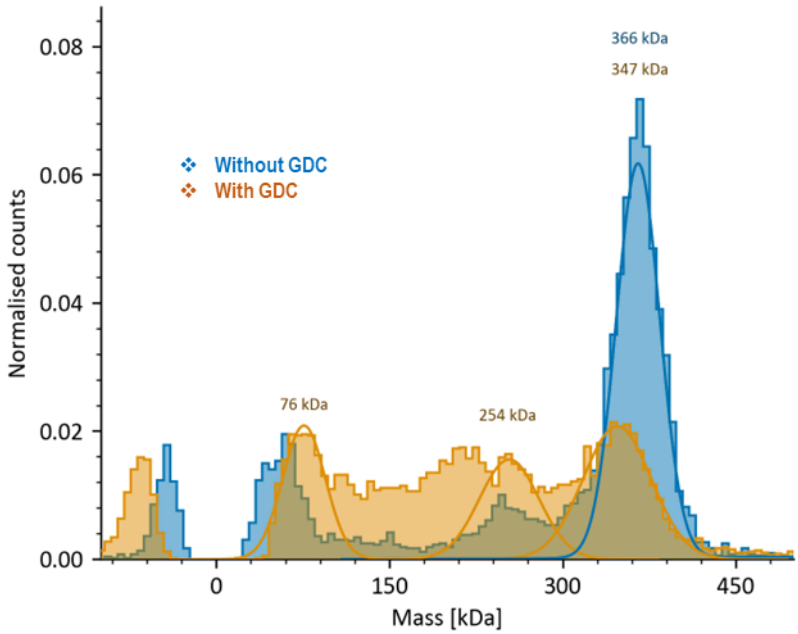


Figure 21: Gradient density separation purification methods result in homogenous Orc1-5

*Refeyn Mass photometry histogram showing populations of fully assembled and partially assembled Orc1-5 complexes before and after gradient separation methods. *GDC = Gradient Density Centrifugation. Protein pools after only affinity chromatography steps contained a heterogenous mix of ORC subunits (indicated by various peaks of molecular weights lower than the expected weight of the Orc1-5 complex) (**orange**), while density gradient separation allowed for pooling of a homogenous prep of protein at the expected molecular weight of the fully assembled Orc1-5 complex (**blue**).*

3.1.1.5 Overall

Overall, through careful optimisation of a range of conditions, Orc1-5 was successfully purified, with all subunits present at their full length, to a scalable yield that had not been previously reported. Additional variables that had to be explored included optimisation of salt concentration, arresting the cells in G1, use of phosphatase inhibitors, and finding the best

protein purification strategies and steps to gain a pure product (Figure 22.A.). All these ultimately lay suggestions to the conditions that ORC is functional in, which will be explored in later chapters.

Analysis of the complex by negative stain electron microscopy also confirmed a fully assembled complex with similar structural architectures as seen in previously published Orc1-5 structures (Figure 22.B).

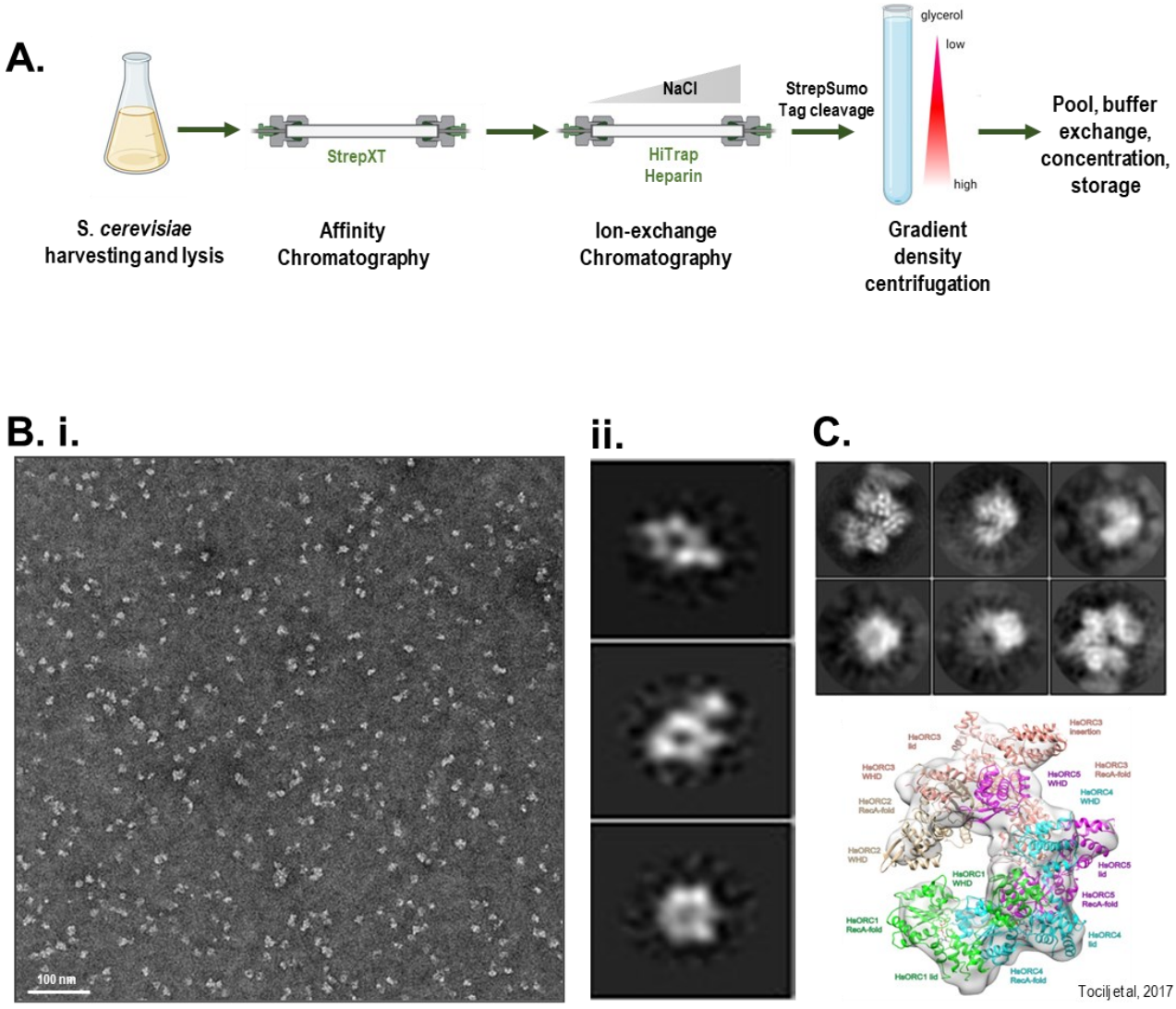


Figure 22: An Optimised Method for human Orc1-5 purification from S. cerevisiae cells

***A**: Schematic for final optimised purification of Orc1-5 **B(i)**: Representative negative stain micrographs of a purified Orc1-5 preparation (left) and **(ii)** corresponding 2D classes generated from averaged individual particles (right) confirming the presence of fully-assembled Orc1-5 complexes. Architecture of generated 2D classes is similar to those of **C:** published human Orc1-5 (Tocilj et al, 2017) 2D classes (above) and 3D model (below), showing a ring-shaped structure with an Orc3 protrusion.*

3.1.2 Cdc6 – Cell-dependent cycle 6

As a single-subunit protein, Cdc6 was expressed in *E. coli* as described in Chapter 2.2.1.2. BL21 cells were transformed with codon-optimised GST-tagged Cdc6 under an IPTG-controlled expression system. Following cell lysis and glutathione agarose affinity binding and PreScission protease mediated cleavage, Cdc6 was released from the beads, with little-to-no GST tag contamination (Figure 23), with a total elution of ~15mg of protein Cdc6 from 5L of culture.

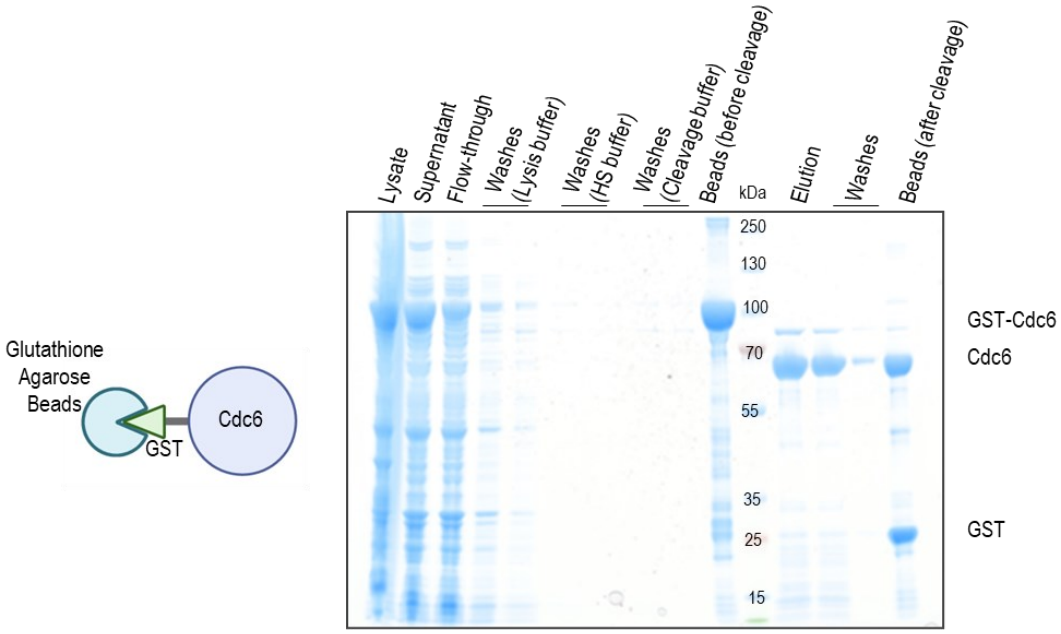


Figure 23: Expression and Affinity purification of GST-Cdc6

SDS-PAGE gel showing elution, binding and lysis of Cdc6 from GST-beads. Cdc6 (a 64 kDa protein) successfully binds and elutes from the glutathione agarose beads after the addition of PreScission protease (cleaves between the N-terminus of Cdc6 and the GST tag) (Elution labelled wells). Some GST-Cdc6 contamination remains following elution from the beads. Gel is representative of optimised GST-affinity purification step, with similar results obtained from independent experimental repeats (n > 5).

3.1.2.1 Removal of contaminants from Cdc6 preparations.

3.1.2.1.1.Size-based separation methods

It was necessary to perform another stage of purification to remove smaller contaminants and degradation products (Figure 23). Because of the success of glycerol density gradients in the production of homogenous ORC1-5, this method was similarly tested for the Cdc6 protein pool post GST-cleavage. However, this method did very little to separate out impurities and instead appeared to promote the presence of degradation products (Figure 24.B.). Similarly, size-exclusion chromatography using a Superdex HiLoad 75 16/50 gel filtration column was not successful in separating impurities or degradation either (Figure 24.A.).

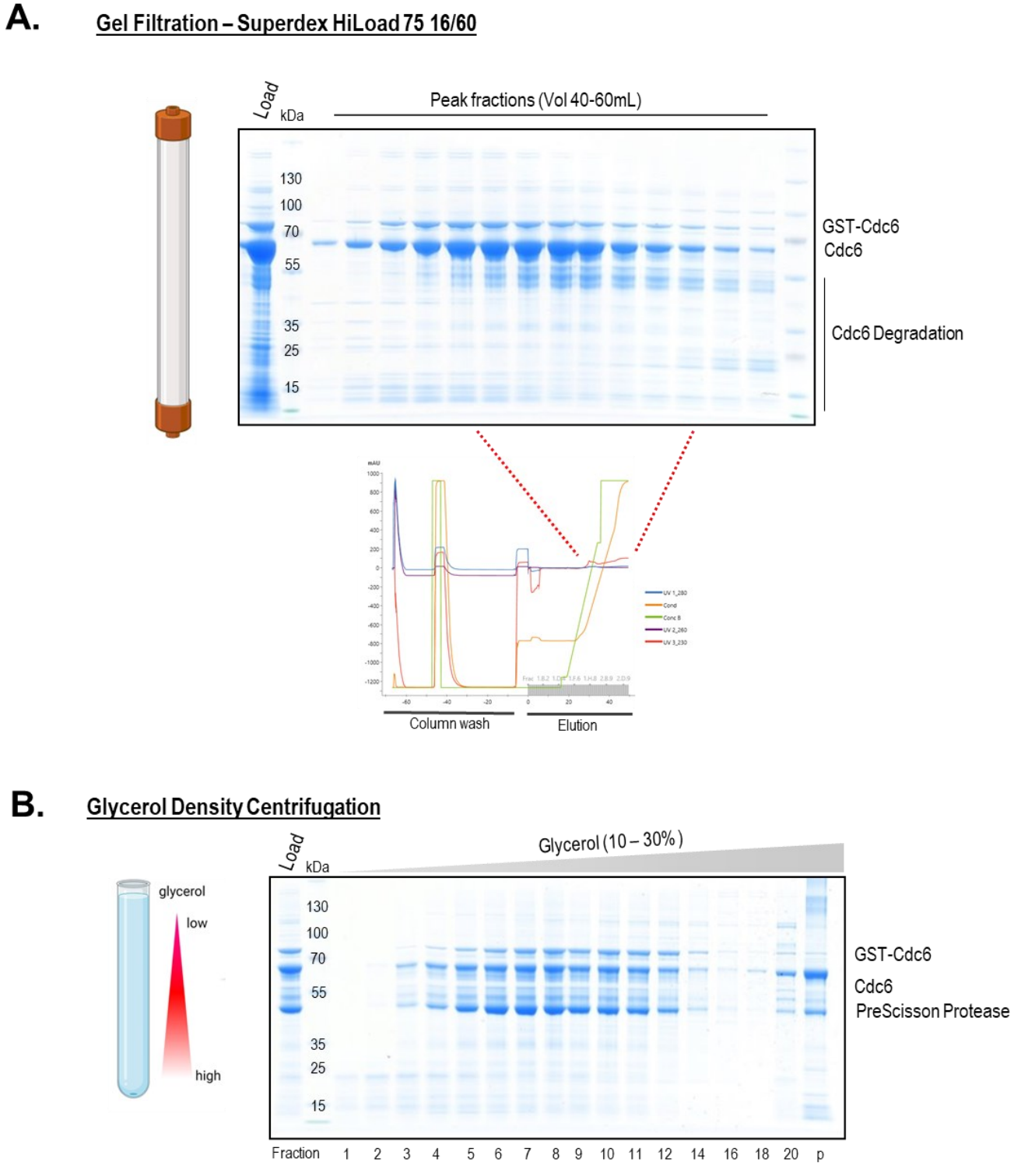


Figure 24: Size-separation methodologies for a homogenous Cdc6 preparation

A: *SDS-PAGE gel and corresponding AKTA trace of Cdc6 following gel filtration method, showing this does not separate out Cdc6 degradation products. Gel is representative of a gel filtration step that was carried out n=2.****B:*** *SDS-PAGE of Cdc6 following gradient density centrifugation on a 10 – 30% glycerol gradient. Note: in this experiment, additional preScission protease (labelled) was added to cleave residual GST-Cdc6, although this method did not separate out the protease proteins, or the contaminants or degradations effectively.*

3.1.2.1.2 Ion Exchange Chromatography

Ion exchange methods were then explored to separate these out. A HiTrap SP HP column was equilibrated with the same buffer used to cleave the GST-tag and separate Cdc6 from the glutathione Sepharose beads, then the pooled and filtered sample was injected into the column using a 50 mL superloop. Following this, the column was washed, then bound protein was eluted by a salt gradient from 200 mM to 1000 mM NaCl/KCl. While some eluted fractions of Cdc6 still contained impurities and degradation products, it was possible to stringently pool and concentrate cleaner fractions (Figure 25) This resulted in a far more successful method for removing unwanted purification byproducts, leading to a homogenous and clean prep of full-length Cdc6 (Figure 26).

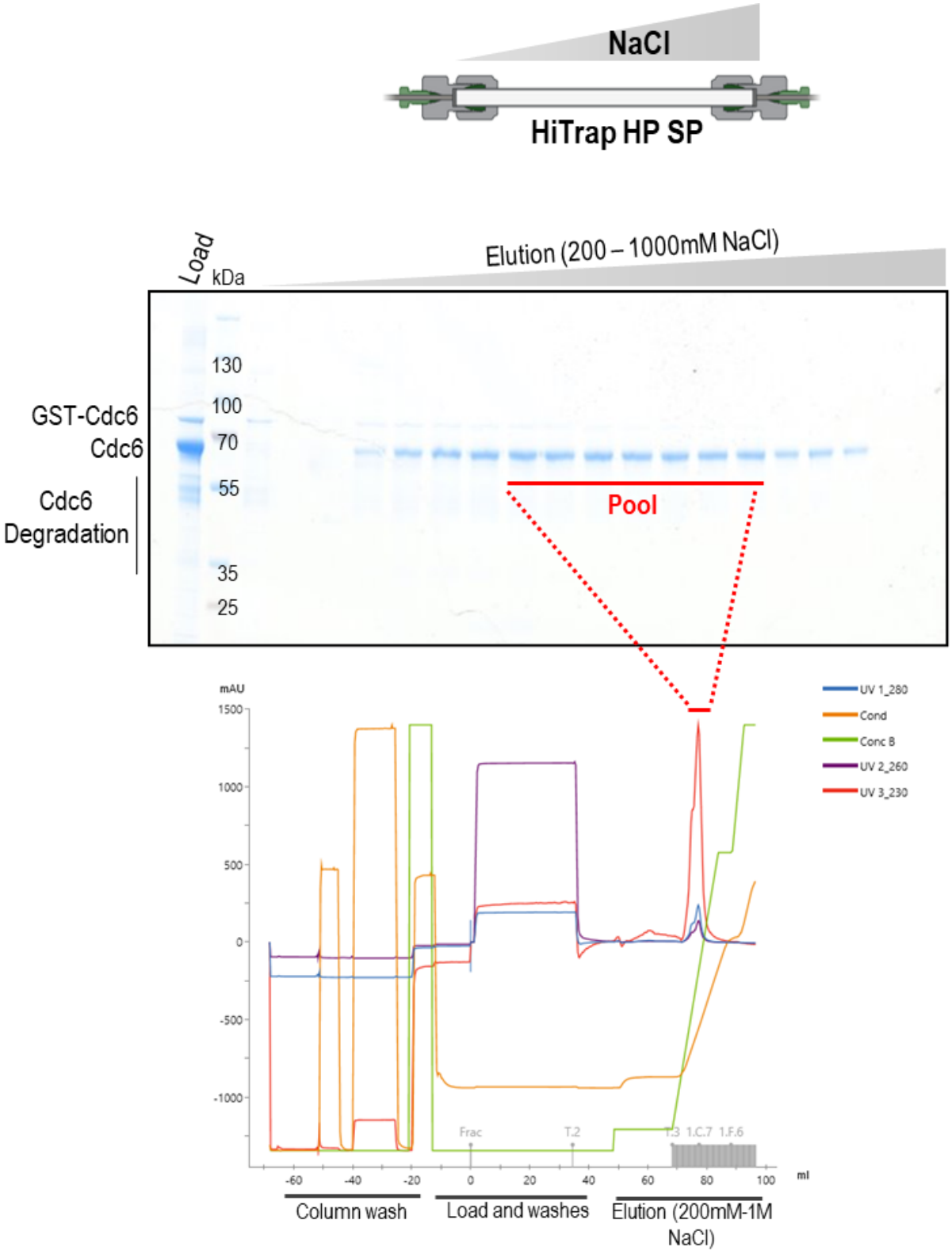


Figure 25: HiTrap HP SP ion exchange chromatography produces homogenous Cdc6

Coomassie-stained SDS-PAGE gel and corresponding AKTA traces of Cdc6 following HiTrap HP SP ion exchange chromatography. 1 mL fractions across a 200 – 1000 mM NaCl gradient were collected, 10 µl per fractions were run on an SDS-PAGE gel and stained. Fractions containing only Cdc6 with minimal degradation products were pooled and concentrated for storage and use in downstream experiments.

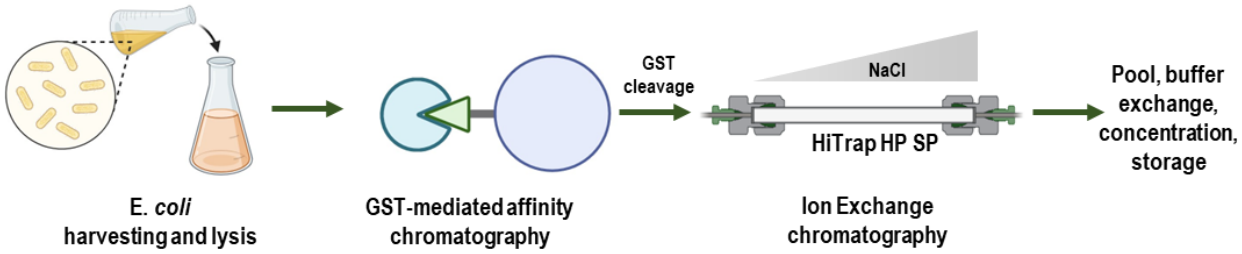


Figure 26: An optimised method for full-length human Cdc6 purification from E. coli

Schematic for final optimised purification of Cdc6. Transformed E. coli BL21 cells are grown and induced with IPTG to produce hCdc6 with a recombinant N-terminal GST tag; the cells are harvested and lysed. Supernatant goes through a GST-mediated affinity purification step that results in an elution of Cdc6, then ion exchange protein purification removes any contaminant and degradation products, leaving a homogenous prep of Cdc6 for further experimental use.

3.1.3 Orc6, Cdt1 and MCM

Expression and purification of these proteins were carried out by other members of our research team. Figure 28 depicts SDS-PAGE analysis of these proteins, showing purified preparations.

DNA Selection and Purification

Since human origin sites do not follow a rigid sequence specificity, multiple different DNA constructs were used to test DNA binding of pre-RC components. As mentioned previously, ORC does not display sequence-specific DNA binding activity – in theory, any sequence of DNA should be compatible with ORC binding. However, to study more physiologically relevant sites of human pre-RC binding, 164-bp segments of DNA were designed according to published ChIP-Seq data showing strong Orc1 or Mcm7 peaks (see chapter 2.2.2 for further detail). In *S. cerevisiae*, the footprint of ORC-Cdc6 is ~48 base pairs (Feng et al, 2021) – to allow for this and any downstream pre-RC protein recruitment, 164 base pairs of ChIP peaks were selected. Chosen sites can be found in Figure 15 and Supplementary Table 2, with a range of AT/GC content and DNA features highlighted. Testing of pre-RC component recruitment to these sites will be discussed in Results Chapter 3.3..

3.2 Amplification and Purification of DNA Sequences

Plasmids containing the sequences of interest described in Supplementary Table 2 were amplified in chemically competent DH5α *E.* coli cells as described in Section 2.2.2. The desired oligonucleotide sequences were then amplified by PCR – optimal conditions including primer melting temperatures (Tm), polymerase type and elongation time were individually characterized per oligo pair before large-scale amplification took place. An example of optimised conditions can be found in Figure 27 and Section 2.2.2.

Following large-scale amplification, the desired sequence was separated from the PCR reaction mix by ion-exchange chromatography on a Poros HQ column combined with a salt-gradient elution (See section X for details). This allowed clean separation of; primers, primer-dimers, plasmid DNA and polymerase in different fractions (Figure 27.B). The fractions containing the desired DNA fragments were pooled and ethanol precipitated, then resuspended in 10 mM Tris-HCl, p.H. 8.5. Typically, from a 10 mL PCR reaction volume, the yield of oligonucleotide ranged from 0.5 – 1.5 mg.

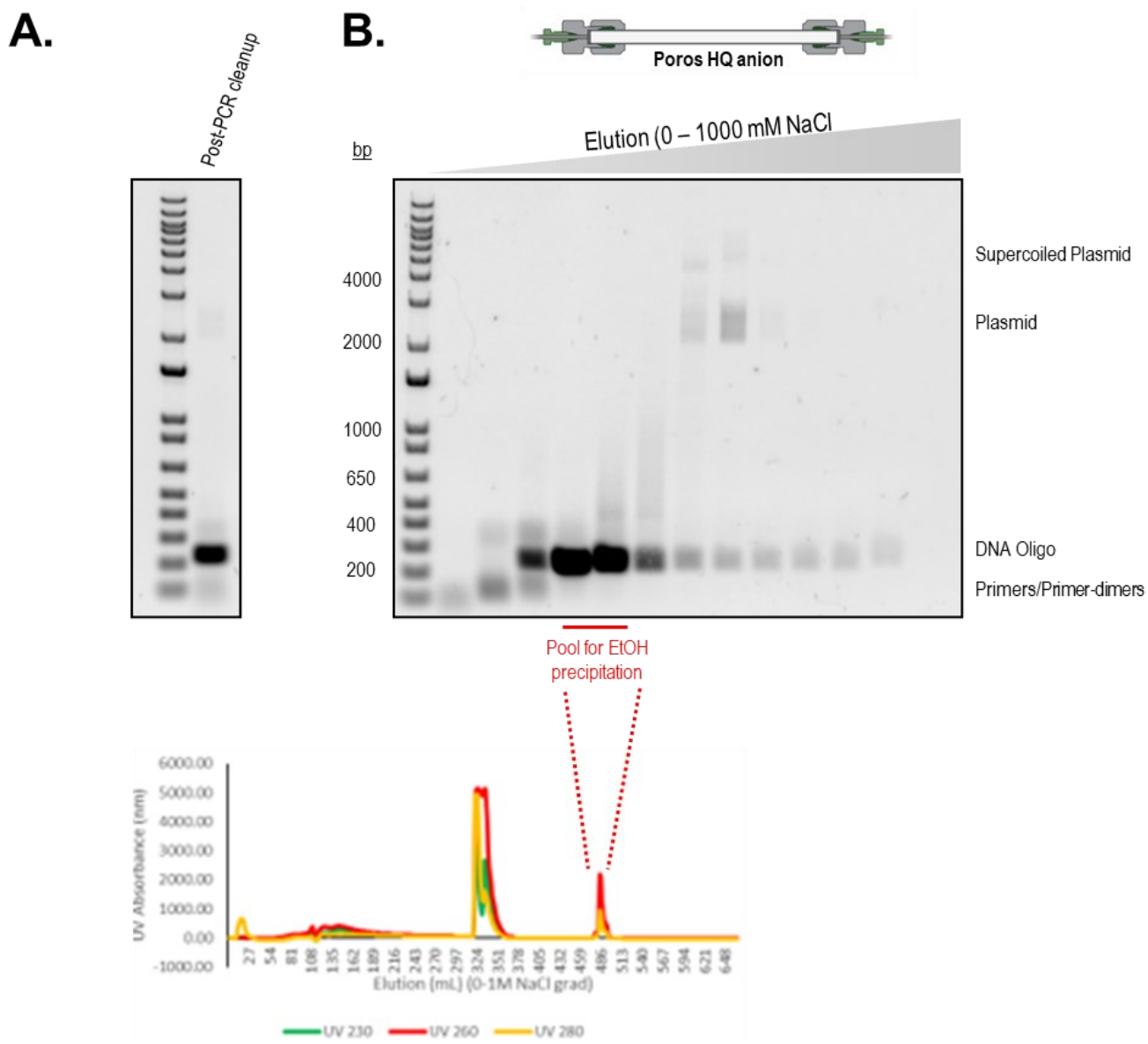


Figure 27: Large Scale PCR and purification methods to isolate DNA fragments.

A: *Example PCR purification of DNA used in ORC-DNA, ORC-Cdc6-DNA and ORC-nucleosome binding experiments (the 147 base pair Widom-601 sequence followed by a short 20-164 base pair sequence of interest identified from published ChIP-Seq experiments that display strong Mcm7 or Orc1 binding). Gel shown is of a PCR-purified Widom-601 sequence followed by 40 bp of a strong Orc1 binding site. Gel shown is representative of purity levels of all DNA constructs post PCR amplification and cleanup with Qiagen PCR cleanup kit.* ***B****: SYBR-safe DNA agarose gel and corresponding PorosHQ trace separating primers and plasmid DNA from amplified DNA fragment across a 200 mM – 1 M NaCl gradient. Fractions containing desired fragment size of DNA were pooled and concentrated via ethanol precipitation and resuspension in Tris Buffer.*

3.3 Reconstitution of the human pre-RC in a bulk biochemical assay

3.3.1 Introduction

Bulk biochemical assays have been fundamental in identifying key intermediates involved in the helicase loading process. These assays have been developed in the past for the *S. cerevisiae* (Coster et al, 2014) and *D. melanogaster* (Schmidt et al, 2020) replicative helicase loading systems and have served to identify the order of protein assembly and key salt-dependent intermediates involved in the process.

As of yet, an equivalent assay has not been published for the *H. sapiens* assembly pathway. Dr. Vera Leber in our research group has been fundamental in successfully reconstituting human helicase loading, with input from myself and other members of the lab for specific optimisation steps. The *in vitro* replicative helicase loading assay presented here is a powerful tool that will enable the research community to further address origin selection, licensing and mutational defects associated with faulty helicase loading.

3.3.2 Assay

As described in Chapter 2.2.3, Orc1-5, Orc6, Cdc6. Cdt1 and MCM2-7 were incubated in a 1:2:2:1:1.5 molar ratio respectively (Figure 28.A.), then DNA conjugated to streptavidin-coated beads via a biotin linker was added. Further information on DNA optimisation will be discussed in Chapter 3.4.2. After removing unbound proteins through multiple low-salt wash steps, unbound DNA was digested with DNAseI resulting in the elution of protein-bound DNA fragments that would then be analysed by SDS-PAGE analysis.

In low salt buffer, recruitment of Orc1-5, Orc6, Cdc6, Cdt1 and MCM – i.e., all proteins involved in the assembly process – was observed (F). Recruitment of all these proteins is highly dependent on the presence of Orc1-5, as shown in Figure 28.B, Lane 2, where a reaction

mixture in the absence of Orc1-5 shows no recruitment of any pre-RC proteins; demonstrating that as in all other eukaryotic helicase loading pathways, binding of ORC to DNA is the first step. The removal of Orc6 from the reaction has little effect on the recruitment of the remaining pre-RC components in low salt buffer (Figure 28.B, Lane 3). However, removal of Cdc6 or Cdt1 appears to reduce the efficiency of loading of the MCM2-7 hexamers, with the absence of Cdc6 showing a more pronounced effect in low salt conditions (Figure 28.B. Lane 4-5).

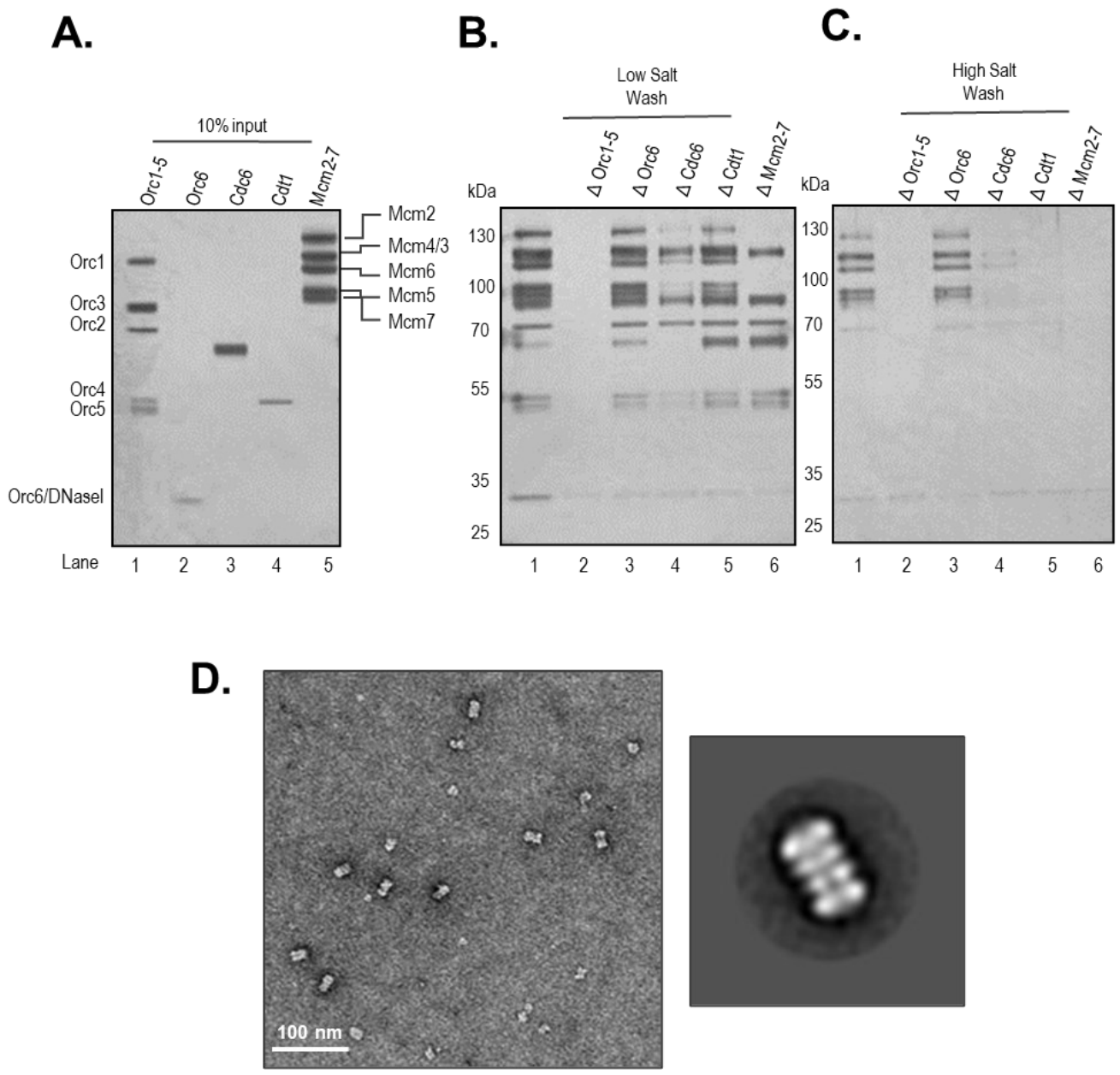


Figure 28: The pre-RC Assay – A method of testing double hexamer assembly

A: *Silver-stained SDS-PAGE gel containing 10% Input of all pre-RC components involved in helicase loading.* ***B:*** *Low salt wash following pre-RC protein incubations with biotinylated DNA. Individual reactions containing a drop-out of an individual component were also run; all components can be recruited to the DNA unless Orc1-5 is absent, showing that association of Orc1-5 to DNA is the first step in pre-RC assembly.* ***C:*** *High salt wash following pre-RC protein incubations with biotinylated DNA. Double hexamer can efficiently form only if all pre-RC components are present or of only Orc6 is absent.* ***D:*** *Negative stain micrograph and corresponding 2D class showing head-to-head Mcm2-7 double hexamer following high-salt wash and elution from biotinylated beads. NS-EM data collected and analysed by Dr. Vera Leber*

One caveat to this analysis is that Cdt1 runs at the same molecular weight as Orc4 on an SDS-PAGE gel, making it difficult to distinguish between the two proteins. While evident in a Cdt1 dropout experiment that MCM2-7 can still be recruited in the absence of Cdt1 (Figure 28.B. Lane 5), it is not possible to determine whether Cdt1 binds to ORC-Cdc6-DNA when no MCM2-7 hexamers are present. One way of potentially studying this could be using fluorescently labelled Cdt1 and identifying the presence of this band though fluorescent excitation of the fluorophore on an SDS-PAGE gel, or by Western Blotting with a Cdt1 antibody.

Overall, these low salt washes provide a valuable insight into the order of assembly of the pathway. This sets the foundation for characterisation of individual complexes and how they bind to each other, which will be discussed in later chapters.

In previous helicase loading experiments in *S. cerevisiae* and *D. melanogaster*, washing the assembled complexes with a high-salt buffer strips away all unbound protein, except for the fully assembled MCM2-7 double hexamer as it encircles DNA. This is also the case with *H. sapiens* proteins, where a 500 mM NaCl wash following incubation of all pre-RC components with biotinylated DNA results in only the MCM2-7 bands visible on an SDS-PAGE gel (Figure 28. C.). Some residual Orc3 remained after the washes, although because of the lack of other ORC subunits and the significantly lower intensity in comparison to the MCM subunits, it is considered a residual artefact of the washes. Characterisation of this salt-stable moiety eluted from the beads by negative-stain EM confirms the presence of a symmetric, head-to-head MCM2-7 structure, suggesting full double hexamer formation (Figure 28.D.).

Double hexamer formation is not possible in the absence of Orc1-5 or Cdt1 (Figure 28.C. Lane 2, Lane 5). Interestingly, it appears that Orc6 is not essential for MCM2-7 loading, with salt-stable MCM2-7 still being observed even in its absence (Figure 28.C. Lane 3). The role of Orc6 has been controversial, since it does not form a complex with Orc1-5 but is still essential for productive DNA replication and cytokinesis in *D. melanogaster* (Balasov et al, 2019). This suggests that – despite its namesake – Orc6 may not be involved in the core helicase

assembly pathway but is required in the context of chromatin or downstream, either during helicase activation or during bidirectional replication itself.

Another interesting observation is the faint presence of MCM2-7 bands present in the high-salt wash of pre-RC samples missing Cdc6 (Figure 28. C. Lane 4). While it is uncertain whether this is double hexamer (this could be confirmed in future experiments by upscaling and negative-stain EM), the presence of this suggests some form of salt-stable MCM2-7 complex. This is especially interesting when we consider the equivalent reaction under low-salt wash conditions (Figure 28. B Lane 4) – here the recruitment of MCM2-7 is relatively low in comparison to samples missing Cdt1 or Orc6. However, the low level of retention during a high-salt wash might indicate some double hexamer formation. Some studies (Shibata et al, 2016) suggest that replication can still occur in some cells even in the absence of Cdc6, although this does affect downstream cellular health and viability. It could be that in the absence of Cdc6, as long as all other pre-RC components are present, some double hexamer may be assembled, allowing for some helicase formation and activation. The lack of double hexamer presence in reaction mixtures missing Orc1-5 and Cdt1 show that this result is not due to chance self-assembly of two single MCM2-7 hexamers. While controversial, this could be explored in the future to identify if Cdc6 is a potential target to reduce helicase loading in therapeutic situations.

3.3.3 Discussion

Overall, this assay has provided a reproducible platform where we can test MCM2-7 helicase recruitment using human proteins. Published literature in the past has established assays with paralogous pre-RC components from different model organisms (as discussed in Section 1.9), notably S. *cerevisiae* and D. *melanogaster*. These assays have been the basis for trapping specific intermediates along the assembly process, such as the yeast ORC-Cdc6-Cdt1-MCM (OCCM) (Miller et al., 2019, Riera et al. 2017, Fernández-Cid et al., 2013) the yeast ORC-MCM (Miller et al, 2019) and the Drosophila ORC-Cdc6-DNA (Schmidt and Bleichert, 2020).

Based off these results, there is now the potential to study these intermediates using human proteins, to identify similarities and differences between species. This assay will be valuable for the identification of stable, long-lived intermediates that can be translated into high-resolution microscopy. It also offers the potential to study clinically relevant mutations that may affect double hexamer assembly, as well as identify key bulk intermediates in the loading pathway. Some of these will be discussed in the following sections.

3.4. Identifying Intermediates in the pre-RC pathway

Because of the establishment of reproducible protein stocks of full-length human ORC1-5, Orc6, Cd6, Cdt1 and MCM2-7 and an assay where MCM2-7 double hexamer formation is testable, the next steps were to identify key intermediate complexes that these proteins form in the recruitment pathway at a structural level.

3.4.1 Single-particle analysis Electron Microscopy of human ORC alone

While there have been published electron microscopy structures of human ORC complexes (Tocilj et al, 2017, Jaremko et al, 2020, Cheng et al, 2020), most feature mutants of the proteins with significant truncations, especially in the N-terminus of Orc1. At the same time as this study another group (Cheng et al, 2020), also focused on same aim of using full length proteins, were able to purify but were not able to resolve to high resolution the remaining missing parts of the ORC complex, particularly the N terminus of Orc1 and key regions involved in WHD flexibility in Orc2. In addition to this, pre-RC activity was not tested in this study. After our own successful optimisation of a reproducible method to generate full-length Orc1-5, we decided to see if it was possible to gain a higher-resolution structure of full-length Orc1-5 alone. This potentially could allow the identification of previously unmapped regions that are key to protein functionality. This process required significant optimisation, as discussed below.

Negative Stain EM

A range of concentrations of Orc1-5 following gradient-density centrifugation purification were applied to carbon-coated copper grids, as described in Chapter 2.2.10.1. Initially, these samples were diluted into the same buffer as in the gradient-density centrifugation experiments, albeit without glycerol. However, upon analysis of the negative stain grids and initial 2D classification, these conditions seemed to result in significant particles dissociation (Figure 29 A). This could potentially be due to the low p.H. of the uranyl acetate stain (4.2) that could be contributing to the denaturing of the complex. To rectify this, samples were cross-linked in 0.1% glutaraldehyde for 15 minutes on ice before being applied to the grids which appeared to solve the dissociation problem, showing monodispersed particles of a constant size and shape (Figure 29 B.).

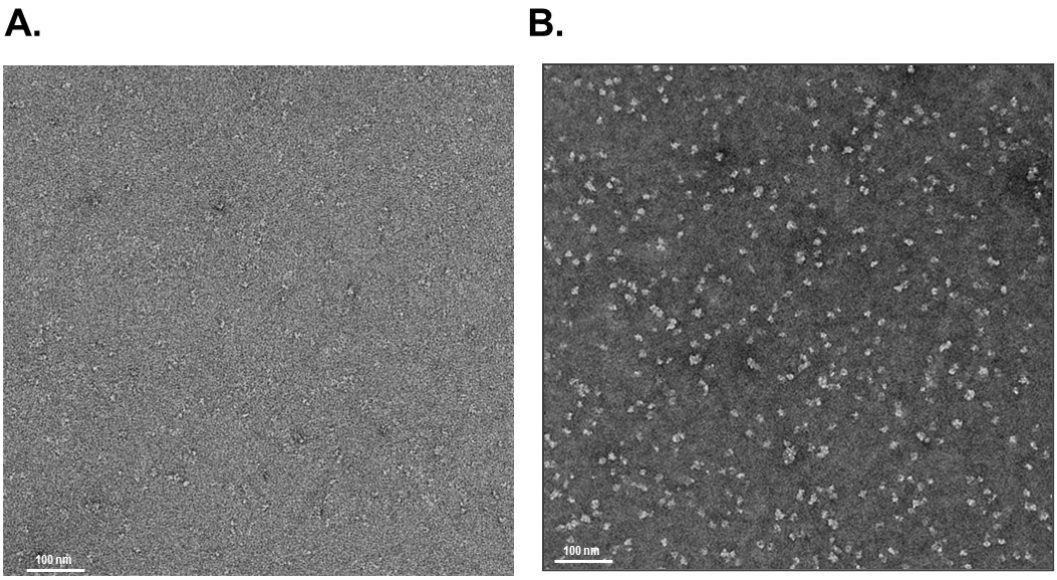


Figure 29: Cross-linking is crucial for ORC1-5 integrity in EM.

A: *85 uM of ORC1-5 applied directly onto negative stain EM grids, without crosslinking. In this condition, it is not possible to identify assembled ORC, and small, likely disassembled protein subunits are seen (white patches). It is likely that Orc1-5 falls apart on the grid upon the addition of the low p.H. uranyl acetate stain (p.H. 4.2) if it is not stabilised in some way.* ***B:*** *85 uM of ORC1-5, cross-linked with 0.1% glutaraldehyde applied directly onto negative stain EM grids. Here, it is possible to distinguish larger particles, assumed to be assembled ORC1-5, suggesting that to maintain structural rigidity in a negative stain environmental condition, the complex must be pre-crosslinked.*

For a small data collection (198 micrographs), negative stain grids were imaged using a TALOS 200kV microscope fitted with a Falcon 3 detector at Imperial College London. The microscope parameters used, and details of the data collection are outlined in Supplementary Table 7.1.

Following dataset collection, the micrographs were processed using RELION-3.1. Micrographs were imported, then underwent CTF correction using CTFFIND4 (Rohou and Grigorieff.,2015); then CTF-corrected micrographs were manually filtered to remove images with large aggregates or showing signs of mechanical drift in the thon rings of the FFT image. The micrographs of good quality were then used for downstream image processing.

First, 1000-1500 particles were manually picked with a box size of 200A in diameter and 2D classified to ensure these could be averaged. Selected 2D classes of the expected correct shape of Orc1-5 were selected and used for reference-based autopicking, then particles were extracted, resulting in ~ 124,000 particles. Extracted particles were once again 2D classified, then underwent filtering to remove junk particles. The 2D class averages displayed a range of particle orientations, featuring a ring-like complex with a protrusion. Comparing this to published structures of human Orc1-5, this protrusion is likely to be the Orc3 subunit, and was used as a basis to predict the arrangement of the other subunits surrounding it (Figure 30). Dimensions of the 2D class averages were measured, which were found to closely agree with the dimensions of previously published structures.

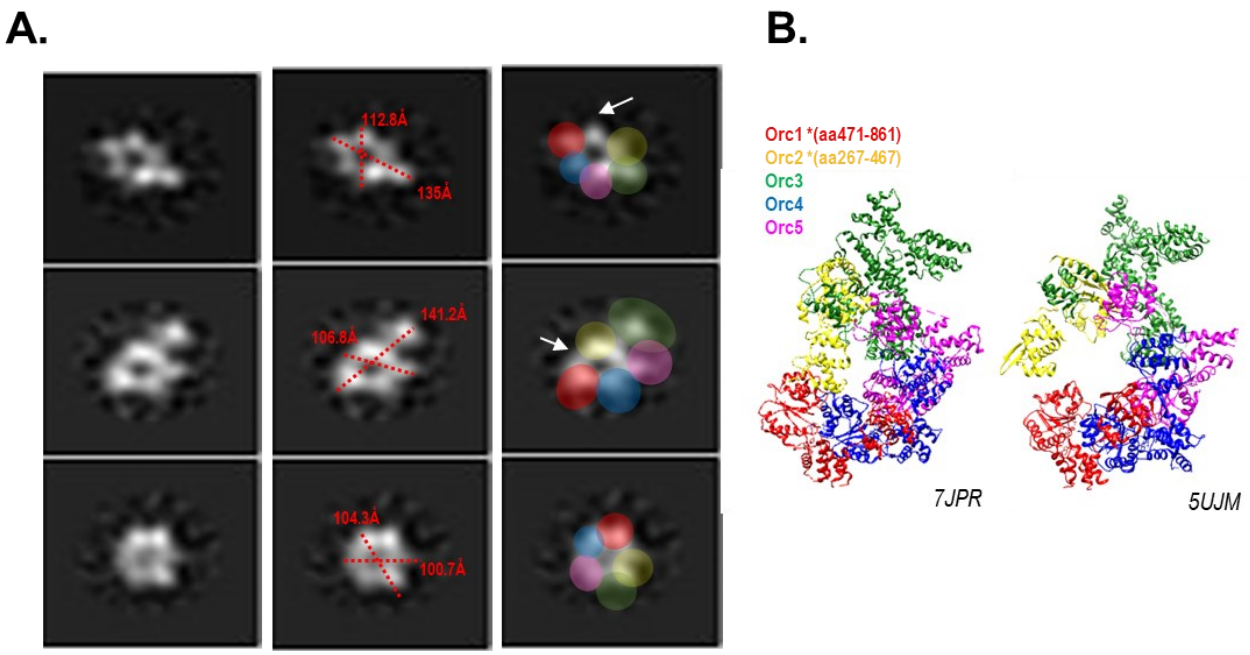


Figure 30: Purified full-length Orc1-5 shows the same architecture as other ORC structures.

A: *2D classes from ORC1-5 negative stain RELION processing, showing that Orc1-5 arranges into a ring structure with similar dimensions to published ORC structures* ***(B)****. Each row indicates a separate 2D classes with corresponding measurements and mapping of predicted subunits labelled by colours. A protrusion that is likely Orc3 (labelled in green) can help to map the rest of the subunits in the 2D classes, with potential extra density in the region of Orc1/2 (white arrows). 7JPR and 5UJM are from Jaremko et al, 2020 and Tocilj et al, 2017 respectively).*

Additional density in the predicted regions of Orc1 and Orc2 (marked with white arrows) showed promising results in potentially allowing further structural data to be elucidated with higher-resolution methods, namely cryo-EM.

Cryo-Electron Microscopy

Because of the promising low-resolution 3D models generated from the negative stain analysis, the next step was to optimise the Orc1-5 sample for cryo-electron microscopy, to try to gain a 3D structure beyond the resolution constraints of negative stain.

Optimisation of cryo-grid preparation

Multiple rounds of optimisation were necessary to find appropriate conditions for data collection of the Orc1-5 sample. Initially, we tried to produce cryogrids using Quantifoil R2/2, 300 mesh (Quantifoil); the motivation behind this was that open-hole grids would potentially provide the lowest level of background against the sample, allowing for a greater single-to-noise ratio (SNR) and thus improved resolution of the final EM map. However, in this case, we found that all of the sample would aggregate on the carbon support between the holes instead of freezing in a vitrified layer in the open holes. Because of this, we decided to explore carbon-supported grids with an either commercial or home-made 2 nm carbon coating. These will be the grids used in all other applications, unless otherwise stated. All samples were prepared with a Mark IV Vitrobot (FEI) system, unless otherwise stated.

