

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER FAKULTÄT FÜR BIOLOGIE
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

The role of Orc6 in the human cell



Mate Ravasz

18. DEZEMBER 2013

Erstgutachter: Prof. Dr. Dirk Eick
Zweitgutachter: Prof. Dr. Wolfgang Enard
Tag der mündliche Prüfung: 19.09.2014.

1. Table of contents

1. TABLE OF CONTENTS	4
2. INTRODUCTION	8
2.1 THE REPLICATION OF DNA	9
2.1.1 THE REPLICON MODEL	10
2.2 THE ORIGIN RECOGNITION COMPLEX PROTEIN FAMILY	14
2.2.1 ORC6	17
2.3 INITIATION OF REPLICATION IN HUMAN CELLS	21
2.4 THE EPSTEIN-BARR VIRUS GENOME AS MODEL FOR HUMAN DNA REPLICATION	24
2.4.1 ANALYZING REPLICATION COMPETENCE OF PROTEINS <i>IN VIVO</i> : THE PLASMID RESCUE ASSAY	26
3. MATERIAL AND METHODS	28
3.1 MATERIAL	29
3.1.1 OLIGONUCLEOTIDES	29
3.1.2 PLASMIDS	30
3.1.3 ANTIBODIES	31
3.1.4 BACTERIAL STRAINS	32
3.1.5 CELL LINES AND CULTURE MATERIAL	33
3.1.6 SOFTWARE	34
3.1.7 EQUIPMENT AND OTHER MATERIAL	34
3.2 METHODS	36
3.2.1 CELL CULTIVATION	36
3.2.2 TRANSFECTION OF HUMAN CELLS	37
3.2.3 PLASMID-RESCUE ASSAY	37
3.2.4 FLUORESCENCE MICROSCOPY	38
3.2.5 PROTEIN EXTRACT PREPARATION	39
3.2.6 WESTERN BLOT	40
3.2.7 PLASMID-BINDING ASSAY	42
3.2.8 PROTEIN EXPRESSION AND PURIFICATION	43
3.2.9 IMMUNOPRECIPITATION	43
3.2.10 PULL-DOWN	44
3.2.11 POLYMERASE CHAIN REACTION	44
3.2.12 BACTERIAL TRANSFORMATION	44
3.2.13 PLASMID PREPARATION	45
3.2.14 AGAROSE GEL ELECTROPHORESIS	45
3.2.15 MASS SPECTROMETRY	45
4. AIM OF THE STUDY	48
5. RESULTS	50
5.1 CHARACTERIZATION OF ORC6 INTERACTION WITH THE REPLICATION MACHINERY	51
5.1.1 ORC6 IS PRESENT IN ABUNDANCE COMPARED TO OTHER ORC SUBUNITS	51
5.1.2 ORC6 INTERACTS WITH ORC1-5 AND Cdc6 IN SOLUTION	53
5.1.3 200BP BUT NOT 36BP OF DNA IS SUFFICIENT FOR ORC TO ATTACH	56
5.1.4 ORC6 SATURATES SIMILARLY TO ORC2-5 WHEN BINDING TO DNA	62
5.1.5 ORC6 IS ABLE TO BIND DNA INDEPENDENTLY OF ORC	65