Initially, the same amount of protein as applied to negative stain grids (50 uM, cross-linked with 0.1% glutaraldehyde for 15 mins) was applied to the cryo-grids, diluted in a buffer containing 25 mM HEPES-KOH p.H. 7.5, 250 mM KOAc, 3 mM ATP and 1 mM DTT. The sample was incubated on the grid for 30 seconds at 4°C temperature in a humidity controlled chamber, before being blotted away for 0.5 seconds with a blot force of 2. The sample was immediately plunge-frozen in liquid ethane and stored in liquid nitrogen until imaging. These conditions ended up producing grids where the sample was in the holes (unlike the open-hole grids), although highly aggregated and without much dispersity across the holes (Figure 31 B). Increasing the concentration to 150 uM under these same conditions only introduced more aggregation (Figure 31 C.). We then decided to introduce a detergent into the buffer to try to reduce the aggregation; the sample was once again diluted to 50 uM in the same buffer supplemented with 0.05% Triton-X100. This appeared to improve the aggregation issue, although the particle density was extremely low with high background, making it difficult to distinguish individual particles (Figure 31 D). We then increased the sample concentration to 150 uM once more, while simultaneously increasing the blot time of the sample to 1 second before plunge-freezing. This improved the contrast between the particles and the background and resulted in some of the particles arranged in a monodisperse layer, although there was

still a significant level of aggregation (Figure 31 E). To reduce this aggregation, we then decided to increase the Triton X-100 concentration to 0.1%. This resulted in cryogrids with majority monodispersed particles with clear distinguishability between the background and the sample (Figure 31 F). These conditions were chosen for large-scale data collection and processing, as described below.

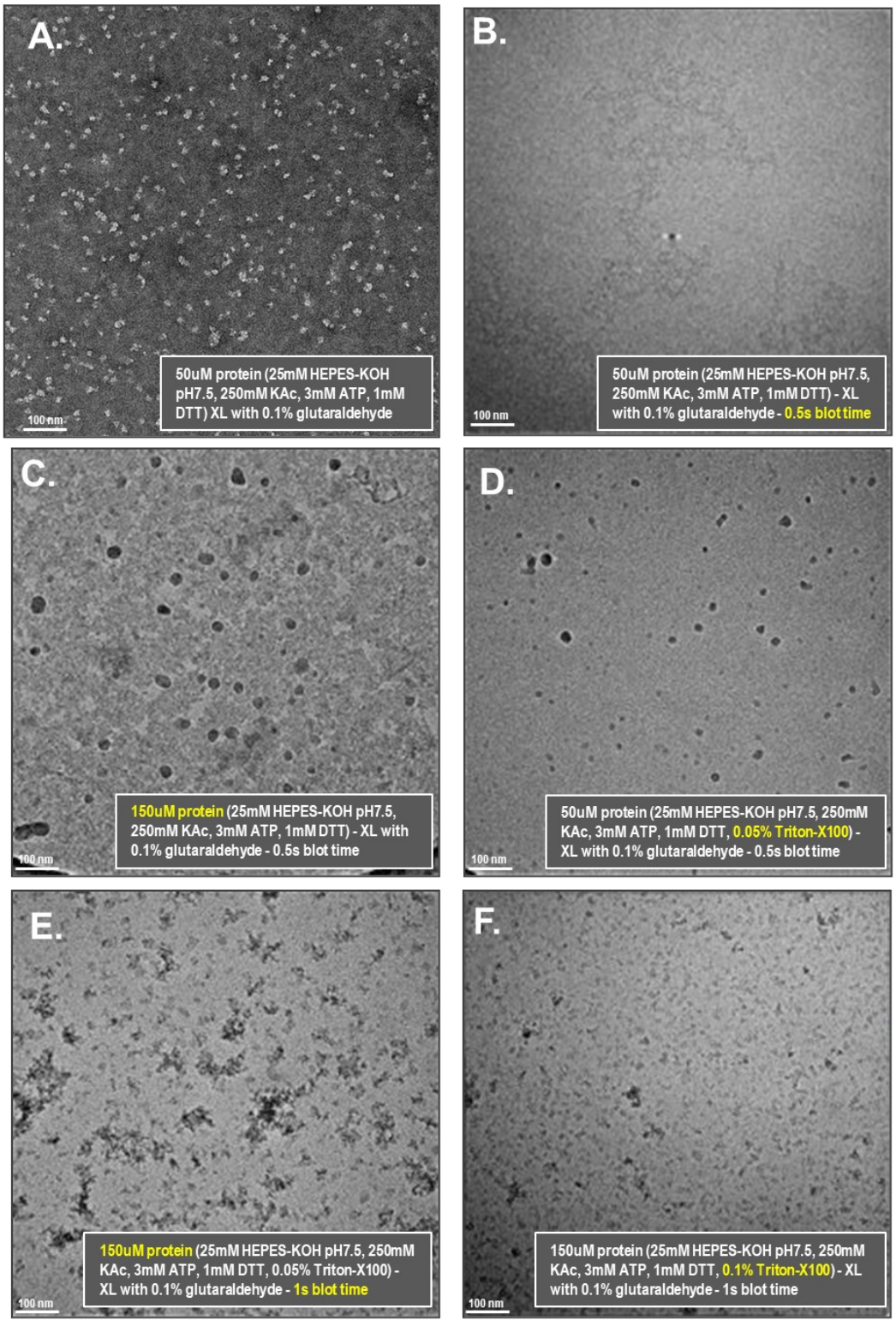


Figure 31: Optimisation of cryogrids – Orc1-5 alone

A-F: *Optimisation conditions for monodisperse and homogenous Orc1-5 proteins on a cryogrid. Individual variables are discussed in the figure. Optimisation of sample from negative stain to cryo-election microscopy involved finding an appropriate protein concentration, grid type, detergent concentration and sample blot time. Optimised conditions (**F**) were used for preparation of future cryogrids with additional protein components.*

Cryo-Electron Microscopy analysis of full-length ORC

Two grids prepared in the optimised manner as described above were used for a large-scale dataset collection at the London Consortium for Cryo-Electron Microscopy (LonCEM) at the Francis Crick Institute, supervised by Dr. Nora Cronin. In total, 21,168 movies were collected on a Titan Krios 300 kV microscope equipped with a Gatan K3 electron detector. Data collection parameters are presented in Supplementary Table 7.2. All movies were processed using RELION-4.0.

Movies were motion-corrected using MotionCorr2 (Zheng et al, 2017), then CTF-corrected using CTFFIND (Rohou and Grigorieff, 2015) – resultant micrographs were used for particle picking. Because of the large size of the dataset, a subset of 5000 micrographs were selected at random to determine the optimum parameters for particle picking. All three particle picking options (Gaussian-based LoG picker, template based autopicking and Topaz training and picking) were explored to find the best option.

Blob-based LoG picking

A minimum – maximum diameter size for the LoG filter between 100 – 250 A was chosen, with a minimum inter-particle distance of 200 A. From a randomised selection of 100 micrographs, a total of 120,843 particles were picked. These were extracted then sent for 2D classification.

This method did not yield valuable results – despite the picking job appearing to select individual particles, 2D averaging showed that most of what was picked could not be averaged into distinguishable particles or was junk (Figure 32.B.).

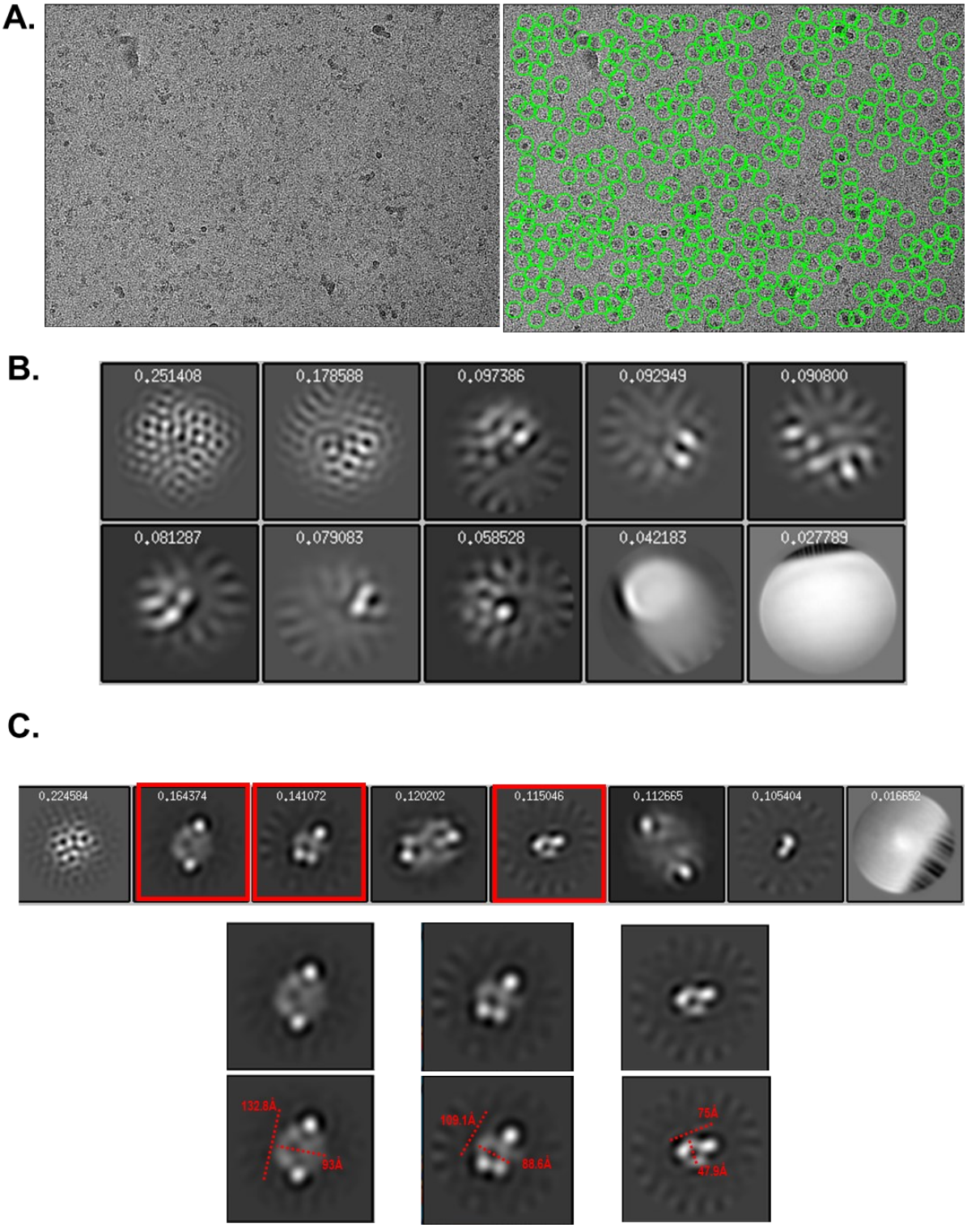


Figure 32: Relion-Integrated Picking parameters for the Orc1-5 dataset

A: *Representative micrograph and associated picking coordinates (green circles) for the Orc1-5 cryo-EM sample, using the RELION-integrated blob-based autopicking parameters.* ***B:*** *2D class averages generated from a total of ~140,000 particles, using the Gaussian-based LoG picker in Relion-4.0. Using these picking parameters did not result in 2D class averages with distinguishable ORC features, with classes averaging junk particles and grid support background.* ***C:*** *2D classifications and particles showing ORC-like dimensions from a total of ~2,000 particles that were manually picked. This improved averaging, with some resultant classes (highlighted with red boxes) showing ORC-like ring with corresponding dimensions. Some of these seemed slightly small (right - ~75 Å across), which could indicate some semi-formed ORC complexes, or ORC molecules with regions with poor signal-to-background noise.*

Reference-Based Autopicking

To explore reference-based picking, 2000 particles were manually picked from the micrograph selection (particle diameter of 200 A), then 2D classified (Figure 32.C.). This method yielded more positive results, with distinguishable classes containing particles of the correct dimensions, although these appeared to contain significant preferential orientation with spots of extremely high intensity (Figure 32. C). Nonetheless, selected classes containing particles with distinguishable ORC features were chosen as templates for reference-based autopicking on a randomised subset of 500 micrographs. ~41,000 particles were picked, then classified into 30 2D classes to try to separate out any preferential orientation and regions of intense density (Figure 33. A.). These classes were subject to three rounds of repeat 2D class selection and filtering to remove junk particles (41,521 - 28,053 particles), to remove particles of featuring sizes too small or too large to be ORC or junk classes. The resultant 2D classes showed an improved selection of particles with distinguishable Orc1-5 features (Figure 33. B.), but these were not deemed good enough to progress to further filtering or initial 3D modelling.

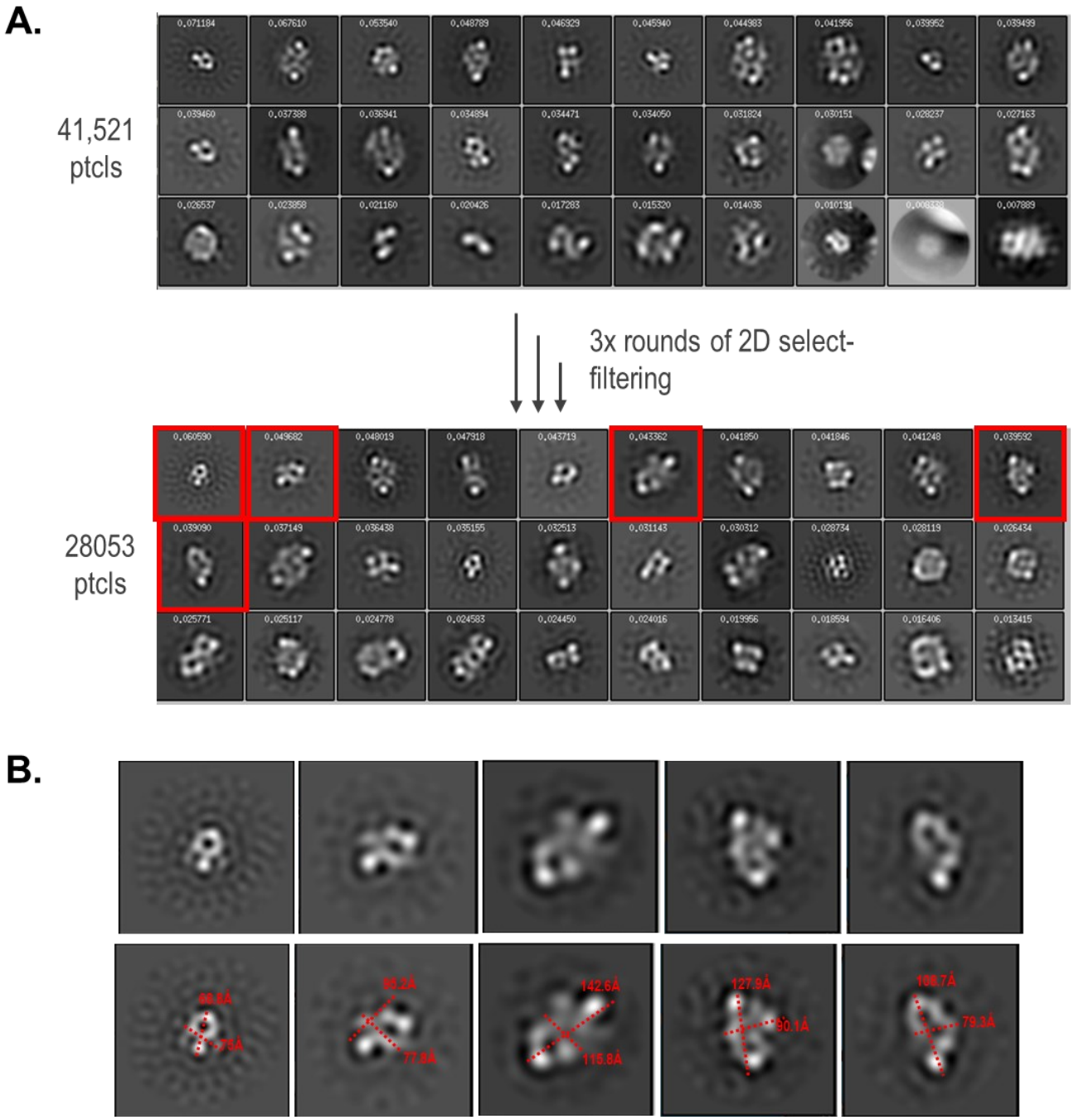


Figure 33: 2D classification and filtering of Reference-based picked particles

A: *RELION 2D classification of autopicked co-ordinates, using 2D classes showing ORC-like features from manually picked particles (Figure 32.C) as templates for autopicking conditions on a random subset of 500 micrographs. A total of 41.5 thousand particles were picked and classified into 30 2D classes to separate out junk classes, then reclassified twice, removing classes that did not contain particles that could be averaged into distinguishable classes, down to 28 thousand particles* ***B:*** *Selected 2D classes from A, showing correct diameters associated with ORC, but unimproved resolution.*

Topaz-trained Autopicking

We also tried using the Topaz autopicker – a trainable neural network that has shown to recognise and correctly pick a large number of particles with a low false-positive rate, especially useful for picking unusually shaped proteins (i.e. asymmetric or non-globular particles (BEPLER ET AL, 2019). A total of 52,510 particles were picked from 1,000 micrographs (Figure 34. A) using pre-train model weights. The particles were subjected to 2D classification, with some classes displaying distinguishable ORC views at an intermediate resolution, with some secondary structure features (Figure 34. B-C).

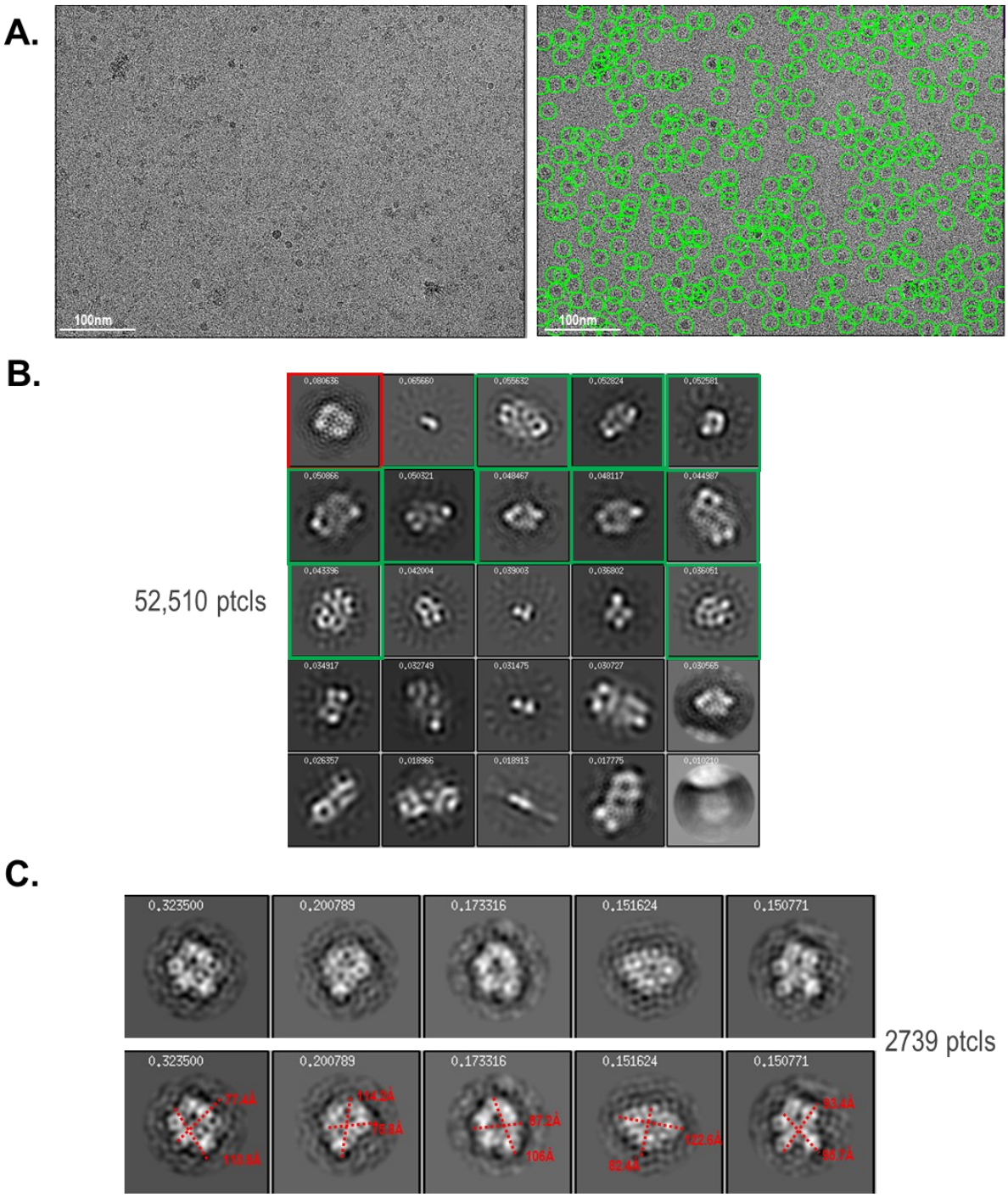


Figure 34: Topaz-trained autopicking

A: *2D micrographs and corresponding picking coordinates using RELION-intergrated neural network software, Topaz. 52.5 thousand particles were picked from a randomised selection of 1,000 micrographs, subjected to 2D classification* ***B:*** *Extracted and 2D classified particles from 1,000 micrographs, with highlighted classes (red and green boxes showing what appears to be secondary structure elements.* ***C:*** *sub-2D classification of the most populous class from B (highlighted in red).*

This was clearly the best method out of those explored, so this method was employed to pick and classify 308,719 particles from 5,000 micrographs (Figure 35 A. The particles were subjected to multiple rounds of 2D classification to remove junk, with the final iterations started to display distinguishable ORC views at an intermediate resolution but did not average to show secondary structure features (Figure 35 A-C). Nonetheless, the particles showing distinguishable ORC features (125,048 particles, Figure 35 B) were stringently filtered and CTF correction was applied. While this did not improve secondary structure features, these particles had the correct dimensions of ORC (Figure 35 C).

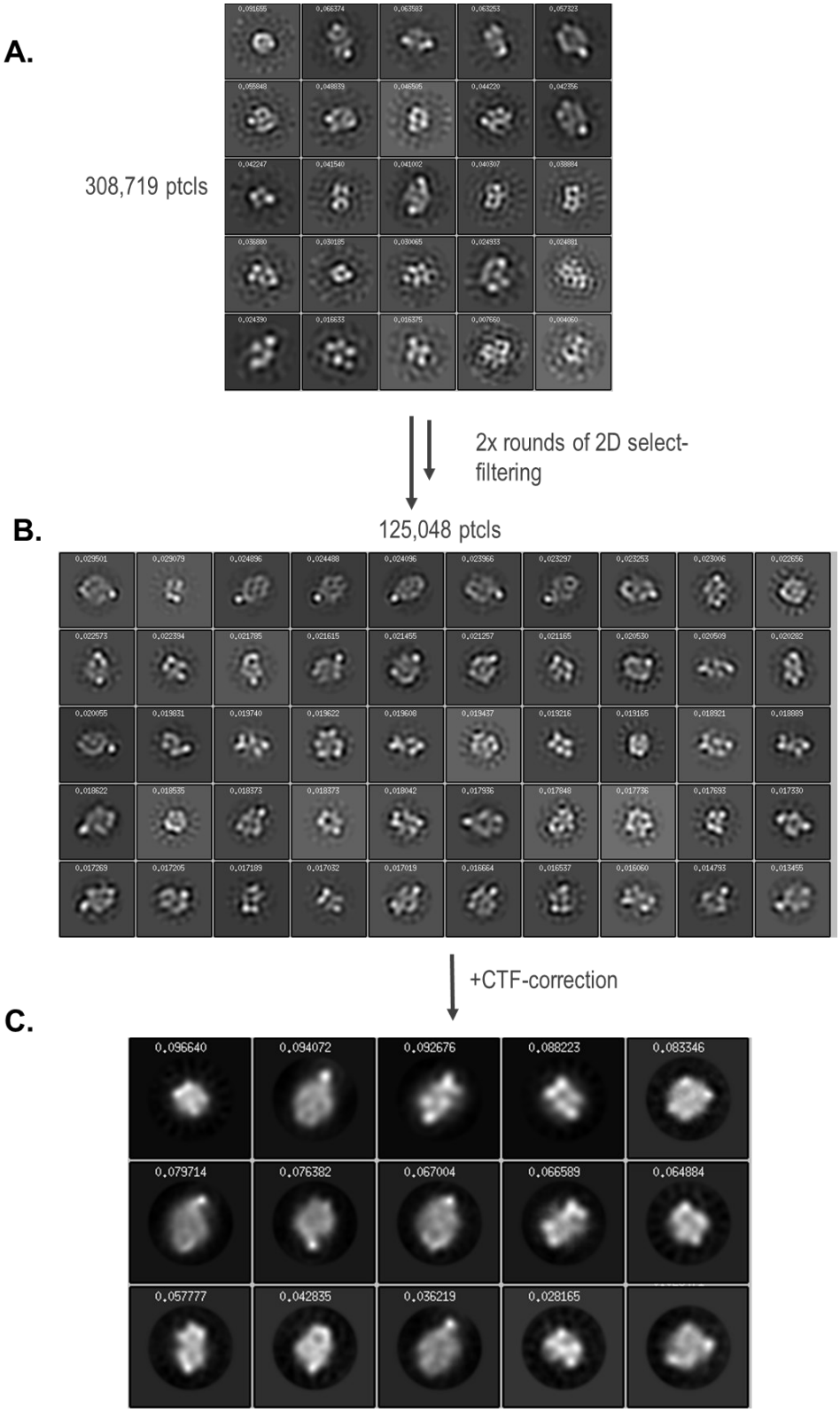


Figure 35: Topaz autopicking on a large scale

A-B: *Extracted, 2D classified and filtered particles from 5,000 micrographs, using Topaz autopicking parameters described above. 308.7 thousand particles were picked and classified into an initial 30 2D classes. These underwent two more rounds of filtering by selecting classes containing distinguishable ORC-shaped densities, removing junk classes and re-classifying the remaining particles until all classes contained ORC ring densities (B)* ***C****: Application of CTF-correction to Classes from B, to attempt to improve secondary structure features.*

Filtered particles were sent for single-class ab-initio 3D modelling (Figure 36. A) with a mask diameter of 200 A. While this resulted in a density map with relatively similar dimensions and low-resolution ring-like features in comparison to published ORC structures, there seemed to be a circular heavy density towards the top of the complex that suggests that there may be preferential orientation of the particles, which is also seen in the 2D classes by the presence of a bright spot protruding from the core ring. Fitting previously published models of ORC into the map suggests that this region may be the Orc3 protrusion on the ring (Figure 36 A.). In an attempt to filter out this region of intense density, ab-initio 3D modelling was performed again, except with 6 classes instead of one (Figure 36 B). This appeared to separate this region out, although only one class (Class 4) seemed to correlate with the correct size and shape of published Orc1-5 models (Figure 36 C.).

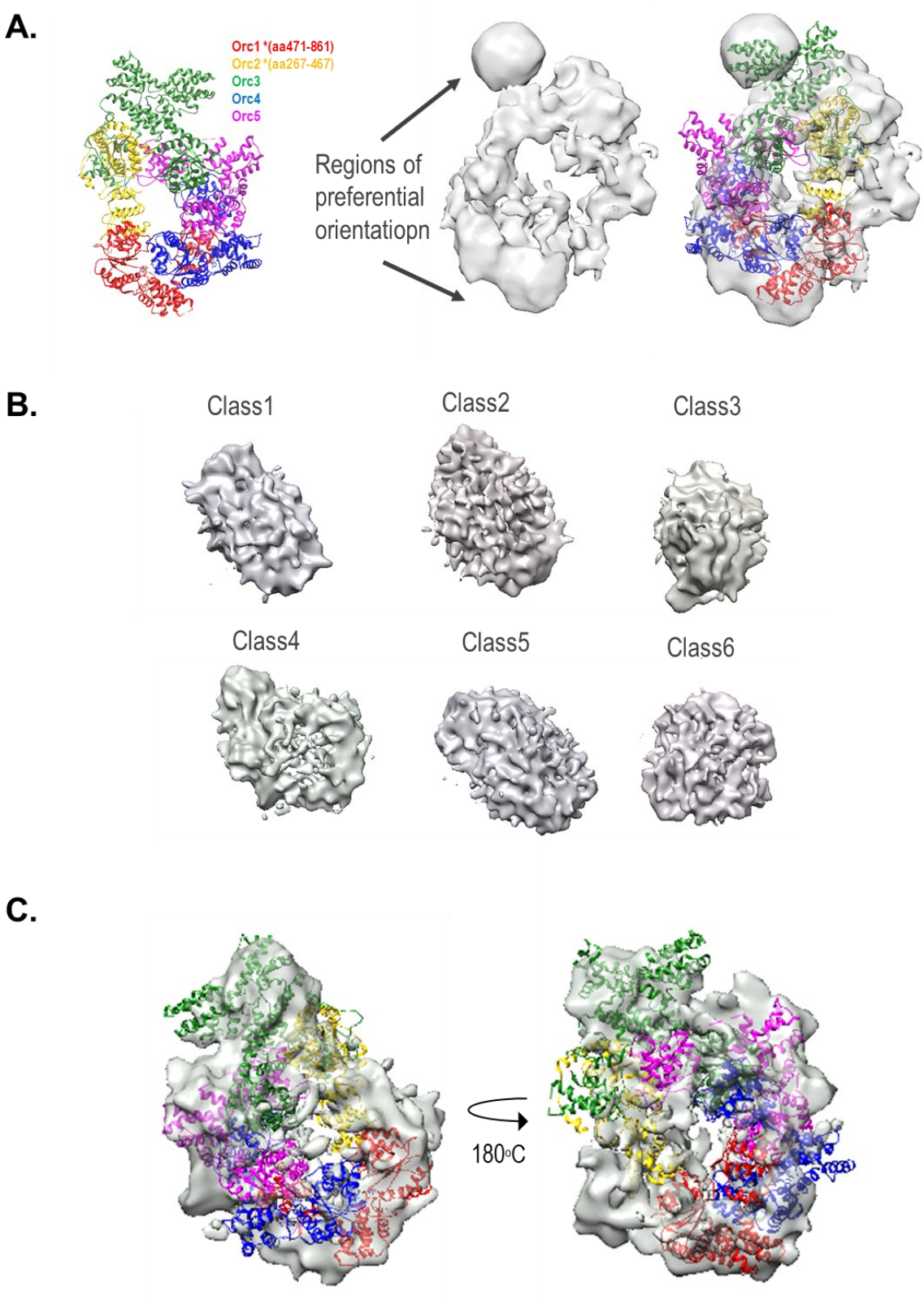


Figure 36: Ab-initio electron density maps of Orc1-5

A: *Ab-initio generated electron density map using particles from Figure 35.B, combined into one class and rigid-body fitted with Orc1-5 published model (7JPR, Jaremko et al, 2020), showing a ring-like structure, but with severe preferential orientation in the Orc3 region (heavy spherical density).* ***B****: Ab-initio generated electron density maps using particles from Figure 35.B, split into six classes, to separate out potential preferential orientation.* ***C.*** *PDB model of Orc1-5 (7JPR, Jaremko et al, 2020) fitted into B) Class 4, seemingly the closest in size and density distribution to published structures.*

Discussion

Overall, despite the large size of this dataset, significant filtering was required before gaining any structural resolution that could be recognisable as Orc1-5. Despite the optimisation of the grids to a level where particles appeared to the naked eye to have a good distribution and limited aggregation, difficulties in data processing showed that this was not necessarily the case. While it could be argued that some extra density is visible in the final 3D ab-initio reconstruction in the region of Orc1 and Orc2, the resolution of the generated density maps is not high enough to characterise these regions at the secondary structure level.

It is likely that flexible regions of Orc1-5 contribute to the lack of resolution both in the 2D classes and the ab-initio models. It was hypothesised that binding the complex to DNA and downstream cofactors may stabilise the protein, potentially providing increased resolution in EM experiments. Because of this, we decided to focussing processing efforts on ORC-DNA samples instead. It is however important to acknowledge that the results described above were obtained from around ¼ of the total dataset collected, due to speed of processing power.

It could be that these results are indicative of the entire dataset – micrographs were randomly selected for all the picking parameters described above, so that we could get an average of the data quality overall. However, optimisation of processing parameters and stringent filtering has the potential to uncover higher resolution features. Increasing the particle number by including the remaining ¾ of the dataset may enable the further separation of high-resolution information from flexible or blurry classes.

3.4.2 Binding of ORC to DNA

The first step in the loading of the double hexamer is the binding of ORC – in the case of human proteins, Orc1-5 – to the DNA. While this is not sequence-specific in most higher order eukaryotes and it remains to be defined what features are involved controlling this process, to keep the system more physiologically relevant we decided to use DNA sequences that have

been shown in a cellular context to promote or allow ORC binding. Four Widom-601 followed by 164-bp sequences were selected from ChIP-Seq studies in human cells, correlating with either strong Orc1 or Mcm7 binding peaks (Figure 15, Supplementary Table 2). These sequences had a range of AT/GC content, to test whether this makes a significant difference in ORC binding; however, identifying precise genomic binding locations along these DNA fragments was beyond the scope of these studies.

ORC is nonspecific in binding to DNA.

As an initial test to see whether the purified Orc1-5 bound to DNA, Orc1-5 was incubated with each DNA construct, titrated from a 1:0.5 – 1:10 molar ratio respectively (Figure 37). Samples were incubated between 15-30 minutes (See Chapter 2.2.8.2 for methodology details), then run on a 0.75% TB-agarose gel at 70 V for 2 hours. These electrophoretic mobility shift assays (EMSAs) would show a shift in the DNA if bound by protein, hopefully allowing the identification of an ideal concentration of ORC required to fully occupy all DNA fragments and to see whether binding was more favourable with specific DNA constructs in comparison to the others.

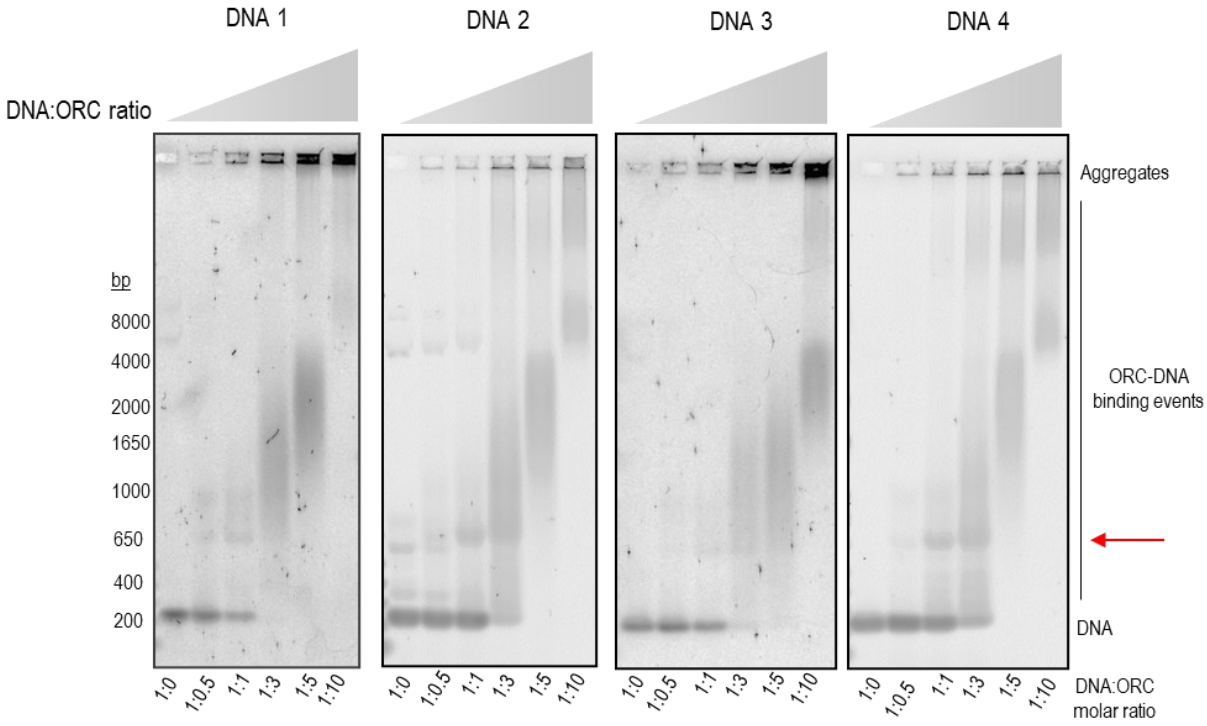


Figure 37: EMSAs show that purified Orc1-5 can bind and remain on DNA

EMSA experiments showing ORC binding to DNA with increasing molar ratios of ORC-DNA DNA. The presence of a band shift indicated protein binding to DNA resulting in slower migration through the agarose gel. A ratio of 1:3 DNA:ORC is sufficient for entire DNA occupancy. The DNA constructs contain the Widom-601sequence, plus 80bp of extranucleosomal DNA, sequences are detailed in Supplementary Table 2.

From Figure 37, it was evident that the purified Orc1-5 complex was competent in binding all the presented DNA sequences, with an optimal DNA:ORC ratio between 1:1 and 1:3. Specific ORC-DNA complexes can be seen for all DNAs, but are more pronounced for DNA 2 and 4 (suggested by the presence of tighter bands), suggesting longer lived intermediates. Except for DNA4, all DNA constructs showed similar binding efficiencies with most of the DNA being shifted at the 1:3 molar ratio. However, even in DNA3, which shows less of a specific band, Orc1-5 could still bind, suggesting at this stage ORC has overall limited preference for a specific DNA. At higher ORC concentrations additional binding events can be observed, as displayed by a larger and more heterogenous shift. In the past, studies have suggested that the ORC footprint is ~48 base pairs – in theory, a ~200-base pair fragment could fit anywhere between 1 to 3 bound ORC molecules. This variability is likely to cause the smearing effect

on the EMSAs, with higher molecular weight regions correlating to more ORC molecules bound to the DNA and vice versa. At the same time, it is important to note that increasing molar excess of ORC results in more of the sample becoming aggregated and stuck in the wells unable to penetrate the gel matrix.

For future ORC binding experiments, a molar ratio of 1:2 DNA:ORC was chosen – this was to ensure mostly full occupancy of the DNA fragments with at least one ORC molecule, but avoided the aggregation seen with higher titrations of protein. While from the EMSAs it does not appear that ORC had a strong binding preference for one specific DNA fragment, thus all subsequent experiments were carried out with DNA1.

Negative Stain Microscopy of ORC binding to DNA

To assess ORC binding to DNA on a structural level, negative stain electron microscopy was initially carried out to ensure the integrity of the complex. Using optimised EMSA conditions described above, DNA and Orc1-5 were incubated in a 1:2 ratio (43:86 nM or 0.03 mg/mL) for 15 minutes at 30°C with shaking at 300 rpm on a tabletop thermomixer, then cross-linked with 0.1% glutaraldehyde for another 15 minutes on ice. Negative stain grids were prepared as described in Chapter 2.2.10.1 and a small dataset was collected to perform an initial 3D analysis of the sample. Collection parameters for the sample are described in Supplementary Table 7.3.

Micrographs were processed in a similar manner as described in Chapter 3.4.1. After CTF correction with CTFFIND4 and manual filtering to remove unsuitable micrographs, ~1500 particles were manually picked and sent for 2D classification (Figure 38. B). While resultant classes were not extremely featureful, the 2D classes appeared to be of roughly the correct size and dimensions, so classes were used for template-based autopicking on the rest of the dataset. A picking threshold of 0.85 and a minimum inter-particle distance of 150 A was chosen, which gave reasonable picking on individual micrographs, avoiding aggregated or overstained areas (Figure 38. A) and giving an overall particle number of ~124,000 paricles.

After filtering junk from the initial 2D classification, it was possible to distinguish the ring-like structure of Orc1-5 with the Orc3 protrusion (Figure 38. C) At this resolution and due to the nature of the staining technique, it was not possible to resolve any DNA-like features - the grain size of the stain limits resolution to ~15-20 Å. Nonetheless, selected 2D classes were used for initial modelling, resulting in a map with distinguishable lobes that could be assigned based on previously published structural data of truncated Orc1-5 encircling DNA (Figure 38 D.). Since the electron density map could fit the model with some extra density (Figure 38 D), it is possible that these are ORC molecules bound to DNA.

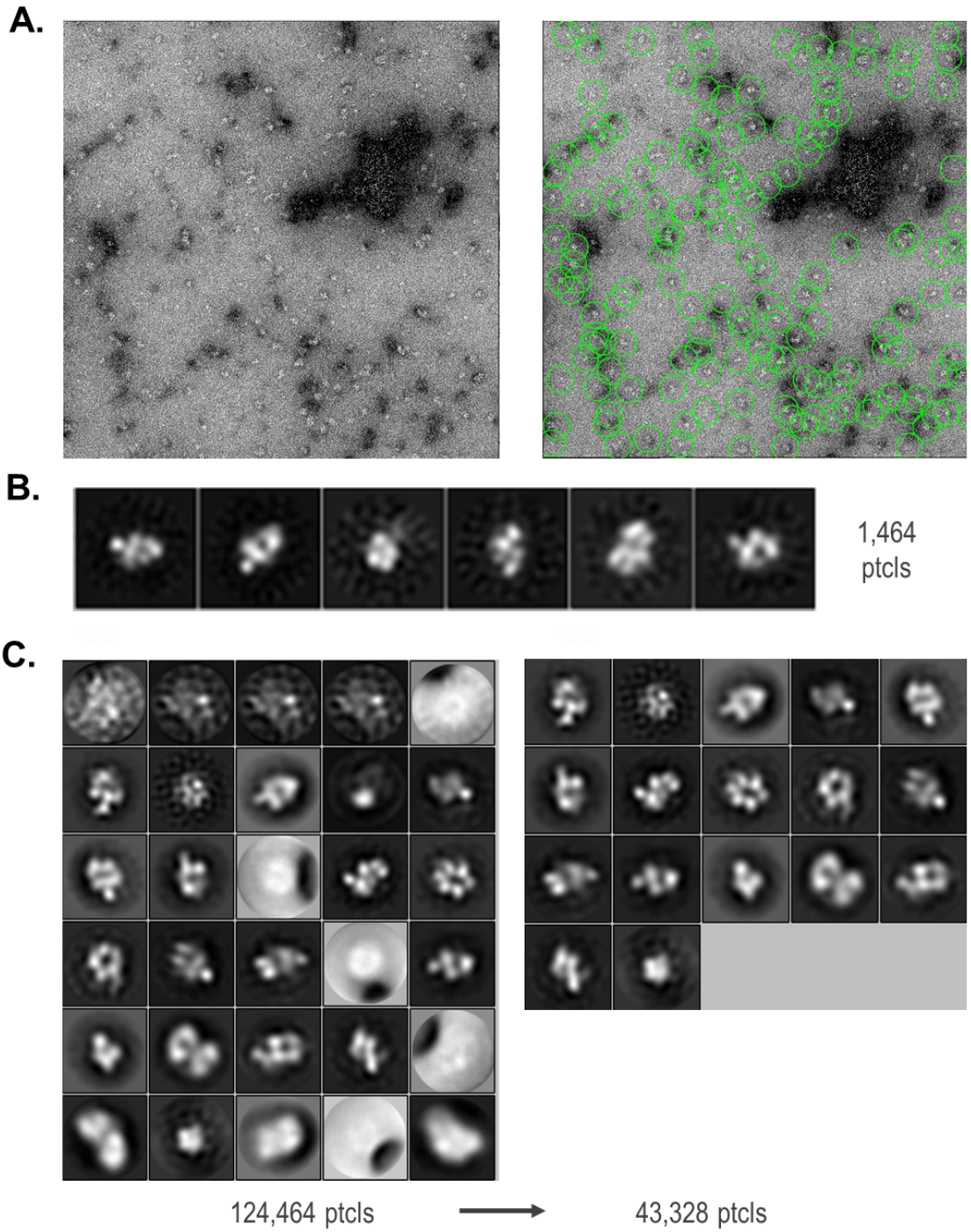

A.
B.
1,464
ptcls
C.
124,464 ptcls
43,328 ptcls
D.
Orc1 *(aa471-861)
Orc2 *(aa267-467)
Orc3
Orc4
Orc5
180°

Figure 38: Negative stain microscopy of ORC bound to DNA.

A: *Negative-stained micrograph and corresponding template-based autopicking of Orc1-5-DNA particles.* ***B:*** *2D classes from a small selection of manually picked particles displaying distinguishable ORC architecture, used for autopicking on the rest of the dataset****. C:*** *autopicked particles, using 2D classes from B as templates, classified into 35 classes. These were filtered to remove junk classes tha did not contain particle features, then reclassified into new 2D classes. Particle number was filtered from 124,464 to 43,328.* ***D:*** *Ab-initio 3D density map generated from cleaned classes from B (43,328 particles), with Orc1-5-DNA (7JPS, Jaremko et al, 2020) rigid-body fitted into the density. This low-resolution fitting manages to fit truncated ORC structures in, with extra densities in the region of Orc1 and Orc2 that may be indicative of newly mapped regions of the protein*

When comparing previously published structures, the best fitting (done via rigid-body fitting of published truncated human ORC structures) was of Orc1-5 bound to DNA (Figure 38 D.), in comparison to Orc1-5 alone, or Orc1-5 with a collapsed Orc2 WHD, suggesting that most of the particles are indeed Orc1-5 bound to DNA. The presence of additional lobes in the region of Orc1 and Orc2 creating a closed ring-like structure suggest that it may be possible with cryo-EM data to characterise extra density of the protein complex that has previously gone unmapped.

Cryo-Electron Microscopy of ORC binding to DNA

The ORC-DNA sample for cryo-EM analysis was prepared as described above, except at a higher concentration (75:150 nM ORC:DNA) in the optimised buffer conditions outlined in Chapter 3.4.1. Cryogrids were prepared in the same manner. Two grids prepared were used for a large-scale dataset collection at the London Consortium for Cryo-Electron Microscopy (LonCEM) at the Francis Crick Institute, supervised by Dr. Nora Cronin. In total, 35,170 movies were collected on a Titan Krios 300 kV microscope equipped with a Gatan K3 electron detector. Data collection parameters are presented in Supplementary Table 7.4. All movies were processed using RELION-4.0.

Movies were motion-corrected using MotionCorr2 (Zheng et al, 2017), then CTF-corrected using CTFFIND4 (Rohou and Grigorieff, 2015) – resultant micrographs were used for particle picking. As with the Orc1-5 only dataset, different autopicking parameters were explored to

find the optimal conditions. Gaussian-based LoG picking was not used since it did not yield usable results in the first dataset.

Reference-based template picking combined with Topaz training.

An increased number of particles (~3,000) were manually picked and sorted into 10 2D classes (figure 39. A). The two classes containing averaged particles where the Orc1-5 lobes were distinguishable were used to train Topaz autopicking on a subset of 100 micrographs. The resultant picking and initial 2D classification (Figure 39 B.) resulted in classes that were recognisable as the ORC shape with potentially a strand of DNA running through the middle (although this cannot be said for certain without further resolution), so these parameters were taken forward for autopicking on a random selection of 1000 micrographs.

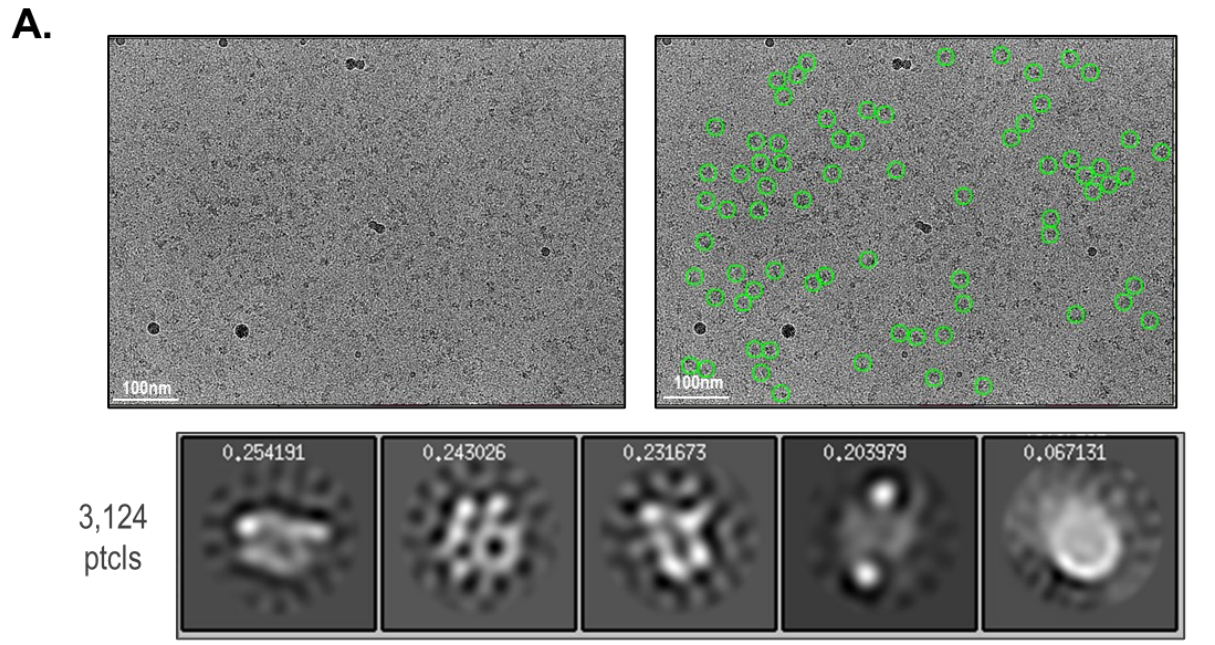

A.
100nm
100nm
0.254191
0.243026
0.231673
0.203979
0.067131
3,124
ptcls
B.
100nm
100nm
100nm
C. i.
24,983
ptcls
ii.
0.156968
0.156968
Potential
DNA
0.147608
0.147608
Potential
DNA
3848
ptcls
3722
ptcls

Figure 39: Reference-based autopicking -optimisation for Orc1-5 on DNA

A: *Manual picking and corresponding 2D classification of 3,124 particles picked from the Orc1-5-DNA dataset. Particles were picked from a randomised selection of micrographs. Highlighted classes were used for Topaz-trained autopicking on a subset of 100 micrographs to give:* ***B: T****opaz trained and FOM0.75-filtered particle picking, extraction and* ***C (i)*** *corresponding 2D classifications, showing classes with ORC-like features* ***(ii),*** *plus a potential DNA-like density running through the central channel (marked in yellow).*

To separate out junk, extracted particles were classified into 100 2D classes, with CTF correction (Figure 40. A). From the ~500,000 particles picked, there was a significant amount of junk, but many classes contained distinguishable Orc1-5 features, including the ring-like structure, individual lobes and the Orc3 protrusion. Crucially, some of the classes contained high-resolution secondary structure features (~150,000 particles) (Figure 40.A-B), which were briefly sub-classified to ensure integrity of the classes (Figure 40 C.).

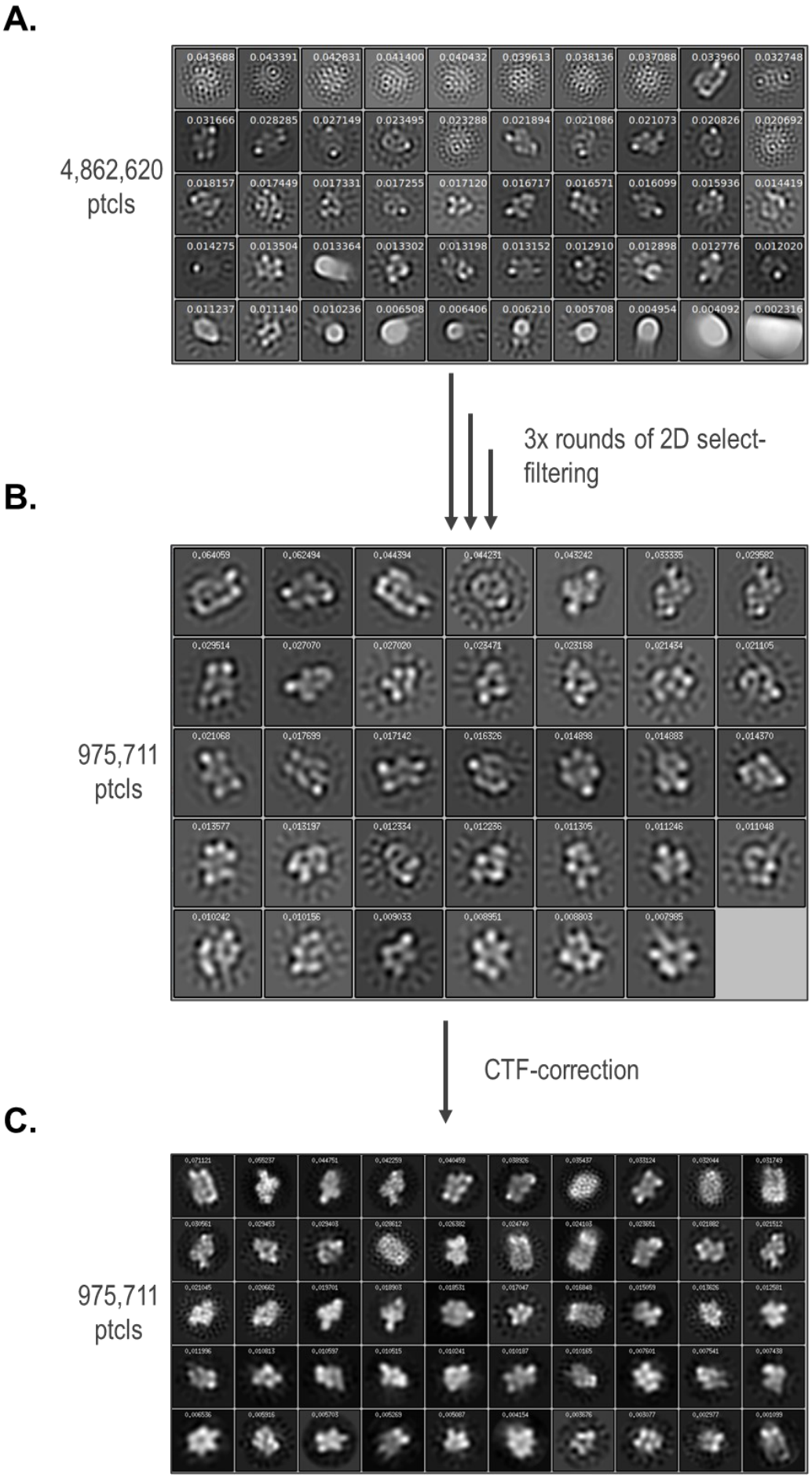

A.
4,862,620 ptcls
3x rounds of 2D select-filtering
B.
975,711 ptcls
CTF-correction
C.
975,711 ptcls

Figure 40: Topaz autopicking and extraction results in particles that can be averaged to show secondary structure elements of Orc1-5-DNA

A: *Particle picking (with CTF correction) and subsequent 2D classes from 1,000 randomly selected micrographs, based on 2D classes from Figure 39. Classes with distinguishable ORC-like features highlighted in green, with classes containing secondary structure elements highlighted in red.* ***B:*** *High-resolution classes from A, reclassified to visualise secondary structure elements* ***C:*** *Highlighted classes from A, extracted and re-classified into a higher number of 2D classes to try to distinguish different particle orientations.*

Selected classes displaying multiple orientations of the particles were then used to re-train Topaz and picked on 1/3 of the dataset (Figure 41). After filtering by FOM0.75 4.8 million particles were extracted and classified into 30 classes as a starting point (Figure 41. A). To remove the junk, this was stringently filtered down to just under 1 million particles and reclassified (Figure 41. B). CTF corrections were reintroduced when classifying, resulting in a mixture of low-resolution classes with distinguishable lobes, but significant unevenness in electron distribution and some higher-resolution classes where some secondary structure elements could be seen (Figure 41. C). All of the classes from Figure 41.C were sent for ab-initio 3D modelling with a mask diameter of 200 A.

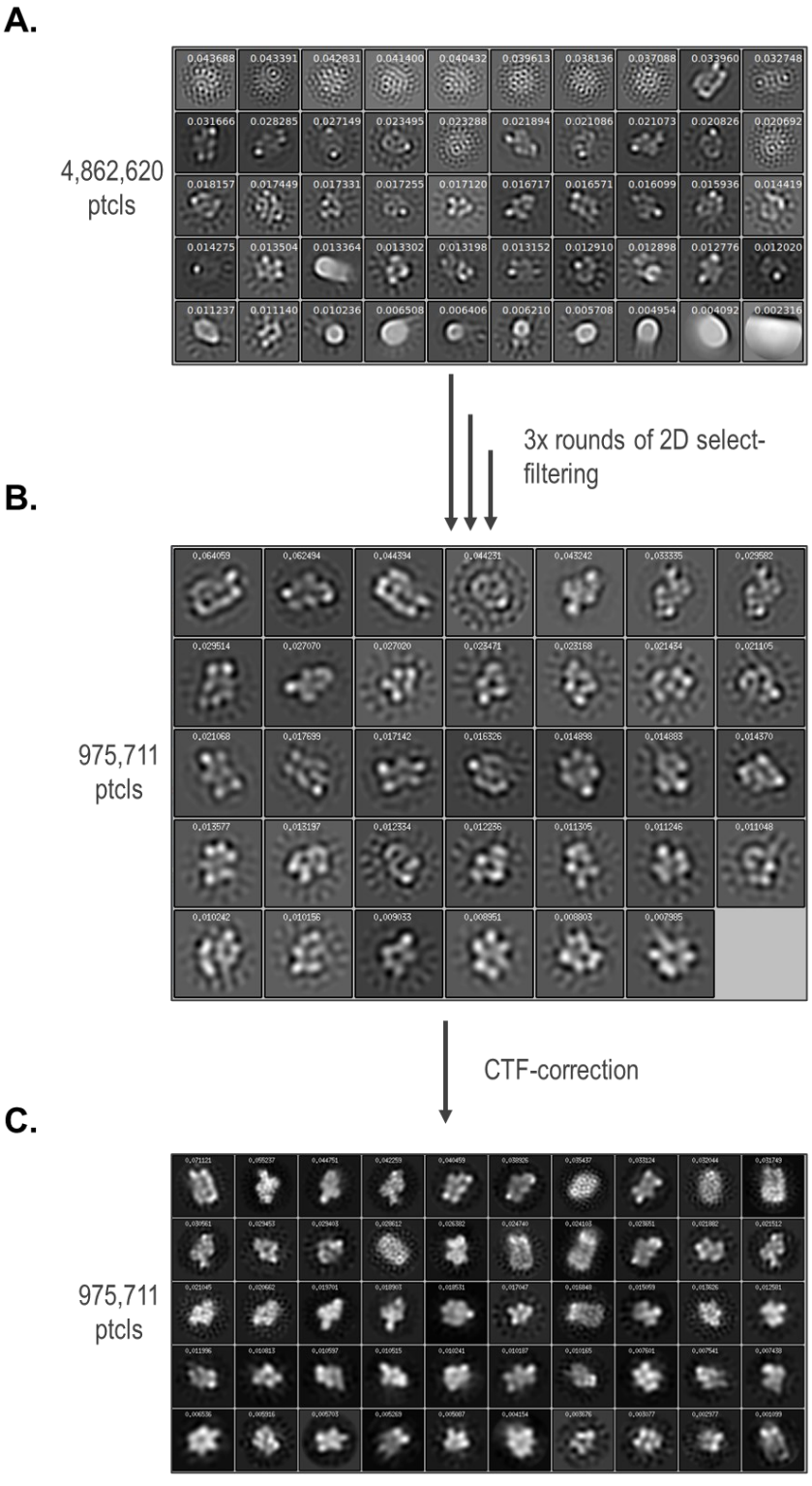


Figure 41: Topaz autopicking on 11,000 micrographs and corresponding 2D classification

*Workflow to show TOPAZ-picked (trained via parameters described in Figure 39), extracted and 2D classified particles from 11,721 micrographs (1/3 of dataset) from **A:** total particles picked (>4 million); 2D classification at this stage shows many classes with high levels of background or junk that were removed, and remaining particles were reclassified. This process was repeated two more times, to give **B:** 2D classes of picked particles that had been filtered to remove junk and high levels of background. Here, all classes contained the correct ORC architecture. **C:** Particles from B were 2D classified again, with the inclusion of CTF correction. Resultant classes contained some secondary structure features. All classes were used for an-initio 3D model generation.*

Ab-initio 3D modelling of Orc1-5-DNA

Selected 2D classes containing multiple orientations of the particles and displaying some secondary structure features (Figure 41.C) were used for ab-initio 3D modelling, split into 5 classes (Figure 42.A). From the resultant initial 3D classes, the optimal candidates for 3D classifications were Classes 1 and Class 5, based on their size, shape, and comparisons with published structures (Figure 42. A-B). These classes had the closest architecture to published ORC models, characterised by rigid-body fitting of published [truncated] Orc1-5-DNA. Particles were then 3D classified into 4 classes using each of these initial models respectively (Figure 42. C). Because it was difficult to distinguish between these 3D classes which could be an optimal density map, different T parameters were explored to place different weighting on the model, to find the best condition to allow for good separation of different particle conformations leading to a more defined EM map. This resulted in a range of 3D density maps; while these were not at a high enough resolution to draw significant structural data, it was possible to distinguish the ring-like structure and the Orc3 protrusion at the top of the complex that could be fitted with published Orc1-5-DNA structures. However, due to the relatively low resolution of these initial maps, it is not possible to state this for certain.

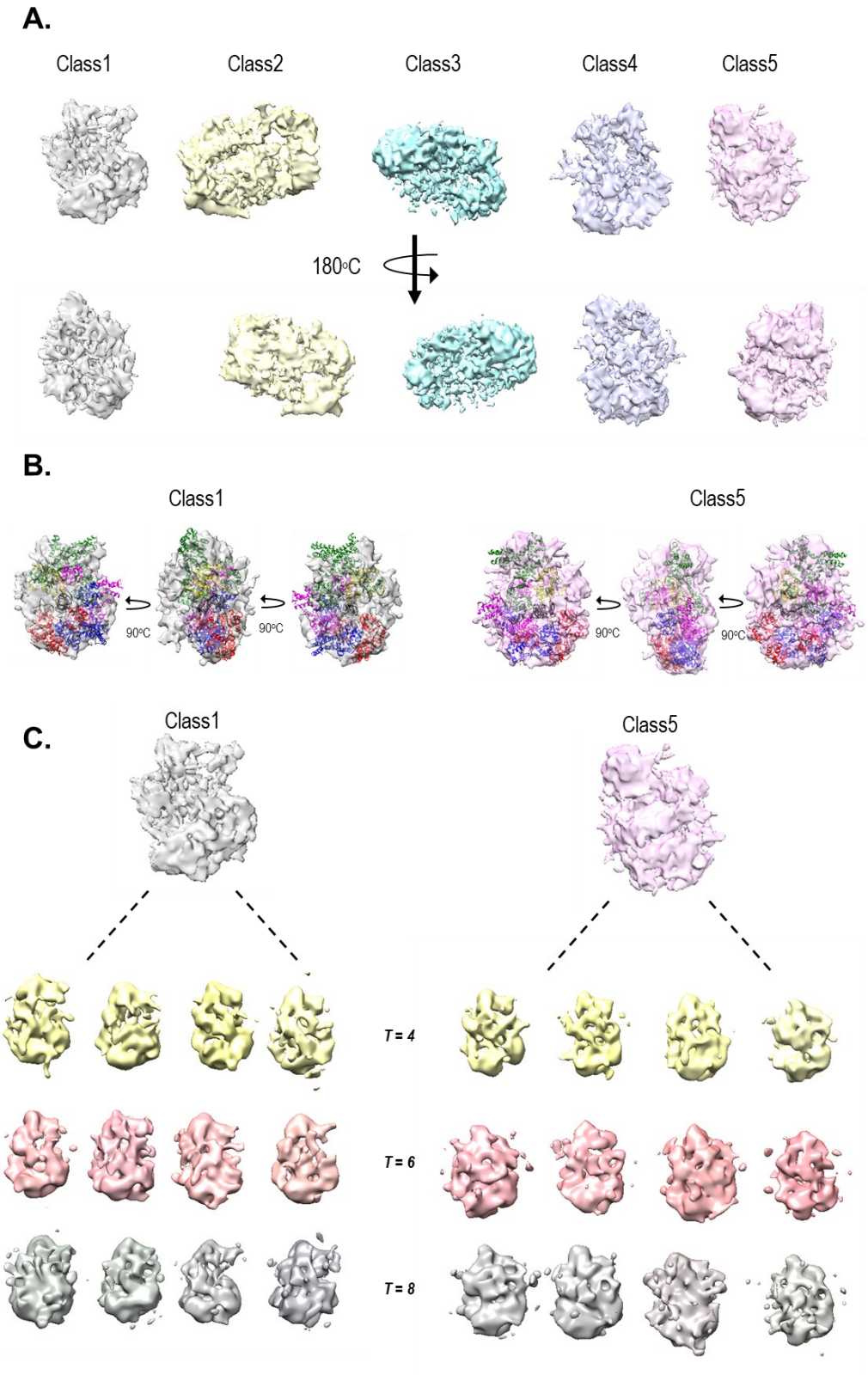


Figure 42 : Ab-initio modelling and 3D electron density of Orc1-5-DNA

A: *ab-initio 3D density maps generated from classes from Fig.41.C.* ***B****: Fitting of Orc1-5-DNA (7JPS, Jaremko et al, 2020) into ab-initio Classes 1/5, the closest in size and fitting to published ORC models.* ***C:*** *3D models generated using Class1/5 as reference models respectively, with varying Regularisation model weighting (T) to try to improve distinguishable structural features. For maps of a low resolution, a higher T parameter can help improve regions, although this can cause overfitting of background and noise. Here, it did not make a significant difference between different T-values.*

Particles were once again refiltered to remove any low-resolution classes or classes that contained significant background levels or noise (Figure 43.A.) to stringently include only those that resulted in 2D classes with distinguishable features with any secondary structure elements (Figure 43 B).

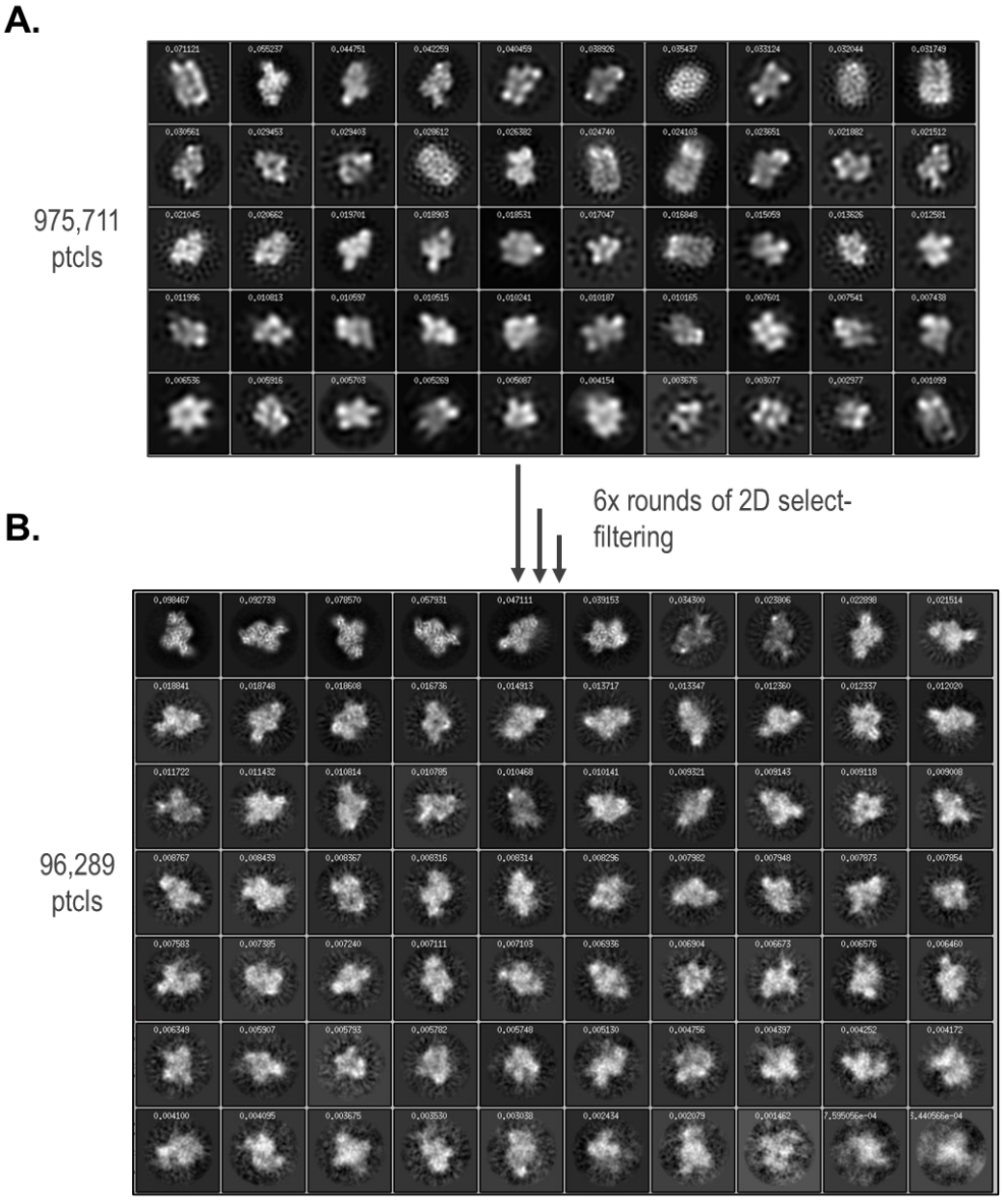


Figure 43: Filtering of 2D classes to ensure particle integrity.

Filtering of 2D classes from Fig.41.C to produce only classes with distinguishable secondary structure elements. Highlighted classes in A were chosen and went through six rounds of stringent re-filtering (via repeated 2D classification, then removal of suboptimal classes) until all classes contained secondary structure features, with minimal background ***(B).*** *The resultant 2D classes after six rounds of filtering containing just under 100,000 particles.*

Only these ~100,000 particles were sent for 3D classification, split into three classes as seen in Figure 44. These produced classes where it was mostly possible to fit the published model of Orc1-5-DNA (Jaremko et al, 2020). Some of these generated density maps were not compatible with fitting the Orc1-5-DNA in its published orientation, but still contained some extra density. This could correlate to some structural rearrangement not previously seen before that the complex adopts in its full-length state.

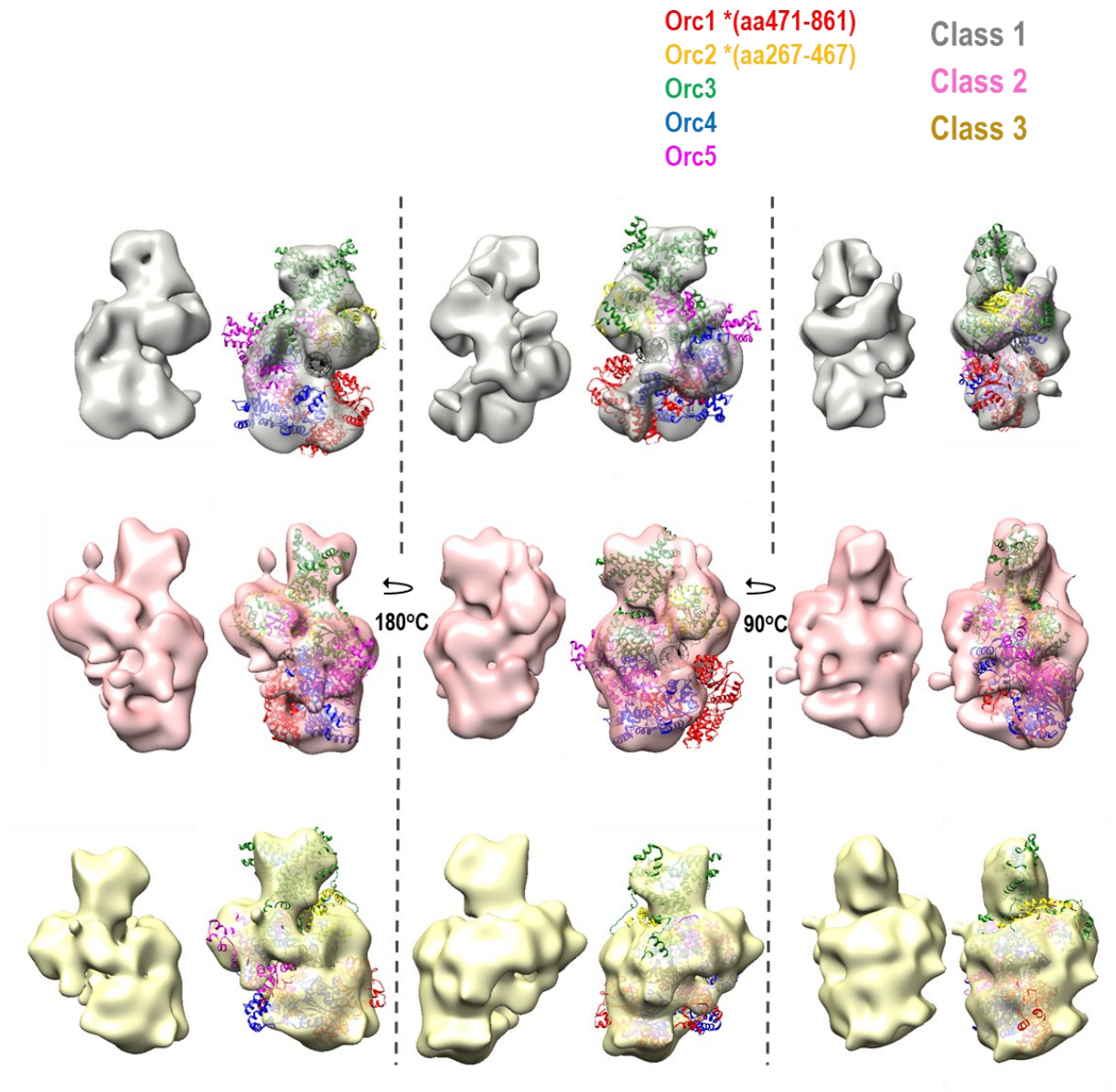


Figure 44: Electron Density Maps of Orc1-5-DNA

Ab-initio 3D electron density maps generated from classes from Fig, 43. All maps contain Orc3 protrusion and an incomplete ring structure, although published Orc1-5 (7JPS, Jaremko et al, 2020) could not fit perfectly. The presence of extra density suggests conformational change or presence of additional flexible regions. However, considering the low resolution of these maps and that they are not refined, they must be treated with caution. Further filtering, plus including more particles from the rest of the dataset will likely improve the resolution.

Due to time constraints, it was not possible to further classify and refine these models, although the results generated are already a promising start to gaining more structural data. The parameters studied enabled isolation of particles that showed clear secondary structure elements. These are of a similar quality with previously published 2D classes, with potentially

extra regions of higher resolution (Figure 45). When considering the 2D classes that contained the most defined secondary structure elements with as little background as possible (Figure 45), it may be possible to re-train Topaz picking on the rest of the dataset for an increased number of good particles for 3D modelling. One particularly interesting result is that these high-resolution 2D classes look similar to published structures of full-length ORC1-5 alone (Cheng et al, 2020). While this might suggest that the most stable particles in this dataset are ORC1-5 that is not bound to DNA, the lack of similar classes in the ORC-only results presented in Section 3.4.1 suggest that these particles have indeed been stabilised by the presence of DNA, with the core adopting a more compact structure while regions of Orc1 and Orc2 remain flexible. However, this cannot be said for certain, and more filtering and classification of the dataset is needed to determine whether this is the case.

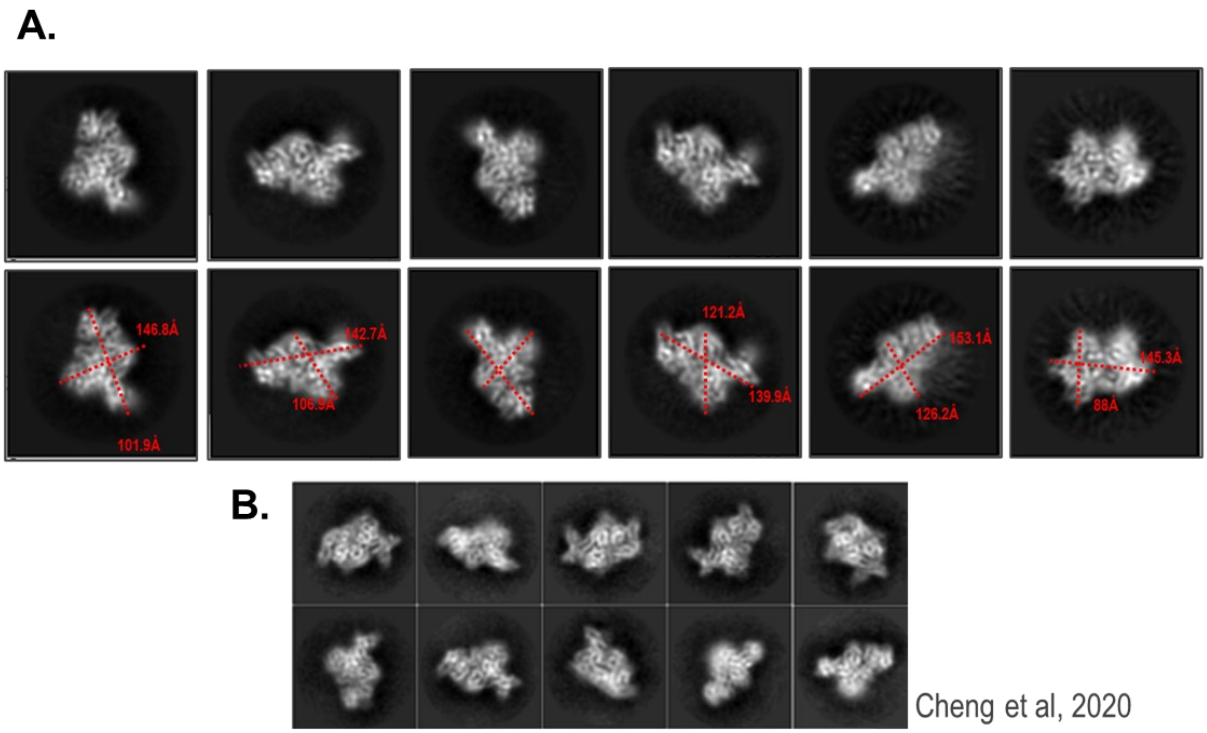


Figure 45: Additional Filtering needed for high resolution structural information

A: *2D classes from the ORC-DNA dataset, with the least amount of background from filtering procedures described above. These classes appear to be similar in structure to* ***B:*** *published full-length Orc1-5 data (Cheng et al, 2020), suggesting that the most stable, high resolution particles in this dataset are Orc1-5 only. However, since we did not see this level of resolution in our own ORC1-5-only datasets, it may be that the addition of DNA stabilised the particles and allowed for more structural features to be observed.*

Discussion

Overall, despite the large size of this dataset, significant filtering was required before gaining any structural resolution, and these were limited to ab-initio maps, rather than validated 3D density maps. Despite the optimisation of the grids to a level where particles appeared to the naked eye to have a good distribution, limited aggregation and reduced background, difficulties in data processing showed that this was not necessarily the case. While it could be argued that some extra density is visible, the resolution of the generated density maps are not high enough to characterise these regions at the secondary structure level. This has the potential to be improved with the addition of using the rest of the dataset – here, only 1/3 of the entire dataset was used to speed up processing time. However, it is likely that the

resolution reached here can be improved upon the addition of more particles and further refinement of the initial 3D maps. 2D classes containing high resolution structural features generated from this dataset can be used as input references for trainable autopicking to facilitate the picking of high-quality particles.

Further processing was done on this dataset post submission of this thesis (processing done by Dr Ricardo Aramayo, LMS), wherein the RELION autopicker was used to pick 2.7 million particles from a new subset of 4,370 micrographs. Resultant 2D classification, filtering via removal of junk and binning of the data 2 times gave 2D classes with 350,000 ORC particles (Figure 46 A), similar to those seen in Figure 43 B. However, even with this larger number of particles, the resultant ab-inito 3D density maps (Figure 46 B) do not show much improvement when rigid-body fitted with published structures of Orc1-5 on DNA (7JPS, Jaremko et al, 2020).

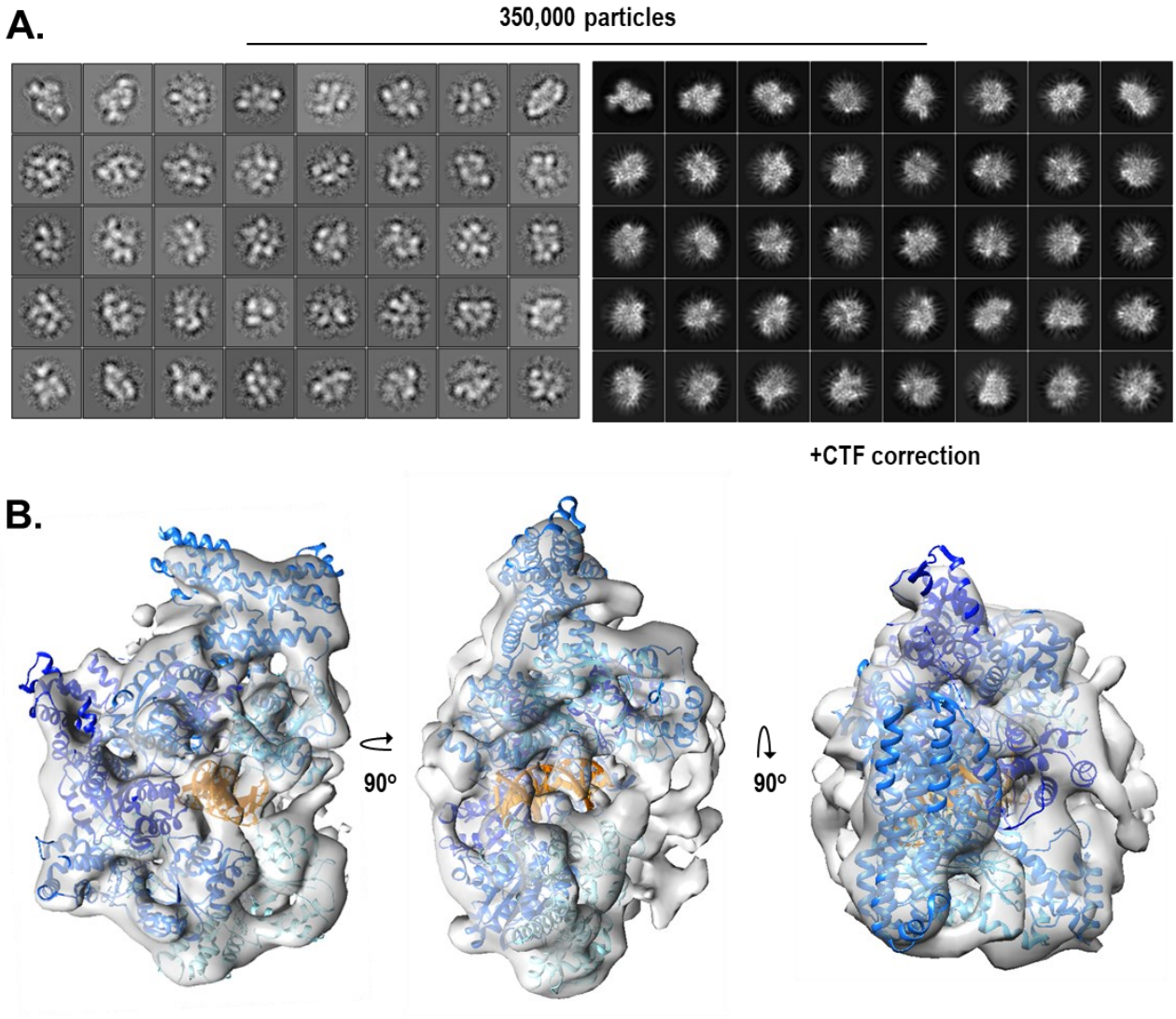


Figure 46: additional processing of the Orc1-5 dataset

A: *Resultant 2D classes from 350,000 particles of Orc1-5-DNA that have been filtered to remove any junk particles. CTF-corrected classes are shown on the right.* ***B****: ab-initio 3D modelling using the 2D classes shown in A as an input, with a rigid-body fitting of Orc1-5-DNA (7JPS, Jaremko et al, 2020). While it was possible to fit in the DNA and there are regions of extra density that could be Orc1/2 flexible regions, resolution was still too poor for further modelling.*

Note: This processing was done by Dr Ricardo Aramayo, LMS, 2024.

Difficulties in generating a high-resolution model is likely due to a combination of factors including increased background from the carbon support, aggregation and the innate full-length protein flexibility. Using carbon as a support on the cryogrids resulted in the increased

level of background that was visible in the 2D classes, even after significant filtering. This may have prevented identification of small, flexible regions due to their inability to be separated and averaged out against the background. We observed, similar to previous studies working on the FL ORC complex (Cheng et al, 2020), that even with cross-linking, previously unmapped regions were very blurry indicating flexibility and thus it would be exceedingly difficult to resolve these regions with the tools and strategy we employed.

Nevertheless, some observations from these datasets have some promise and pick up on features that were not emphasised previously – it appears that ORC bound to DNA may be more stable than ORC alone, since an increased level of promising 2D averages were able to be sorted with more ease in comparison to the ORC-only dataset. The presence of secondary structural features in the 2D classes with low background have the potential to uncover more 3D data. Unfortunately, with the time constraints this was not possible in the scope of this project, but additional filtering of these promising classes and using these to re-train particle picking and classification may still result in extra electron density being elucidated in new maps. Particle picking was primarily done by Topaz in RELION-4.0 due to its increased ability in identifying asymmetric particles. However, it may be worth using other picking software – such as CryoSPARC or CisTEM to identify if these algorithms are better at sorting particles from junk/background.

Another potential solution could be to optimise cryo-EM conditions on open-hole grids to really reduce the level of background on the micrographs. Since this was unsuccessful initially, this will require heavy optimisation of protein concentration, buffer components and cryo-grid preparation parameters to ensure the protein sample freezes in a vitreous layer in the grid holes and avoids aggregation on the grid support. It may also be that using a different support, such as gold may promote sufficient particle density in the gridholes. If optimisation of a reduced-background sample is possible, it may be extremely valuable in generating better results for poorly averaged areas and would improve the contrast on the grid and this will help particularly at the stage of determining by eye if the grids feature mono-disperse particles.

Overall, these optimisations could provide a wealth of new data that has the potential to unlock new information about the structure of full-length ORC. As discussed in the introduction (and will also be discussed in further results chapters, the regions of ORC that are not considered in structural studies are of high importance for the regulation of ORC and will therefore affect protein function. Having a high-resolution model of this has the potential to correlate its structure to its *in vivo* function.

3.4.3 Introduction of Cdc6 to form an ORC-Cdc6-DNA complex

Once Orc1-5 has bound to DNA, the next step in the pre-RC assembly process is the binding of Cdc6, thought to close the ORC ring around DNA to form a docking platform for Cdt1-MCM2-7. Previously in the group, the structure of the *S. cerevisiae* ORC bound to Cdc6 had been solved (Yuan et al, 2020), and at the same time, *D. melanogaster* (Schmidt et al, 2020). While not introducing significant structural rearrangements in either of these studies, Cdc6 is still crucial for productive double hexamer loading and DNA replication.

Cdc6 enhances ORC binding to DNA

Cdc6 was initially added to ORC in an EMSA experiment, as described in Section 2.2.8. Full-length Cdc6 was incubated alongside Orc1-5 and DNA for 15-30 minutes at 30°C, shaking at 300 rpm, then the samples were run on 0.75% TB-agarose gels for at 70 V 2 hours. In all samples, double the molar concentration of Cdc6 relative to ORC was used.

As with the ORC alone samples, Cdc6 did not introduce DNA specificity in terms of binding to the DNA (Figure 46). It did, however, enhance the binding of the proteins to DNA, with a molecular weight shift of the DNA happening at lower ORC concentrations. In addition to this, there is also an even clearer presence of a tight band at the ~650 bp mark (Figure 47 A.), suggesting a population of protein-DNA complexes that remain stably bound. The presence of larger molecular weight smears also indicates the presence of multiple ORC molecules bound to DNA, although these are primarily limited of 1:3:6 DNA:ORC:Cdc6 ratios and above.

The presence of protein-DNA aggregates also increases with an increase in concentration, thus a molar ratio of 1:1:2 DNA:ORC:Cdc6 was chosen for further experiments.

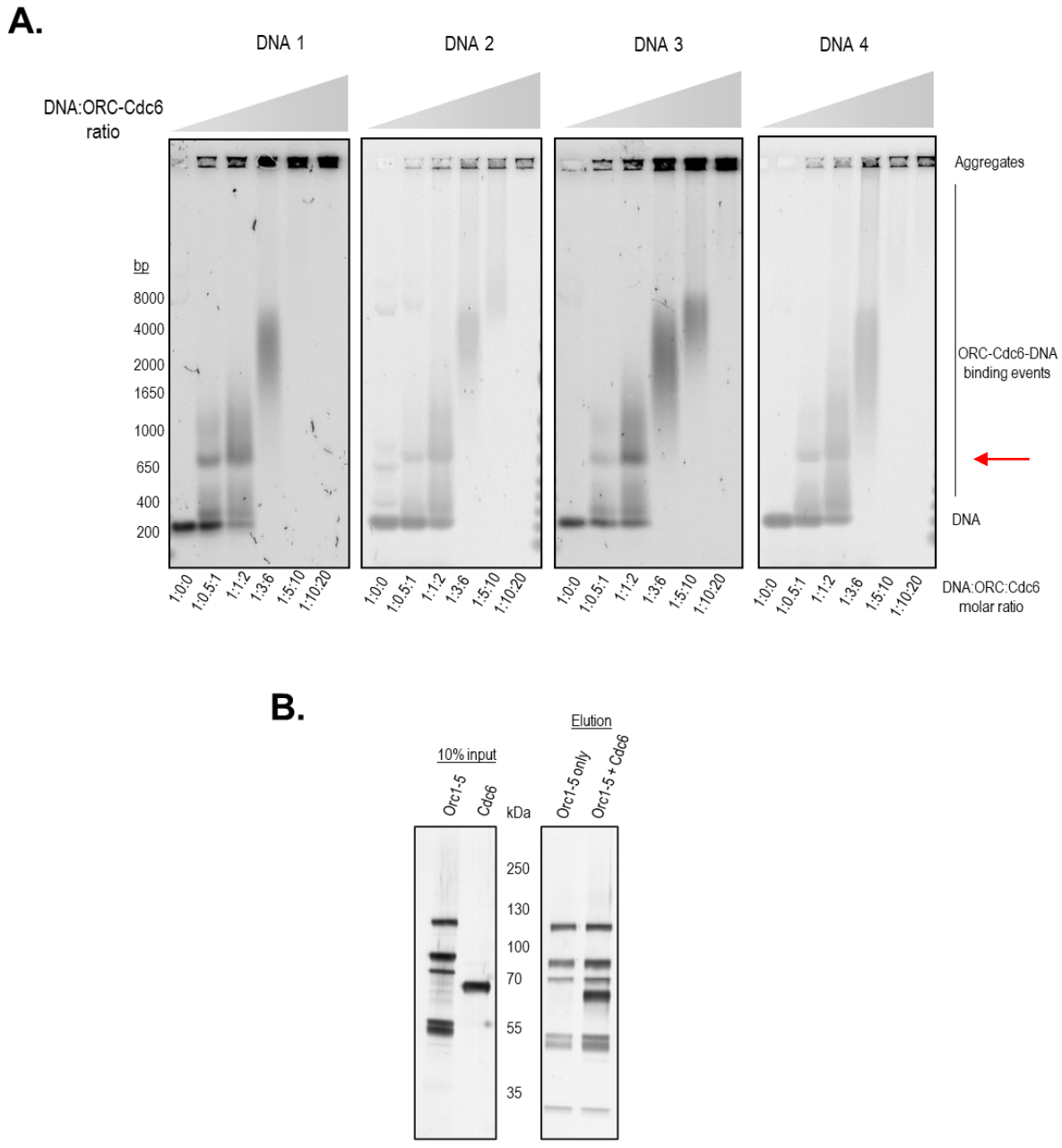


Figure 47: Cdc6 enhances binding of ORC to DNA

A: *EMSA experiments of ORC and Cdc6, incubated with the DNA constructs described above (Section 2.2.2) then run on an agarose gel. Gel shows shift of DNA with increasing molar ratios of protein, suggesting an increased molecular weight, likely corresponding to protein binding. Gels shown are representative of repeated experiments (n=3 per DNA).* ***B****: pulldown-style experiment with biotinylated DNA1 conjugated onto magnetic beads. Incubation of biotinylated DNA with ORC in the presence of Cdc6 shows an enhancement of ORC on DNA when compared to incubation with ORC alone. Gel is representative of n=3 experiments.*

This increased efficiency of ORC binding in the presence of Cdc6 is also visible in pulldown style experiments. When incubating biotinylated DNA with either ORC alone or ORC + Cdc6, an increased intensity of protein bands when both proteins are present suggests that the addition of Cdc6 increases ORC binding efficiency or stabilising of the ORC complex allowing for better binding (Figure 47. B).

Cryo-Electron Microscopy of in-solution assembly of OCD

To prepare ORC-Cdc6 bound DNA (OCD) samples for cryo-EM, the proteins were incubated in a 1:1:2 ORC:Cdc6:DNA ratio, as determined by optimal EMSA conditions for protein binding, with minimal aggregation (Figure 47). The reaction conditions were the same as the cryogrids prepared for the ORC only and ORC-DNA samples, with the addition of 0.1% Triton X-100 into the core buffer, then crosslinking of the sample in 0.1% glutaraldehyde for 15 minutes on ice before application onto the cryogrids.

A small dataset of 515 micrographs was collected in-house on a TALOS F200i TEM fitted with a Falcon 3EC detector. The full collection parameters are available in Supplementary Table 7.5. These micrographs were imported into RELION-4.0 beta and CTF corrected using CTFFIND4.

A small subset of ~1,200 particles were manually picked from a random selection of 40 micrographs (Figure 48 A.) and averaged into 5 2D classes (Figure 48 B.). Two of the resultant classes contained ORC-like features and corresponding diameters; as well as this, one class had a distinguishable DNA strand visible in the averaged images (labelled).

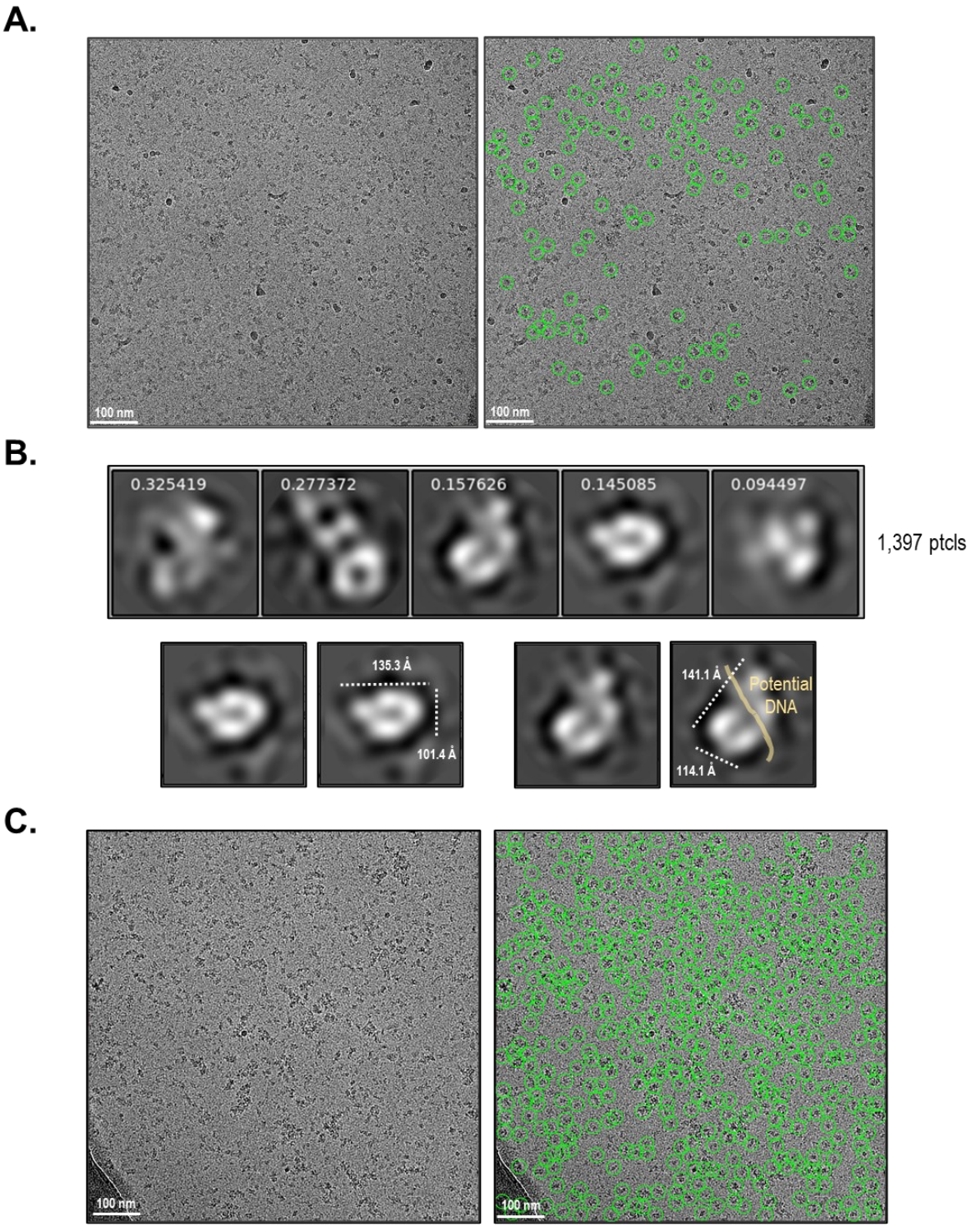
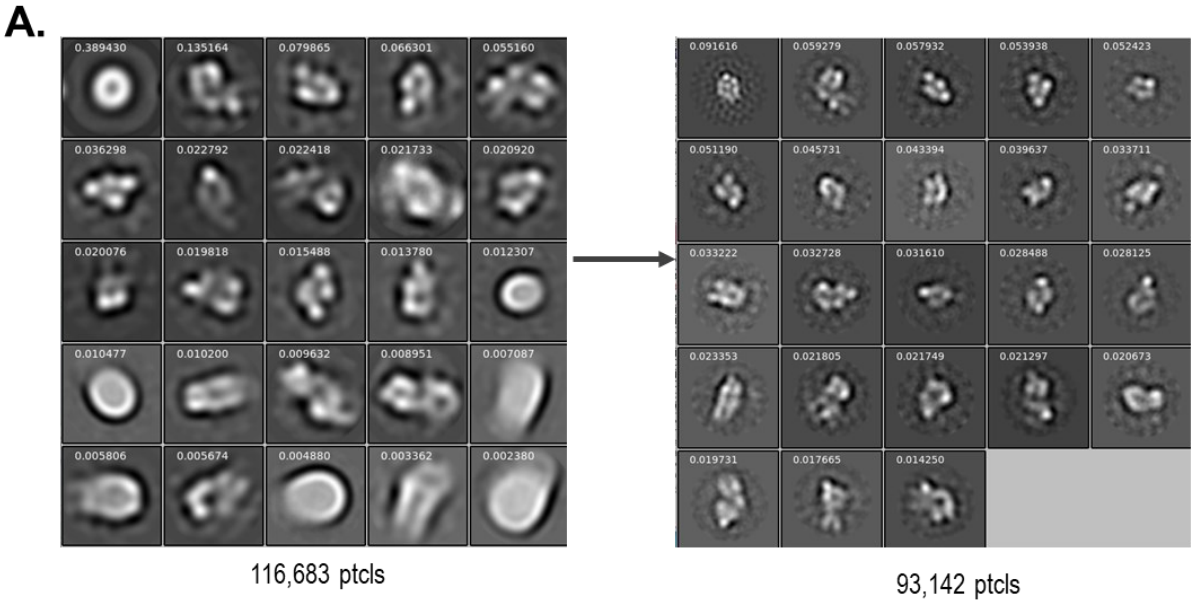
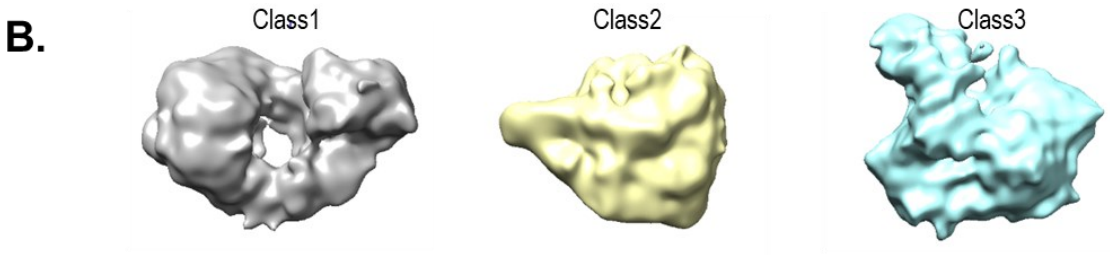


Figure 48: Cryo-electron microscopy picking on ORC-Cdc6-DNA

A: *Cryo-electron micrograph and corresponding manual picking of OCD in solution* ***B:*** *2D classes generated from manual picking of 1.3 thousand particles, with selected classes (below) used as references for autopicking. These classes contain a DNA-like density running through the central core, labelled in yellow* ***C:*** *autopicking micrograph based on reference-based template picking using 2D classes from B.*

These classes were selected and used as templates for reference-based template matching and particle picking on the rest of the dataset (Figure 48 C.). In total, 116683 particles were extracted and sent for initial 2D classification (Figure 49. A), with resultant classes being filtered out for junk and sent for 3D initial modelling (Figure 49. B). Class 3 was used as a reference map for 3D modelling, due to its closest diameters to Orc1-5 with additional density that could correlate to Cdc6. From the 5 classes generated, it appeared that the two most populous classes (in total, ~54% of the particles) constituted heterologous, flexible models, with extended densities that could not be mapped. The next most populous classes contained a tight, smaller ring structure, closer in size to Orc1-5 bound to DNA only (Figure 49. C).