Table of contents

5.1.6	ORC ENHANCES ORC6 DNA BINDING	68
5.1.7	ORC6 IS NOT REQUIRED FOR ORC DNA BINDING, BUT IT ENHANCES THE PROCESS IF PRESENT	70
5.1.8	ORC6 BINDS VIA ITS C TERMINUS TO ORC <i>IN VITRO</i>	71
5.1.9	ORC6 NEEDS A NUCLEAR LOCALIZATION SIGNAL OR ITS C TERMINUS TO ENTER THE NUCLEUS	77
5.2	FUNCTIONAL ANALYSIS OF ORC6 AT ORIGINS OF REPLICATION	81
5.2.1	ORC6 VARIANTS ARE ABLE TO CREATE ORIGINS OF REPLICATION	81
5.2.2	PR-SET7 IS ABLE TO CREATE ORIGINS OF REPLICATION	84
5.2.3	PR-SET7 ENHANCES EXISTING REPLICATION ORIGINS	88
5.3	ORC6 INTERACTS WITH VARIOUS OTHER FACTORS IN THE CELL	90
5.3.1	FUNCTIONAL CLUSTERING OF ORC6 INTERACTING PROTEINS	97
6.	<u>DISCUSSION</u>	104
6.1	ORC6 IS AN ABUNDANT ORC COMPONENT	105
6.2	ORC6 HAS TWO DISTINCT WAYS OF ATTACHING TO ORC1-5	107
6.3	ORC6 HAS DOMAINS THAT CONTROL CELLULAR LOCALIZATION AND ORC INTERACTION	108
6.4	DERIVATIVES OF ORC6 ARE SUFFICIENT TO CREATE ORIGINS OF REPLICATION	112
6.5	ORC6 IS A VERSATILE CELL CYCLE REGULATION FACTOR	114
7.	<u>SUMMARY</u>	118
8.	<u>ZUSAMMENFASSUNG</u>	119
9.	<u>REFERENCES</u>	122
10.	<u>APPENDIX</u>	144
10.1	RAW MICROSCOPY IMAGES	146
10.2	GLOSSARY	148
10.3	PUBLICATIONS	150
10.4	ACKNOWLEDGEMENTS	151
10.5	ERKLÄRUNG	152

2. Introduction

2.1 The replication of DNA

DNA replication is the process of copying the entire genome of a cell during proliferation. This extraordinary task is carried out with remarkable efficiency in living organisms: a fertilized egg goes through 5 trillion cell divisions to develop into an adult human, creating 20 billion kilometers of DNA in the process (DePamphilis 2006).

The workhorse behind this process is the DNA polymerase. Bacterial polymerases copy up to 100,000 bases of DNA each minute, allowing them to replicate their entire genome in minutes with error rates as low as 1 in 100 million basepairs (Kornberg & Baker 2005; Kunkel 2004). In eukaryotes, genome sizes are drastically increased, and the sophisticated eukaryotic replication machinery only handles about 2000 bases per minute. Therefore, with a single polymerase it would take 20 years for a human cell to divide, whereas in reality this task is accomplished in about 8 hours. To achieve this, eukaryotic cells have developed a way to parallelize replication by starting it in multiple locations. These start sites are named replication origins, which are functional genomic locations that allow the recruitment of the replication machinery.

This thesis deals with the protein Orc6, a subunit of the human origin recognition complex (ORC). This complex is the first to recognize and bind to replication origins during the cell cycle. To place Orc6 into the context of the human cell and DNA replication, first replication initiation and origins are discussed in more detail.

2.1.1 The replicon Model

The existence of origins was first suggested in the 1960s by François Jacob (Jacob et al. 1963). The replicon model, based on studies in bacteria, proposed a cis-acting replicator element, and a trans initiator element, the recruitment of the latter triggering replication initiation (Figure 1). The replicon itself was defined as a stretch of DNA, which is replicated from a single origin.



Figure 1. The replicon model. This drawing in the sand represents the model proposed by François Jacob, Sydney Brenner and Francois Cuzin to explain how DNA replication initiates. It depicts the circular bacterial genome with a box as the replicator element (Jacob et al. 1963). In order for replication to start, the trans acting initiator element (indicated by the arrow) is recruited. Image: (Thomae 2008; Skarstad et al. 2003)

Such a single replicon system is typical for bacteria, and at the time it was postulated that the larger genomes of eukaryotes simply have several replicons next to each other, which are activated in a pre-determined order (Gilbert 2004). More recent studies however found eukaryotic replication initiation to be more complex.

2.1.1.1 The replicator

The replicator element is the cis-acting structure, which defines the locus where DNA replication commences. Different organisms have developed different replicators, which are referred to as origins of replication.

Bacteria have origins similar to the structure proposed in Figure 1, where a single, genetically defined origin sequence serves as the replicator. In *Escherichia coli* this is the 250bp *oriC* element, which is sufficient to replicate the circular, 4.6mb long prokaryotic genome (Mott & Berger 2007). *OriC* consists of three, AT rich 13-mers and four 9-mer repeats, which are all crucial for replication (Leonard & Grimwade 2010). Bacterial plasmids, being independent, circular pieces of DNA in prokaryotes, also contain a replicator sequence. These have been well characterized for use in genetic engineering, such as the pUC origin sequence, which I also took advantage of for creating bacterial plasmids, as this allows stable, high copy number presence of plasmids in *E. coli* (Herman et al. 1994).

Archaea, despite being prokaryotes, have replication machineries related to eukaryotes. They use multiple origins to replicate their circular genome. These replicators are poorly characterized to date, have no known consensus motifs, but are generally AT rich, contain inverted repeats and are often in close proximity of initiator genes (Kelman & Kelman 2003; Kelman & Kelman 2004).

The first origins in eukaryotes were found in the budding yeast *Saccharomyces cerevisiae*. These were initially described as sequences which initiate replication if introduced into plasmids, and were therefore named autonomous replicating sequences (Stinchcomb et al. 1979). ARS elements have two conserved regions: A and B, the latter often divided into B1 and B2, while some ARS elements also contain a B3 sequence. ARS1, the first yeast origin studied in detail contained an eleven basepair consensus motive 5'-WTTTAYRTTTW-3' in the A element, the mutation of which lead to the

complete loss of origin function. Introducing mutations in the B1, B2 and B3 sequences only caused a reduction in origin activity, and later studies found that most ARS sequences of the budding yeast genome are missing some of their B elements (Rao et al. 1994; Lee & Bell 1997; Marahrens & Stillman 1992).

Schizosaccharomyces pombe, another yeast species used as model organism, already differs from *cerevisiae* as it does not have ARS motif consensus sequences. Instead it uses 500-1000bp long AT rich stretches of DNA, which were found with the same method as the ARS sequences in budding yeast. *S. pombe* origins, like metazoan origins, do not have consensus sequences, however they often contain stretches of pure A or T, and are located in intergenic regions (Dai et al. 2005; Clyne & Kelly 1995).

In metazoans, replicators are even less defined. In *Drosophila* and humans, any DNA sequence is able to recruit the initiator complex (Vashee et al. 2003; Remus et al. 2004). Also, plasmids containing random eukaryotic sequences larger than 5kb, and having a size greater than 15kb are able to replicate in human cells as evidenced by both short- and long-term plasmid maintenance assays (Heinzel et al. 1991; Krysan et al. 1993). Despite this fact, initiation events in the human genome are not random. As estimated, approximately 30.000 origins are activated during each cycle in human cells (Nasheuer et al. 2002; Aladjem 2007). However, the number of potential origins is estimated to be significantly higher, and from this pool of potential start sites only a subset is activated during each replication. A study published by my group shows that AT content, nucleosome density, MNase sensitivity and epigenetic marks all contribute to regulating where replication starts (Papior et al. 2012; Cayrou et al. 2011). This makes it difficult to adapt the replicon model to metazoans, as replicator elements are not defined by sequence, are not necessarily activated during each division, instead of discrete loci they can form initiation zones, and finally replicon sizes and locations vary between divisions (Schepers & Papior 2010).

2.1.1.2 The initiator

Replicators presented in the previous chapter are responsible for pinpointing future sites of replication initiation. Initiators are the trans elements needed to trigger the initiation event.

In bacteria, activation commences with *dnaA* proteins attaching to the 9-mer sequences, bending the origin DNA in the process. This allows the *dnaB* helicase to bind to the 13-mers, along with its cofactor, *dnaC*. The helicase encircles single stranded DNA, therefore two units are loaded - one on each strand - allowing bidirectional replication at the *oriC* locus. As the helicase further opens up the origin, the bacterial polymerase along with its primase is recruited. The latter, also called *dnaG*, synthesizes an RNA primer, which is required for the polymerase to function. Once the complete replisome is assembled, the origin fires and DNA synthesis initiates (Kaguni 2011).