A.
0.389430
0.135164
0.079865
0.066301
0.055160
0.036298
0.022792
0.022418
0.021733
0.020920
0.020076
0.019818
0.015488
0.013780
0.012307
0.010477
0.010200
0.009632
0.008951
0.007087
0.005806
0.005674
0.004880
0.003362
0.002380
116,683 ptcls
0.091616
0.059279
0.057932
0.053938
0.052423
0.051190
0.045731
0.043394
0.039637
0.033711
0.033222
0.032728
0.031610
0.028488
0.028125
0.023353
0.021805
0.021749
0.021297
0.020673
0.019731
0.017665
0.014250
93,142 ptcls


B.
Class1
Class2
Class3

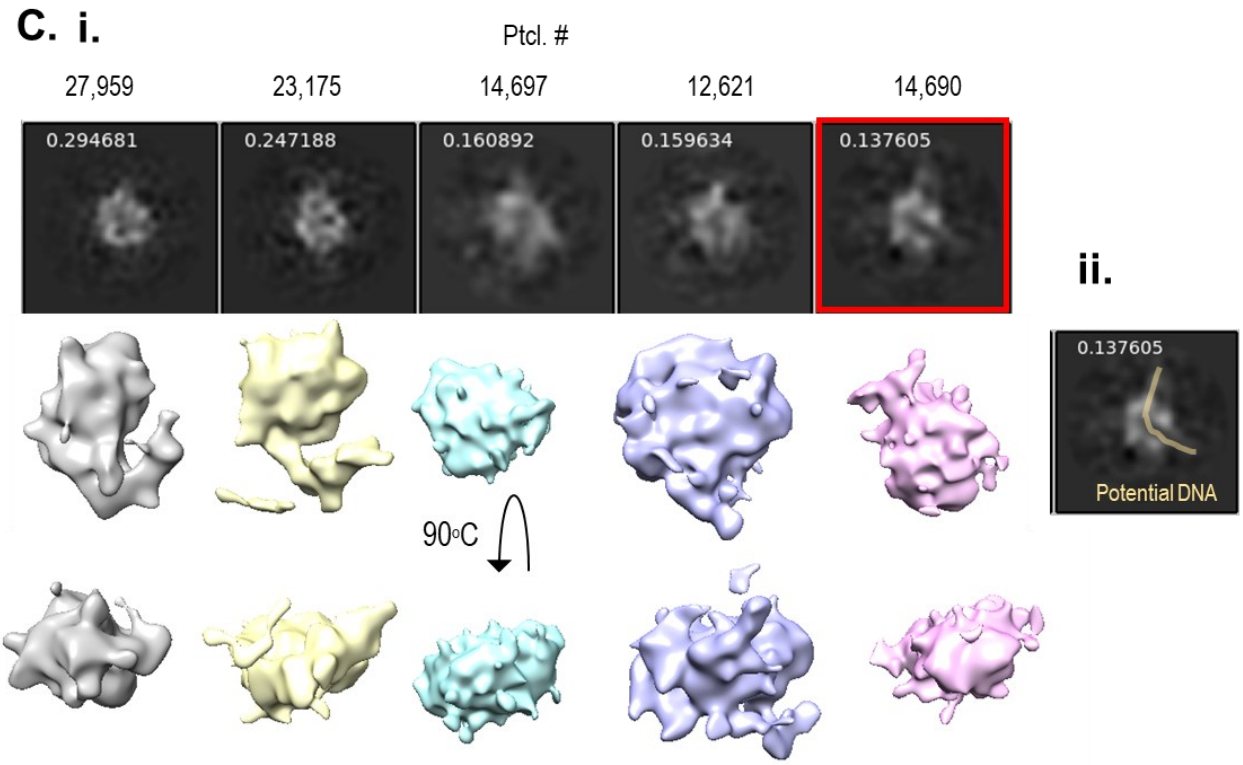

C. i.
Ptcl. #
27,959
23,175
14,697
12,621
14,690
0.294681
0.247188
0.160892
0.159634
0.137605
90°C
ii.
0.137605
Potential DNA

Figure 49: ORC-Cdc6-DNA in solution results in a heterogenous population of particles

A: *2D classification and filtering of picked particles, picked via parameters shown in Figure 47. Classes were filtered to remove junk particles and 93 thousand particles from cleaned classes (right) sent for ab-initio 3D map generation* ***B:*** *ab-initio 3D maps of 2D classes from A. Class 3 appeared to be the most similar to a distinguishable ORC structure, so this was used as a reference map to generate* ***Ci****: further 3D electron density maps, with corresponding slice view (above) using Class 3 from B as a template.* ***Cii****: Classes containing potential DNA densities.*

The class with the fewest particles contained a ring-like structure with extra density – this could be Cdc6 bound to ORC on DNA. In addition to this, the 2D average of this class contains an apparent bent DNA strand, like that seen in *S. cerevisiae* and *D. melanogaster* models (Figure 49.Cii). Particles from this class were sent for 3D model refinement, resulting in 3D density map that could fit published Orc1-5 bound to DNA, with additional extra density between Orc1 and Orc2, closing the ring (Figure 50). This density appears to be more flexible and sits further away from the core of the ring structure when compared to published structures of ORC and Cdc6 (Figure 9, Section 1.4.2.1) – this is likely due to the current low resolution of these maps; this could be improved upon with further data collection and analysis. Cdc6 itself also contains significant flexible regions which may be contributing to the large surface seen in this 3D density map, although this cannot be said for certain without further structural analysis.

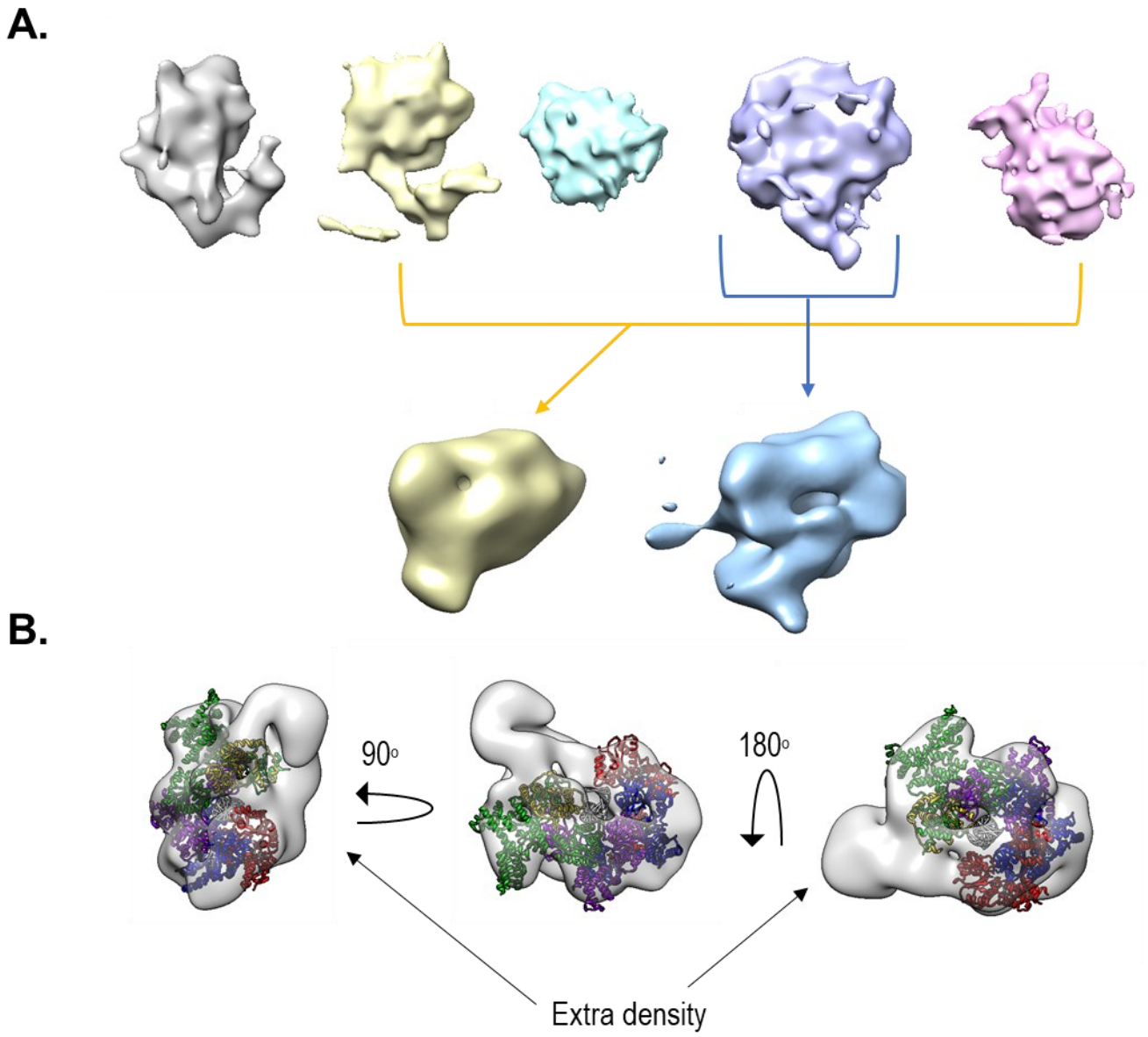


Figure 50: ORC-Cdc6-DNA makes up a small population of particles in solution

A: *3D modelling refinement of the ab-initio models using the most compact densities from Figure 49.C. resulting in* ***B:*** *an electron density map containing the correct orientations of Orc1-5 bound to DNA, with extra density in the region between Orc1 and Orc2 (where Cdc6 sits in other eukaryotes). This region is relatively large in comparison to previous structures, although this can be attributed to the combination of Cdc6 flexibility and the low resolution of these density maps. Overall, only 14% of the particles averaged out to give an initial 3D model with potential Cdc6 density – increasing this population of particles will increase resolution.*

While it is promising that the sample does indeed contain ORC-Cdc6 bound to DNA, the heterogeneity in the distribution of the particles is unideal. Instead of going forward with these

sample conditions, attempts were made to produce a homogeneous sample for cryogrids, as discussed below.

Preparing a homogenous prep of ORC-Cdc6-DNA

Because of the success of using density gradient centrifugation to separate protein complexes of differing size and constitution (Saleh et al, 2022; Chapter 2.2.5), these were employed to separate free ORC and ORC-DNA complexes from ORC-Cdc6-DNA. ORC, Cdc6 and DNA were incubated in a buffer containing 25 mM HEPES-KOH, 250 mM KOAc, 3mM ATP, 1mM $MgCl_2$ and 1 mM DTT- for 30 minutes at 30°C with shaking at 300 rpm, then applied onto a 10-30% glycerol density gradient of the same buffer. These samples were run as described in Chapter 2.2.5, then manually fractionated and silver stained to identify which fractions contained the assembled complex.

Cdc6 does not stably associate with ORC over time

Interestingly, over the course of centrifuge run, Cdc6 did not appear to stably associate with ORC-DNA complexes at all. When incubated without DNA, Cdc6 does not appear to associate with ORC, with the Cdc6 peak in earlier fractions of the gradient (F. 3-5, with some streaking into later fractions) and Orc1-5 peaking in F. 7-10 (Figure 51. A). Upon the addition of DNA, the shift of some Orc1-5 into later fractions (11-13) alongside a DNA band suggests that ORC binds to DNA to form a complex with a higher molecular weight. The absence of Cdc6 in these fractions shows that Cdc6 does not remain bound. (Figure 51.B)

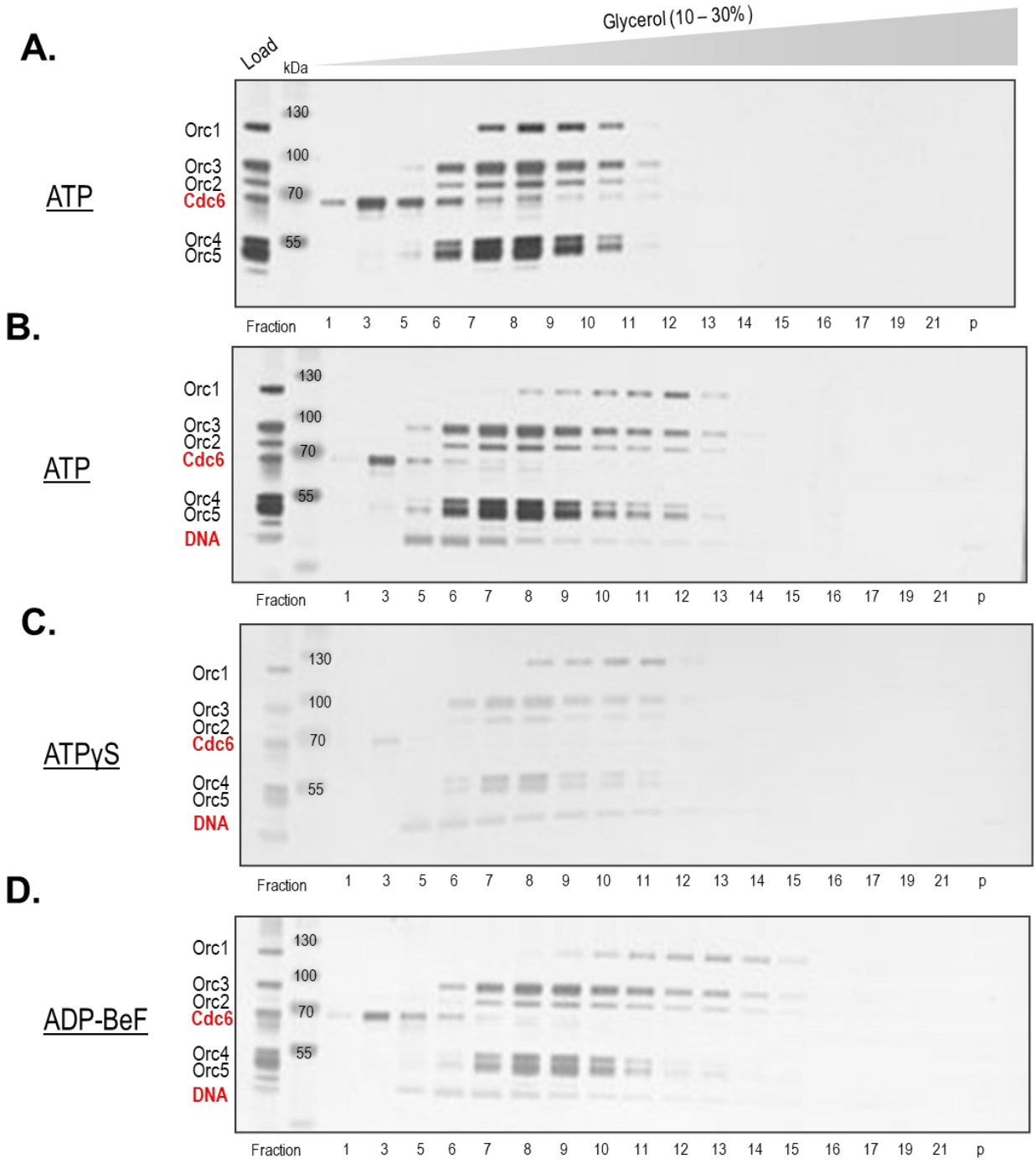


Figure 51: Cdc6 does not remain bound to ORC-DNA in the long term.

A: *ORC-Cdc6 only on a 10-30% glycerol density gradient, showing that ORC fractionates primarily in F7-10, and Cdc6 mostly in F.3-5. Some streaking of Cdc6 is seen in later fractions (6-8) but it does not appear to bind to the ORC complex in the absence of DNA (no higher-density fractions containing all subunits).* ***B:*** *ORC-Cdc6-DNA in on a gradient in the presence of ATP. Here, there are fractions later in the gradient (F10-13) presumed to be higher density, although the bands suggest only Orc1-5 and DNA are present here, suggesting Cdc6 cannot form a stable, long lived complex with ORC and DNA. The Cdc6 peak remains in F3-5. These gels are representative of 3 experiments. Similar results are shown for ORC-Cdc6-DNA in the presence of* ***C****: ATPγS and* ***D:*** *ADP-BeF.(n=2 each condition). Cdc6 does not co-fractionate alongside ORC-DNA.*

This could be because Cdc6 is either loosely associated and can freely disassociate; or interacts in a transient manner with ORC bound to DNA. The separation of Cdc6 into earlier fractions in the gradients due to either reason results in a complete separation in space of DNA-bound ORC and Cdc6, meaning that it can no longer associate with ORC-DNA, even in a short-lived interaction. The results presented in Figure 46, where Cdc6 is pulled down alongside ORC-DNA upon incubation supports the theory that as long as the two proteins are in close proximity, Cdc6 can associate with ORC-DNA and even enhances ORC binding to DNA. However, this interaction may be transient or short lived, considering the lack of a stable ORC-Cdc6-DNA complex across a gradient over time. If Cdc6 were to disassociate from ORC-DNA on the gradient, the smaller density protein would diffuse and fractionate separately, meaning it could not reassociate with ORC and DNA. The complex may also be sliding off of DNA; while so far there is not any significant evidence of human ORC sliding on DNA, single-molecule studies of S. *cerevisiae* ORC show it can diffuse across DNA (Sánchez et al. 2023) - caging effects from the agarose matrix in EMSAs may stabilise this in comparison to a glycerol density gradient.

Cdc6 and its ATPase activity

To study whether the disassociation or release of Cdc6 from ORC-DNA is ATP dependent, the same gradient experiments were carried out, except with ATPγS (a slowly hydrolysable analogue) or ADP-BeF (a non-hydrolysable analogue) in place of ATP in the reaction and gradient buffers. Despite this, no significant change was observed, with Cdc6 continuing to fractionate separately from any fractions containing both ORC and DNA (Figure 51 C-D.). The reason for this could be that release of Cdc6 from ORC-DNA is not ATP dependent and Cdc6's hydrolase activity is required further downstream in the interaction pathway upon binding of other pre-RC components.

Another important factor to consider is that both Cdc6 and Orc1-5 require ATP during their purification processes to remain stable and reduce complex disassociation or aggregation.

Because of this, despite buffer-exchange steps being taken to concentrate the protein preps and remove the nucleotides at the end of the purification process and long-term storage, some residual ATP may be left bound to the protein stocks. If the proteins are already in an ATP-bound state during the reaction process, it could be this hydrolysis driving disassociation of ORC from Cdc6 even in the presence of slowly- or non-hydrolysable nucleotide analogues. To ensure that this is not the case, one potential solution could be to perform the purifications of these proteins in the presence of ADP-Bef instead of ATP, although this may require further optimisation.

Alternatively, a Cdc6 ATPase mutant could be used to explore this relationship further – this is the route we decided to examine. Cdc6 with a mutation in the Walker B motif (E285Q) (Figure 52) was expressed and purified, as previously described in Chapter 2.2.1.2.

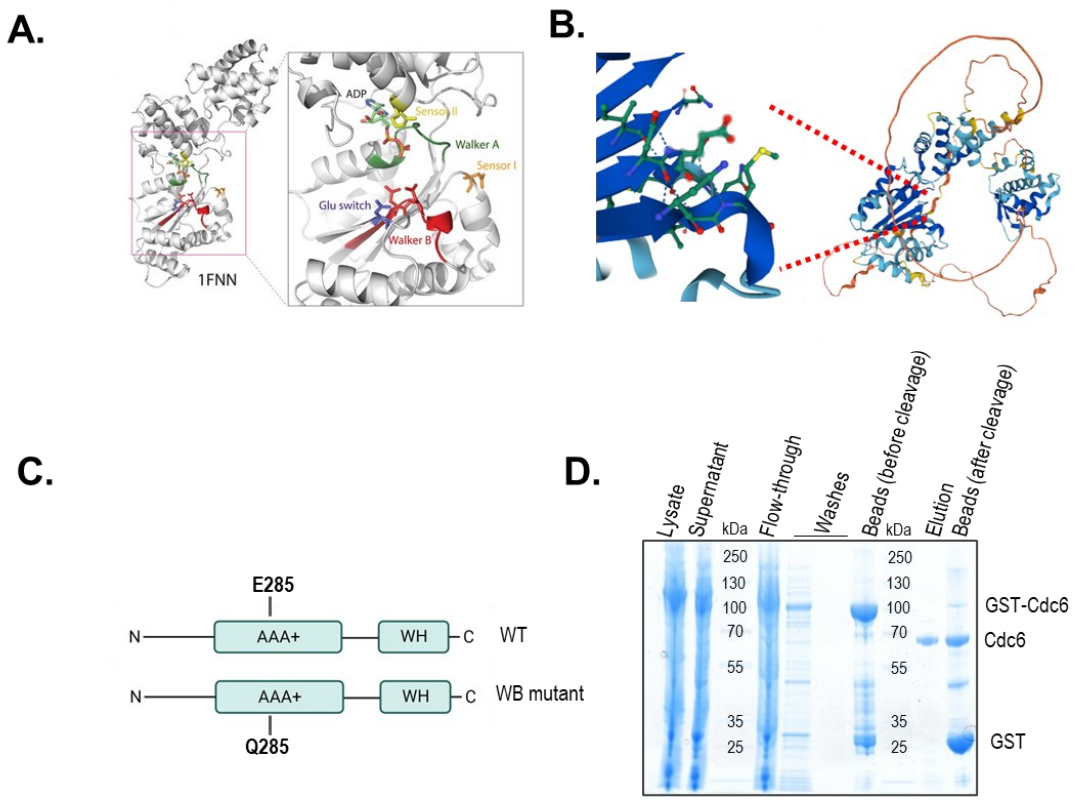


Figure 52: An ATPase mutant of Cdc6

A-B: *a close-up of the catalytic site of Cdc6, with glutamate in the ATPase domain, mutated to remove the ATPase activity of Cdc6.* ***C:*** *Schematic of mutation in the context of the domains. Mutation from a glutamate to a glutamine removes the functional group responsible for catalysing the hydrolysis of ATP.* ***D:*** *Purification and GST-mediated elution of Cdc6 (E285Q) from agarose beads, following purification procedure described in Section 3.1.2).*

This mutant was subject to the same gradient experiments, as described above. Surprisingly, upon fractionation of the sample, a high-molecular weight species containing Orc1-5, Cdc6 (E285Q) and DNA was independently isolated in Fractions 12-14, suggesting the presence of an ORC-Cdc6-DNA complex (Figure 53 A.). This suggest that in part, ATP hydrolysis could be responsible for releasing Cdc6 from the ORC-DNA complex.

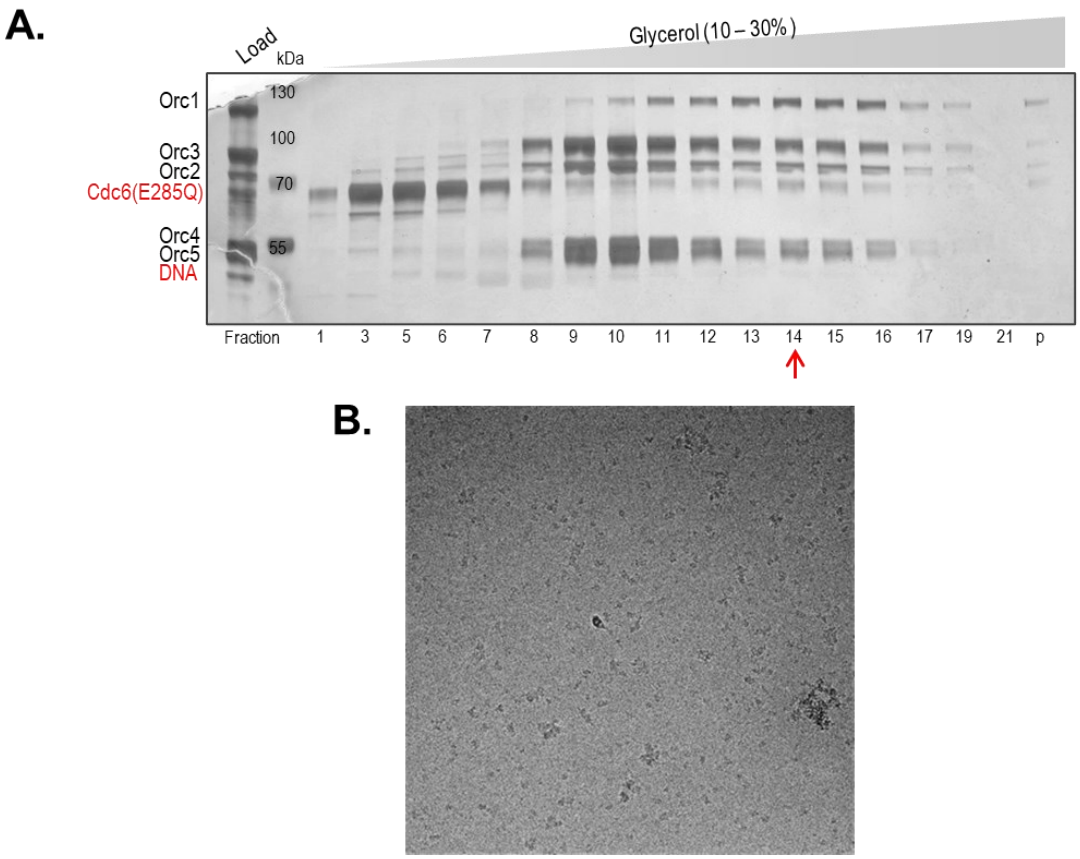


Figure 53: Using an ATPase-dead mutant of Cdc6 traps OCD on a gradient in the long term.

A: *16-hour glycerol gradient fractions showing that OC(E285Q)D can be isolated (F.12-14) in the long term. Here, all subunits (Orc1-5, Cdc6 and DNA) are visible in the higher density columns, suggesting a stable, long-lived complex between the three can be formed.*
B: *Cross-linked and buffer-exchanged samples from Fraction 14 in A, applied to cryogrids. Here there are monodisperse particles seen on the grid, further data collection and processing may result in a structure of ORC-Cdc6 bound to DNA.*

The usage of an ATPase mutant of Cdc6 may serve to stabilise the complex for structural analysis; using this in combination with density gradient centrifugation may offer a way to isolate a homogenous sample of ORC-Cdc6-DNA. To prepare cryo-electron microscopy grids,

Fraction 14 was cross-linked with 0.1% glutaraldehyde post gradient fractionation, then buffer exchanged into the optimised cryogrid buffer described in Section 2.2.5. Grids were prepared as described in chapter 2.2.10.2, then screened to check for sample dispersity (Figure 53 B). Due to time constraints, I was not able to collect a dataset for processing – however, the success in producing this homogenous sample is promising for future analysis.

3.4.4 The Orc2-5-DNA complex

As well as the results discussed above, the results from the ORC-Cdc6-DNA gradients also provided another surprising result – the presence of an Orc2-5 where it appears that Orc1 has dissociated. While Orc2-5 subunits are present in the ORC-Cdc6 alone gels shown in Figure 51.A (that are not expected to bind to each other in the absence of DNA), this is greatly enhanced in the samples where DNA is present, with Fractions 6-9 containing reduced or no Orc1 (Figure 51. B-D).

Over time, ORC1 can dissociate.

To test whether the release of Orc1 is Cdc6 induced, ORC alone and ORC-DNA samples were run in the same gradient conditions as described above. While some Orc2-5 was present in Fraction 6 when running Orc1-5 alone on a gradient, this was greatly enhanced upon the addition of DNA. As well as the population of Orc1-5 bound to DNA present in later fractions (11-14), earlier fractions from 6-10 contain either no or greatly reduced Orc1 in comparison to the other subunits (Figure 54). This also does not appear to be nucleotide-dependent, as the same trend is seen when replacing ATP with ATPγS or ADP-BeF (Figure 54 A-C) although this may also be due to residual ATP in protein stocks. To test whether Orc2-5 is bound to DNA, the samples were also run on an EMSA; the shift in the DNA band suggested that the DNA was occupied by protein, in an Orc2-5-DNA complex (Figure 56).

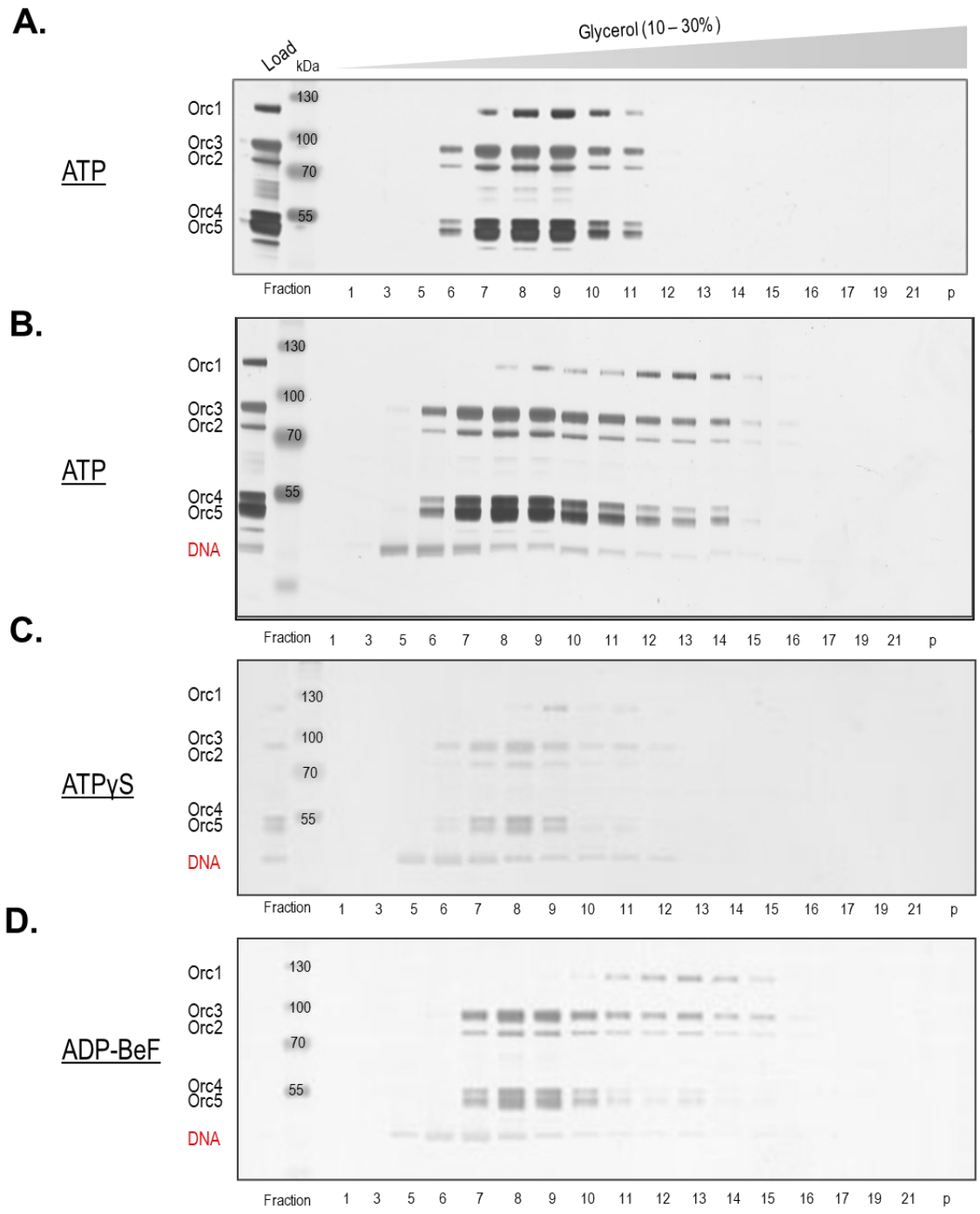


Figure 54: Binding of Orc1-5 to DNA releases Orc1.

A: *ORC1-5 only on a 10-30% glycerol density gradient, showing that the complex fractionates at primarily F.7-10.* ***B:*** *ORC-DNA in on a gradient in the presence of ATP, showing a shift in the ORC and DNA subunits. The presence of Orc1-5 with DNA in later fractions (F.11-14) indicate a higher molecular weight moiety, likely Orc1-5 bound to DNA. In this condition, there appears to be release of Orc1 in earlier fractions, (F7-8).* ***C:*** *Same conditions as described for Panel B, except in the presence of ATPγS. This gel was stained less than previously and was only repeated once, but it is possible here to also see the release of Orc1 in F8-9.* ***D:*** *Same experimental conditions as above, except in the presence of ADP-BeF. In all conditions where DNA is present, there are two distinct populations: Orc1-5 bound to DNA (F.11-14) and a population of ORC2-5 lacking the Orc1 subunit (F7-9). This does not occur when ORC is not incubated with DNA, as seen in Panel 1. All gels are representative of experiments with 4 repeats, except the ATPγS sample, which was only carried out once.*

Because of the long time frame of these gradient experiments (16 hours), it was hypothesised that Orc1 is released slowly over time, since Orc1 release is not observed in biotinylated DNA pulldown experiments (Figure 47. B), where the timescale is much shorter (30 minutes – 1 hour). To test this, Orc1-5 and Cdc6 were incubated together on biotinylated DNA for a range of timescales, then protein-bound DNA was released by DNaseI digestion and run on an SDS-PAGE gel. Between 30 minutes to 1 hour, no loss of Orc1 was observed, but at the 2, 4 and 8 hour incubation time points, there was a gradual decrease in the amount of Orc1 present (Figure 55). While it must be considered that increasing incubation times may increase aggregation of proteins to the beads (for this reason, the 16 and 8 hour time points were not considered reliable), the decrease in Orc1 stoichiometry in comparison to the rest of the subunits do indeed hint at release of Orc1 over time.

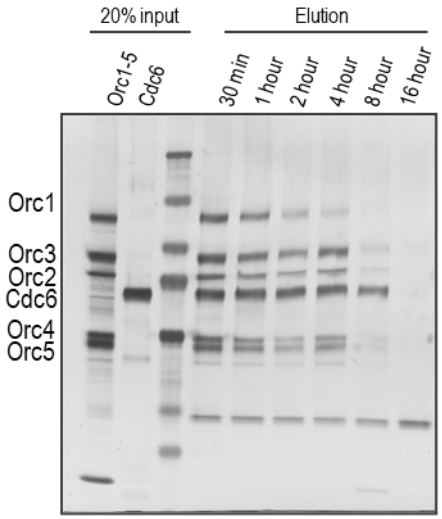


Figure 55: Orc1 is released over time

Pulldown-style experiment incubating Orc1-5 with biotinylated DNA, eluting from DNA with DNAse1 after different incubation periods as labelled. Orc1 is released with increasing incubation times, seen from the 30 minute timepoint to 4 hours. A visible decrease is seen from the 2 hour timepoint. Interestingly, Cdc6 remains associated suggesting it may be a transient interaction, although it could also be nonspecifically sticking to the beads. At the 8 and 16 hr timepoint, it was assumed that the complex had aggregated to the tubes and the beads, hence limited to no protein bands in the elution

Interestingly, Cdc6 appears to remain at a constant level during these pulldown experiments – this may be due to aggregation to the beads since the 8 hour time point contains a

stoichiometric excess of Cdc6 in comparison to Orc2-5 (Figure 55). While Cdc6 may transiently interact with the rest of the DNA-bound complex, further optimisations of this process is necessary to ensure that the results observed are not due to protein aggregation.

Negative Stain Microscopy of Orc2-5 on DNA

The Orc2-5 species bound to DNA has not been observed structurally. While previous studies have managed to isolate Orc2-5 alone from cells (Figure 6), the tightly compact structure is not associated with DNA.

To begin structurally characterising this, Fraction 7 of an ORC-DNA gradient (Figure 56) was cross-linked with 0.1% glutaraldehyde, then applied onto negative stain grids as described in Chapter 2.2.10. A small dataset of 385 micrographs was collected for processing; data collection parameters are available in Supplementary Table 7.6.

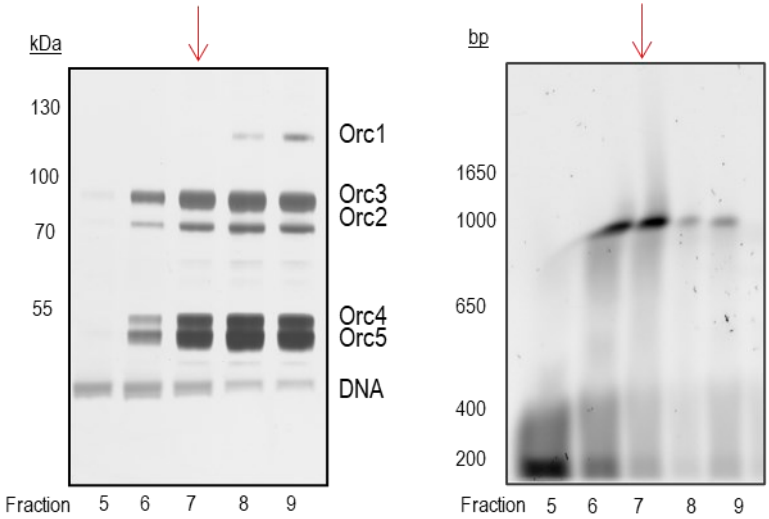


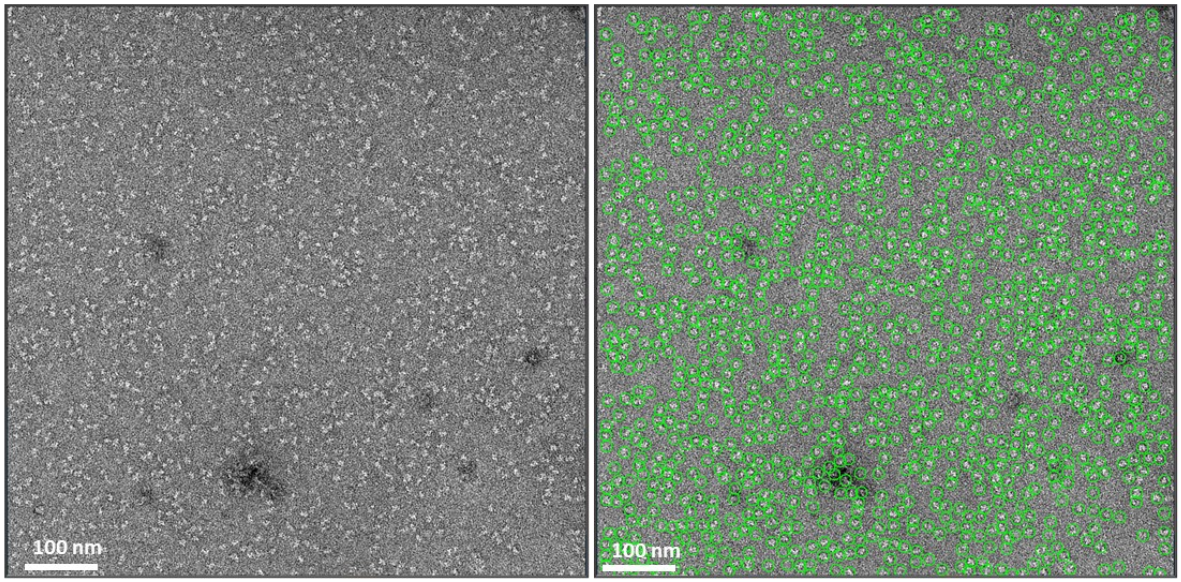
Figure 56: SDS-PAGE and corresponding agarose gel fraction of Orc2-5 bound to DNA

A: *Silver-stained SDS-PAGE segment showing fractions of Orc2-5, potentially bound to DNA (Segment from whole gel shown in Figure 53.B).* ***B:*** *Corresponding fractions run on a 1.5% agarose gel, showing a shift in the DNA molecular weight, indicating protein binding, likely Orc2-5 on DNA.. Highlighted fraction (F.7) cross-linked and prepared for NS-EM analysis, as described in Section 2.2.10.*

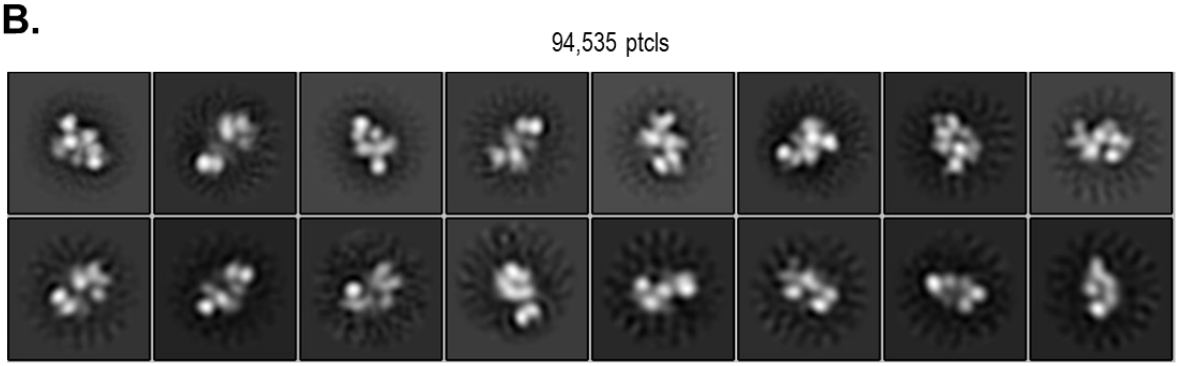
A small subset of 1000 particles were manually picked, before using the generated 2D classes to set up reference-based template picking (Figure 57 A) on the rest of the dataset. Resultant particles were extracted and sorted into 2D classes (Figure 57 B.).

While the resolution limits imposed by negative stain microscopy do not allow for precise DNA identification, some of the 2D classes contain visible DNA-like strands running through the structure (Figure 57 B.-C, yellow strands). However, since this is negative stain microscopy so this cannot be claimed for certain – with further 3D ab initio modelling, this density is averaged out. Lobes correlating to those of the ORC complex could be identified, especially those aligned with the Orc3 protrusion (Figure 57 C.). However, the overall dimensions of the 2D classes were smaller than those of Orc1-5 and Orc1-5 bound to DNA and did not appear to conform to the ring-like structures previously observed, suggesting the absence of a subunit, likely Orc1 (Figure 57 C).

A.


100 nm
100 nm

B.


94,535 ptcls

C.

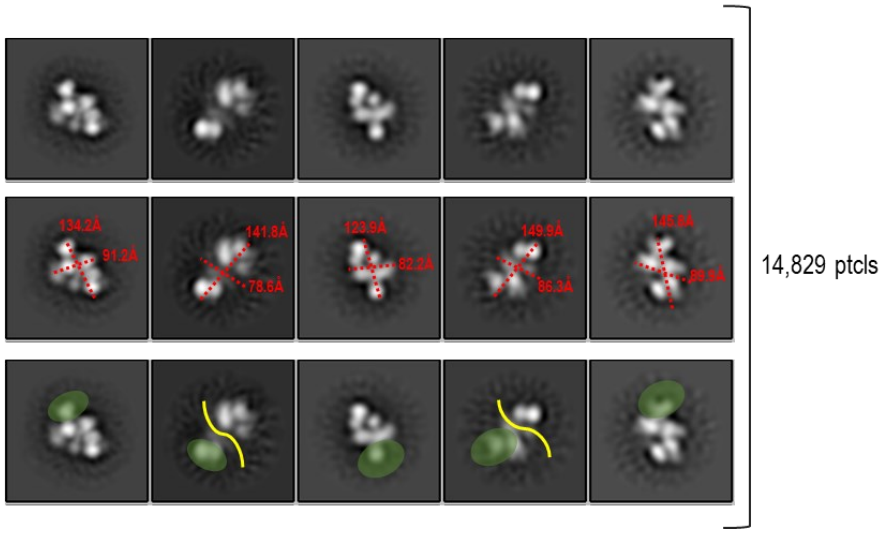

134.2Å
91.2Å
141.8Å
78.6Å
123.9Å
82.2Å
149.9Å
86.3Å
14,829 ptcls

Figure 57: The Orc2-5-DNA complex NS processing

***A**: Negative-stained electron micrograph and associated reference-based particle picking, based on manually picking one thousand particles and using these as autopicker templates. **B**: Resultant 2D classes from ~95,00 particles, showing an incomplete-ring structure, missing density near the Orc1 region. **C**: measurements and features of specific 2D classes from B suggesting that these classes may be Orc2-5 bound to DNA. These classes have a diameter across smaller than Orc1-5 or Orc1-5 bound to DNA (as seen in Sections 3.4.1-3.42, with potential missing density in the region where Orc1 should be. Negative stain imaging cannot image DNA, but predicted DNA binding positions are shown in yellow, following possible density that cannot be averaged to a higher resolution.*

These 2D classes were used to produce initial 3D density maps to determine the overall shape and size of the complex. When carrying out rigid body fitting of published 3D structures into the low-resolution map, it was evident that the model that fitted best was the Orc1-5 bound to DNA with the Orc1 subunit removed (Figure 58 A/C.) in comparison to the published Orc2-5 alone 3D model (Figure 58 B.). While unable to visualise any DNA density in the generated maps (which is not unusual for negative stain datasets) this improved fitting over the Orc2-5 3D model may suggest that the Orc2-5 is bound to DNA.

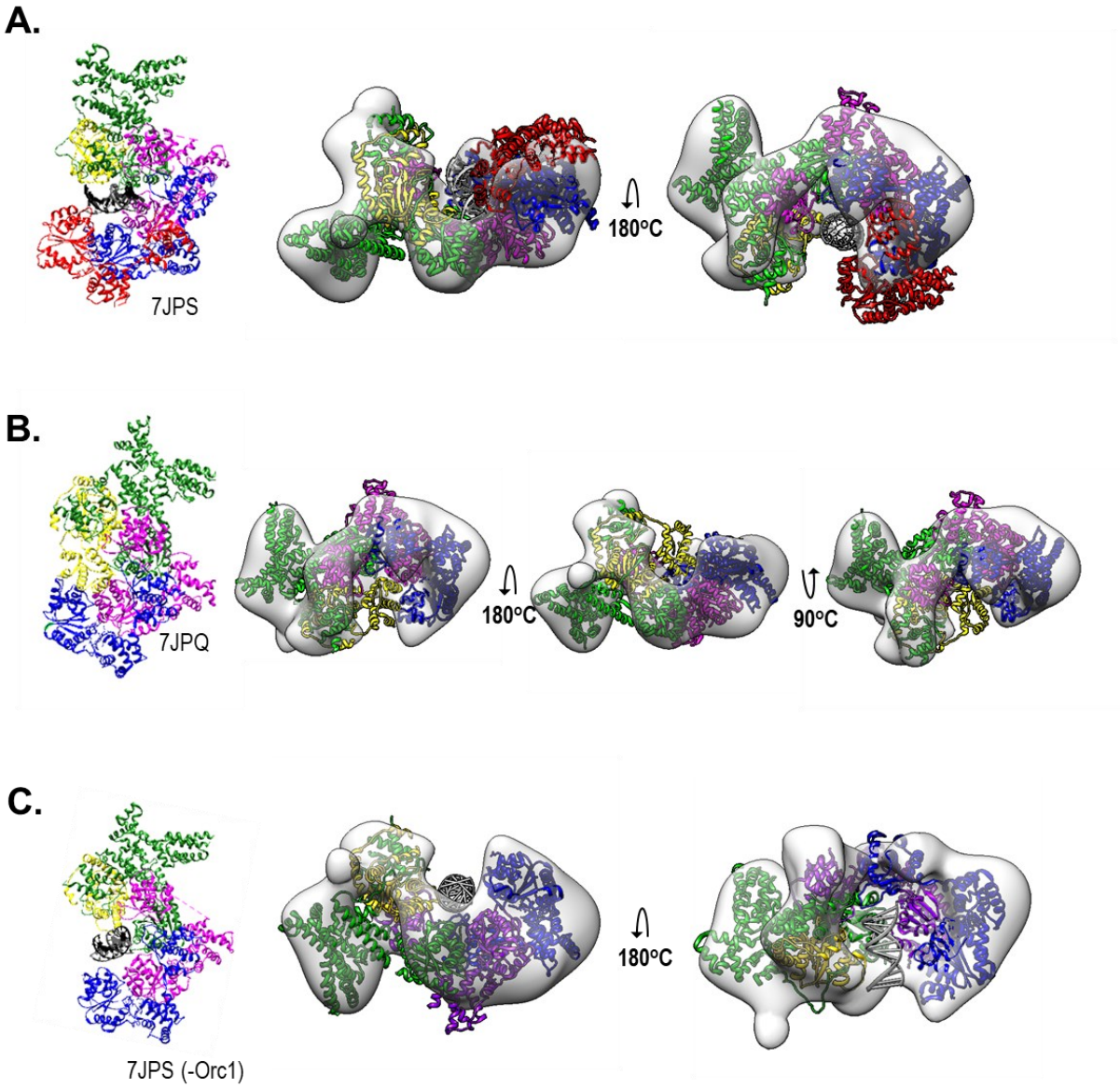


Figure 58: The Orc2-5-DNA complex

Resultant ab-initio 3D electron density map, generated from 2D classes shown in Figure 56.B. ***A:*** *Map with a rigid-body fitting of published Orc1-5-DNA (7JPS, Jaremko et al, 2020), showing that the Orc1 subunit (red) cannot fit into the mapped density.* ***B:*** *The same electron density map as A, except fitted with published Orc2-5 alone, with no DNA (7JPQ, Jaremko et al, 2020). Here, the winged helix domain of Orc2 (yellow) is collapsed into the central core of the model, but this density is not visible in our electron density maps, thus it is predicted that Orc2-5 is not in the same conformation as previously published data* ***C****: The same electron density map as above, fitted with the same Orc1-5-DNA model as described in A, with the Orc1 subunit hidden. This seems to be the closest fitting, with Orc2-5 seemingly bound in a similar orientation as Orc1-5 on DNA, but missing the Orc1 density. Negative stain microscopy does not average out DNA densities, but further cryo-EM can validate whether this Orc2-5 is bound to DNA.*

To structurally determine whether this is the case, samples were optimised for cryo-EM analysis, since this would remove the resolution limits of the stain grain size that resulted in any DNA in the samples being averaged out. Fractions containing what we assumed to be

Orc2-5 bound to DNA were cross-linked with 0.1% glutaraldehyde then buffer-exchanged into the optimised cryogrid buffer (Chapter 3.4.1). While time constraints did not allow for cryo-EM data collection and analysis, initial screening of the grids showed positive results, with good particle distribution (Figure 59). Further work in this direction has the potential to lead to exciting results involving a newly identified Orc2-5-DNA structure.

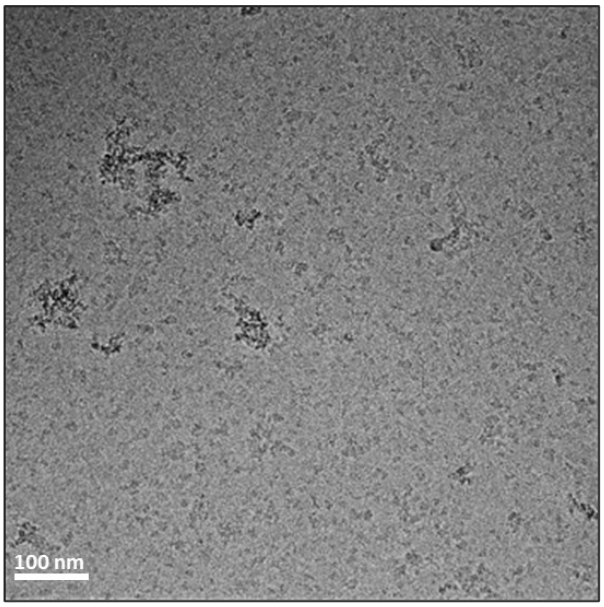


Figure 59: Cryogrid preparation of Orc2-5-DNA

The same sample described in Figures 56 and 57 of Orc2-5 on DNA, cross-linked and buffer exchanged into optimised cryogrid buffer, then applied onto carbon-coated cryogrids. Monodispersed particle are visible, further processing will be needed for generation of 2D classes, ab-initio and refined models,

3.4.5 Introduction of ORC1-ATPase mutations

While using slowly- or non-hydrolysable ATP analogues during the gradient experiments did not show a difference in Orc1 release, experiments showing decreased Cdc6 release when a Cdc6 ATPase mutant was used suggesting that an ORC ATPase mutant could reveal the role of ATP-hydrolysis on complex formation and stability.

Due to the Orc1-Orc4 interface being the only composite ATPase active site, this was targeted to produce an ORC ATPase mutant. An Orc1 Walker B and Orc4 arginine finger double mutant (Orc1 E621Q and Orc4 R209A) was expressed and purified as described in Chapter 2.2.1.1. (Figure 60).

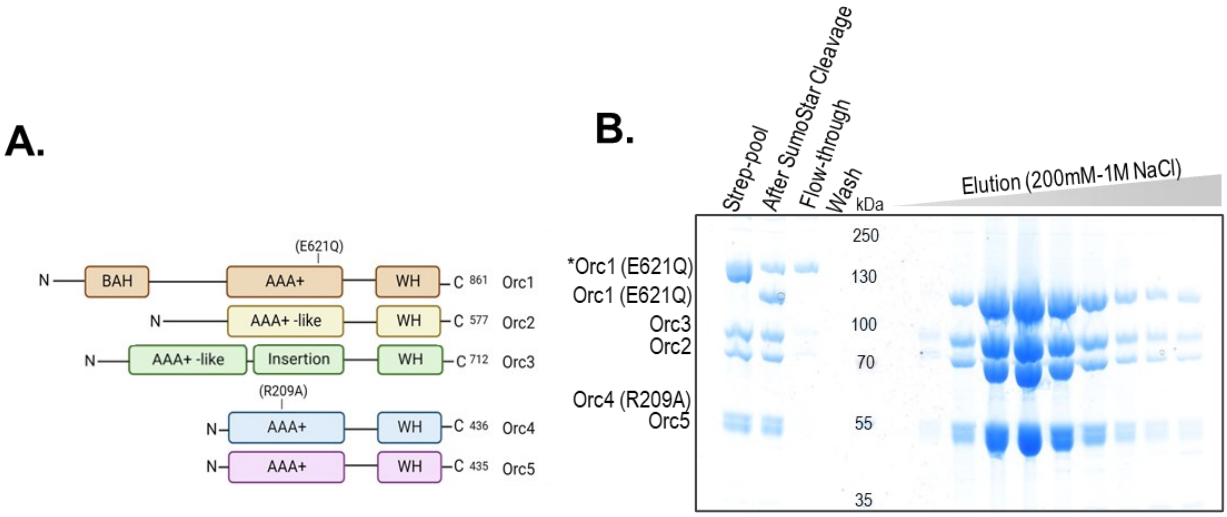


Figure 60: Design and purification of an ORC1-5 ATPase mutant

*A: Schematic representation of domains with catalytic site formed from the Orc1-Ocr4 interface mutated to remove ATPase activity. the arginine finger of Orc4 and the catalytic glutamate in Orc1 have been mutated to an alanine and glutamine respectively, removing the functional groups required for catalytic ATPase activity.. B: Coomassie-stained SDS-PAGE of the Orc1/Orc4-ATPase mutant ORC complex, following successful Streptavidin-mediated affinity, then ion exchange purification (as described in Section 3.1). All five subunits are present (note the * denotes the Strep-tag on Orc1 that was cleaved following affinity purification. Fractions containing clean protein were pooled for future experiments.*

This mutant construct was then subject to ORC alone, ORC-DNA and ORC-Cdc6-DNA gradient experiments, as previously described. Surprisingly, this did not completely block Orc1 release – like Wild-type ORC, upon addition of DNA, there appeared to be some release of the Orc1 subunit, in addition to the ORC-DNA population (Figure 61). However, in comparison to Chapter 3.4.3 and 3.4.4., the increased presence of higher molecular weight Orc1-5-DNA in fractions 10-15 suggests there may be some stabilisation effect when removing the hydrolytic capability of ORC (Figure 61.B in comparison to Figure 51).

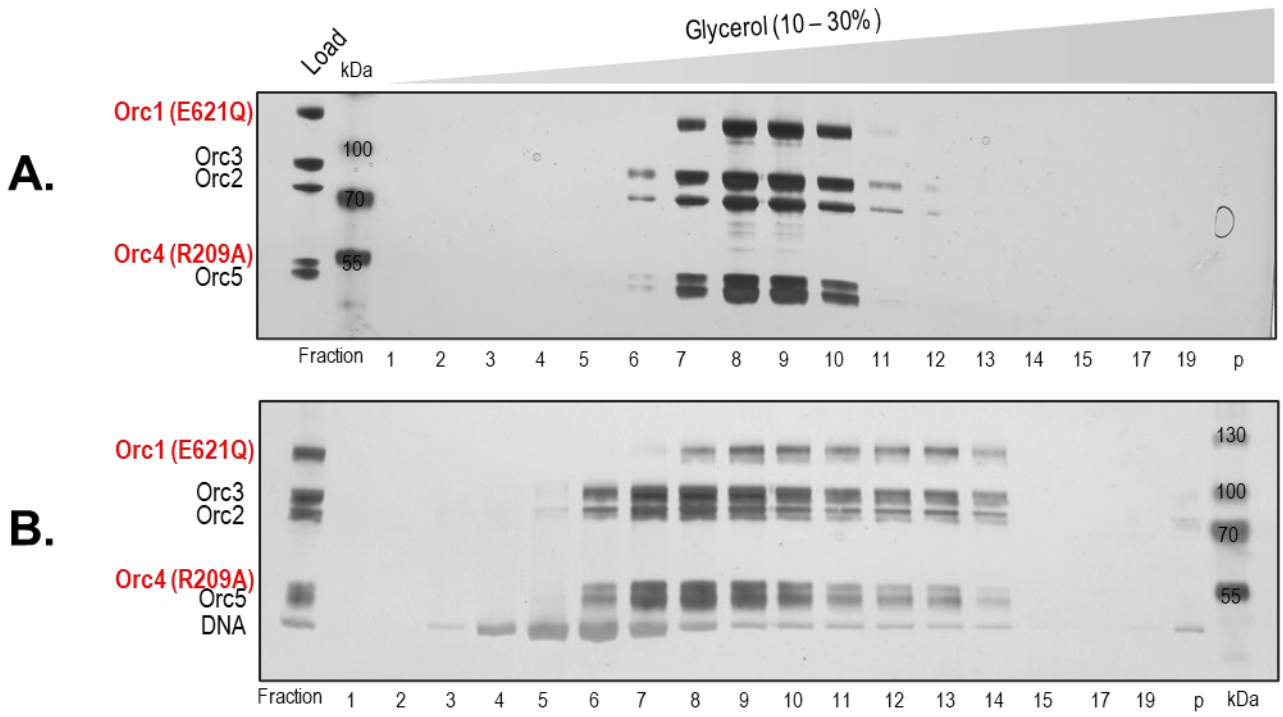


Figure 61: Blocking ORC ATPase activity does not completely stop Orc1 release upon DNA addition but does decrease it.

*10-30% glycerol density gradients of **A**: ORC alone and **B**: ORC-DNA gradients using **ATPase-dead mutant** of ORC1-5 as described in Figure 59. Addition of DNA still resulted in a population of Orc2-5 bound to DNA (F.6-7), but this did not completely block Orc1 release, as there is still a higher-density population of particles (F11-14, suggesting that reducing the hydrolytic capability of ORC can stabilise it long-term on DNA.*

Interestingly, upon the addition of Cdc6, there appeared to be a population of fractions containing ORC (ATPase mutant), Cdc6 (WT) and DNA (Fractions 8 – 10) (Figure 62 A). What is particularly striking is that this is not the same pattern as ORC(ATPase mutant)-Cdc6(E285Q)-DNA, where the full OCD complex fractionates in Fs. 12 -14 (Figure 62.B). This suggests at the possibility of multiple OCD molecules stably bound to DNA in the absence of both ORC and Cdc6 hydrolase activity, resulting in a population of increased molecular weight moieties but a reduced population in the presence of the ORC mutant only. Glycerol density gradients also fractionate according to molecule density as well as size – it could be that the double ATPase mutant samples result in a more dense structure that fractionates differently. However, in order to state this with certainty, more repeats of these experiments must be

carried out, as well as structural studies at a high enough resolution to determine whether this is the case.

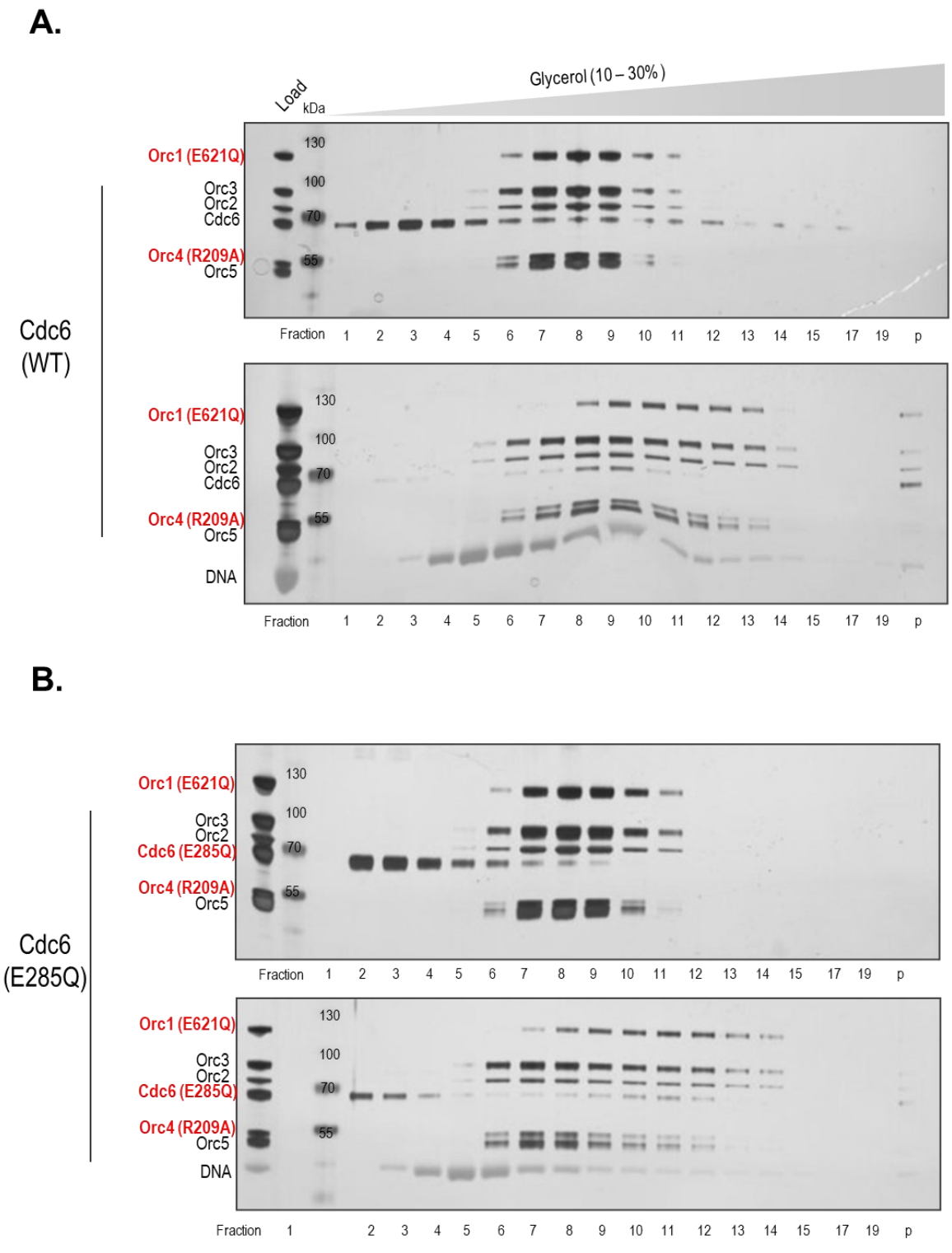


Figure 62: ATPase dead ORC mutants can stabilise Cdc6 binding

A*: Glycerol density gradients of ORC ATPase mutants with and without wild-type Cdc6 and DNA (top and bottom panel respectively). Cdc6 can only bind to the rest of the ORC in the presence of DNA, with a long-lived complex containing both proteins and DNA forming in F8-10. This was not observed in previous ORC-DNA or ORC-Cdc6-DNA gradient experiments previously and must be repeated for validation.* ***B****: Glycerol density experiments with ATPase-dead Cdc6 in combination with ATPase-dead ORC, without and with DNA (top and bottom panel respectively). Fractions 9-12 show a population containing all ORC-Cdc6-DNA units, which has not been seen in previous ORC-Cdc6 DNA binding studies.*

3.4.6 ATPase Mutations in the ORC-Cdc6 complex affects double hexamer formation

While density gradients were initially employed as a method to produce homogenous complexes for electron microscopy, they ultimately uncovered some interesting biochemical behaviours of ORC subunits and Cdc6 in the presence of DNA. However, the interaction of these proteins with DNA does not happen in isolation in the cell – during ORC-Cdc6 binding to origin DNA, they are constantly in the presence of the other pre-RC components. To explore this, the pre-RC assay described in Chapter 2.2.3 was used to study the effects of ORC and Cdc6 ATPase mutations on double hexamer formation. As of yet, it is unclear what the precise role of these proteins' hydrolytic capability is.

It does not appear that these proteins require ATP hydrolytic activity to bind to the DNA – this is evident in low salt washes of the pre-RC assembly, where all pre-RC components are recruited to the DNA, even when ATPase mutants were included (Figure 63 – Low Salt Wash). Notably, the presence of a Cdc6 hydrolysis mutant appears to stabilise itself on the intermediate complexes (Figure 63, Low Salt Wash, Lane 5), which may indicate that ATPase activity is required for disassociation of Cdc6 from other pre-RC components.

Crucially, in the presence of both ORC and Cdc6 ATPase mutants, high salt-stable DNA-bound species consisting of all pre-RC components are observed. This is more pronounced in case of the Cdc6 mutant in comparison to the ORC mutant (Figure 63 , High Salt Wash, Lanes 2-3), potentially correlating to the increased binding of Cdc6 in the low-salt washes. Interestingly, using both of these mutants in a single reaction results in a combined effect, with an increased intensity of ORC subunits (Figure 63, High Salt Wash, Lane 4).

Overall, this suggests that ATPase activity in both ORC and Cdc6 is responsible for release of these proteins upon MCM2-7 recruitment. Without the ability to hydrolyse ATP, these proteins remain associated with the DNA, even upon the recruitment of Cdt1 and MCM2-7.

This is perhaps an adjacent mechanism to that observed in S. cerevisiae, where ATPase activity of Cdc6 is hindered when binding to canonical origin sites, while hydrolysis occurs to release it upon binding to non-origin DNA (Speck and Stillman, 2007). While these results offer this theory, the precise nature of the interaction between these proteins is not known. It could be that these proteins interact in an OCCM-like manner and ATP hydrolysis is required to release them. However, this cannot be said for certain without structural characterisation of the complexes. Electron microscopy of these complexes once eluted from the beads has exciting potential in the understanding of these complexes and the role ATP hydrolysis plays in ORC and Cdc6.

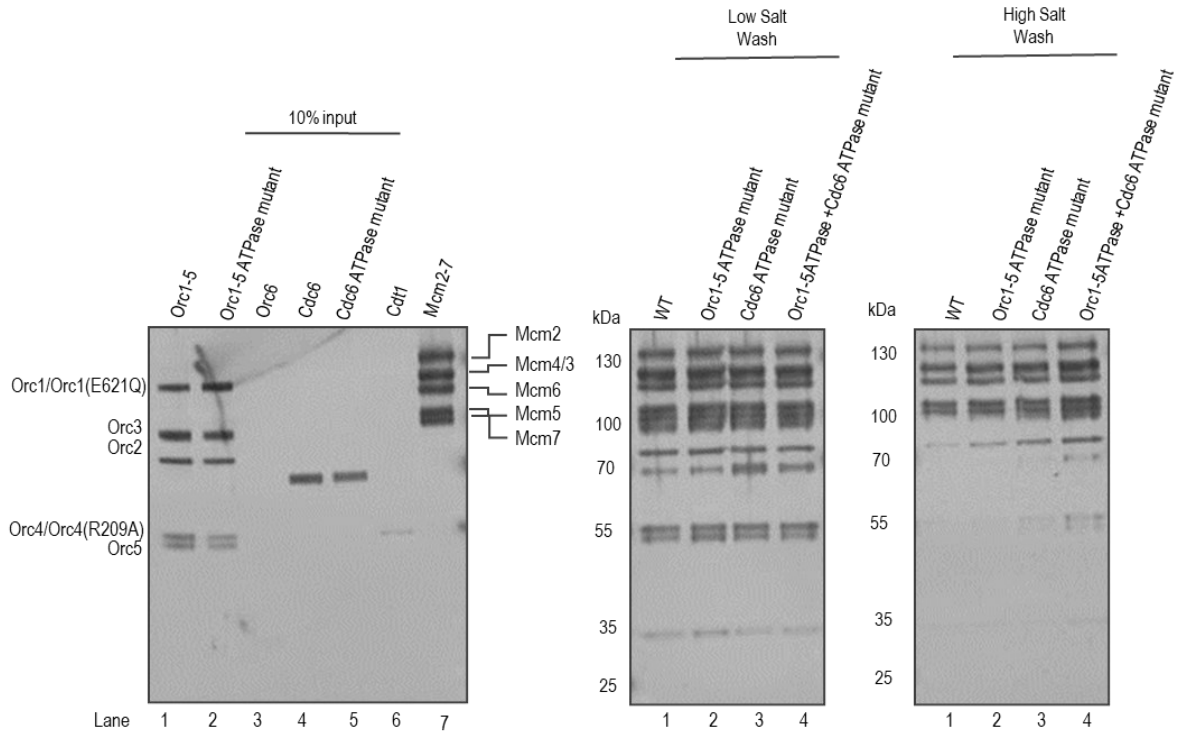


Figure 63: pre-RC loading reaction with ATPase mutants stabilise intermediate complexes

*Silver-stained pre-RC assay using ATPase-dead mutants of Orc1-5 and/or Cdc6 (inputs shown in Panel **A**). B: Low-salt stable pre-RC intermediates bound to DNA. Using each mutant still allows for binding and recruitment of all pre-RC components to DNA, although the use of a Cdc6 ATPase mutant increases its retention (Lane 3). All other subunits are present in what appears to be equal amounts, judging by band density. **C**: high-salt stable complexes bound to DNA, in the presence of ATPase-dead mutants of ORC1-5 and/or Cdc6. The presence of these mutants appears to introduce a stabilisation effect of all pre-RC components when used in combination, even after a high-salt wash (Lane 4). Using the ORC mutant only results in double hexamer loading, similar to wild-type pre-RC loading (Lane 2). Using only the Cdc6 ATPase mutant (Lane 3) resulted in some retention of early proteins, although not to the extent of using both mutants, suggesting that a concerted ATPase hydrolysis between ORC and Cdc6 is required for effective release of these proteins for double hexamer loading.*

3.4.7 N-terminal Deletion of Orc1

Because of the difficulties in structural characterisation of full-length ORC, a truncated form of ORC from amino acids 429-861 (removing the N-terminal BAH domain and linker) was expressed and purified as described above (Figure 64). While likely that this would improve structural characterisation since this domain is highly flexible and responsible for the averaging difficulties in the experiments described above, it is not known what effect this has on the protein's functionality. It's also of note that previous structural studies involving truncated human ORC have not studied its influence on pre-RC loading. Preparations of these proteins were then subject to the same gradient and pre-RC assay conditions as described below, to identify if the presence of the N-terminal domain of Orc1 decreases stability of the entire ORC complex at a biochemical level.

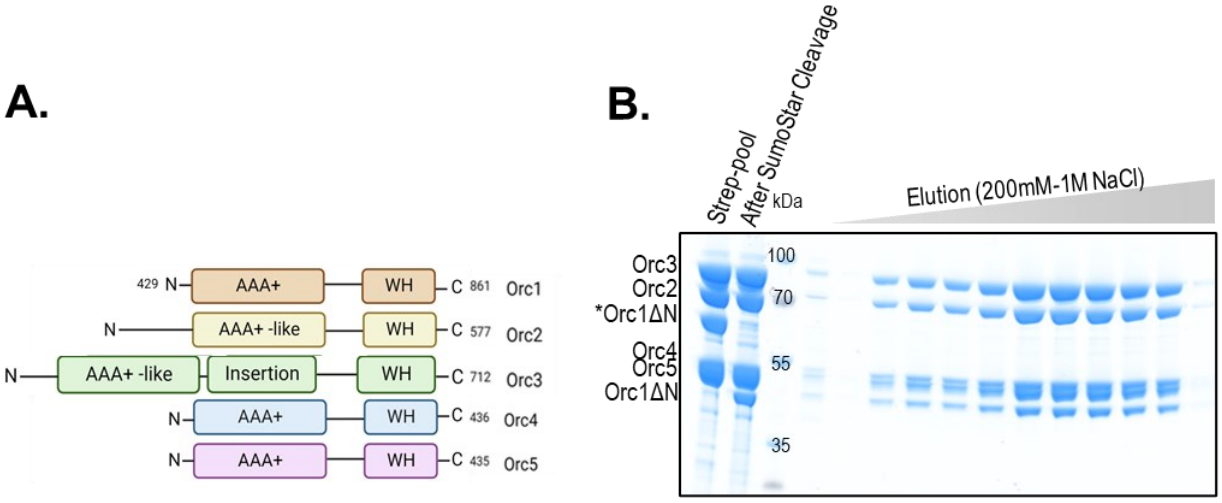


Figure 64: Deletion of the N-terminus of Orc1

***A**: Schematic representation of Orc1 expressed with a truncated Orc1 BAH-linker domain. Amino acids 1-428 were removed from the expression construct, with the new protein still containing a 2x-Strep-Sumo tag on the N-terminus. The truncated complex was produced in the same manner as wild-type ORC, as described in Section 2.2.1. B: Coomassie-stained SDS-PAGE gel following Streptavidin-dependent affinity and ion-exchange purification showing all subunits present in a cleaned preparation. Note: The * denotes 2xStrep-Sumo tagged Orc1ΔN, which was cleaved post affinity purification.*

Orc1-5(Orc1ΔN) on a gradient

At ~40 kDa Orc1ΔN is much smaller than full-length Orc1, so it was assumed that the truncated Orc1-5 complex alone would settle in earlier fractions during a density gradient centrifugation experiment. However, this was not the case and like full-length Orc1-5, the peak fractions were 9-11 (Figure 65 A.). Density gradients can separate based on shape as well as size, so this suggests that the conformation of the core Orc1-5 complex, regardless of the large, flexible N-terminal domain of Orc1 is primarily what contributes to its position across a density gradient.

Including DNA on a Orc1ΔN-5 gradient causes a shift into fractions 11-13 (Figure 65 C.). Despite the DNA running at the same molecular weight on the gel (although DNA appears to stain differentially to proteins) as the Orc1ΔN, this shift suggests that the truncated ORC still retains its DNA-binding capability. It also appears to display the same activity of Orc1 release upon the addition of DNA, judged by the difference in staining of the DNA versus the Orc1 protein band (Figure 65 A-C). Like wild-type Orc1-5, it cannot associate with Cdc6 unless in the presence of DNA (Figure 65 B). However, unlike full-length ORC, truncated ORC seems to be able to form an ORC-Cdc6-DNA complex that can be isolated over the course of the 16-hour gradient (Figure 65 E.). This complex surprisingly fractionates earlier than the main Orc1ΔN-5 peak (peaking in Fractions 7-9) – while this could be due to human error in fractionation, it may also suggest that in total it forms a compact structure that is more buoyant than Orc1-5-DNA or Orc1-5-Cdc6-DNA. Upon the addition of an ATPase-dead mutant of Cdc6, the entire complex, like in previous experiments appears to be stabilised, with potentially multiple OCD molecules bound to DNA fractionating later in the gradient (Figure 65 F, Fractions 11-13). It is unclear whether a population of Orc2-5-Cdc6-DNA exists – in future, experimental repeats and electron microscopy of each of these fractions will be necessary to determine the complex population in each of them.

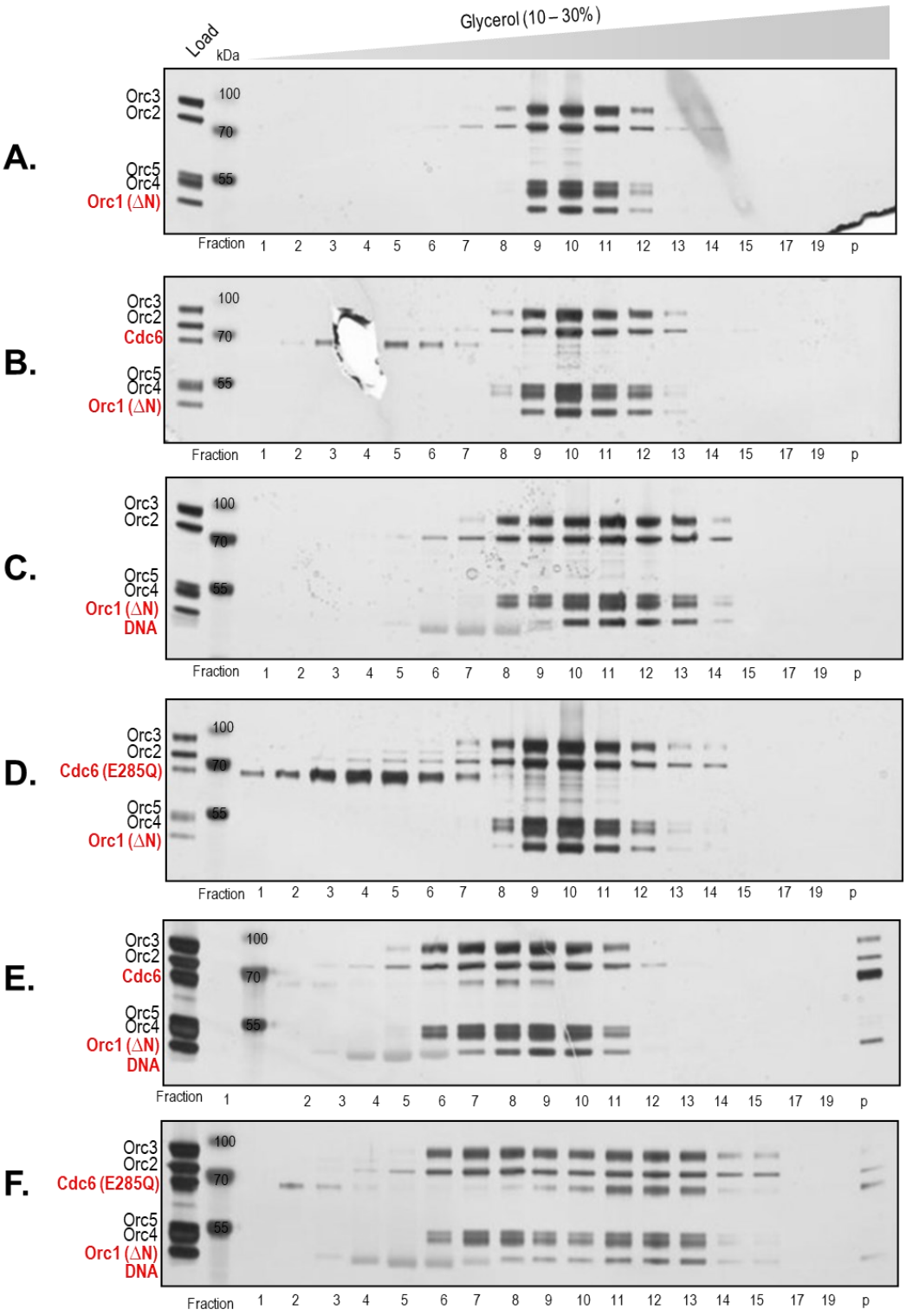

Load
kDa
Glycerol (10 – 30%)
A.
Orc3
Orc2
Orc5
Orc4
Orc1 (ΔN)
100
70
55
Fraction 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 19 p
B.
Orc3
Orc2
Cdc6
Orc5
Orc4
Orc1 (ΔN)
C.
Orc3
Orc2
Orc5
Orc4
Orc1 (ΔN)
DNA
D.
Orc3
Orc2
Cdc6 (E285Q)
Orc5
Orc4
Orc1 (ΔN)
E.
Orc3
Orc2
Cdc6
Orc5
Orc4
Orc1 (ΔN)
DNA
F.
Orc3
Orc2
Cdc6 (E285Q)
Orc5
Orc4
Orc1 (ΔN)
DNA

Figure 65: Orc1ΔN-5 can still bind to DNA and appears to stabilise OD and OCD interactions.

A: *Orc1ΔN-5 on a 10-30% glycerol density gradient alone. This complex fractionates further into the gradient (F9-11) than full-length Orc1-5, despite being lower in molecular weight.* ***B:*** *Orc1ΔN-5-Cdc6(WT) incubated together and run on a gradient. These two proteins cannot bind without the presence of DNA; Cdc6 fractionates completely separately (in F.3-6) to ORC* ***C:*** *Orc1ΔN-5 can associate with DNA, and this appears to facilitate Orc1 dissociation (Orc2-5 on DNA in F.8-9). It also appears to stabilise a population of ORC on DNA (F11-13).* ***D:*** *Orc1ΔN-5-Cdc6(E285Q) cannot associate with each other in the absence of DNA, with each protein fractionating separately.* ***E:*** *Orc1ΔN-5-Cdc6-DNA can be stabilised in the long term (Fraction 7-9), and can be even further stabilised by the addition of* ***F:*** *Cdc6 (E285Q), ATPase-dead Cdc6m which forms a stable truncated ORC-Cdc6 mutant – DNA complex in Fractions 11-13.*

Note: this experiment has only been carried out once and needs repeating for validation of these results. The unexpected fractionation pools due to the size of protein need validation – the proteins may be more/less dense than their wild-type counterparts, which could be causing them to fractionate differently across a 10-30% gradient.

Orc1-5(Orc1ΔN) can still recruit the double hexamer.

The presence of a truncated form of Orc1 does not affect double hexamer assembly – in fact, when comparing high salt washes following pre-RC assembly of full-length and truncated ORC, truncated ORC appears to result in ~~–~~clean~~,~~ MCM2-7 bands without any residual ORC subunits (Figure 66 A.).

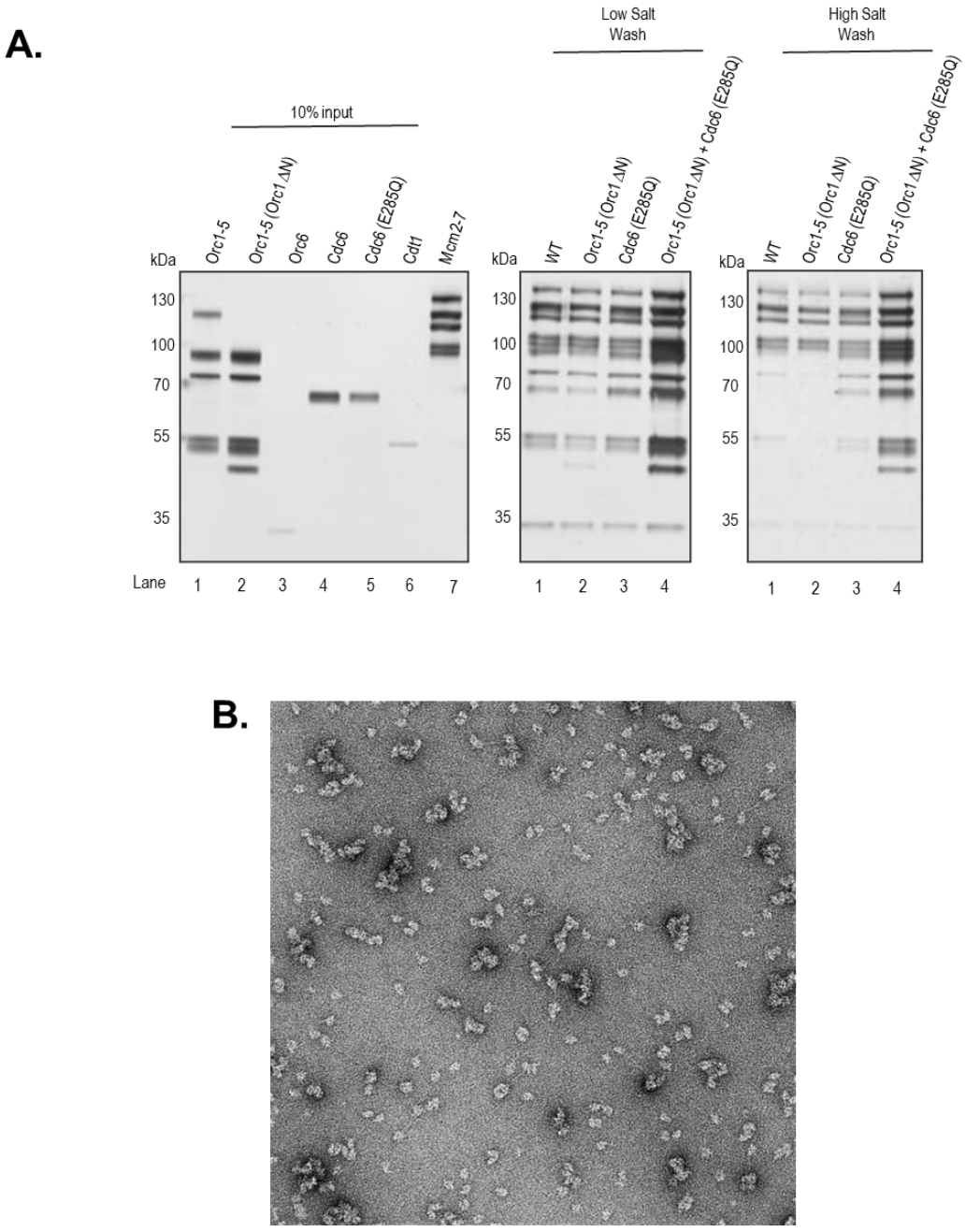


Figure 66: Orc1ΔN-5 does not hinder double hexamer formation.

A: *Pre-RC assay showing that Orc1ΔN-5 can still load double hexamer, with a massive stabilisation of both low- and high-salt stable intermediates (Lanes 4, right panels) upon using ATPase-dead Cdc6, suggesting the N-terminus of Orc1 may play regulatory roles in double hexamer formation and pre-RC formation release.* ***B:*** *Negative stain micrograph of Orc1ΔN-5-Cdc6-DNA of F.12 from Figure 64 F, which could be the truncated ORC-Cdc6-DNA complex.*

Orc1-5(Orc1ΔN) in association with a Cdc6 ATPase mutant

Interestingly, using combination of truncated Orc1 and Cdc6 ATPase mutants results in significant stabilisation of intermediary and high-salt stable complexes (Figure 66 A, High Salt washes Lane 3-4). Additionally, the gradients containing the Cdc6 ATPase mutant also show a stabilised population of ORC-Cdc6-DNA in Fractions 11-14 (Figure 64 F.). Considering the population of OCD with wild-type Cdc6 fractionates primarily in Fs. 7-9, this shift may suggest a stable population of DNA that contains multiple ORC-Cdc6 bound complexes per DNA. This late-fractionating population is also seen when using full-length ORC and Cdc6 ATPase mutants in combination, suggesting that this produces a complex that is stable over time.

Negative Stain Electron Microscopy of Orc1ΔN-Cdc6-DNA

This seemingly increased stabilisation of ORC-Cdc6-DNA when using an N-terminal mutant is extremely promising for electron microscopy experiments. Fraction 12 from Figure 65 F. was collected, and negative stain grids were prepared as previously described (Figure 66 B.). While time limitations in time did not allow for the processing of this sample, it would be interesting to identify whether removal of the N-terminal domain results in a more rigid structure, allowing for higher resolution in 2D classes and 3D modelling.

Discussion

Overall, these results have begun to uncover that the N-terminal domain of Orc1 affects ORC functionality. While it may be more favourable in structural studies to remove this region, it is evident from biochemical assays that this region may be involved in the regulation of double hexamer assembly.

While initially, the residual subunits of ORC remaining post high-salt was dismissed as an artefact (Figure 28), the lack of these residual subunits when carrying out the same

experiments with truncated ORC suggests that the N-terminal region of Orc1 may be involved in regulating double hexamer formation. This N-terminal region of Orc1, as well as its reported BAH domain involved in chromatin localisation, is highly regulated via phosphorylation modifications that controls its role throughout the cell cycle. While studying the phosphorylation state of proteins is outside the scope of this thesis, these results show that it may be worth studying the effects of this. Previous work in the lab has identified specific sites involved in post-translational modifications, particularly phosphorylation (Figure 67)

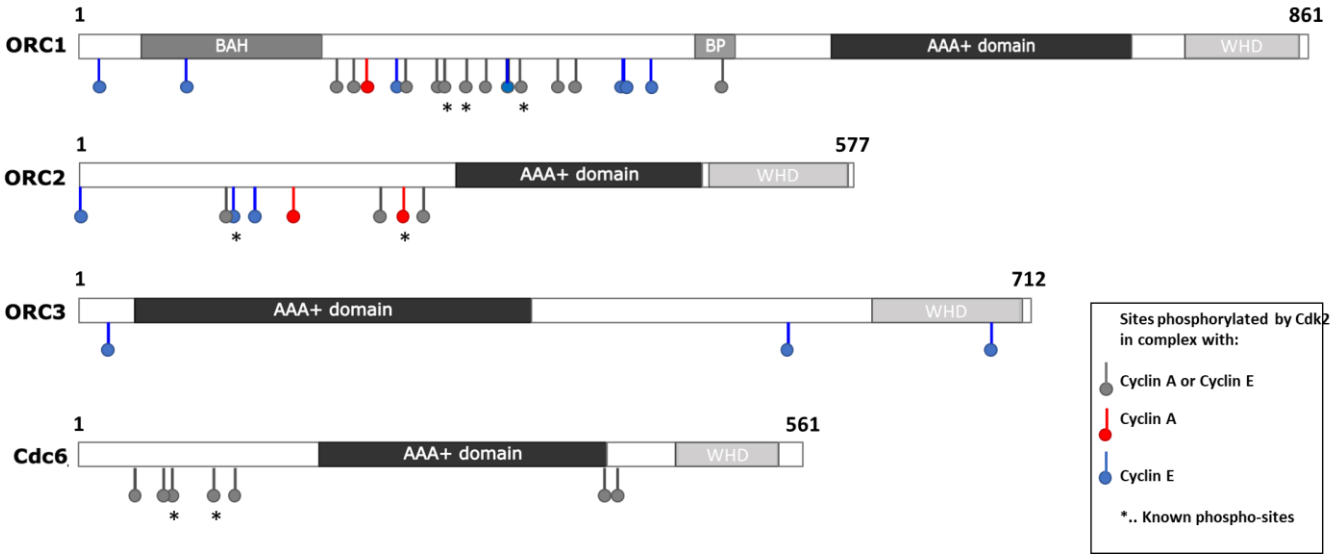


Figure 67: The intrinsically disordered regions of ORC and Cdc6 are heavily modified by phosphorylation

Schematic representation of Orc1, Orc2, Orc3 and Cdc6 with marked phosphorylation sites as determined by in vitro phosphorylation with purified proteins and mass spectrometry. In total 37 sites were identified. 21 are phosphorylated by both cyclin A/cdk2 or cyclin E/cdk2 (grey); a majority of the sites (19) are located on the N-terminal half of Orc1 which includes the Bromo Adjacent Homology (BAH) domain, the basic patch (BP), as well as intrinsically disordered regions. 8 sites are located on the N-terminal part of Orc2, which is also disordered and poorly resolved in structural studies. Only 3 sites are located on Orc3 and 7 sites were found on Cdc6. No sites were found on Orc4 and Orc5. Experiment carried out by Dr. Vera Leber and the LMS-MRC Mass Spectrometry facility

The improved yield upon use of phosphatase inhibitors in the purification process also indicates that the phosphorylation state of ORC contributes to its stability. Producing phosphorylation mutants of ORC and testing these in the helicase loading assay may be useful in determining the precise roles of specific modifications on ORC stability throughout the cell cycle. Since phosphorylation is key in regulating activation and deactivation of proteins throughout the cell cycle, combinations of modifications may be a finely tuned way of

repurposing active ORC into an autoinhibitory state that can still bind to DNA, but does not produce double hexamer that can be activated for bidirectional replication.

It may be interesting to discover whether the release of Orc1 can be reduced or inhibited with specific phosphorylation states. Previous studies have shown that non-chromatin bound Orc1 is subject to ubiquitination post S-phase onset (Li and DePamphilis 2002), presumed to facilitate the inactivation of ORC outside of G1. The observed release of Orc1 across the gradients over time upon addition of DNA could be an intrinsic property of the protein, resulting in ORC that can bind and recruit the double hexamer in G1, but can be inactivated by it's disassociation over time (G1 lasts on average 11 hours in proliferating human cells (Wang, 2021), resulting in residual chromatin-bound Orc2-5. It could be that chromatin-bound Orc2-5 serves to reduce the transcriptional energy required in daughter cells to synthesize a completely new population of ORC molecules, or marks replication sites in daughter cells.

One particularly interesting future experiment could be testing whether Orc1 can reassociate with Orc2-5 bound to DNA. Attempts were made to purify Streptavidin-tagged Orc1 alone with the intention of performing immunoprecipitation experiments of Tagged Orc1 with Orc2-5 produced from density gradient experiments, although poor yield of Orc1 alone limited further experiments. Optimisation of these experiments could provide an answer to whether this is possible, or if a conformational change is required upon the M/G1 transition to enable reassociation of Orc1 with the rest of the ORC subunits.

3.4.8 Beyond the ORC-Cdc6-DNA complex: An attempt at OCCM

Despite the difficulty in obtaining a stable ORC-Cdc6-DNA complex, we decided to try to form later intermediates, with the potential that some of these may stabilise flexible regions to allow easier structural determination. With the differences in intensities between eluted complexes from low salt and the high salt washes shown in Figure 28, it is evident that alongside the salt-

stable double hexamer, there are a number of low-salt intermediate complexes where all pre-RC components are recruited to the DNA.

Initial work by Dr. Jen Wells and Dr. Tonia Kefala Stavridi in the Speck lab showed that electron microscopy of eluted material resulted in what appears to be a range of different intermediary complexes (Figure 68).