Archaea employ a different system, which resembles a simplified form of the eukaryotic initiation mechanism (Kelman & Kelman 2003). During the archaeal DNA replication initiation process, origins are first bound by origin binding proteins (OBP), which are homologous to the eukaryotic *Orc1/Cdc6* (MacNeill 2001; Dueber et al. 2007). These allow the recruitment of minichromosome maintenance factors (MCMs), which form a double-hexamer at the origin site, and unwind DNA. Once the MCMs are loaded, the recruitment of single stranded DNA binding proteins is triggered, which stabilize the single DNA strands and act as a platform to recruit the multi-subunit polymerase-primase complex. The entire process is similar to that of eukaryotes, but it is simpler and involves fewer elements. Interestingly, the number of OBP genes varies between different Archaea, whereas there are always six subunits of it in Eukarya (Myllykallio & Forterre 2000). Similarly, MCM genes may have 1-4 copies in Archaea, but assemble into a double hexameric complex like the six distinct subunits of eukaryotes (Tye 2000).

Eukaryotes have developed an even more sophisticated initiator system based on the principles established in Archaea. Briefly, replicators are first recognized by the six-subunit origin recognition complex (ORC). Once ORC has assembled on DNA, it recruits *Cdc6*, *Cdt1* and the MCM2-7 proteins to form the pre-replication complex (pre-RC) (Bell & Dutta 2002). The MCM2-7 proteins form head-to-head double hexamers like their counterparts in Archaea, and are also required to melt the DNA for the replication machinery (Evrin et al. 2009; Remus et al. 2009). Once the pre-RC is established, it further develops into the pre-initiation complex (pre-IC) by the addition of MCM10, *Sld2*, *3*, *7* proteins, the GINS complex and *Cdc45*. This assemblage is finally able to attach the polymerase to initiate replication (MacNeill 2010; Takisawa et al. 2000).

This thesis deals with the role of the smallest subunit of ORC, Orc6. In order to understand its role in replication initiation, the origin recognition complex and eukaryotic replication initiation is introduced in more detail.

2.2 The Origin Recognition Complex protein family

The first eukaryotic initiators were identified in 1992 while conducting DNaseI footprinting experiments on ARS1 sequences in *S. cerevisiae* (Bell & Stillman 1992). It was found that a protein complex of six subunits binds specifically to ARS. This was named the origin recognition complex (ORC), and the six subunits were labeled Orc1-6 in the order of their sizes, with Orc6 being the smallest subunit of 50 kilodaltons. Shortly after this discovery, the genes corresponding to each subunit were identified, and were found to be located on separate chromosomes (Li & Herskowitz 1993; Bell et al. 1993; Foss et al. 1993; Bell et al. 1995). Using these sequences it was quickly established that ORC components are highly conserved in all eukaryotes, and are key factors of replication initiation from yeast to man.

Five of the six subunits have similar structures, the exception being Orc6, which is unlike any other member of the complex. Interestingly, the large, conserved consensus domain of Orc1-5 is also present in the Cdc6 protein, another prominent member of the pre-replication complex. This indicates that these six proteins are paralogs (Duncker et al. 2009).