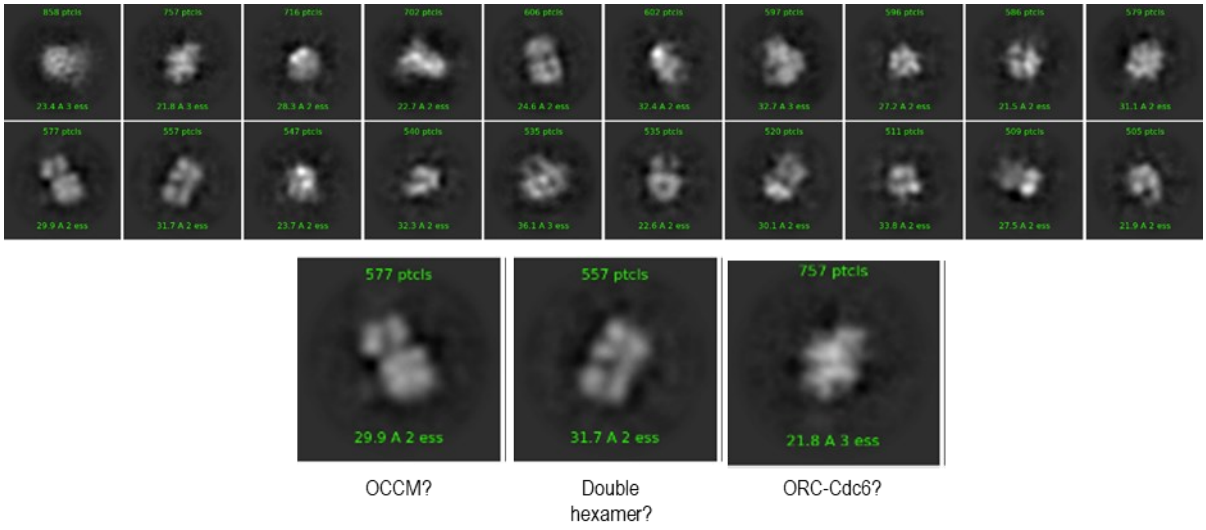


Figure 68: Pre-RC intermediates in solution are variable.

2D classes showing a range of pre-RC intermediates upon elution from biotin-conjugated DNA beads with a low salt wash. Classes show a number of distinct particle dimensions that could correlate to different intermediates, such as the OCCM, double hexamer and Orc-Cdc6. Data collected and processed in CryoSPARC by Dr Jen Wells and Dr Tonia Kefala Stavridi

While this shows that the double hexamer assembly process is dynamic with multiple intermediate processes, it makes structural characterisation of individual intermediates more complicated. To try to target this, density gradient experiments were used to separate intermediates by size and shape. This process has been used in the lab in the past to successfully separate *S. cerevisiae* double hexamer from double hexamer bound to activating kinases (Saleh et al, 2022)

Glycerol Gradients of Orc-Cdc6-Cdt1-MCM bound to DNA

To separate larger molecular weight intermediates, glycerol density gradients were prepared as previously described, except with a range of 15 – 45% glycerol. In addition to this, ATPγS

was used in the buffers instead of ATP, to attempt to preserve intermediates over the course of the gradient that rely on ATP hydrolysis for component release. Pre-RC components were added in the same ratios as used in the double hexamer assembly assay (Chapter 2.2.3)

In these conditions, ORC-DNA alone fractionated earlier, in F.5-7 (Figure 69 A.), with the MCM2-7 single hexamers fractionating across F.7-14, although with a peak at F.12 (Figure 69 B.). This does not change upon the addition of DNA, Cdc6 and Cdt1, showing that without ORC, these components cannot bind to DNA (as shown biochemically in the pre-RC assay in Figure 28). There does exist a single fraction (F.10) in which it seems MCM2-7-Cdt1 can bind to each other (Figure 69 C.).

Interestingly, upon addition of all pre-RC components (Figure 69 D.), the core MCM2-7 peaks do not shift in size, suggesting they do not assemble into a double hexamer or associate in the long term with the other pre-RC components. However, a shift in the ORC, Cdc6 and DNA bands is observed, further into the gradient in Fractions 7-11. It may also be that Cdt1 follows this shift, although it cannot be conclusively determined, since it runs at the same molecular weight as Orc4.

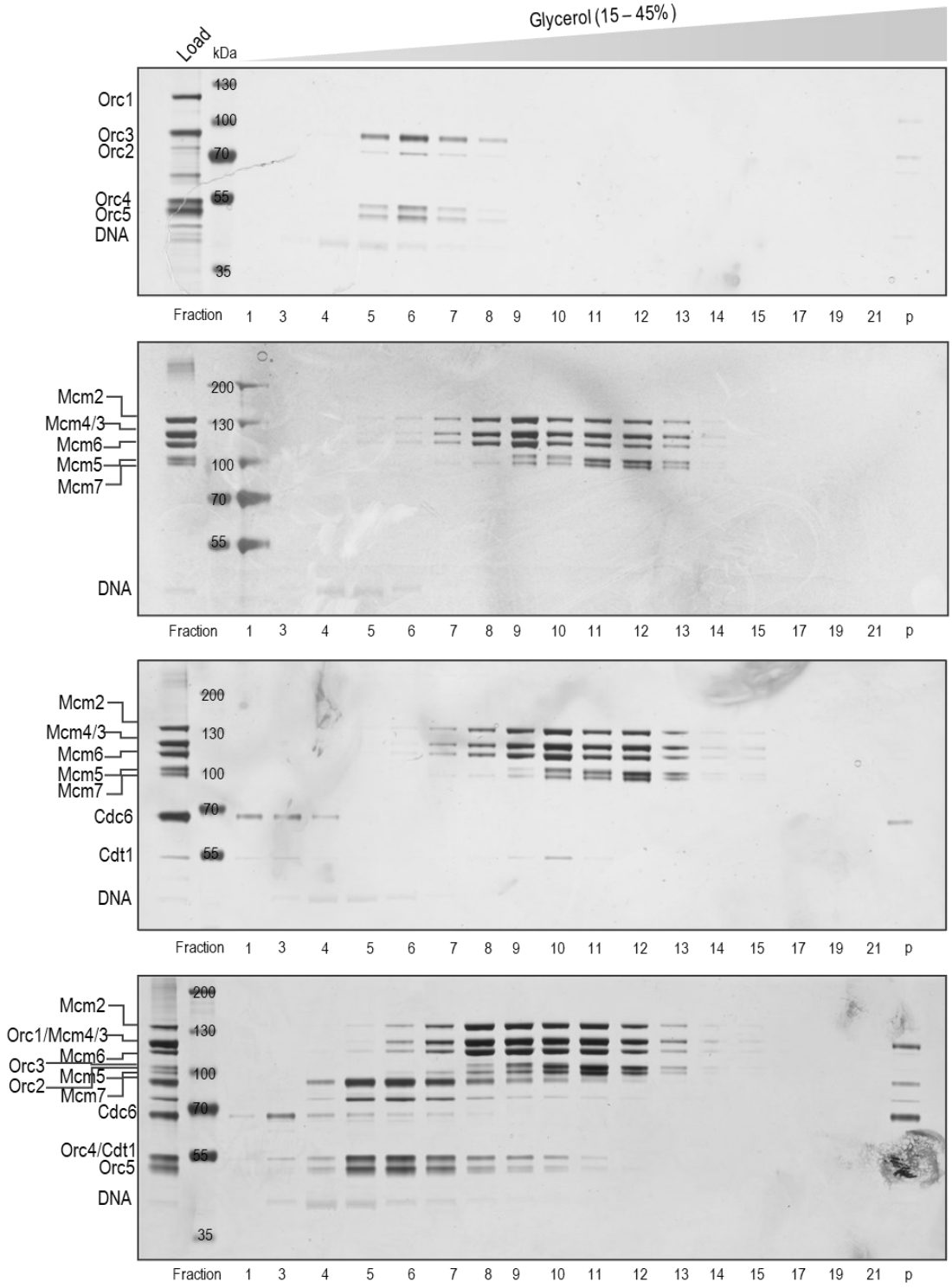


Figure 69: ORC-Cdc6-Cdt1 and MCM on a 15-45% glycerol density gradient

A: *Orc1-5 and DNA on a 15-45% gradient, showing a majority Orc2-5-DNA complex fractionating in F5-7.* ***B:*** *Similar gradients with MCM2-7 and DNA incubated together. MCM2-7 cannot bind to DNA alone, with the MCM2-7 subunits streaking across the gels, although with a peak in F.11/12 and the DNA fractionating separately in F4-6.* ***C:*** *Cdc6-Cdt1-MCM2-7 cannot bind to DNA without the presence of Orc1-5, with all proteins and DNA fractionating separately (Cdc6 in F1-4, DNA in F.4-6, Cdt1 in F1-3 and alongside MCM in F10, MCM2-7 in F9-13).* ***D****: Gradient displaying all subunits, with only Orc1-5 subunits shifting further into the gel (F.5-9), although no shift of the DNA is observed. No shift of MCM2-7 is seen, suggesting that a higher molecular weight moiety is not formed. .*

Negative stain Electron Microscopy of potential OCCM

Despite the lack of MCM2-7 shift, Fractions 8 and 10 were cross-linked with 0.1% glutaraldehyde, then negative stain grids were prepared as described previously. These two fractions were chosen since Fraction 8 appeared to contain all pre-RC components, while Fraction 10 showed a shift in the ORC subunits but does not appear to contain Cdc6 – potentially an intermediate of ORC-MCM bound to DNA as described in (Figure. 70) (Gupta et al, 2021). Small datasets were acquired, the details of which can be found in Supplementary Table 7.7 and 7.8 respectively.

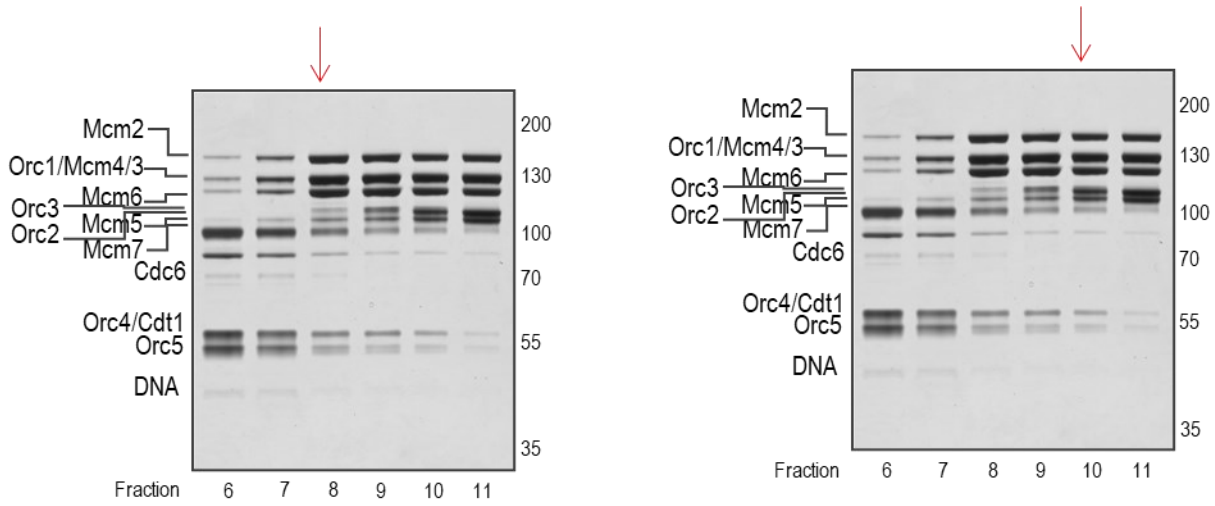


Figure 70: ORC-Cdc6-Cdt1 and MCM on a 15-45% glycerol density gradient – NS EM analysis

Fractions from gradients shown in Figure 69.D chosen for negative stain analysis. Fractions 8 and 10 were chosen since they contained all protein subunits, although only Fraction 8 had a visible Cdc6 and DNA band. Fractions were cross-linked and negative stain grids were prepared according to Section 2.2.10.

However, upon particle picking and 2D classification, it was evident from resultant classes that the particles present in both datasets were primarily MCM2-7 subunits alone (Figure 71). Some potential ORC subunits were present in Fraction 8, but the predominant classes contained Mcm2-7 single subunits. Interestingly, there were some classes present with 'double hexamer-like' particles, where two MCM2-7 hexamers are positioned in a seemingly

head-to-head manner, but not in a perfectly symmetrical configuration as seen in Figure 28, released by cleavage of surrounding DNA after a high salt wash. Populations of these asymmetric particles were even more prevalent in Fraction 10, with a majority of the 2D classes displaying this configuration.

The presence of these asymmetric double hexamers hint at the potential of two MCM hexamers that can associate, but do not form the salt-stable DNA-bound head to head hexamer that remains at origin sites until activation at the onset of S-phase. While Chapter 3.4 shows that MCM hexamers cannot bind to DNA unless ORC1-5 is present, a recent structural study into human MCM2-7 hexamers (Li et al, 2023) show that they can associate into loose double-hexamer like structures, without DNA (Figure 71. C). It is likely that this is what is happening across the density gradient.

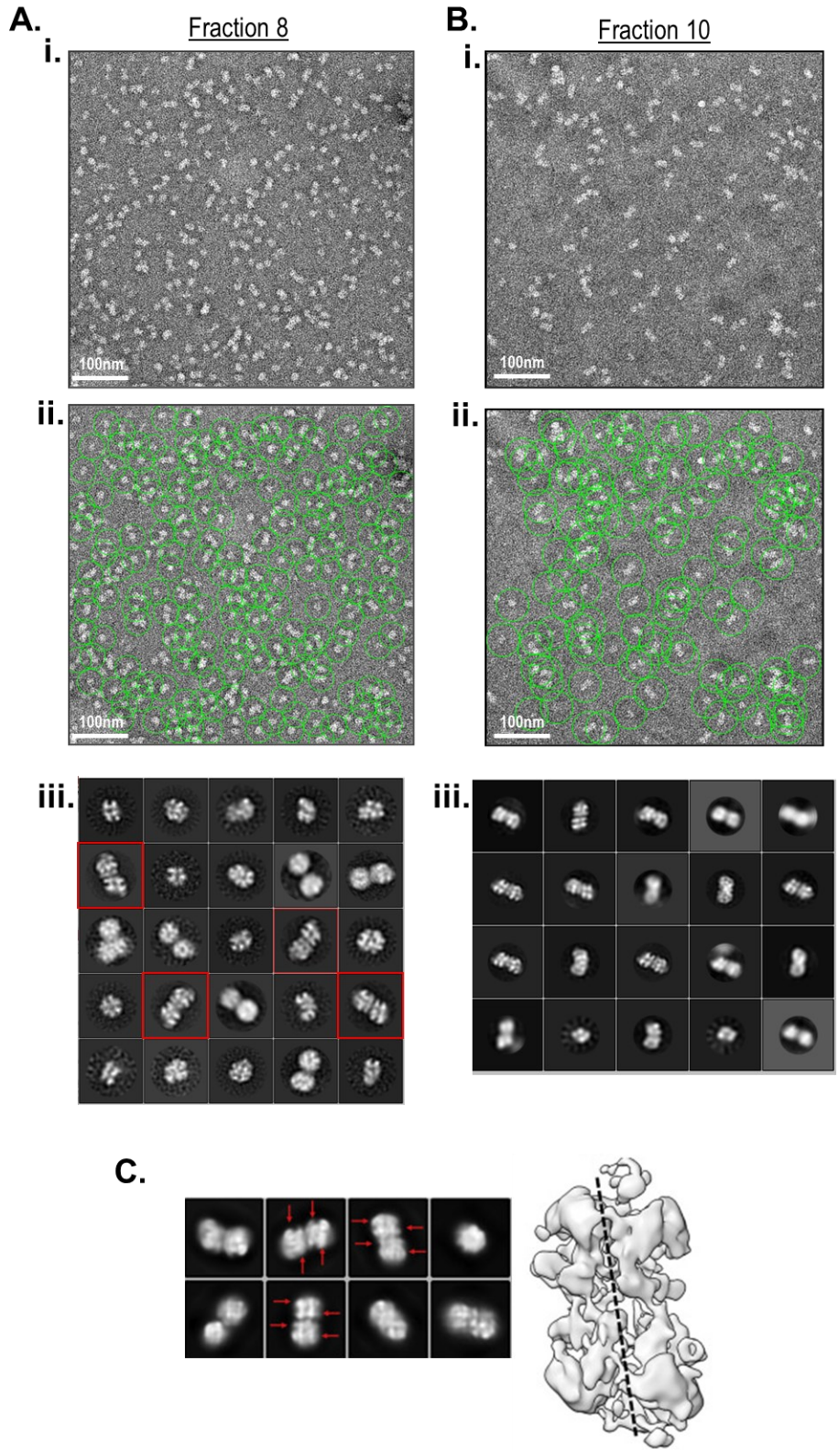


Figure 71: MCM2-7 can self-associate into MCM2-7 asymmetric double hexamers.

A: i: *negative stain micrograph of F.8* ***ii.*** *Corresponding picked particles from reference-based particle picking i****ii:*** *corresponding 2D classes generated, showing mostly solo MCM2-7 hexamers and asymmetric DH* ***B: i****: negative stain micrograph of F.10* ***ii****. Corresponding picked particles from reference-based particle picking* ***iii:*** *corresponding 2D classes generated, showing mostly asymmetric DH.* ***C:*** *comparison with 2D classification and 3D modelling results from Li et al, 2023 showing that two MCM2-7 can associate into an asymmetric double hexamer without binding to DNA*

While this is an interesting insight into the behaviour of MCM, it is not possible to determine any OCCM or OM-like intermediates bound to DNA with these experimental conditions. It is also entirely possible that assembled double hexamer can slide off DNA – it may be that later fractions of the gradient contains these structures. In this case, it may be valuable to use DNA constructs with structural 'roadblocks' at each end that still allows pre-RC formation along the DNA, but assembled complexes cannot slide off of DNA and potentially fall apart into their individual components.

These reaction conditions may also need optimisation to identify and separate any intermediate complexes. Like in Chapter 3.4.3, where an ORC-Cdc6-DNA complex could only be isolated in the case of an ATPase mutant despite the usage of ATPγS or ADP-BeF, it could be that residual ATP in the protein preparations are leading to release of individual protein complexes across the course of the gradient.

Discussion – The recruitment and regulation of pre-RC components to DNA to facilitate DNA replication.

Overall, these results have been a step forward into uncovering the mechanisms behind pre-RC assembly using human proteins.

Much progress has been made regarding the purification of full-length human pre-RC proteins, with a particular emphasis on ORC. In combination with optimised purifications of the other pre-RC components, it has allowed for the development of a reproducible method to test double hexamer loading onto DNA and the steps leading up to its assembly at origin sites.

Like previously established eukaryotic systems, effective loading is highly dependent on the presence of Orc1-5, Cdc6, Cdt1 and Mcm2-7. Supporting findings from previous cellular and biochemical studies, Orc6 is not crucial for helicase loading, although further research will be required to identify where its role lies in the replication pathway.

Genomic studies have long shown that human ORC does not require a specific sequence to bind to DNA. While specific feature analysis was out of the scope of these studies, it was evident that human ORC1-5 has an innate ability to bind to any sequence of DNA. This binding is enhanced upon the addition of Cdc6, although gradient density results indicate that this interaction may be transient. Indeed, in the cell, ORC-Cdc6 is faced with an excess of Cdt1 and MCM2-7 so it is likely that an ORC-Cdc6 interaction is not long-lived in vivo and is ultimately required to disassociate for efficient double hexamer loading. This is further supported by results presented here that show that the ATP hydrolysis activity of ORC and Cdc6 is required for disassociation from the DNA to allow for efficient double hexamer loading. When using mutants that do not hold any hydrolytic capability, salt-stable double hexamer assembly is hindered by reduced dissociation of ORC and Cdc6.

In addition to these findings, the Orc1 subunit has shown some extremely interesting activity in vitro that may be correlated to its function in a cellular context. Full length Orc1-5 has the capability to bind to DNA and recruit double hexamer, but over time, it can disassociate from the core, leaving a stable Orc2-5 complex bound to DNA. Previous work has shown that levels of Orc1 can fluctuate throughout the cell cycle, which supports this finding (Kara et al, 2015). In a cellular context, this may be important for reducing the energetic cost of synthesising completely new ORC subunits upon the onset of a new G1 phase, or as an epigenetic marker of origin sites in progeny cells. It remains unclear whether Orc1 can reassociate with Orc2-5 on DNA – this cannot be concluded from the experiments carried out on a glycerol gradient level, but future experiments in purifying a stable Orc1 alone will be crucial in probing whether this can happen. Initial experiments with ORC ATPase mutants and an Orc1 truncation mutant show that these features can marginally reduce Orc1 release, but there remains a significant population of Orc2-5 bound to DNA post gradient fractionation.

This activity may be dependent on the phosphorylation states of these complexes - indeed, phosphorylation of proteins is a key method of regulating protein function throughout the cell cycle. Both ORC and Cdc6 contain regions subject to variable levels of protein

phosphorylation; while Orc1 contains the most phospho-variable sites, Orc2 and Orc3 can also be modified in this way, which may explain the decreased, but still present Orc1 release. Studying precise phosphorylation patterns was out of the scope of this thesis, but it is an exciting idea for future experiments. Combinations of phospho-mimetic mutants can be purified and studied for its ability to bind DNA, recruit double hexamer and whether it promotes or reduces Orc1 release. Correlating these phospho-mutants and their in vitro biochemical activity to phosphorylation states observed at different stages in the cell cycle will allow an in-depth insight into the way ORC is regulated both in and out of replication licensing.

More work is required to gain high resolution structural data for the identified intermediates discussed in these results. These experiments have provided a good starting point in identifying conditions that these complexes are most stable in and have managed to trap short-lived intermediates such as ORC-Cdc6-DNA using specific ATPase mutants to allow for homogenous sample generation. High sample flexibility has been a major obstacle in preventing high-resolution structural generation – published structural studies have, understandably, had to truncate these proteins or model only part of them in the electron microscopy density data to allow for high-resolution structural information. Truncation of flexible linkers may provide better structural resolution but comes with the caveat of potential pseudo-structures that do not exist in vivo when full-length proteins are subject to functional – and structural - modifications.

3.5 – ORC and Chromatin

Introduction

Another aim of this project was to understand how the entire ORC structure interacts with chromatin and nucleosomes, as this can offer insight into how origins are specified across the diverse landscape of the genome. ORC must be distributed across the genome – including nucleosome-dense regions, to ensure complete replication during S-phase. Although current

data shows that ORC does not possess an intrinsic DNA sequence specificity, multiple protein-protein interactions between ORC and chromatin/chromatin associated factors are likely utilised to stabilise ORC binding across the genome to regulate origin licensing.

Because of the success in the purification of full-length ORC, with an intact Orc1-BAH domain, interactions with nucleosomes were studied, with the intention of elucidating whether there are specific requirements needed for recruitment of ORC to chromatin. A range of histone modifications were investigated – unmodified nucleosomes, H4K20me2 modified nucleosomes, typically associated with open chromatin and H3K9me3 modified nucleosomes, typically found in heterochromatic regions.

The potential establishment of this ORC-nucleosome construct is not restricted to the study of replication initiation. Previous studies in *S. cerevisiae* have shown that while the N terminus of Orc1 is not involved in DNA binding, absence of this domain leads to aberrant silencing (Zhang et al, 2000). Since ORC has been shown to be involved in multiple non-replicative functions across the genome, particularly in heterochromatic regions (reviewed Popova et al, 2017), the ORC-nucleosome complex can be used as a template to explore high resolution mechanisms of novel non-replicative ORC-related functions.

Reconstitution of nucleosomes by salt-gradient dialysis

To increase the understanding of how ORC interacts with chromatin, mononucleosomes were assembled by mixing either unmodified, H4K20me2 or H3K9me3 octamers with a DNA construct of the Widom-601 147-bp nucleosome positioning sequence, fused with varying levels of DNA1 (See Supplementary Table 2) to serve as extranucleosomal flanking sequences.

The nucleosomal core particle (NCP) was assembled by using a salt gradient dialysis as described in Chapter 2.2.6 and Luger et al, 1997. This involved a titration of histone octamers to determine the optimal DNA:histone ratio for fully occupied and assembled mononucleosomes with minimal free DNA. Eventually, a 1:1.5 DNA:histone ratio was deemed

as optimal, although some free DNA remained, which may compete with ORC-nucleosome binding (Figure 72 A-B).

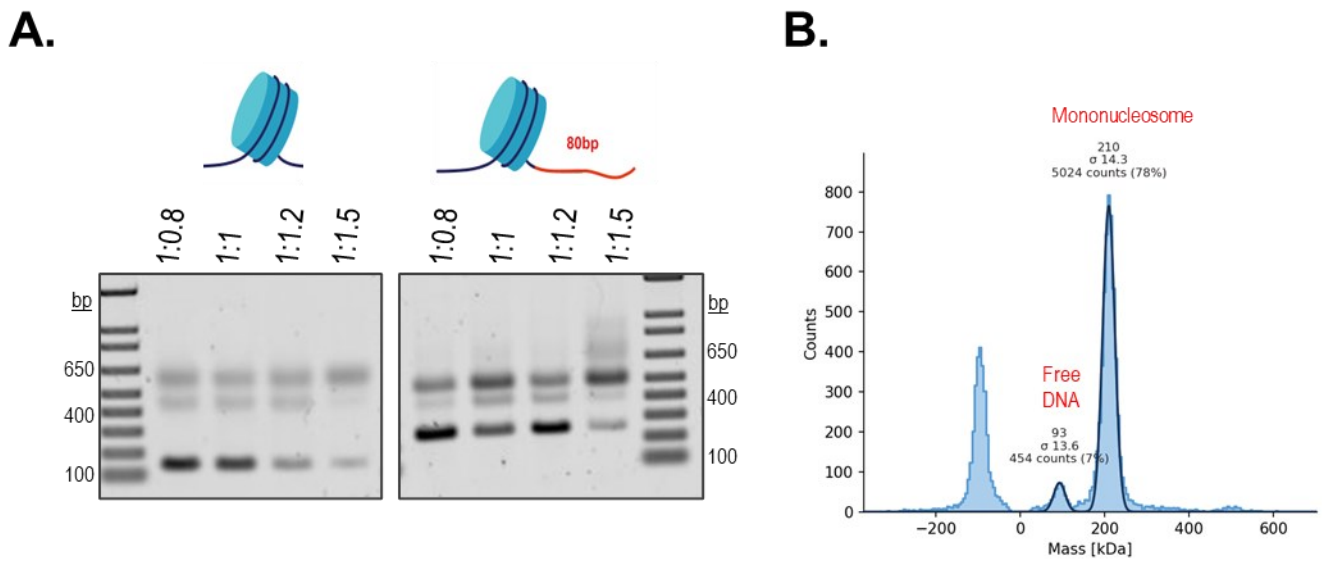


Figure 72: Reconstitution of nucleosomes by salt gradient dialysis

A: *The Widom 601 sequence (with or without additional extranucleosomal DNA) was incubated alongside a titration of nucleosomes in a 2M – 150 mM KCl salt gradient dialysis experiment and corresponding EMSAs were run to determine the optimal DNA:nucleosome ratio (1:1.5). The shift in DNA side from 150bp to ~500bp indicates nucleosome binding to the DNA.* ***B:*** *Mass photometry experiments showing a mostly homogenous mononucleosome preparation, with only a small amount of free DNA*

ORC does not bind to nucleosomes unless extranucleosomal DNA is present

Despite the presence of some free DNA, Orc1-5 was incubated alongside the nucleosome preparations in a range of ratios from 0.5:1 to 5:1 (ORC:nucleosome). Reactions were incubated for 15-30 minutes, then run on an EMSA agarose gel, as described previously (Figure 73).

In the presence of nucleosomes without any extranucleosomal fragments, there were no significant shifts in the nucleosome bands, suggesting that there were no stable ORC-nucleosome interactions formed (Figure 73, Row A). The faint bands seen are likely due to

residual DNA fragments from the nucleosomal preparations. Overall, this suggests that ORC, despite its intact BAH region, cannot bind directly to nucleosomes – it may be that additional adjacent DNA is required for stability.

With the addition of extranucleosomal DNA fragments, there is a slight increase in the shifts seen in the nucleosomal bands. With an extra 20 base pairs of extranucleosomal DNA, shifts are seen from 1:3 nucleosome:ORC ratios upwards, although with increasing aggregation. At the same time, there is not a significant difference between the nucleosomes containing different modifications, suggesting that any shift seen may not be affected by these mutations (Figure 73, 3rd and 4th column). Additional base pairs of extranucleosomal DNA (either 40 or 80 bp) cause an increased shift with lower nucleosome:ORC ratios, although once again, this is not significantly different between different modifications. It is also important to note that the band shifts seen are similar in size when comparing to a DNA-only EMSA (Figure 73, 1st column). This may suggest displacement of the nucleosomes for ORC binding, although it is difficult to conclude this from a gel-shift assay alone.

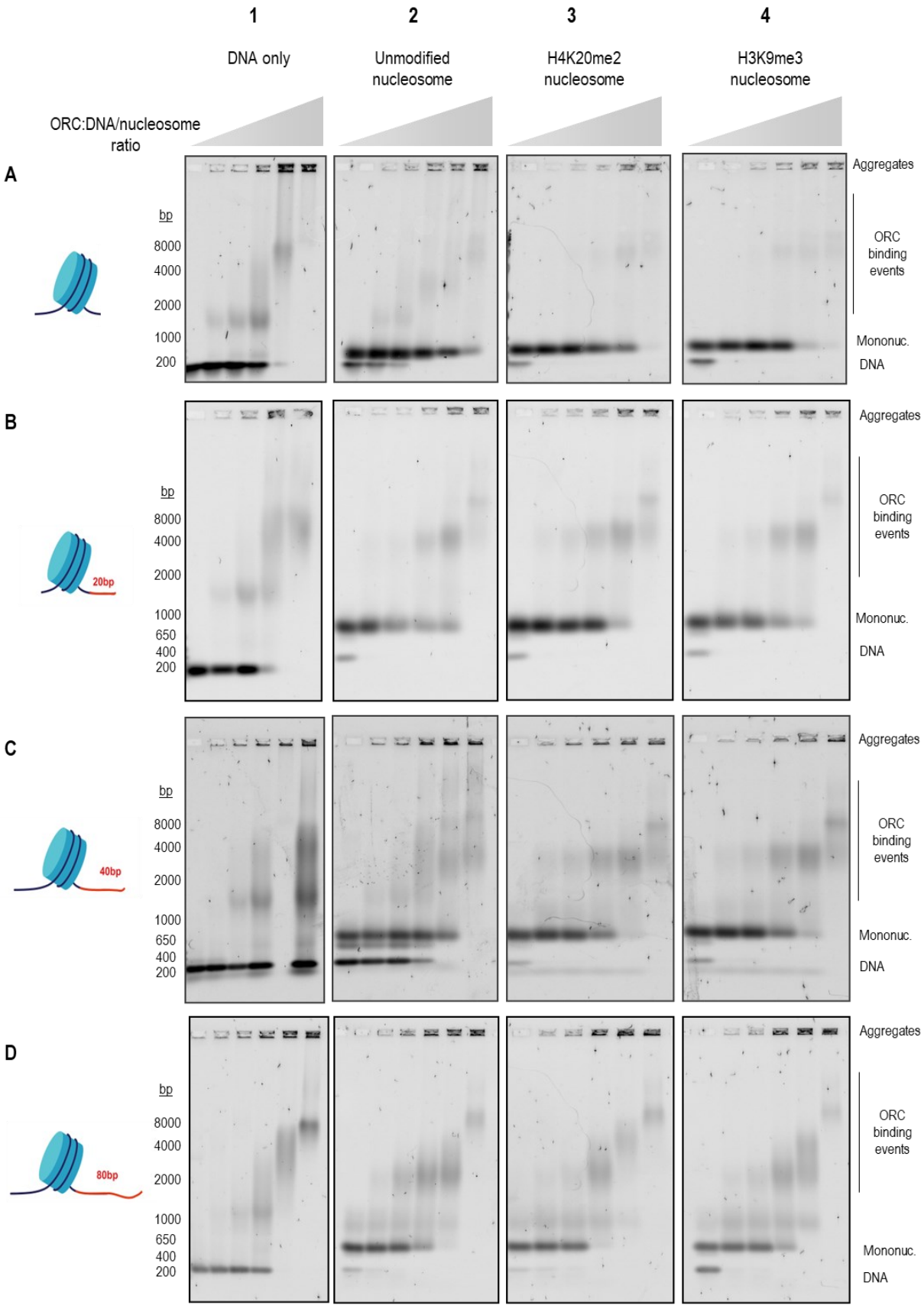

1
2
3
4
DNA only
Unmodified nucleosome
H4K20me2 nucleosome
H3K9me3 nucleosome
ORC:DNA/nucleosome ratio
A
B
C
D
bp
8000
4000
2000
1000
650
400
200
20bp
40bp
80bp
Aggregates
ORC binding events
Mononuc. DNA

Figure 73: ORC can only bind to mononucleosomes when flanking DNA is present

*SYBR-safe EMSA assays of ORC binding to different nucleosome constructs, containing different nucleosome modifications (**Columns 1-4, 1:** Widom-601 DNA only, **2**: Unmodified nucleosome, **3**: H4K20me2 nucleosome, **4**: H3K9me3 nucleosomes) and different extranucleosomal (beyond the Widom-601 DNA) DNA lengths (**Rows A-D,** where **A**: no extranucleomal DNA, **B**: Widom-601 + 20bp, **C**: Widom-601 + 40bp, **D**: Widom-601 + 80bp). When no extranucleosomal DNA is present, there is no band shift in the DNA weight, indicating that ORC cannot stably bind to a nucleosome alone. Upon introduction of 20- 40- or 80- bp of flanking sequence, ORC can bind, with multiple bands suggesting ORC binding events (Rows 2-4). There did not appear to be any significant difference when using different nucleosome modifications.*

It is important here to note that these experiments were carried out with preparations of ORC that had not been optimised for fully homogenous Orc1-5 (Chapter 3.1). Therefore, there is a chance that some subunits present in the protein stocks were not fully assembled into Orc1-5, despite having no other impurities. This could be why the shifts are not as pronounced, especially in the DNA-only EMSAs in comparison to experiments carried out in Chapter 3.4.2 and required higher amounts of ORC to see an equivalent shift.

Negative Stain Microscopy of ORC-nucleosomes

Since the EMSA experiments did not yield conclusive results on whether ORC molecule were binding to the nucleosomes and being stabilised by the extranucleosomal DNA - or if they were binding only to the extranucleosomal DNA without forming any nucleosome-specific contacts – negative stain electron microscopy was used to screen samples to see if we could see stabilisation of ORC-nucleosome molecules.

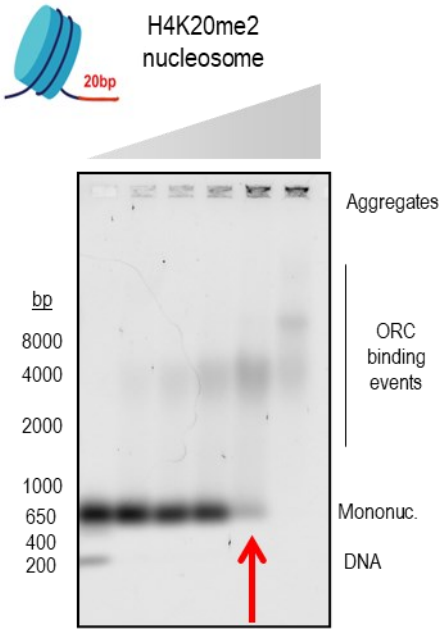


Figure 74: H4K20me2 mononucleosomes with 40bp flanking DNA can bind ORC.

Sample used for NS analysis to test whether ORC has a visible association with DNA when stabilised by a small fragment of extranucleosomal DNA. +20 base pairs were chosen as the construct with a 1:3 ratio of nucleosome:ORC, as this was shown to shift the weight of the DNA, assuming protein binding. H4K20me2 nucleosomes were chosen, in keeping with published crystallography literature of the BAH domain binding to this modification.

H4K20me2 modified nucleosomes with an extranucleosomal DNA length of 40 base pairs were chosen for negative stain grid preparation. This modification specifically was chosen, since previous crystallographic experiments has shown direct binding of the BAH domain with a histone peptide segment containing this modification (Kuo et al, 2019). A nucleosome:ORC ratio of 1:5 was chosen, since this showed a significant shift in the nucleosome band (Figure 74), while avoiding too much potential aggregation associated with higher ratios. Nucleosomes and ORC were incubated as described as above, and negative stain grids were prepared as described previously. A small dataset was collected, the details of which can be found in Supplementary Table 7.9.

A particle diameter size of 300 A was chosen, to accommodate for both the nucleosome and ORC diameters, in addition to any intermediate DNA length. Upon 2D classification of picked

particles (142,785), nucleosome structures were extremely clear, with a stable ring-like structure (Figure 75 B). However, it was difficult to say whether any ORC molecules truly were bound to the nucleosome structures or are just situated adjacent to the nucleosomes on the grid. It is also likely that the stabilisation effect that the extranucleosomal DNA offers enables the ORC molecule to bind and slide at different lengths along the DNA. This results in the variability seen in the 2D classes, where a presumed ORC molecule (although extremely flexible and at low resolution) is situated at different distances from the nucleosome (Figure 75 C).

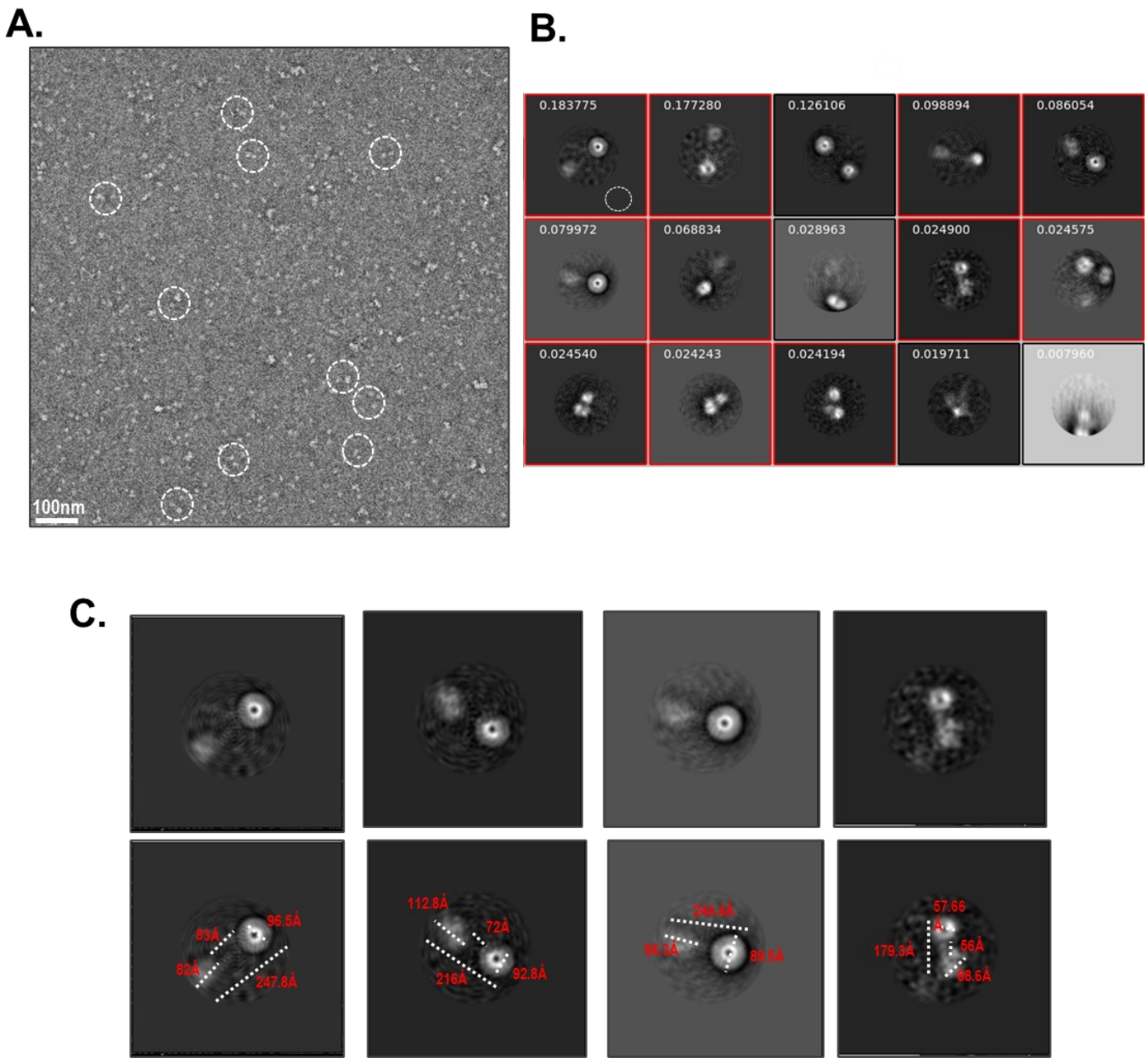


Figure 75: NS-EM of ORC-H4K20me2-40bp mononucleosomes

A: *A representative negative stain micrograph of the-ORC-nucleosome EM dataset, with manually picked coordinates* ***B:*** *represented 2D classes of* ***142,785*** *picked and extracted particles, showing nucleosomes (round ring-like particles, i.e. DNA wrapped around the core histone octamer) and flexible ORC smears* ***C:*** *Representative example of 2D classes and the variability in length between the ORC and mononucleosome core particle. The variability in distance between the ORC and the nucleosome indicates no rigid binding between the two, even in the case of a small exposed DNA.*

Since ORC's BAH domain is situated on a long and flexible linker, away from the core DNA-binding domain of the main complex, it is not possible to say whether ORC is interacting directly with the nucleosome or if it is merely binding to extranucleosomal DNA.

It could be that the BAH domain does allow for initial localisation of ORC towards nucleosomes, but this interaction is relatively weak, with stabilisation of ORC by flanking DNA being more influential in sustaining the complex in nucleosome-rich regions.

Electron microscopy is not yet at a level where it can elucidate extremely flexible and dynamic interactions – structural studies of the ORC-BAH domain has relied on heavily truncated segments of the structure that can interact stably to allow crystallisation. From these negative stain results, it does not appear that the rest of the ORC interacts in a stable manner with the nucleosome structure. If indeed the BAH-nucleosome interaction is occurring on a dynamic basis, it may be more suitable to utilise single-molecule experiments to track these transient interactions in real-time. This could involve fluorescent labelling of a region near the BAH domain and the predicted interaction site on the nucleosome, then seeing if FRET can occur between the two in the instance that they become closely associated. Measuring this on a timescale level will enable an insight into protein dynamics and could answer whether this interaction is transient, or happens over a longer period of time, but the flexible linker allows ORC to slide on nucleosome flanking DNA without disassociating completely from chromatin.

ORC-HP1

While the results discussed above show that full-length ORC is not necessarily stabilised in the presence of nucleosomes, in vivo this interaction does not happen in isolation. In the cell, there are several of chromatin-binding partners of ORC. One such factor is HP1, a chromatin-associated protein that exists in multiple isoforms and serves to condense chromatin into its silenced state. It has been shown that ORC and HP1 isoforms can co-immunoprecipitate from *D. melanogaster* cells, and disrupting one dysregulates the other and results in aberrations in chromatin formation (Prasanth et al, 2002, Prasanth et al, 2004).

While the precise molecular mechanism behind this remains unknown, it is possible that in cells, ORC-chromatin interactions are mediated in part by HP1. To validate the existence of an ORC-HP1 interaction, biochemical assays then negative stain EM was used. This will

hopefully serve as a starting point for future experiments to study ORC's interaction with chromatin-associated factors and correlate this not only with replication, but any non-replicative functions it may have throughout the cell cycle.

HP1 purification

HP1α, β and γ were each recombinantly expressed with a GST tag in E. coli, under IPTG induction, as described in SECTIONX. Following lysis, the GST-tagged protein was successfully separated from the rest of the cell lysate via glutathione-sepharose beads and subsequently eluted via PreScission protease cleavage of the GST tag. SDS-PAGE analysis showed a high concentration of expressed protein, with relatively few low molecular weight impurities bar some protein degradation (Figure 76 A). The fractions with high protein concentrations were pooled and further purified by gel filtration (Figure 76 B.), then concentrated with 10 kDa MWCO Amicon concentrator to remove low molecular weight degradations. All three isoforms were purified in the same manner (Figure 76 C.), with an average of 500 mg per 1L of culture.

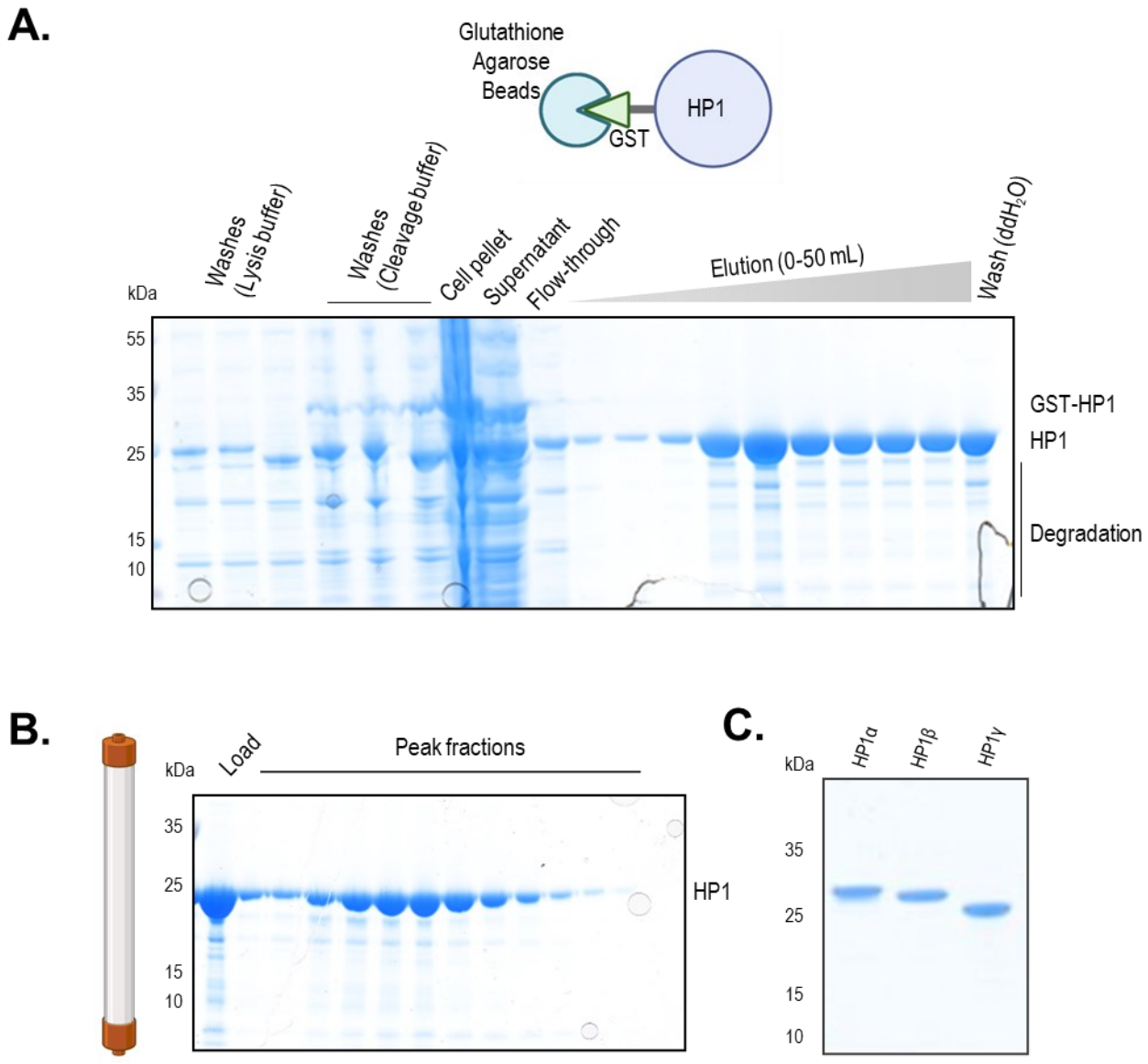


Figure 76: Purification of HP1

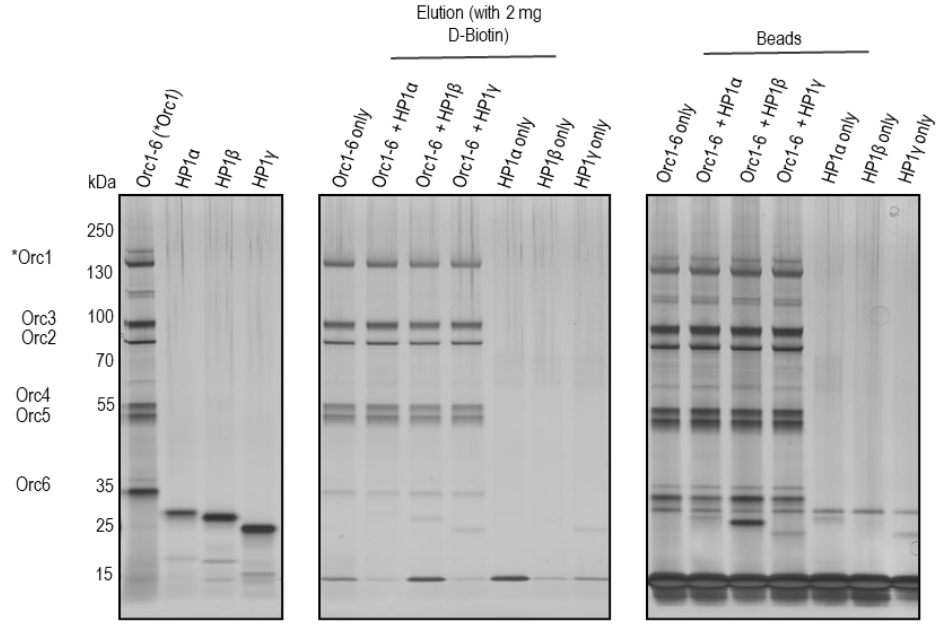
A: Coomassie-stained SDS-PAGE gel showing *GST-mediated affinity pulldown and elution of HP1 using glutathione agarose beads, following transformed E. coli BL21 cell lysis. The elution gave a relatively clean peak of HP1, with some degradation products.* ***B:*** *Gel filtration of pooled sample from A to remove degradation products. Cleaned fractions were pooled and concentrated for experimental use. For simplicity, only the purification of HP1β is shown although the process was the same for each isoform (**C**).*

Biochemical analysis of the HP1-ORC interaction.

Pulldown assays using purified proteins were conducted between Orc1-5 (with a 2xStrepSumo-tag on Orc1) and all three isoforms of HP1, to determine whether these proteins could associate with each other *in vitro*. Optimal buffer conditions were used from the pre-RC

assay (Chapter 3.3 and 2.2.3), with magnetic Strep beads used to separate Orc1 associated complexes from the reaction mixture. Only very weak interactions were observed for all subunits, indicated by faint HP1 bands (Figure 77 A). To ensure that the magnetic beads or protein tags were not either sterically hindering or facilitating any interaction, untagged ORC and HP1β were analysed by density gradient centrifugation as described in Chapter 2.2.5. SDS-PAGE gels of gradients indicate HP1β and ORC fractionate distinctly (Figure 77 B.i. and B.ii respectively), with a higher weight molecular assembly consisting both proteins (Figure 77 B.iii). This suggests the possibility of an ORC-HP1β complex, although co-fractionation of ORC and HP1β with different band intensities may indicate the presence of non-stochiometric complex populations, with ORC potentially interacting with multiple or oligomerized HP1β proteins.

A.


Elution (with 2 mg D-Biotin)
Beads
kDa
Orc1-6 (*Orc1)
HP1α
HP1β
HP1γ
Orc1-6 only
Orc1-6 + HP1α
Orc1-6 + HP1β
Orc1-6 + HP1γ
HP1α only
HP1β only
HP1γ only
*Orc1
Orc3
Orc2
Orc4
Orc5
Orc6
250
130
100
70
55
35
25
15

B.

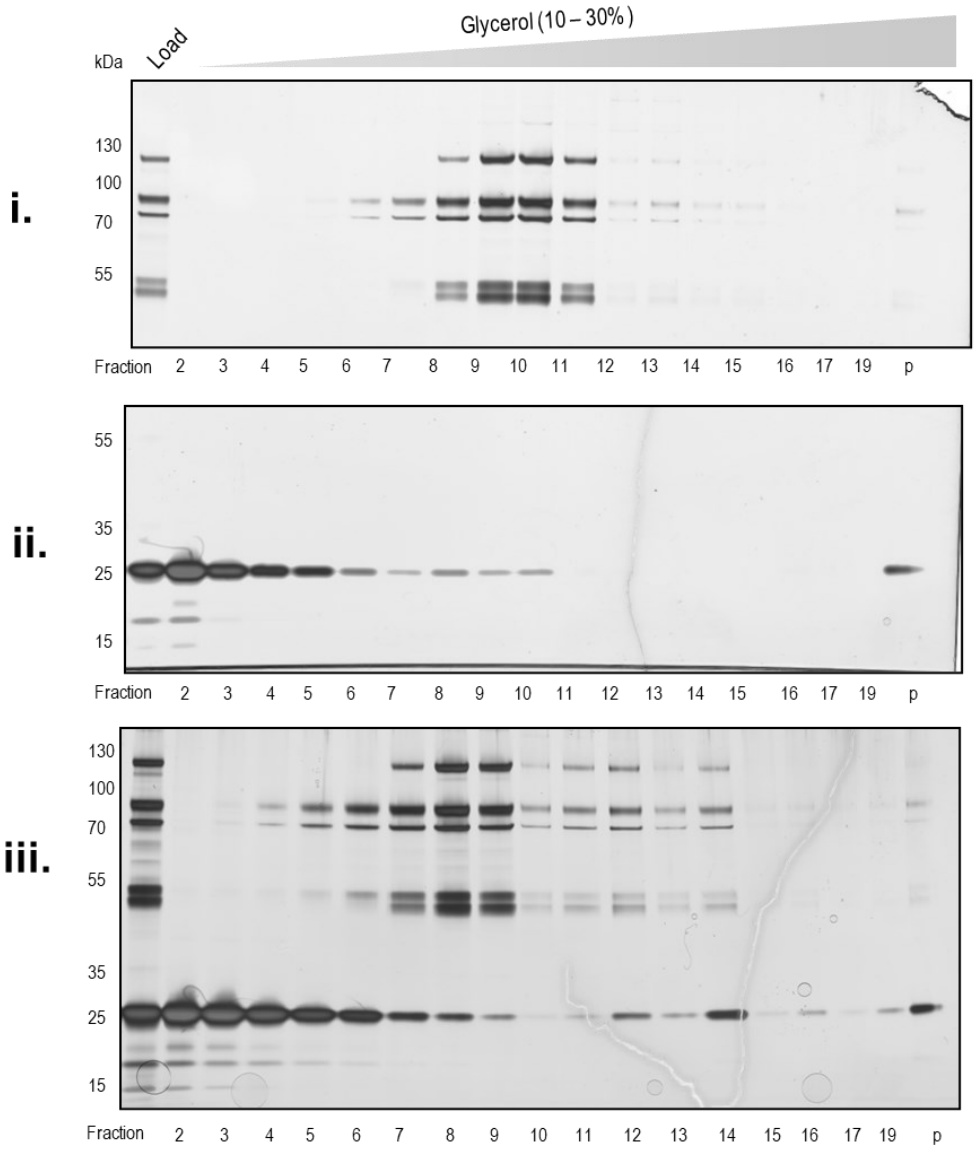

Load
Glycerol (10 – 30%)
kDa
i.
ii.
iii.
130
100
70
55
35
25
15
Fraction 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 p

Figure 77: The ORC-HP1 interaction is relatively weak.

A: *Silver stained gel showing 20% input and D-biotin elution of ORC-HP1 complexes from magnetic Strep beads. Minimal interaction observed in the case of all isoforms, with HP1β and γ being the strongest.* ***B****: Glycerol gradient of ORC1-5 (i) and HP1β (ii) independently and together (iii). Some ORC-HP1β co-fractionates in in F.12-14, suggesting a possible interaction between the two proteins.*

Negative Stain Microscopy of ORC-HP1

Samples from Fraction 14 of the density gradient experiments were cross-linked with 0.1% glutaraldehyde and negative stain grids were prepared as described previously. A small dataset was acquired, the details of which are available in Supplementary Table 7.10. (Figure 78)

Following 2D class generation, all micrographs showed populations of small particles - presumably HP1β - are present in addition to the ORC ring structure (Figure 78 B). However, due to the small size of HP1 (~25 kDa, although it dimerizes in solution), any extra density on the ORC is difficult to distinguish when comparing to ORC-only 2D classes (Figure 78 B.ii.). It may also be that ORC that is not bound to DNA is insufficient for a stable ORC-HP1β interaction. HP1 is typically associated with H3K9Me3 modified nucleosomes. Therefore, preparing H3K9Me3 modified nucleosomes or nucleosome arrays may stabilise the complex and provide an improved insight on direct ORC-HP1 association, or if this interaction is mediated by heterochromatin. Furthermore, this may reveal structural changes in ORC or chromatin that could either activate or inhibit origin licensing at H3K9Me3 modified chromatin.

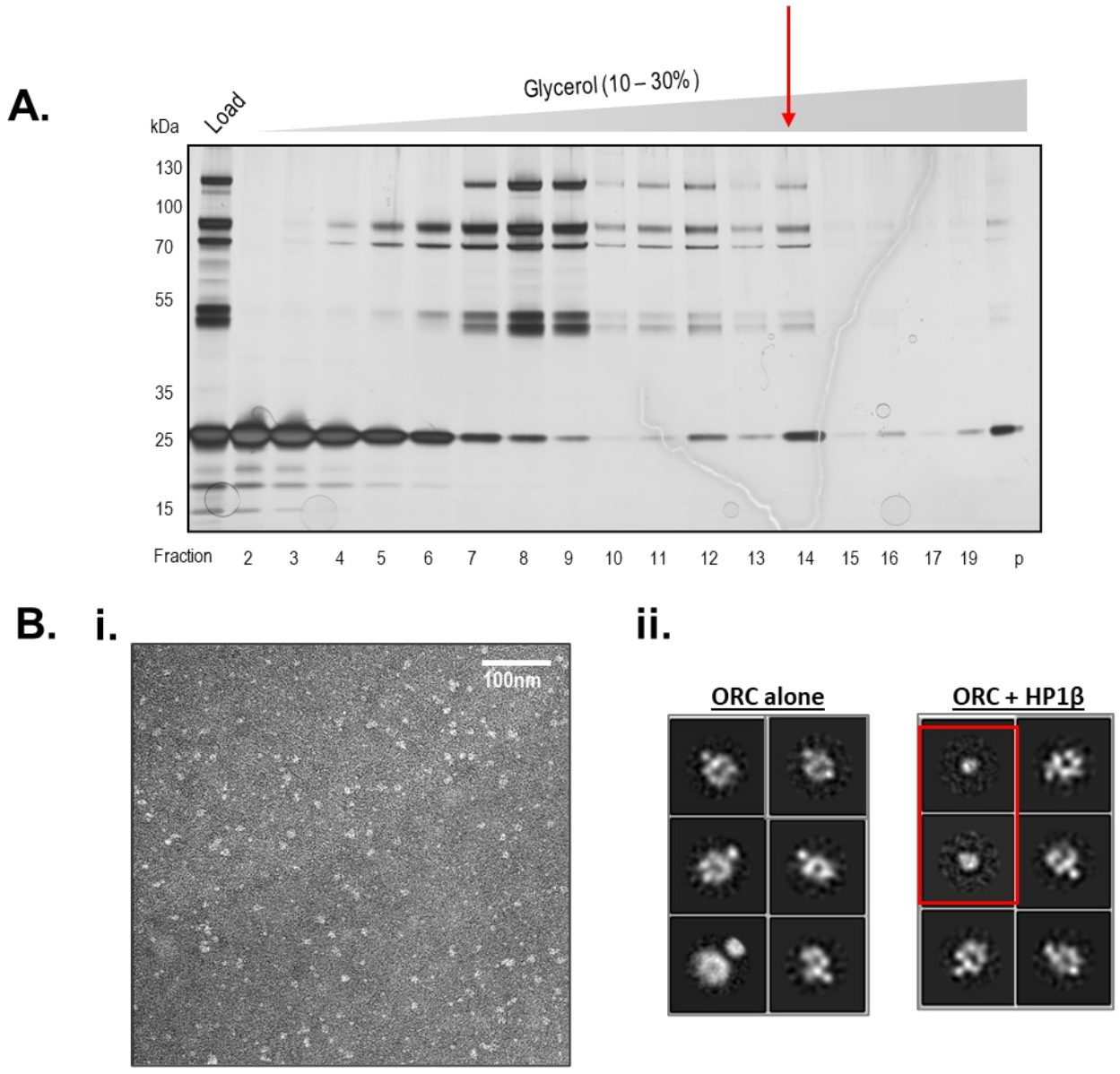


Figure 78: ORC-HP1β NS EM analysis

***A**: 10-30% glycerol gradient of ORC1-5-HP1β from Figure 77 Biii. F14 was used for NS grid preparation; samples were cross-linked and applied onto negative stain grids, as described in Section 2.2.10. **B(i).** Representative negative stain micrograph and corresponding 2D classes (**ii**) contained mostly ORC-like particles with small proteins not associated (labelled in red), presumed to be HP1.*

The ORC interactome

Exploring the proteins that ORC associates with on chromatin at stages of the cell cycle may help to elucidate the full complement of proteins participating in various functions. Currently, the ORC interactome is not well defined outside of pre-RC assembly, since it is difficult to

purify intact ORC-interaction factor complexes. ChIP-MS (Chromatin-immunoprecipitation – Mass Spectrometry) is a useful tool that allows isolation of chromatin-bound tagged proteins alongside their *in vivo* binding partners.

This could be done with cells arrested at each stage of the cell cycle to explore how ORC's role changes during different phases. Interesting factors, with a focus on chromatin binding proteins could be further investigated and will be useful as a basis for future studies to generate high resolution models of ORC and its interactors. In the long term, this could be especially interesting when focussing on factors that contain mutations associated with disease, since ORC mutations have been implicated in several cancers and Meier-Gorlin syndrome.

So far, HEK-293 cells have been engineered to express tagged ORC subunits that can be used as bait for immunoprecipitation reactions. These cells will downstream be used to generate ChIP-MS data and resolve cellular ORC-DNA and ORC-protein interactions.

Producing a tagged ORC cell line for ChIP-MS

Orc2 was chosen as the subunit to introduce a tag onto since it remains associated with the rest of the complex throughout the cell cycle and remains functional (Prasanth, Prasanth et al. 2004). This will allow exploration of the fluctuations in chromatin associated proteins that ORC associates with at different stages of the cell cycle, potentially allowing an insight into roles outside of origin recognition.

To begin this, the Orc2 subunit was cloned into a ligation-independent vector containing a Flag-2xStrep tag. Sequencing of the plasmid confirmed the presence of the Orc2 gene in the correct location. 1ug, 3ug, and 5ug of this plasmid was transfected into $1x10^7$ HEK-293T cells. Cells were selected two days post transfection with puromycin. Cell populations were tested via western blot to confirm the presence of Orc2 and the Flag tag (Figure 79 A). The different samples all showed the presence of the Flag tag and had two bands for the Orc2 protein, indicating the presence of both the endogenous and the tagged version. However, the 3- and

5ug-transfected populations show a higher overexpression of Orc2, which may result in assembly of nonspecific protein complexes. Monoclonal cells from the 1ug transfected population were isolated by limited dilution to ensure cells used for downstream ChIP-MS express identical levels of tagged Orc2. Clones were once again tested by western blot to find specific ones that display similar levels of tagged and endogenous Orc2 (Figure 79 B).

These Clones E, G, H, I and J (Fig.79 B) have been expanded and nuclear and cytoplasmic extracts have been prepared. Due to time constraints, this section of the project has not evolved beyond this, but the next stages involve small scale immunoprecipitations to ensure the tagged Orc2 is viable and integrates into the ORC complex. Once an appropriate clone has been identified, a larger scale cross-linked ChIP-MS experiment will allow identification of ORC-associated chromatin factors. While this will be initially done with asynchronous cells, in the future synchronised cell populations can be used to allow identification of different interactions that occur depending on cell cycle position.

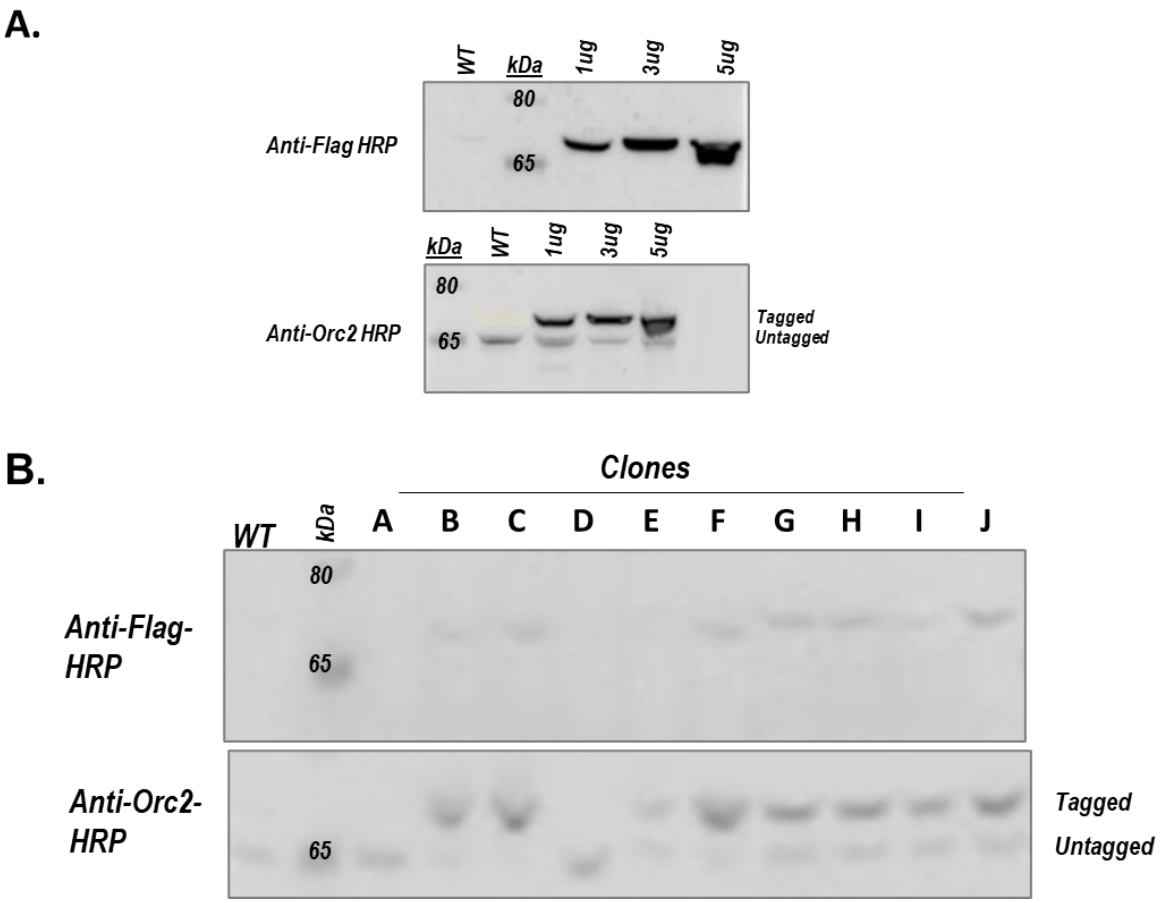


Figure 79: Preparation of a Tagged Orc2 cell line for ChIP-MS studies

***A**: Western blot of HEK-293 cells transfected with 1, 3, and 5ug of Flag-2xStrep-Orc2. All cells transfected with the tagged Orc2 vector displayed bands when probed with an anti-Flag and anti-Orc2 antibody, suggesting transformation and production of the tagged construct was successful **B:** Western blot of single clones transfected with 1 ug of vector. Clones expressing sufficient tagged Orc2 (E, H-J) were expanded for further use. F was not expanded because of the intensity of the tagged Orc2 band – high levels of overexpression may affect function of the cells.*

Discussion

Overall, these initial results show that an interaction between ORC and chromatin or chromatin binding partners is not as simple as a single interaction interface. While these results do not eliminate the possibility of a direct interaction between ORC and nucleosomes, it is highly likely that these interactions are context-dependent, highly dynamic and regulated by other cellular factors.

The ChIP-MS analysis will provide an insight into ORC associated factors in cells. These factors can be purified and added to the ORC-nucleosome complexes, to further explore their influence on ORC-nucleosome structure and function. Using cross-linking as part of the ChIP-MS method will stabilise protein-protein interactions, allowing the identification of specific chromatin-associated factors that ORC binds to within a cellular context.

In future, the *in vitro* biochemical assays described here could be used in conjunction with results from ChIP-MS studies to find factors that associate with both ORC and specific nucleosome modifications. Using these in combination may provide a more stabilised structure of ORC's association with nucleosomes, and whether the interaction surfaces vary with differing scaffold proteins and different nucleosome modifications.

4: General Discussion, Study Limitations and Future Perspectives

This project set out to establish the molecular mechanisms of recognition of origin DNA and downstream recruitment of human origin licensing proteins. This is a complex, multistep process that has historically been studied primarily in the single-celled eukaryotic S. *cerevisiae* system and more recently, in higher order eukaryotes such as X. *laevis* and D. *melanogaster*. These studies have undoubtedly been crucial in understanding the processes behind the recruitment of the DNA replication machinery – indeed, as such a crucial step for cellular proliferation, there are many core components and mechanics that are conserved across all domains of life.

However, uncovering the precise mechanisms underlying the H. *sapiens* pre-replicative process and how this differs to other organisms in its regulation has the potential to be beneficial for understanding human health and disease. Aberrations in the replicative process has downstream effects that can accumulate into genomic instability, eventually evolving into defects in growth and uncontrolled cellular proliferation. Pre-replication complex components are often used as oncogenic biomarkers, with the potential to be used as therapeutic targets themselves. To do this, it is paramount to understand their assembly and roles in forming functional DNA replicative complexes to find suitable steps in the process that can be targeted by therapeutic agents. As a testament to this, many research groups have pivoted in recent years from using model organisms into using human proteins to study pre-RC formation, although complexities in protein production and functional assay development have slowed progress.

Purification of human pre-RC proteins

The findings presented in this thesis have established a reproducible method of producing scalable yields of both human ORC and Cdc6, that have been successfully manipulated to produce mutations and variants (Materials & Methods, Section 2.2.1; Results, Section 3.1). Just prior to, and over the course of developing these methods, other groups published several structures of human Orc1-5 (Tocilj et al. 2017, Cheng et al. 2020, Jaremko et al, 2020) isolated and purified from recombinant methods. These published methods had all used the BigBac Baculovirus expression system in insect cells to express recombinant ORC, with significant N-terminal truncations in the Orc1, Orc2 and Orc5 to improve sample stability and reduce flexibility for structural studies. This expression system has the benefit of having similar post-translational protein processing to mammalian systems, which is why we initially tested Sf9 expression alongside yeast cell expression (Results Section 3.3.1). However, this system comes with the challenges of time-consuming vector production and more demanding culture conditions, making scaling up difficult. Using the S. *cerevisiae* expression system with extrachromosomal vectors expressing human ORC subunits under galactose-inducible promoters allowed for a fine balance between having an expression system that contained inherent eukaryotic post-translational modification processing, but also the ability to easily scale expression volumes up to >24 litres with relatively simple media and expression requirements. Optimisation of this purification method has allowed for a reproducible way of producing high quality, full-length protein on a large scale, and was repurposed to produce mutants of interest in the same way (Results Section 3.4.5).

The dynamic nature of the Origin Recognition Complex

One of the primary goals of this project was to establish a reproducible method of purifying full-length human Orc1-5 and Cdc6, to study their interactions on a structural scale. This had not previously been reported in the literature, where all structural studies involving these complexes have been heavily truncated for ease of expression. These truncated regions contain certain functional features, responsible for the regulation of the complex roles *in vivo.*

The success in producing full length protein in this study only highlighted these complexities – optimisations in the purification requirements hint at how variable and dynamic the state of ORC can be in a cell. As discussed in Figure 67, the phosphorylation state of ORC contributes to its stability, and ORC subunits are variable throughout the cell cycle (Kara et al. 2015, DePamphillis et al, 2005) . Combinations of post-translational modifications may be a finely tuned way of regulating activation and deactivation, as discussed in Section 3.3. While testing the effects of precise modification states of ORC on pre-RC assembly was out of the scope of this thesis, it is a potential further project of interest, especially when correlating this to modification states typical at different stages in the cell cycle. The release of the Orc1 subunit as discussed in Results Section 3.4 may also be affected by protein modification. The observed release across the gradients over time upon addition of DNA may be an intrinsic property of the protein, but this may change upon phosphorylation at different stages of the cell cycle. Producing phosphorylation mutants have the potential to study the plasticity that ORC has across the cell cycle.

Limitations of using cryo-EM to study ORC and Cdc6

While attempts to uncover further structural information of full-length ORC not seen in previous studies have so far not been successful, it must be noted that the electron microscopy datasets presented in this thesis still have the potential for further processing to elucidate increased resolution of previously unmapped regions of the protein. The processing presented in this thesis was carried out on randomised partitions of the datasets to speed up processing time. Including the entirety of the datasets during processing may result in more structural detail. 2D classes containing high resolution structural features were generated (Results Section 3.4.2) and it was only due to time constraints that these could not be developed further beyond ab-initio 3D modelling and very preliminary 3D refinement. Given the extra time, these classes could be expanded to include particles picked from the remainder of the datasets and filtered to remove junk, leaving only particles suitable for developing high-resolution information. The

initial 3D models discussed here would have benefitted from additional structural refinement, which would give statistically validated information on resolution limits.

Masking regions of the protein that have not been structurally resolved before for specific sub-complex structural processing may provide some initial density information. For example, initial classification of the ORC-Cdc6-DNA dataset resulted in additional density that could be ascribed to Cdc6, forming a closed ring around ORC-DNA (Results Section 3.4.3). Producing the complex stabilisation ATPase mutants (as discussed above) has the downstream potential to produce a homogenous cryo-EM dataset capable of providing high resolution structural information of the core of the ORC-Cdc6-DNA complex.

However, cryo-EM is likely not a suitable method to study the N-terminal domains of some ORC subunits.and Cdc6. Previous structural studies removed these regions for tractability (Jaremko et al. 2020, Tocilj et al. 2017), since they did not affect the overall architecture of the core complexes. These domains have an inherent flexibility, and it is unlikely that they are fixed in any single conformation at any stage in the cell cycle, thus challenging conventional structure-function paradigms. This flexibility is likely conducive to cycle-dependent modifications, and thus these regions cannot be rigidly fixed or buried within the core; doing so may be unfavourable for effective modification in the cellular environment and could have knock-on effects for complex disassembly or inactivation post-G1. This has the potential to result in over-replication and dysregulated pre-RC formation outside of G1.

As discussed above, the EM datasets containing full length protein may provide minor additional structural information but the relevance of this may be debatable. Since the heterogenous nature and conformational fluctuations may span various timescales, it's imperative to adopt a multiparametric approach for a full, functional exploration of their roles as complexes. Machine learning aids in modelling protein flexibility and can be used to explore the energetics, structural conformations and transitions in an *in silico* environment. This can provide valuable data and capabilities are consistently being improved upon but comes with the caveat of being limited in environmental testing conditions (i.e. does not contextualise

within the cellular environment) and simulations of large complexes can often be resource-heavy (reviewed in Maiti et al. 2024).

Other experimental techniques could include smFRET – this has been used previously to study yeast pre-RC dynamics around nucleosomes (Sánchez et al, 2023). Instead of using bulk biochemical EMSAs and electron microscopy to study human ORC interactions with nucleosomes as described in Results Section 3.5, energy transfer between a donor and acceptor fluorophore on the ORC and nucleosome respectively may provide sub-millisecond dynamic measurements that can infer insights into the kinetics of this interaction. The presence of a nucleosome-binding domain on Orc1 that remains conserved across species indicates that this is a functional region of high importance, but the interaction between the two exists beyond a static protein-protein interface.

Overall, ORC and Cdc6 exist as highly dynamic members of the pre-replication complex, and one technique alone is not enough to fully understand their roles. These change over the course of the cellular timeline and using a range of techniques can help to piece together its biological functions.

The establishment of an assay to test human pre-replication complex assembly.

As discussed in Section 1.9, a bulk biochemical assay to isolate the salt-stable protein-protein and protein-DNA intermediates involved in the assembly of the human pre-RC is yet to be published (as of Dec. 2023), although paralogous experiments using model organism pre-RC proteins have been used extensively and provided a wealth of data on the process in other eukaryotes (Evrin et al. 2009, Schmidt and Bleichert. 2020) The pre-RC assembly assay described here, facilitated by the work carried out in this thesis and by other referenced members of the DNA Replication Group has provided a reproducible platform where helicase loading onto DNA can be tested using human proteins. On editing of this thesis (June 2024), three scientific publications – including the method described here – have been made public,

all of which contain similar protocols to study pre-RC assembly on a biochemical scale (Wells et al. 2024, Weissmann et al. 2024, Yang et al. 2024) These independently published experimental methodologies and results underscore the reproducibility and significance of this work.

While the assay in this capacity was used to study the order of pre-RC component loading and its dependence on ATP hydrolysis, it has the potential to be used in future applications to study mutants with specific mutations relevant to human health. The Catalogue of Somatic Mutations in Cancer (COSMIC) database (cancer.sanger.ac.uk; Tate et al. 2019) lists over two thousand individual mutations reported in cancer cases found in ORC alone. While these mutations in patients are likely found alongside other dysregulated proteins with their cumulative effects resulting in cancerous phenotypes, the methodologies described in this thesis may be valuable in future experiments to produce disease-associated mutations to study how they affect double hexamer assembly and downstream replication firing. The combination of easily producing these mutations in a scalable S. *cerevisiae* expression system then testing in the pre-RC assay may provide an insight into how dysregulation at the protein sequence level leads to over- or under- replication.

Limitations of the pre-RC assay

The pre-RC assay described here has been extremely useful as an initial test for double hexamer loading, although it comes with the caveat that it cannot provide quantifiable or kinetic data. Results in this thesis – and in previous studies – have emphasised that DNA replication is a highly dynamic process that cannot be captured through identification of stable intermediates in single-particle electron microscopy alone. While further high-resolution structural analysis will be paramount to identifying core interaction surfaces and potential druggable binding pockets, it must be supplemented with experiments that take into consideration the dynamics of the assembly process. Work in the DNA Replication Group has already begun to reconstitute the pre-RC system using single-molecule methods to elucidate

binding kinetics and dwell times of double hexamer assembly on DNA. This has involved the production of fluorescently labelled and active pre-RC components that can be tested in a C-trap setup, where binding and movement of proteins to individual DNA strands can be visualised and measured on a nanosecond level by their fluorescent emission upon association with DNA.

However, the basis of this experiment could be modified into a high-throughput assay where DNA bound pre-RC proteins can be tethered to a DNA slide or multi-well plate and readouts based on fluorescence, luminescence, or absorption changes upon introduction of other proteins can be measured. Optimisation and automation of this process has the potential to be expanded into a high-throughput fragment-based drug discovery assay, where potential drug fragment candidate libraries can be tested against bound pre-RC complexes.

It is important to consider that the pre-RC assembly does not exist in a vacuum *in vivo.* The described ChIP-MS design described in Section 3.5. above will provide an insight into associated factors in cells. Using cross-linking as part of the ChIP-MS method will stabilise protein-protein interactions, allowing the identification of specific chromatin-associated factors within a cellular context.

5. Concluding Remarks

This project set out with three main aims – to purify fully-length human pre-RC proteins (specifically ORC and Cdc6), study the complexes that these form with DNA on an *in vitro* biochemical and structural scale, then contextualise this in the presence of nucleosome-bound DNA, to explore the roles of accessory domains on the proteins.

In typical scientific fashion, chasing these goals resulted in an exponential number of new questions and avenues for future projects. Successes in producing full length pre-RC proteins resulted in difficulties in structural determination – but has also opened the door for future experiments involving accessory domains that may be crucial for protein function *in vivo*.

Unexpected results, such as the Orc2-5 complex bound to DNA has the potential to evolve into a future project combining high resolution microscopy and dynamic studies to probe the role of Orc1 on a molecular scale. The pre-RC proteins do not exist in a vacuum in the cell and there are a multitude of interactions it plays a part in to keep the cell cycle functioning smoothly.

It's extremely evident from the results presented in this thesis – and indeed, by the rest of the field – that the assembly of the replicative helicase is a highly variable and complex process. The field has made an enormous amount of progress in understanding replication mechanics using model organisms and has only recently begun to break the ice on the same process in humans. These results and methodologies presented in this thesis, in combination with new and upcoming scientific publications, provide an exciting basis to explore human pre-RC formation on a foundational level, with the potential to be expanded into therapeutic or translational studies.

-

6: Supplementary Information

Supplementary Table 1: Plasmids

Name	DNA Rep Group Identifier	Gene	Plasmid Backbone	Notes
PLAS1	1026	*hOrc1 (2x Strep-Sumo on N-term Orc1) and hOrc5 Uracil selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS2	1099	Orc2 Tryptophan selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS3	1103	Orc3 and Orc4 Leucine selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS4	1176	*hOrc1 (2x Strep-Sumo on N-term Orc1) and hOrc5 Uracil selection marker Ampicillin resistance gene	pMYA	Integratable
PLAS5	1181	Orc2 Tryptophan selection marker Ampicillin resistance gene	pMYA	Integratable
PLAS6	1178	Orc3 and Orc4 Leucine selection marker Ampicillin resistance gene	pMYA	Integratable
PLAS7	1237	hOrc1 ΔN (2x Strep-Sumo on N-term of truncated Orc1) and hOrc5 Uracil selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS8	1366	*hOrc1 (E621Q) (2x Strep-Sumo on N-term Orc1) and hOrc5 Uracil selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS9	1365	Orc3 and Orc4 (R209A) Leucine selection marker Ampicillin resistance gene	pESC	Non-Integratable
PLAS10	536	Cdc6 (GST tagged on N-term) Ampicillin resistance marker	pGEX	
PLAS11	1339	Cdc6 E285Q mutant (GST tagged on N-term) Ampicillin resistance marker	pGEX	
PLAS12	1271	Widom601-DNA1 Ampicillin resistance marker	pUC57	See Supplemetary table 2
PLAS13	1272	DNA2 Ampicillin resistance marker	pUC57	
PLAS14	1269	DNA3	pUC57	

		Ampicillin resistance marker		
PLAS15	1270	DNA4 Ampicillin resistance marker	pUC57	
PLAS16	1268	Flag-2xStrep-Orc2 Ampicillin resistance marker	pCAGEN	For expressing tagged Orc2 for HEK293T transfection. Backbone from Tianyi Zhang (LMS-MRC)
PLAS17		GST-HP1α (GST on N-term)	pGEX	
PLAS18		GST-HP1β (GST on N-term)	pGEX	
PLAS19		GST-HP1γ (GSTon N-term)	pGEX	

Supplementary Table 2: DNA Sequences

	DNA sequence	Gene/Region of DNA/chromosome	AT/GC content
1	GCGAGGATATCTGGAAGAAATTCGAGCTGCTGCCCA CCCCGCCCCTGTCCCCTAGCCGCCGCTCCGGGCTCT GCTCGCCCTCCTACGTTGCGGTCACACCCTTCTCCC TTCGGGGAGACAACGACGGCGGTGGCGGGAGCTTC TCCACGGCCGACCAGCTGGAG	Chromosome 8: MYC gene (128,750,645-128,750,808)	46:54
2	GGGGTATCTCATTGTTTTAATTTGCATTTCCCTAATGA CATATGCTATGGAGCATTTTTATATGCTTTTTTGCTAG CTATATATCTATATATAGCTATATAAATATATATTTTTT ATTTCAATAGTTTTTGGGGTTCAGTTGGTTTTTGGTTA AATGGATGAAT	Chromosome 7: (105,540,017-105,540,180)	41:59 with multiple polyT tracts
3	CCTGACCCCAAAAGTAATTCTTTCCTCTTTAAAAAAAA AAAAATTATAAATACATTATTTTAGTCCAGGGGCAGT GGTTTACACCTGTAATCCCAGCACTTTGGGAGGCCA AGGCGGCTGGATCACTTGAGGTCAAGAGTTCAAGAC CAGCCTGGCCAACATGG	Chromosome 5: KDM3B intron (137,701,344-137,701,507)	38:62 with polyA tract
4	GCCTCCTTGAGGCGGGCCCGGGCGGGGCGGTTGTA TATCAGGGCCGCGCTGAGCTGCGCCAGCTGAGGTG TGAGCAGCTGCCGAAGTCAGTTCCTTGTGGAGCCGG AGCTGGGCGCGGATTCGCCGAGGCACCGAGGCACT CAGAGGAGGTGAGAGAGCGGCGG	Chr.6 – CDKN1A intron/exon overlap (36,646,427-36,646,590)	50:50 with multiple polyG tracts.
5	ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTC GTAGACAGCTCTAGCACCGCTTAAACGCACGTACGC GCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACT CCCTAGTCTCCAGGCACGTGTCAGATATATACATCCG AT- GCGAGGATATCTGGAAGAAATTCGAGCTGCTGCCCA CCCCGCCCCTGTCCCCTAGCCGCCGCTCCGGGCTCT GCTCGCCCTCCTACGTTGCGGTCACACCCTTCTCCC TTCGGGGAGACAACGACGGCGGTGGCGGGAGCTTC TCCACGGCCGACCAGCTGGAG	WIDOM601-Sequence fused with DNA1	For mononucleosome studies

Supplementary Table 3: Primers

Name	Sequence (5'-3')	Use
PRIM1	GCGAGGATATCTGGAAGAAATTCGAGCTGCTG	Amplification of DNA1 (fwd)
PRIM2	CTCCAGCTGGTCGGCCG	Amplification of DNA1 (rv)
PRIM3	GGGGTATCTCATTGTTTTAATTTG	Amplification of DNA2 (fwd)
PRIM4	CATTCATCCATTTAACCAAAAACC	Amplification of DNA2 (rv)
PRIM5	CCTGACCCCAAAAGTAATTCTTTCCTC	Amplification of DNA3 (fwd)
PRIM6	CCATGTTGGCCAGGCTGG	Amplification of DNA3 (rv)
PRIM7	GCCTCCTTGAGGCGGGC	Amplification of DNA4 (fwd)
PRIM8	GCCTCCTTGAGGCGGGC	Amplification of DNA4 (rv)
PRIM9	Biotin-GCGAGGATATCTGGAAGAAATTCGAGCTGCTG	Amplification of Biotin-DNA1 (fwd)
PRIM10	CATCGAGAATCCCGGTGCCG	Amplification of Widom601 (fwd)
PRIM11	TTTCTTCCAGATATCCTCGCATCGG	Amplification of Widom601-DNA1 20bp (rv)
PRIM12	GGGGTGGGCAGCAGCTC	Amplification of Widom601-DNA1 40bp (rv)
PRIM13	GAGGGCGAGCAGAGCCC	Amplification of Widom601-DNA1 80bp (rv)

Supplementary Table 4: Yeast Strains

Name	Genotype	Use	Notes
YC658	MATa, lys2::pGAL1 GAL4::LYS2, pep4::HIS3, bar1::hisG derived from W30	Expression of hORC1-5	Derived from Matheson et al, 2017.

Supplementary Table 5: Bacteria Strains

Name	Use	Notes
DH5α competent	Plasmid Amplification	Invitrogen
BL21 Codon+ competent	Protein Overexpression (Cdc6)	New England Biolabs

Supplementary Table 6: Mammalian Cells

Name	Use	Notes
HEK293-T	Transfection with Flag-2x-Strep-Orc2 for ChIP-MS	From Cell Stress Group (MRC-LMS)

Supplementary Tables 7: Electron Microscopy Parameters

7.1 ORC1-5 NS

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 75,000
Micrograph pixel size	2.1 A/pix
Exposure (e-/A^2)	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	198

7.2 ORC1-5 Cryo

Microscope Information and Data Acquisition Parameters (collected by Dr. Nora Cronin)	
Location	London Consortium for Electron Microscopy LonCEM, the Francis Crick Institute (May 2023)
Microscope	Titan Krios
Camera	Gatan K3
Spherical Abberation	2.7 mm
Voltage	300 kV
Magnification	81,000X
Micrograph pixel size	1.1 A
Exposure (e-/A^2)	50

Defocus values	-0.5 to -3.5
Total # of micrographs collected	21,168

7.3 ORC-DNA NS

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 75,000
Micrograph pixel size	2.1 A/pix
Exposure (e-/A^2)	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	297

7.4 ORC-DNA Cryo

Microscope Information and Data Acquisition Parameters (collected by Dr. Nora Cronin)	
Location	London Consortium for Electron Microscopy LonCEM, the Francis Crick Institute (June 2023)
Microscope	Titan Krios
Camera	Gatan K3
Spherical Abberation	2.7 mm
Voltage	300 kV
Micrograph pixel size	1.1 A
Exposure (e-/A^2)	50
Defocus values	-0.5 to -3.5
Total # of micrographs collected	35,170

7.5 OCD Cryo

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 92,000
Micrograph pixel size	1.58 A/pix
Exposure	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	515

7.6 Orc2-5 NS

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 92,000
Micrograph pixel size	1.58 A/pix
Exposure	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	385

7.7 OCCM – F8

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm

Voltage	200 kV
Magnification	X 73,000
Micrograph pixel size	2.1 A/pix
Exposure	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	354

7.8 OCCM – F9

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London/ MRC LMS
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 73,000
Micrograph pixel size	2.1 A/pix
Exposure	50
Defocus values	-3.5 to -1.9
Total # of micrographs collected	385

7.9 ORC-Nucleosome

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London
Microscope	TALOS F200i
Camera	Falcon 3EC
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 73,000
Micrograph pixel size	2.1 A
Exposure e-/Å $^{-2}$	100
Defocus values	-3.2 to -1.5
Total # of micrographs collected	846

7.10 ORC-HP1

Microscope Information and Data Acquisition Parameters	
Location	Imperial College London
Microscope	TALOS F200i
Camera	Falcon 3
Spherical Abberation	2.7 mm
Voltage	200 kV
Magnification	X 73,000
Micrograph pixel size	2.1 A
Exposure e-/Å $^{-2}$	100
Defocus values	-3.2 to -1.5
Total # of micrographs collected	421

6. References

AMASINO, A. L., GUPTA, S., FRIEDMAN, L. J., GELLES, J. & BELL, S. P. 2023. Regulation of replication origin licensing by ORC phosphorylation reveals a two-step mechanism for Mcm2-7 ring closing. Proceedings of the National Academy of Sciences, 120, e2221484120.

ANDREAS, M., THEODOROS P, V., EVMORFIA, B., GEORGIA, L., MARIA K, A., IRENE, T., MARIE-CHRISTINE, K., VASSILIKI, P., OLGA, T., PANAYIOTIS, P., GERASSIMOS A, P., PHOTIS, B., EFSTRATIOS, P. & PENELOPE, K. 2011. Immunohistochemical Expression and Prognostic Significance of CCND3, MCM2 and MCM7 in Hodgkin Lymphoma. Anticancer Research, 31, 3585.

ANG, X. L. & WADE HARPER, J. 2005. SCF-mediated protein degradation and cell cycle control. Oncogene, 24, 2860-2870.

ARIAS, E. E. & WALTER, J. C. 2006. PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication. Nature Cell Biology, 8, 84-90.

AUSTIN, R. J., ORR-WEAVER, T. L. & BELL, S. P. 1999. Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. Genes & Development, 13, 2639-2649.

BALASOV, M., HUIJBREGTS, R. P. & CHESNOKOV, I. 2009. Functional analysis of an Orc6 mutant in Drosophila. Proc Natl Acad Sci U S A, 106, 10672-7.

BARIS, Y., TAYLOR, M. R. G., ARIA, V. & YEELES, J. T. P. 2022. Fast and efficient DNA replication with purified human proteins. Nature, 606, 204-210.

BARRY ELIZABETH, R. & BELL STEPHEN, D. 2006. DNA Replication in the Archaea. Microbiology and Molecular Biology Reviews, 70, 876-887. BARTHOLDY, B., MUKHOPADHYAY, R., LAJUGIE, J., ALADJEM, M. I. & BOUHASSIRA, E. E. 2015. Allele-specific analysis of DNA replication origins in mammalian cells. Nature Communications, 6, 7051.

BEPLER, T., MORIN, A., RAPP, M., BRASCH, J., SHAPIRO, L., NOBLE, A. J. & BERGER, B. 2019. Positive-unlabeled convolutional neural networks for particle picking in cryoelectron micrographs. Nature Methods, 16, 1153-1160.

BELL, S. & LABIB, K. 2016. Chromosome Duplication in Saccharomyces cerevisiae.

BELL, S. P. & STILLMAN, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature, 357, 128-134.

BICKNELL, L. S., WALKER, S., KLINGSEISEN, A., STIFF, T., LEITCH, A., KERZENDORFER, C., MARTIN, C.-A., YEYATI, P., AL SANNA, N., BOBER, M., 231 JOHNSON, D., WISE, C., JACKSON, A. P., O'DRISCOLL, M. & JEGGO, P. A. 2011. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. Nature Genetics, 43, 350-355.

BLEICHERT, F., BOTCHAN, M. B. & BERGER, J. M. 2017. Mechanisms for initiating cellular DNA replication. Science, 355.

BLEICHERT, F., BOTCHAN, M. R. & BERGER, J. M. 2015. Crystal structure of the eukaryotic origin recognition complex. Nature, 519, 321-6.

BLEICHERT, F., LEITNER, A., AEBERSOLD, R., BOTCHAN, M. R. & BERGER, J. M. 2018. Conformational control and DNA-binding mechanism of the metazoan origin recognition complex. Proc Natl Acad Sci U S A, 115, E5906-E5915.

BLOW, J. J., GE, X. Q. & JACKSON, D. A. 2011. How dormant origins promote complete genome replication. Trends in Biochemical Sciences, 36, 405-414.

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, 5865-5873.

BOCHMAN, M. L. & SCHWACHA, A. 2008. The Mcm2-7 Complex Has In Vitro Helicase Activity. Molecular Cell, 31, 287-293.

BONGERS, E., VAN KAMPEN, A., VAN BOKHOVEN, H. & KNOERS, N. 2005. Human syndromes with congenital patellar anomalies and the underlying gene defects. Clinical Genetics, 68, 302-319.

BONGERS, E. M. H. F., OPITZ, J. M., FRYER, A., SARDA, P., HENNEKAM, R. C. M., HALL, B. D., SUPERNEAU, D. W., HARBISON, M., POSS, A., BOKHOVEN, H. V., HAMEL, B. C. J. & KNOERS, N. V. A. M. 2001. Meier-Gorlin syndrome: Report of eight additional cases and review. American Journal of Medical Genetics, 102, 115-124.

CALLAWAY, E. 2015. The revolution will not be crystallized: a new method sweeps through structural biology. Nature, 525, 172-174.

CALLEBAUT, I., COURVALIN, J.-C. & MORNON, J.-P. 1999. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. FEBS Letters, 446, 189-193.

CANNON, R. D., JENKINSON, H. F. & SHEPHERD, M. G. 1990. Isolation and nucleotide sequence of an autonomously replicating sequence (ARS) element functional in Candida albicans and Saccharomyces cerevisiae. Molecular and General Genetics MGG, 221, 210-218.

CAYROU, C., BALLESTER, B., PEIFFER, I., FENOUIL, R., COULOMBE, P., ANDRAU, J.-C., VAN HELDEN, J. & MÉCHALI, M. 2015. The chromatin environment shapes DNA 232 replication origin organization and defines origin classes. Genome Research, 25, 1873-1885.

CAYROU, C., COULOMBE, P., PUY, A., RIALLE, S., KAPLAN, N., SEGAL, E. & MÉCHALI, M. 2012. New insights into replication origin characteristics in metazoans. Cell Cycle, 11, 658-667.

CAYROU, C., COULOMBLE, C., VIGNERON, A., STANOJCIC, S., GARNIER, O., PFFEIFER, I., RIVALS, E., PUY, A., LAURENT-CHABALIER, S., DESPRAT, R. & MECHALI, M. 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Research, 21.

CHAKRABORTY, A., SHEN, Z. & PRASANTH, S. G. 2011. “ORCanization” on heterochromatin. Epigenetics, 6, 665-670.

CHAMPERIS TSANIRAS, S., KANELLAKIS, N., SYMEONIDOU, I. E., NIKOLOPOULOU, P., LYGEROU, Z. & TARAVIRAS, S. 2014. Licensing of DNA replication, cancer, pluripotency and differentiation: An interlinked world? Seminars in Cell & Developmental Biology, 30, 174-180.

CHANG, F., RIERA, A., EVRIN, C., SUN, J., LI, H., SPECK, C. & WEINREICH, M. 2015. Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. eLife, 4, e05795.

CHEN, S. & BELL, S. P. 2011. CDK prevents Mcm2–7 helicase loading by inhibiting Cdt1 interaction with Orc6. Genes & Development, 25, 363-372.

CHEN S, JIN Z, XIN L, LV L, ZHANG X, GONG Y, LIU J. 2021. Expression and Clinical Significance of Origin Recognition Complex Subunit 6 in Breast Cancer - A Comprehensive Bioinformatics Analysis. Int J Gen Med, 14: 9733-9745

CHEN, Y.-H., KEEGAN, S., KAHLI, M., TONZI, P., FENYÖ, D., HUANG, T. T. & SMITH, D. J. 2019. Transcription shapes DNA replication initiation and termination in human cells. Nature Structural & Molecular Biology, 26, 67-77.

CHENG, J., LI, N., WANG, X., HU, J., ZHAI, Y. & GAO, N. 2020. Structural insight into the assembly and conformational activation of human origin recognition complex. Cell Discovery, 6, 88.

CHENG, J., LI, N., HUO, Y., DANG, S., TYE, B.-K., GAO, N. & ZHAI, Y. 2022. Structural Insight into the MCM double hexamer activation by Dbf4-Cdc7 kinase. Nature Communications, 13, 1396.

CHENG, K.-Y., NOBLE, M. E. M., SKAMNAKI, V., BROWN, N. R., LOWE, E. D., KONTOGIANNIS, L., SHEN, K., COLE, P. A., SILIGARDI, G. & JOHNSON, L. N. 2006. The Role of the Phospho-CDK2/Cyclin A Recruitment Site in Substrate Recognition*. Journal of Biological Chemistry, 281, 23167-23179. 233

CHESNOKOV, I., REMUS, D. & BOTCHAN, M. 2001. Functional analysis of mutant and wildtype Drosophila origin recognition complex. Proceedings of the National Academy of Sciences, 98, 11997-12002.

CHOY, B., LALONDE, A., QUE, J., WU, T. & ZHOU, Z. 2016. MCM4 and MCM7, potential novel proliferation markers, significantly correlated with Ki-67, Bmi1, and cyclin E expression in esophageal adenocarcinoma, squamous cell carcinoma, and precancerous lesions. Human Pathology, 57, 126-135.

COHEN, A., MULAS, R., SERI, M., GAIERO, A., FICHERA, G., MARINI, M., BAFFICO, M. & CAMERA, G. 2002. Meier-Gorlin syndrome (ear–patella–short stature syndrome) in an Italian patient: Clinical evaluation and analysis of possible candidate genes. American Journal of Medical Genetics, 107, 48-51.

COMOGLIO, F., SCHLUMPF, T., SCHMID, V., ROHS, R., BEISEL, C. & PARO, R. 2015. High-Resolution Profiling of Drosophila Replication Start Sites Reveals a DNA Shape and Chromatin Signature of Metazoan Origins. Cell Reports, 11, 821-834.

COSTAS, C., DE LA PAZ SANCHEZ, M., STROUD, H., YU, Y., OLIVEROS, J. C., FENG, S., BENGURIA, A., LÓPEZ-VIDRIERO, I., ZHANG, X., SOLANO, R., JACOBSEN, S. E. & GUTIERREZ, C. 2011. Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. Nature Structural & Molecular Biology, 18, 395-400.

COSTER, G., FRIGOLA, J., BEURON, F., MORRIS, EDWARD P. & DIFFLEY, JOHN F. X. 2014. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. Molecular Cell, 55, 666-677.

COVERLEY, D., PELIZON, C., TREWICK, S. & LASKEY, R. A. 2000. Chromatin-bound Cdc6 persists in S and G2 phases in human cells, while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process. Journal of Cell Science, 113, 1929-1938.

COVERLEY, D., LAMAN, H. & LASKEY, R. A. 2002. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. Nature Cell Biology, 4, 523-528.

COULOMBE, P., NASSAR, J., PEIFFER, I., STANOJCIC, S., STERKERS, Y., DELAMARRE, A., BOCQUET, S., & MÉCHALI, M. 2019. The ORC ubiquitin ligase OBI1 promotes DNA replication origin firing. Nature communications vol. 10,1 2426.

CVETIC, C. & WALTER, J. C. 2005. Eukaryotic origins of DNA replication: could you please be more specific? Seminars in Cell & Developmental Biology, 16, 343-353.

D'IMPRIMA, E., FLORIS, D., JOPPE, M., SÁNCHEZ, R., GRININGER, M. & KÜHLBRANDT, W. 2019. Protein denaturation at the air-water interface and how to prevent it. eLife, 8, e42747.

DAVEY, M. J., INDIANI, C. & O'DONNELL, M. 2003. Reconstitution of the Mcm2-7p Heterohexamer, Subunit Arrangement, and ATP Site Architecture *. Journal of Biological Chemistry, 278, 4491-4499. 234

DE IOANNES, P., LEON, V. A., KUANG, Z., WANG, M., BOEKE, J. D., HOCHWAGEN, A. & ARMACHE, K.-J. 2019. Structure and function of the Orc1 BAH-nucleosome complex. Nature Communications, 10, 2894.

DE MUNNIK, S. A., BICKNELL, L. S., AFTIMOS, S., AL-AAMA, J. Y., VAN BEVER, Y., BOBER, M. B., CLAYTON-SMITH, J., EDREES, A. Y., FEINGOLD, M., FRYER, A., VAN HAGEN, J. M., HENNEKAM, R. C., JANSWEIJER, M. C. E., JOHNSON, D., KANT, S. G., OPITZ, J. M., RAMADEVI, A. R., REARDON, W., ROSS, A., SARDA, P., SCHRANDER-STUMPEL, C. T. R. M., SCHOOTS, J., TEMPLE, I. K., TERHAL, P. A., TOUTAIN, A., WISE, C. A., WRIGHT, M., SKIDMORE, D. L., SAMUELS, M. E., HOEFSLOOT, L. H., KNOERS, N. V. A. M., BRUNNER, H. G., JACKSON, A. P. & BONGERS, E. M. H. F. 2012. Meier–Gorlin syndrome genotype–phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. European Journal of Human Genetics, 20, 598-606.

DEEGAN, T. D. & DIFFLEY, J. F. X. 2016. MCM: one ring to rule them all. Current Opinion in Structural Biology, 37, 145-151.

DEVAULT, A., VALLEN, E. A., YUAN, T., GREEN, S., BENSIMON, A. & SCHWOB, E. 2002. Identification of Tah11/Sid2 as the Ortholog of the Replication Licensing Factor Cdt1 in Saccharomyces cerevisiae. Current Biology, 12, 689-694.

DHAR, M. K., SEHGAL, S. & KAUL, S. 2012. Structure, replication efficiency and fragility of yeast ARS elements. Research in Microbiology, 163, 243-253.

DIFFLEY, J. F. X. 2011. Quality control in the initiation of eukaryotic DNA replication. Philosophical Transactions of the Royal Society B: Biological Sciences, 366, 3545- 3553.

DIFFLEY, J. F. X., COCKER, J. H., DOWELL, S. J. & ROWLEY, A. 1994. Two steps in the assembly of complexes at yeast replication origins in vivo. Cell, 78, 303-316.

DOUGLAS, M. E., ALI, F. A., COSTA, A. & DIFFLEY, J. F. X. 2018. The mechanism of eukaryotic CMG helicase activation. Nature, 555, 265-268.

DRURY, L. S., PERKINS, G. & DIFFLEY, J. F. X. 1997. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. The EMBO Journal, 16, 5966-5976.

DUBOCHET, J. & MCDOWALL, A. W. 1981. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. Journal of Microscopy, 124, 3-4.

DUDERSTADT, K. E. & BERGER, J. M. 2008. AAA+ ATPases in the Initiation of DNA Replication. Critical Reviews in Biochemistry and Molecular Biology, 43, 163-187.
DUNCKER, B. P., CHESNOKOV, I. N. & MCCONKEY, B. J. 2009. The origin recognition complex protein family. Genome Biology, 10, 214. 235

EATON, M. L., PRINZ, J. A., MACALPINE, H. K., TRETYAKOV, G., KHARCHENKO, P. V. & MACALPINE, D. M. 2011. Chromatin signatures of the Drosophila replication program. Genome Research, 21, 164-174.

EKHOLM-REED, S., MÉNDEZ, J., TEDESCO, D., ZETTERBERG, A., STILLMAN, B. & REED, S. I. 2004. Deregulation of cyclin E in human cells interferes with prereplication complex assembly. Journal of Cell Biology, 165, 789-800.

EKUNDAYO, B. & BLEICHERT, F. 2019. Origins of DNA replication. PLOS Genetics, 5, e1008556.

ELADL, A., YAMAOKI, Y., KAMBA, K., HOSHINA, S., HORINOUCHI, H., KONDO, K., WAGA, S., NAGATA, T. & KATAHIRA, M. 2023. NMR characterization of the structure of the intrinsically disordered region of human origin recognition complex subunit 1, hORC1, and of its interaction with G-quadruplex DNAs. Biochemical and Biophysical Research Communications, 683, 149112.

ENEMARK, E. J. & JOSHUA-TOR, L. 2008. On helicases and other motor proteins. Current Opinion in Structural Biology, 18, 243-257.

ERZBERGER, J. P. & BERGER, J. M. 2006. EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS. Annual Review of Biophysics and Biomolecular Structure, 35, 93-114.

EVRIN, C., CLARKE, P., ZECH, J., LURZ, R., SUN, J., UHLE, S., LI, H., STILLMAN, B. & SPECK, C. 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, 106, 20240-20245.

EVRIN, C., FERNÁNDEZ-CID, A., ZECH, J., HERRERA, M. C., RIERA, A., CLARKE, P., BRILL, S., LURZ, R. & SPECK, C. 2013. In the absence of ATPase activity, pre-RC formation is blocked prior to MCM2–7 hexamer dimerization. Nucleic Acids Research, 41, 3162-3172.

FENG, X., NOGUCHI, Y., BARBON, M., STILLMAN, B., SPECK, C. & LI, H. 2021. The structure of ORC–Cdc6 on an origin DNA reveals the mechanism of ORC activation by the replication initiator Cdc6. Nature Communications, 12, 3883.

FERNÁNDEZ-CID, A., RIERA, A., TOGNETTI, S., HERRERA, M. C., SAMEL, S., EVRIN, C., WINKLER, C., GARDENAL, E., UHLE, S. & SPECK, C. 2013. An ORC/Cdc6/MCM2- 7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM DoubleHexamer Assembly. Molecular Cell, 50, 577-588.

FORSBURG S.L. 2004. Eukaryotic MCM Proteins: Beyond Replication Initiation. Microbiology and Molecular Biology Reviews, 68, 109-131. 236

FRIGOLA, J., HE, J., KINKELIN, K., PYE, V. E., RENAULT, L., DOUGLAS, M. E., REMUS, D., CHEREPANOV, P., COSTA, A. & DIFFLEY, J. F. X. 2017. Cdt1 stabilizes an open MCM ring for helicase loading. Nature Communications, 8, 15720.

FRIGOLA, J., REMUS, D., MEHANNA, A. & DIFFLEY, J. F. X. 2013. ATPase-dependent quality control of DNA replication origin licensing. Nature, 495, 339-343. FUJITA, M. 2006. Cdt1 revisited: complex and tight regulation during the cell cycle and consequences of deregulation in mammalian cells. Cell Division, 1, 22.

FURSTENTHAL, L., KAISER, B. K., SWANSON, C. & JACKSON, P. K. 2001. Cyclin E Uses Cdc6 as a Chromatin-Associated Receptor Required for DNA Replication. Journal of Cell Biology, 152, 1267-1278.

GANAPATHI, M., PALUMBO, M. J., ANSARI, S. A., HE, Q., TSUI, K., NISLOW, C. & MORSE, R. H. 2011. Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Research, 39, 2032-2044.

GAO, F. & ZHANG, C.-T. 2007. DoriC: a database of oriC regions in bacterial genomes. Bioinformatics, 23, 1866-1867.

GAVIN, E. J., SONG, B., WANG, Y., XI, Y., & JU, J. 2008. Reduction of Orc6 expression sensitizes human colon cancer cells to 5-fluorouracil and cisplatin. *PloS one*, *3*(12), e4054.

GE, X. Q., JACKSON DA FAU - BLOW, J. J. & BLOW, J. J. 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress.

GIORDANO-COLTART, J., YING, C. Y., GAUTIER, J. & HURWITZ, J. 2005. Studies of the properties of human origin recognition complex and its Walker A motif mutants. Proceedings of the National Academy of Sciences, 102, 69-74.

GIRI, S., AGGARWAL, V., PONTIS, J., SHEN, Z., CHAKRABORTY, A., KHAN, A., MIZZEN, C., PRASANTH, K. V., AIT-SI-ALI, S., HA, T. & PRASANTH, S. G. 2015. The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. Elife, 4.

GIRI, S. & PRASANTH, S. G. 2015. Association of ORCA/LRWD1 with repressive histone methyl transferases mediates heterochromatin organization. Nucleus, 6, 435-41.

GONZALEZ, M. A., TACHIBANA, K. E., LASKEY, R. A., & COLEMAN, N. 2005. Control of DNA replication and its potential clinical exploitation. Nature reviews. Cancer, 5(2), 135–141.

GORLIN RJ FAU - CERVENKA, J., CERVENKA J FAU - MOLLER, K., MOLLER K FAU - HORROBIN, M., HORROBIN M FAU - WITKOP, C. J., JR. & WITKOP, C. J., JR. Malformation syndromes. A selected miscellany. 237

GOWRISHANKAR, J. 2015. End of the Beginning: Elongation and Termination Features of Alternative Modes of Chromosomal Replication Initiation in Bacteria. PLOS Genetics, 11, e1004909.

GRANT, T., ROHOU, A. & GRIGORIEFF, N. 2018. cisTEM, user-friendly software for singleparticle image processing. eLife, 7, e35383.

GREIWE, J. F., MILLER, T. C. R., LOCKE, J., MARTINO, F., HOWELL, S., SCHREIBER, A., NANS, A., DIFFLEY, J. F. X. & COSTA, A. 2022. Structural mechanism for the selective phosphorylation of DNA-loaded MCM double hexamers by the Dbf4- dependent kinase. Nature Structural & Molecular Biology, 29, 10-20.

GUAN, B., WANG, X., YANG, J., ZHOU, C. & MENG, Y. 2015. Minichromosome maintenance complex component 7 has an important role in the invasion of papillary urothelial neoplasia. Oncol Lett, 10, 946-950.

GUERNSEY, D. L., MATSUOKA, M., JIANG, H., EVANS, S., MACGILLIVRAY, C., NIGHTINGALE, M., PERRY, S., FERGUSON, M., LEBLANC, M., PAQUETTE, J., PATRY, L., RIDEOUT, A. L., THOMAS, A., ORR, A., MCMASTER, C. R., MICHAUD, J. L., DEAL, C., LANGLOIS, S., SUPERNEAU, D. W., PARKASH, S., LUDMAN, M., SKIDMORE, D. L. & SAMUELS, M. E. 2011. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. Nature Genetics, 43, 360-364.

GUPTA, S., FRIEDMAN, L. J., GELLES, J. & BELL, S. P. 2021. A helicase-tethered ORC flip enables bidirectional helicase loading. eLife, 10, e74282. HAN, L. & LI, F. Origin recognition complex subunit 1 (ORC1) augments malignant behaviors of lung adenocarcinoma cells via targeting Wnt signaling.

HAN, L., & LI, F. 2022. Origin recognition complex subunit 1 (ORC1) augments malignant behaviors of lung adenocarcinoma cells via targeting Wnt signaling. Bioengineered, 13(5), 13520–13533.

HAYASHI, M. T., TAKAHASHI, T. S., NAKAGAWA, T., NAKAYAMA, J.-I. & MASUKATA, H. 2009. The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. Nature Cell Biology, 11, 357-362.

HIRAGAMI-HAMADA, K., SOEROES, S., NIKOLOV, M., WILKINS, B., KREUZ, S., CHEN, C., DE LA ROSA-VELAZQUEZ, I. A., ZENN, H. M., KOST, N., POHL, W., CHERNEV, A., SCHWARZER, D., JENUWEIN, T., LORINCZ, M., ZIMMERMANN, B., WALLA, P. J., NEUMANN, H., BAUBEC, T., URLAUB, H. & FISCHLE, W. 2016. Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin. Nat Commun, 7, 11310.

HIRATANI, I., RYBA, T., ITOH, M., YOKOCHI, T., SCHWAIGER, M., CHANG, C.-W., LYOU, Y., TOWNES, T. M., SCHÜBELER, D. & GILBERT, D. M. 2008. Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. PLOS Biology, 6, e245. 238

HOSSAIN, M., BHALLA, K. & STILLMAN, B. 2021. Multiple, short protein binding motifs in ORC1 and CDC6 control the initiation of DNA replication. Molecular Cell, 81, 1951- 1969.e6.

HU, Y., TAREEN, A., SHEU, Y. J., IRELAND, W. T., SPECK, C., LI, H., JOSHUA-TOR, L., KINNEY, J. B. & STILLMAN, B. 2020. Evolution of DNA replication origin specification and gene silencing mechanisms. Nature Communications, 11, 5175.

HU, Y., WANG, L., LI, Z., WAN, Z., SHAO, M., WU, S., & WANG, G. 2019. Potential Prognostic and Diagnostic Values of CDC6, CDC45, ORC6 and SNHG7 in Colorectal Cancer. OncoTargets and therapy, *12*, 11609–11621.

HUANG, D. W., FANTI, L., PAK, D. T. S., BOTCHAN, M. R., PIMPINELLI, S. & KELLUM, R. 1998. Distinct Cytoplasmic and Nuclear Fractions of Drosophila Heterochromatin Protein 1: Their Phosphorylation Levels and Associations with Origin Recognition Complex Proteins. Journal of Cell Biology, 142, 307-318.

HYRIEN, O. 2016. How MCM loading and spreading specify eukaryotic DNA replication initiation sites [version 1; peer review: 4 approved]. F1000Research, 5.

HYRIEN, O., MARHEINEKE, K. & GOLDAR, A. 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. BioEssays, 25, 116- 125.

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, 105, 8956-8961.

ILVES, I., PETOJEVIC, T., PESAVENTO, J. J. & BOTCHAN, M. R. 2010. Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. Molecular Cell, 37, 247-258.

JACOB, F., BRENNER, S. & CUZIN, F. O. 1963. On the Regulation of DNA Replication in Bacteria. Cold Spring Harbor Symposia on Quantitative Biology, 28, 329-348.

JAREMKO, M. J., ON, K. F., THOMAS, D. R., STILLMAN, B. & JOSHUA-TOR, L. 2020. The dynamic nature of the human origin recognition complex revealed through five cryoEM structures. eLife, 9, e58622.

JONES, M. L., BARIS, Y., TAYLOR, M. R. G. & YEELES, J. T. P. 2023. Structure of a human replisome shows the organisation and interactions of a DNA replication machine. The EMBO Journal, n/a, e115685.

KANG, S., WARNER, MEGAN D. & BELL, STEPHEN P. 2014. Multiple Functions for Mcm2–7 ATPase Motifs during Replication Initiation. Molecular Cell, 55, 655- 665.

KARA, N., HOSSAIN, M., PRASANTH, S. G., & STILLMAN, B. 2015. Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. The Journal of biological chemistry, 290(19), 12355–12369.

KATAYAMA, T., KASHO, K. & KAWAKAMI, H. 2017. The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. Frontiers in Microbiology, 8. 239

KELMAN, L. M. & KELMAN, Z. 2013. DNA Replication Fork, Eukaryotic. In: LENNARZ, W. J. & LANE, M. D. (eds.) Encyclopedia of Biological Chemistry (Second Edition). Waltham: Academic Press.

KLEMM, R. D., AUSTIN, R. J. & BELL, S. P. 1997. Coordinate Binding of ATP and Origin DNA Regulates the ATPase Activity of the Origin Recognition Complex. Cell, 88, 493-502.

KLINGSEISEN, A. & JACKSON, A. P. 2011. Mechanisms and pathways of growth failure in primordial dwarfism. Genes & Development, 25, 2011-2024.

KOUSHOLT, A. N., MENZEL, T. & SØRENSEN, C. S. 2012. Pathways for Genome Integrity in G2 Phase of the Cell Cycle. Biomolecules, 2, 579-607.

KRAFT, C., HERZOG, F., GIEFFERS, C., MECHTLER, K., HAGTING, A., PINES, J. & PETERS, J.-M. 2003. Mitotic regulation of the human anaphase-promoting complex by phosphorylation. The EMBO Journal, 22, 6598-6609.

KUENG, S., OPPIKOFER, M. & GASSER, S. M. 2013. SIR Proteins and the Assembly of Silent Chromatin in Budding Yeast. Annual Review of Genetics, 47, 275-306.

KUMAR, A. & KONO, H. 2020. Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. Biophysical Reviews.

KUO, A. J., SONG, J., CHEUNG, P., ISHIBE-MURAKAMI, S., YAMAZOE, S., CHEN, J. K., PATEL, D. J. & GOZANI, O. 2012. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature, 484, 115-9.

LABIB, K. & DIFFLEY, J. F. X. 2001. Is the MCM2–7 complex the eukaryotic DNA replication fork helicase? Current Opinion in Genetics & Development, 11, 64-70.

LAU, K. M., CHAN, Q. K. Y., PANG, J. C. S., LI, K. K. W., YEUNG, W. W., CHUNG, N. Y. F., LUI, P. C., TAM, Y. S., LI, H. M., ZHOU, L., WANG, Y., MAO, Y. & NG, H. K. 2010.

Minichromosome maintenance proteins 2, 3 and 7 in medulloblastoma: overexpression and involvement in regulation of cell migration and invasion. Oncogene, 29, 5475-5489.

LEI, M., KAWASAKI, Y. & TYE, B. K. 1996. Physical Interactions among Mcm Proteins and Effects of Mcm Dosage on DNA Replication in Saccharomyces cerevisiae. Molecular and Cellular Biology, 16, 5081-5090.

LEONARD, A. C. & MECHALI, M. 2013. DNA Replication Origins. Cold Spring Harbour Perspecives in Biology, 5.

LEWIS, J. S. & COSTA, A. 2020. Caught in the act: structural dynamics of replication origin activation and fork progression. Biochemical Society Transactions, 48, 1057-1066. LI, A. & BLOW, J. J. 2004. Non-proteolytic inactivation of geminin requires CDK-dependent ubiquitination. Nature Cell Biology, 6, 260-267. 240

LI, C.-J., VASSILEV, A. & DEPAMPHILIS, M. L. 2004. Role for Cdk1 (Cdc2)/Cyclin A in Preventing the Mammalian Origin Recognition Complex's Largest Subunit (Orc1) from Binding to Chromatin during Mitosis. Molecular and Cellular Biology, 24, 5875-5886.

LI, H. & STILLMAN, B. 2012. The Origin Recognition Complex: A Biochemical and Structural View. In: MACNEILL, S. (ed.) The Eukaryotic Replisome: a Guide to Protein Structure and Function. Dordrecht: Springer Netherlands.

LI, N., LAM, W. H., ZHAI, Y., CHENG, J., CHENG, E., ZHAO, Y., GAO, N. & TYE, B.-K. 2018. Structure of the origin recognition complex bound to DNA replication origin. Nature, 559, 217-222.

LI, N., ZHAI, Y., ZHANG, Y., LI, W., YANG, M., LEI, J., TYE, B.-K. & GAO, N. 2015. Structure of the eukaryotic MCM complex at 3.8 Å. Nature, 524, 186-191.

LI, P.-C., CHRETIEN, L., COTE, L., KELLY, T. J. & FORSBURG, S. L. 2011. S. pombe replication protein Cdc18 (Cdc6) interacts with Swi6 (HP1) heterochromatin protein. Cell Cycle, 10, 323-336.

LI, Q. & CHANG, Z. 2012. Linking DNA replication to heterochromatin silencing and epigenetic inheritance Acta Biochimica et Biophysica Sinica, 44, 3–13.

LIKU, M. E., NGUYEN, V. Q., ROSALES, A. W., IRIE, K. & LI, J. J. 2005. CDK Phosphorylation of a Novel NLS-NES Module Distributed between Two Subunits of the Mcm2-7 Complex Prevents Chromosomal Rereplication. Molecular Biology of the Cell, 16, 5026-5039.

LIU, C., WU, R., ZHOU, B., WANG, J., WEI, Z., TYE, B. K., LIANG, C. & ZHU, G. 2012. Structural insights into the Cdt1-mediated MCM2–7 chromatin loading. Nucleic Acids Research, 40, 3208-3217.

LIU, G., XING, Y., ZHAO, H., CAI, L. & WANG, J. 2018. The implication of DNA bending energy for nucleosome positioning and sliding. Scientific Reports, 8, 8853.

LIU, J., SMITH, C. L., DERYCKERE, D., DEANGELIS, K., MARTIN, G. S. & BERGER, J. M. 2000. Structure and Function of Cdc6/Cdc18: Implications for Origin Recognition and Checkpoint Control. Molecular Cell, 6, 637-648.

LIU, S., BALASOV, M., WANG, H., WU, L., CHESNOKOV, I. N. & LIU, Y. 2011. Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. Proceedings of the National Academy of Sciences, 108, 7373-7378.

LIU, Y., CHEN, P., LI, M., FEI, H., HUANG, J., ZHAO, T., & LI, T. 2021. Comprehensive Analysis of the Control of Cancer Stem Cell Characteristics in Endometrial Cancer by Network Analysis. Computational and mathematical methods in medicine, 6653295.

LUDTKE, S. J., BALDWIN, P. R. & CHIU, W. 1999. EMAN: Semiautomated Software for HighResolution Single-Particle Reconstructions. Journal of Structural Biology, 128, 82-97. 241 LYUMKIS, D. 2019. Challenges and opportunities in cryo-EM single-particle analysis. Journal of Biological Chemistry, 294, 5181-5197.

MACHIDA, S., TAKIZAWA, Y., ISHIMARU, M., SUGITA, Y., SEKINE, S., NAKAYAMA, J. I., WOLF, M. & KURUMIZAKA, H. 2018. Structural Basis of Heterochromatin Formation by Human HP1. Mol Cell, 69, 385-397 e8.

MADINE, M. A., SWIETLIK, M., PELIZON, C., ROMANOWSKI, P., MILLS, A. D. & LASKEY, R. A. 2000. The Roles of the MCM, ORC, and Cdc6 Proteins in Determining the Replication Competence of Chromatin in Quiescent Cells. Journal of Structural Biology, 129, 198-210.

MAIORANO, D., KRASINSKA, L., LUTZMANN, M. & MECHALI, M. 2005. Recombinant Cdt1 Induces Rereplication of G2 Nuclei in *Xenopus* Egg Extracts. Current Biology, 15, 146-153.

MAIORANO, D., MOREAU, J. & MÉCHALI, M. 2000. XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature, 404, 622-625.

MAITI, S., SINGH, A., MAJI, T., SAIBO, N. V., & DE, S. 2024. Experimental methods to study the structure and dynamics of intrinsically disordered regions in proteins. Current research in structural biology, 7, 100138.

MAO, R., WANG, Z., ZHANG, Y., CHEN, Y., LIU, Q., ZHANG, T., & LIU, Y. 2020. Development and validation of a novel prognostic signature in gastric adenocarcinoma. Aging, 12(21), 22233–22252.

MARAHRENS, Y. & STILLMAN, B. 1992. A Yeast Chromosomal Origin of DNA Replication Defined by Multiple Functional Elements. Science, 255, 817-823.

MCGARRY, T. J. & KIRSCHNER, M. W. 1998. Geminin, an Inhibitor of DNA Replication, Is Degraded during Mitosis. Cell, 93, 1043-1053. MECHALI, M. 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. Nature Reviews Molecular Cell Biology, 11.

MECHALI, M., YPSHIDA, K., COULOMBE, P. & PASERO, P. 2013. Genetic and epigenetic determinants of DNA replication origins, position and activation. Current Opinion in Genetics and Development, 23.

MÉNDEZ, J. & STILLMAN, B. 2000. Chromatin Association of Human Origin Recognition Complex, Cdc6, and Minichromosome Maintenance Proteins during the Cell Cycle: Assembly of Prereplication Complexes in Late Mitosis. Molecular and Cellular Biology, 20, 8602-8612.

MÉNDEZ, J., ZOU-YANG, X. H., KIM, S.-Y., HIDAKA, M., TANSEY, W. P. & STILLMAN, B. 2002. Human Origin Recognition Complex Large Subunit Is Degraded by Ubiquitin Mediated Proteolysis after Initiation of DNA Replication. Molecular Cell, 9, 481-491.

MESNER, L. D., VALSAKUMAR, V., KARNANI, N., DUTTA, A., HAMLIN, J. L. & BEKIRANOV, S. 2011. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res, 21, 377-89. 242

MILLER, T. C. R., LOCKE, J., GREIWE, J. F., DIFFLEY, J. F. X. & COSTA, A. 2019. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. Nature, 575, 704-710.

MIOTTO, B., JI, Z. & STRUHL, K. 2016. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc Natl Acad Sci U S A, 113, E4810-9.

MIZUSHIMA, T., TAKAHASHI, N. & STILLMAN, B. 2000. Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. Genes & Development, 14, 1631-1641.

NAKANE, T., KOTECHA, A., SENTE, A., MCMULLAN, G., MASIULIS, S., BROWN, P. M. G. E., GRIGORAS, I. T., MALINAUSKAITE, L., MALINAUSKAS, T., MIEHLING, J., UCHAŃSKI, T., YU, L., KARIA, D., PECHNIKOVA, E. V., DE JONG, E., KEIZER, J., BISCHOFF, M., MCCORMACK, J., TIEMEIJER, P., HARDWICK, S. W., CHIRGADZE, D. Y., MURSHUDOV, G., ARICESCU, A. R. & SCHERES, S. H. W. 2020. Single particle cryo-EM at atomic resolution. Nature, 587, 152-156.

NEELSEN KJ, ZANINI IM, MIJIC S, HERRADOR R, ZELLWEGER R, RAY CHAUDHURI A, CREAVIN KD, BLOW JJ, LOPES M. 2013. Deregulated origin licensing leads to chromosomal breaks by rereplication of a gapped DNA template. Genes Dev, 27(23):2537-42.

NELSON, H. C. M., FINCH, J. T., LUISI, B. F. & KLUG, A. 1987. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. Nature, 330, 221-226. NGUYEN, V. Q., CO, C. & LI, J. J. 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature, 411, 1068-1073.

NISHITANI, H., LYGEROU, Z., NISHIMOTO, T. & NURSE, P. 2000. The Cdt1 protein is required to license DNA for replication in fission yeast. Nature, 404, 625-628.

NISHITANI, H., TARAVIRAS, S., LYGEROU, Z. & NISHIMOTO, T. 2001. The Human Licensing Factor for DNA Replication Cdt1 Accumulates in G_1 and Is Destabilized after Initiation of S-phase *. Journal of Biological Chemistry, 276, 44905- 44911.

NOGUCHI, K., VASSILEV A FAU - GHOSH, S., GHOSH S FAU - YATES, J. L., YATES JL FAU - DEPAMPHILIS, M. L. & DEPAMPHILIS, M. L. The BAH domain facilitates the ability of human Orc1 protein to activate replication origins in vivo.

OCAÑA-PALLARÈS, E., VERGARA, Z., DESVOYES, B., TEJADA-JIMENEZ, M., ROMEROJURADO, A., GALVÁN, A., FERNÁNDEZ, E., RUIZ-TRILLO, I. & GUTIERREZ, C. 2020. Origin Recognition Complex (ORC) Evolution Is Influenced by Global Gene 243 Duplication/Loss Patterns in Eukaryotic Genomes. Genome Biology and Evolution, 12, 3878-3889.

OHI, M., LI, Y., CHENG, Y., & WALZ, T. 2004. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biological procedures online, *6*, 23–34.

PAK, D. T. S., PFLUMM, M., CHESNOKOV, I., HUANG, D. W., KELLUM, R., MARR, J., ROMANOWSKI, P. & BOTCHAN, M. R. 1997. Association of the Origin of Replication Complex with Heterochromatin and HP1 in Higher Eukaryotes. Cell, 91, 311-323.

PAQUIN, K. L. & HOWLETT, N. G. 2018. Understanding the Histone DNA Repair Code: H4K20me2 Makes Its Mark. Mol Cancer Res, 16, 1335-1345.

PARKER, M. W., BOTCHAN, M. R. & BERGER, J. M. 2017. Mechanisms and regulation of DNA replication initiation in eukaryotes. Critical Reviews in Biochemistry and Molecular Biology, 52, 107-144.

PERERA, H. M. & TRAKSELIS, M. A. 2019. Amidst multiple binding orientations on fork DNA, Saccharolobus MCM helicase proceeds N-first for unwinding. eLife, 8, e46096.

PERKINS, G., DRURY, L. S. & DIFFLEY, J. F. X. 2001. Separate SCFCDC4 recognition elements target Cdc6 for proteolysis in S phase and mitosis. The EMBO Journal, 20, 4836-4845. PETERS, J.-M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. Nature Reviews Molecular Cell Biology, 7, 644-656.

PETERSEN, B. O., WAGENER, C., MARINONI, F., KRAMER, E. R., MELIXETIAN, M., DENCHI, E. L., GIEFFERS, C., MATTEUCCI, C., PETERS, J.-M. & HELIN, K. 2000. Cell cycle– and cell growth–regulated proteolysis of mammalian CDC6 is dependent on APC–CDH1. Genes & Development, 14, 2330-2343.

PIATTI, S., LENGAUER, C. & NASMYTH, K. 1995. Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast Saccharomyces cerevisiae. The EMBO Journal, 14, 3788-3799.

POON, R. Y. C. 2016. Cell Cycle Control: A System of Interlinking Oscillators. In: COUTTS, A. S. & WESTON, L. (eds.) Cell Cycle Oscillators: Methods and Protocols. New York, NY: Springer New York.

POZO, P. N. & COOK, J. G. 2017. Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. Genes [Online], 8.

POZO, P. N., MATSON, J. P., COLE, Y., KEDZIORA, K. M., GRANT, G. D., TEMPLE, B. & COOK, J. G. 2018. Cdt1 variants reveal unanticipated aspects of interactions with cyclin/CDK and MCM important for normal genome replication. Molecular Biology of the Cell, 29, 2989-3002. 244

PRASANTH, S. G., PRASANTH, K. V., SIDDIQUI, K., SPECTOR, D. L. & STILLMAN, B. 2004. Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. EMBO Journal, 23.

PRASANTH, S. G., PRASANTH, K. V. & STILLMAN, B. 2002. Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. Science, 297, 1026-31.

PRASANTH, S. G., SHEN, Z., PRASANTH, K. V. & STILLMAN, B. 2010. Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proceedings of the National Academy of Sciences, 107, 15093-15098.

PRIOLEAU, M.-N. & MACALPINE, D. M. 2016. DNA replication origins—where do we begin? Genes and Development, 30, 1683-1687.

PROROK, P., ARTUFEL, M., AZE, A., COULOMBE, P., PEIFFER, I., LACROIX, L., GUÉDIN, A., MERGNY, J.-L., DAMASCHKE, J., SCHEPERS, A., CAYROU, C., TEULADEFICHOU, M.-P., BALLESTER, B. & MÉCHALI, M. 2019. Involvement of G-quadruplex regions in mammalian replication origin activity. Nature Communications, 10, 3274.

PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. & BRUBAKER, M. A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods, 14, 290-296.

QUENTIN, D. & RAUNSER, S. 2018. Electron cryomicroscopy as a powerful tool in biomedical research. Journal of Molecular Medicine, 96, 483-493.

RANDELL, J. C. W., BOWERS, J. L., RODRÍGUEZ, H. K. & BELL, S. P. 2006. Sequential ATP Hydrolysis by Cdc6 and ORC Directs Loading of the Mcm2-7 Helicase. Molecular Cell, 21, 29-39.

RAO, H. & STILLMAN, B. 1995. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proceedings of the National Academy of Sciences, 92, 2224-2228.

REMUS, D., BEALL, E. L. & BOTCHAN, M. R. 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC–DNA binding. The EMBO Journal, 23, 897- 907.

REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J. D., MORRIS, E. P. & DIFFLEX, J. F. X. 2009a. Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell, 139, 719-730.

REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J. D., MORRIS, E. P. & DIFFLEY, J. F. X. 2009b. Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell, 139, 719-730.

REMUS, D., BLANCHETTE, M., RIO, D. C. & BOTCHAN, M. R. 2005. CDK Phosphorylation Inhibits the DNA-binding and ATP-hydrolysis Activities of the Drosophila Origin Recognition Complex*. Journal of Biological Chemistry, 280, 39740-39751. 245

REMUS, D. & DIFFLEY, J. F. X. 2009. Eukaryotic DNA replication control: Lock and load, then fire. Current Opinion in Cell Biology, 21, 771-777.

RIERA, A., BARBON, M., NOGUCHI, Y., REUTER, L. M., SCHNEIDER, S. & SPECK, C. 2017. From structure to mechanism— understanding initiation of DNA replication. Genes and Development, 31, 1073-1088.

ROWLEY, A., COCKER, J. H., HARWOOD, J. & DIFFLEY, J. F. 1995. Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. The EMBO Journal, 14, 2631-2641.

RUSINIAK, M. E., KUNNEV, D., FREELAND, A., CADY, G. K. & PRUITT, S. C. 2012. Mcm2 deficiency results in short deletions allowing high resolution identification of genes contributing to lymphoblastic lymphoma. Oncogene, 31, 4034-4044.

RYBA, T., HIRATANI, I., LU, J., ITOH, M., KULIK, M., ZHANG, J., DALTON, S. & GILBERT, D. M. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Research.

SALEH, A., NOGUCHI, Y., ARAMAYO, R., IVANOVA, M. E., STEVENS, K. M., MONTOYA, A., SUNIDHI, S., CARRANZA, N. L., SKWARK, M. J. & SPECK, C. 2022. The structural basis of Cdc7-Dbf4 kinase dependent targeting and phosphorylation of the MCM2-7 double hexamer. Nature Communications, 13, 2915.

SAMEL, S. A., FERNÁNDEZ-CID, A., SUN, J., RIERA, A., TOGNETTI, S., HERRERA, M. C., LI, H. & SPECK, C. 2014. A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2–7 onto DNA. Genes & Development, 28, 1653- 1666.

SÁNCHEZ, H., LIU, Z., VAN VEEN, E., VAN LAAR, T., DIFFLEY, J. F. X. & DEKKER, N. H. 2023. A chromatinized origin reduces the mobility of ORC and MCM through interactions and spatial constraint. Nature Communications, 14, 6735.

SÁNCHEZ, H., MCCLUSKEY, K., VAN LAAR, T., VAN VEEN, E., ASSCHER, F. M., SOLANO, B., DIFFLEY, J. F. X. & DEKKER, N. H. 2021. DNA replication origins retain mobile licensing proteins. Nature Communications, 12, 1908.

SANTOCANALE, C., SHARMA K FAU - DIFFLEY, J. F. & DIFFLEY, J. F. 1999. Activation of dormant origins of DNA replication in budding yeast.

SAXENA, S., YUAN, P., DHAR, S. K., SENGA, T., TAKEDA, D., ROBINSON, H., KORNBLUTH, S., SWAMINATHAN, K. & DUTTA, A. 2004. A Dimerized Coiled-Coil Domain and an Adjoining Part of Geminin Interact with Two Sites on Cdt1 for Replication Inhibition. Molecular Cell, 15, 245-258. 246

SCARFF, C. A., FULLER, M. J. G., THOMPSON, R. F., & IADANZA, M. G. 2018. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. Journal of visualized experiments : JoVE, (132), 57199.

SCHERES, S. H. W. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology, 180, 519-530.

SCHMIDT, J. M. & BLEICHERT, F. 2020. Structural mechanism for replication origin binding and remodeling by a metazoan origin recognition complex and its co-loader Cdc6. Nature Communications, 11, 4263.

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, 1093-1104.

SCOTT, B. & GHEORGHE, C. 2022. Visualizing the Dynamics of DNA Replication and Repair at the Single-Molecule Molecule Level. bioRxiv, 2022.10.22.513350.

SEO, Y.-S. & KANG, Y.-H. 2018. The Human Replicative Helicase, the CMG Complex, as a Target for Anti-cancer Therapy. Frontiers in Molecular Biosciences, 5.

SHECHTER, D., YING, C. Y. & GAUTIER, J. 2004. DNA Unwinding Is an MCM Complexdependent and ATP Hydrolysis-dependent Process*. Journal of Biological Chemistry, 279, 45586-45593.

SHEN, P. S. 2018. The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. Analytical and Bioanalytical Chemistry, 410, 2053-2057.

SHIBATA, E., KIRAN, M., SHIBATA, Y., SINGH, S., KIRAN, S. & DUTTA, A. 2016. Two subunits of human ORC are dispensable for DNA replication and proliferation. eLife, 5, e19084.

SIDDIQUI, K., ON, K. F. & DIFFLEY, J. F. X. 2013. Regulating DNA Replication in Eukarya. Cold Spring Harbour perspective in biology, 5.

SIMON, N. E. & SCHWACHA, A. 2014. The Mcm2-7 Replicative Helicase: A Promising Chemotherapeutic Target. BioMed Research International, 2014, 549719.

SPECK, C., CHEN, Z., LI, H. & STILLMAN, B. 2005. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat Struct Mol Biol, 12, 965-71.

SPECK, C. & STILLMAN, B. 2007. Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. Journal of Biological Chemistry, 282.

STANOJCIC, S., LEMAITRE, J.-M., BRODOLIN, K., DANIS, E. & MECHALI, M. 2008. In Xenopus Egg Extracts, DNA Replication Initiates Preferentially at or near Asymmetric AT Sequences. Molecular and Cellular Biology, 28, 5265-5274.

STINCHCOMB, D. T., STRUHL, K. & DAVIS, R. W. 1979. Isolation and characterisation of a yeast chromosomal replicator. Nature, 282, 39-43. 247

SUN, J., FERNANDEZ-CID, A., RIERA, A., TOGNETTI, S., YUAN, Z., STILLMAN, B., SPECK, C. & LI, H. 2014. Structural and mechanistic insights into Mcm2–7 double-hexamer assembly and function. Genes and Development, 28, 2291-2303.

TADA, S., LI, A., MAIORANO, D., MÉCHALI, M. & BLOW, J. J. 2001. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. Nature Cell Biology, 3, 107-113.

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, 437-444.

TANAKA, S. & DIFFLEY, J. F. X. 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2–7 during G1 phase. Nature Cell Biology, 4, 198-207.

TANAKA, T., UMEMORI, T., ENDO, S., MURAMATSU, S., KANEMAKI, M., KAMIMURA, Y., OBUSE, C. & ARAKI, H. 2011. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. The EMBO Journal, 30, 2019-2030.

TANG, M., CHEN, J., ZENG, T., YE, D.-M., LI, Y.-K., ZOU, J. & ZHANG, Y.-P. 2023. Systemic analysis of the DNA replication regulator origin recognition complex in lung adenocarcinomas identifies prognostic and expression significance. Cancer Medicine, 12, 5035-5054.

TARDAT, M., BRUSTEL, J., KIRSH, O., LEFEVBRE, C., CALLANAN, M., SARDET, C. & JULIEN, E. 2010. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nautre Cell Biology, 12.

THEIS, J. F. & NEWLON, C. S. 1997. The ARS309 chromosomal replicator of Saccharomyces cerevisiae depends on an exceptional ARS consensus sequence. Proceedings of the National Academy of Sciences, 94, 10786-10791.

THOME, K. C., DHAR, S. K., QUINTANA, D. G., DELMOLINO, L., SHAHSAFAEI, A. & DUTTA, A. 2000. Subsets of human origin recognition complex (ORC) subunits are expressed in non-proliferating cells and associate with non-ORC proteins. J Biol Chem, 275, 35233-41.

THOMPSON, R. F., WALKER, M., SIEBERT, C. A., MUENCH, S. P. & RANSON, N. A. 2016. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods, 100, 3-15.

TICAU, S., FRIEDMAN, L. J., CHAMPASA, K., CORRÊA, I. R., GELLES, J. & BELL, S. P. 2017. Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. Nature Structural & Molecular Biology, 24, 309-315.

TOCILJ, A., ON, K. F., YUAN, Z., SUN, J., ELKAYAM, E., LI, H., STILLMAN, B. & JOSHUATOR, L. 2017. Structure of the active form of human origin recognition complex and its ATPase motor module. Elife, 6. 248

TRIOLO, T. & STERNGLANZ, R. 1996. Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature, 381, 251-253.

TUBBS, A., SRIDHARAN, S., VAN WIETMARSCHEN, N., MAMAN, Y., CALLEN, E., STANLIE, A., WU, W., WU, X., DAY, A., WONG, N., YIN, M., CANELA, A., FU, H., REDON, C., PRUITT, S. C., JASZCZYSZYN, Y., ALADJEM, M. I., APLAN, P. D., HYRIEN, O. & NUSSENZWEIG, A. 2018. Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. Cell, 174, 1127-1142.e19.

VALVERDE, L. D. F., DE FREITAS, R. D., PEREIRA, T. D. A., DE RESENDE, M. F., AGRA, I. M. G., DOS SANTOS, J. N., DOS REIS, M. G., SALES, C. B. S. & GURGEL ROCHA, C. A. 2018. MCM3: A Novel Proliferation Marker in Oral Squamous Cell Carcinoma. Applied Immunohistochemistry & Molecular Morphology, 26.

VASHEE, S., CVETIC, C., LU, W., SIMANCEK, P., KELLY, T. J. & WALTER, J. C. 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev, 17, 1894-908.

VAZIRI, C., SAXENA, S., JEON, Y., LEE, C., MURATA, K., MACHIDA, Y., WAGLE, N., HWANG, D. S. & DUTTA, A. 2003. A p53-Dependent Checkpoint Pathway Prevents Rereplication. Molecular Cell, 11, 997-1008.

WALKER, J. E., SARASTE, M., RUNSWICK, M. J. & GAY, N. J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO Journal, 1, 945-951.

WANG, B., FENG, L., HU, Y., HUANG, S. H., REYNOLDS, C. P., WU, L. & JONG, A. Y. 1999. The Essential Role of *Saccharomyces cerevisiae CDC6* Nucleotidebinding Site in Cell Growth, DNA Synthesis, and Orc1 Association *. Journal of Biological Chemistry, 274, 8291-8298.

WANG XK, WANG QQ, HUANG JL, ZHANG LB, ZHOU X, LIU JQ, CHEN ZJ, LIAO XW, HUANG R, YANG CK, ZHU GZ, HAN CY, YE XP, PENG T. 2020. Novel candidate biomarkers of origin recognition complex 1, 5 and 6 for survival surveillance in patients with hepatocellular carcinoma. J Cancer. 11(7):1869-1882.

WANG, Z. 2021. Regulation of Cell Cycle Progression by Growth Factor-Induced Cell Signaling. Cells [Online], 10.

WELLS, J.N., LEBER, V., EDWARDES, L.V., ALLYJAUN, S., PEACH, M., TOMKINS, J., KEFALA-STAVRIDI, A., FAULL, S.V., ARAMAYO, R., PESTANA, C.M., RANJHA, L., SPECK, C. 2024. Reconstitution of human DNA licensing and the structural and functional analysis of key intermediates. bioRxiv 2024.04.11.589023.

WEISSMANN, F., GREIWE, J.F., PÜHRINGER, T., MILLER, T.C.R., DIFFLEY, J.F.X., COSTA, A. 2024. MCM Double Hexamer Loading Visualised with Human Proteins. bioRxiv 2024.04.10.588848.

WENDLER, P., CINIAWSKY, S., KOCK, M. & KUBE, S. 2012. Structure and function of the AAA+ nucleotide binding pocket. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1823, 2-14.

WILMES, G. M. & BELL, S. P. 2002. The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. Proceedings of the National Academy of Sciences, 99, 101-106. 249

WOHLSCHLEGEL, J. A., DWYER, B. T., DHAR, S. K., CVETIC, C., WALTER, J. C. & DUTTA, A. 2000. Inhibition of Eukaryotic DNA Replication by Geminin Binding to Cdt1. Science, 290, 2309-2312.

WOODWARD , A. M., GÖHLER , T., LUCIANI , M. G., OEHLMANN , M., GE , X., GARTNER , A., JACKSON , D. A. & BLOW , J. J. 2006. Excess Mcm2–7 license dormant origins of replication that can be used under conditions of replicative stress. Journal of Cell Biology, 173, 673-683.

WU, W., WANG, X., SHAN, C., LI, Y. & LI, F. 2018. Minichromosome maintenance protein 2 correlates with the malignant status and regulates proliferation and cell cycle in lung squamous cell carcinoma.

XIONG W, XIE C, QIU Y, TU Z, GONG Q. 2020. Origin recognition complex subunit 1 regulates cell growth and metastasis in glioma by altering activation of ERK and JNK signaling pathway. Mol Cell Probes, 49:101496.

XU, J., YANAGISAWA, Y., TSANKOV, A. M., HART, C., AOKI, K., KOMMAJOSYULA, N., STEINMANN, K. E., BOCHICCHIO, J., RUSS, C., REGEV, A., RANDO, O. J., NUSBAUM, C., NIKI, H., MILOS, P., WENG, Z. & RHIND, N. 2012. Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biology, 13, R27.

XU, N., LIN, Q., TIAN, H., LIU, C., WANG, P., SUEN, C. M., YANG, H., XIANG, Y. & ZHU, G. 2022. Cryo-EM structure of human hexameric MCM2-7 complex. iScience, 25, 104976.
YANG, N. & XU, R. M. 2013. Structure and function of the BAH domain in chromatin biology. Crit Rev Biochem Mol Biol, 48, 211-21.

YANG, R., HUNKER, O., WISE, M., BLEICHERT, F. 2024. Multiple pathways for licensing human replication origins. bioRxiv 2024.04.10.588796

YEELES, J. T. P., DEEGAN, T. D., JANSKA, A., EARLY, A. & DIFFLEY, J. F. X. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. Nature, 519, 431-435.

YELLA, V. R., VANAJA, A., KULANDAIVELU, U. & KUMAR, A. 2020. Delving into Eukaryotic Origins of Replication Using DNA Structural Features. ACS Omega, 5, 13601-13611.

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, 4334-4344.

YOU, Z. & MASAI, H. 2008. Cdt1 Forms a Complex with the Minichromosome Maintenance Protein (MCM) and Activates Its Helicase Activity *. Journal of Biological Chemistry, 283, 24469-24477.

YU, S., WANG, G., SHI, Y., XU, H., ZHENG, Y., & CHEN, Y. 2020. MCMs in Cancer: Prognostic Potential and Mechanisms. Analytical cellular pathology (Amsterdam), 3750294.

YUAN, Z., RIERA, A., BAI, L., SUN, J., NANDI, S., SPANOS, C., CHEN, Z. A., BARBON, M., RAPPSILBER, J., STILLMAN, B., SPECK, C. & LI, H. 2017. Structural basis of Mcm2–250 7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nature Structural and Molecular Biology, 24, 316-324.

YUROV, Y. B. & LIAPUNOVA, N. A. 1977. The units of DNA replication in the mammalian chromosomes: Evidence for a large size of replication units. Chromosoma, 60, 253- 267.

ZENG, W., BALL JR, A. R. & YOKOMORI, K. 2010. HP1: Heterochromatin binding proteins working the genome. Epigenetics, 5, 287-292.

ZHAI, Y., CHENG, E., WU, H., LI, N., YUNG, P. Y. K., GAO, N. & TYE, B.-K. 2017. Openringed structure of the Cdt1–Mcm2–7 complex as a precursor of the MCM double hexamer. Nature Structural & Molecular Biology, 24, 300-308.

ZHANG, Y., TSENG, J. T., LIEN, I. C., LI, F., WU, W., & LI, H. 2020. mRNAsi Index: Machine Learning in Mining Lung Adenocarcinoma Stem Cell Biomarkers. Genes, 11(3), 257

ZHENG, S. Q., PALOVCAK, E., ARMACHE, J.-P., VERBA, K. A., CHENG, Y. & AGARD, D. A. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature Methods, 14, 331-332.

ZOU, L. & STILLMAN, B. 2000. Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Molecular and Cellular Biology, 20, 3086-3096.