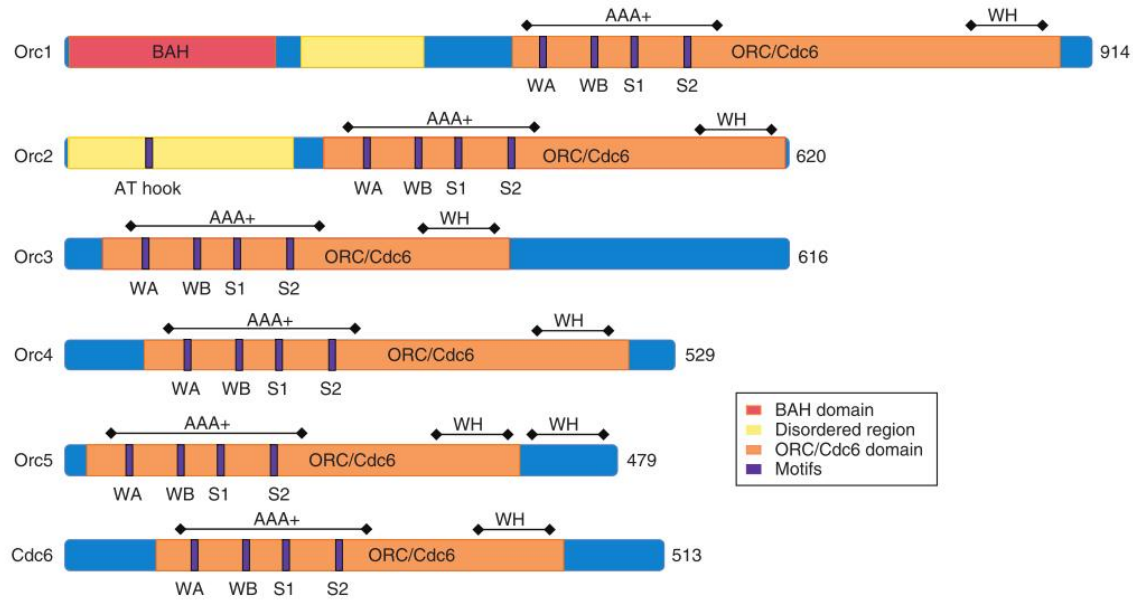


Figure 2. Comparison of domains for Orc1-5 and Cdc6 from *S. cerevisiae*. Each of the six proteins depicted and labeled on the left are crucial for replication initiation. ORC proteins form a steady complex at origins, to which Cdc6 is promptly recruited in G1. The similar motifs found in all members are presented in orange, and are named the ORC/Cdc6 domain. This contains the AAA+ type ATPase domain and a winged helix responsible for DNA binding. The purple marks indicate the consensus Walker A and B, and Sensor 1 and 2 motifs of the ATPase domain, as well as the AT hook found in ScOrc2. The ATPase domain is complete in Orc1, Orc2, Orc5 and Cdc6, while being less conserved in Orc3 and Orc4 based on studies in yeast. Orc1 also contains a Bromo-adjacent-homology (BAH) domain which is used in heterochromatin formation. Orc6 is not depicted as it does not share any similar domains with any other ORC components. (image from: Duncker et al. 2009).

The conserved ORC/Cdc6 domain contains an AAA+ type ATPase domain and a winged helix domain, as presented in Figure 2. ATPase activities of the different subunits vary among eukaryotes, but it is generally observed that ATP binding, but not hydrolysis is required for complex assembly and DNA binding. While Orc1, 4 and 5 have recognizable AAA+ sites, Orc2 and 3 only structurally resemble this motif. The most active subunit in this respect is Orc1, which binds ATP upon DNA attachment in yeast, *Drosophila* and humans, but its hydrolysis only occurs after Cdc6 recruitment to attach this factor to DNA. Cdc6 ATP hydrolysis is also required for this step (Speck et al. 2005; Chesnokov et al. 2001; Vashee et al. 2003). The presence of nucleotides is essential for complex

formation when using purified human proteins, but here ATP hydrolysis does not occur (Ranjan & Gossen 2006; Siddiqui & Stillman 2007). Orc5 was also observed binding ATP in yeast, but it does not hydrolyze it, whereas this ability is not a requirement in *Drosophila* (Klemm et al. 1997).

The winged helix motif, the other conserved feature in the ORC/Cdc6 domain is required for DNA binding. This helix-turn-helix type domain is found in all known ORCs, however, binding characteristics are diverse among eukaryotes. For example, *S. cerevisiae* ORC recognizes ARS DNA, the AT hook of Orc4 in *S. pombe* prefers AT stretches, and metazoan ORCs bind any DNA sequence as discussed in chapter 2.1.1.1. (Tada et al. 2008). Although both the AAA+ and the winged helix domains are able to function on their own, they are likely to act in concert when the origin recognition complex is assembled.

The architecture of eukaryotic ORCs has been studied using archaeal, yeast and human proteins. The archaeal Orc1/Cdc6 protein first forms a dimer after which it is able to recognize DNA with its winged helix, and stabilize the interaction by inserting a helix-turn-helix motif from its AAA+ domain into the minor groove. DNA is bent, twisted and melted in the process (Gaudier et al. 2007; Dueber et al. 2007). In eukaryotes, the six subunit complex forms a ring-like structure in solution in yeast, and with the addition of Cdc6, the complex contains six AAA+ motifs, which loosely resembles the MCM- and replication factor C complexes (Speck et al. 2005). This assemblage also bends DNA upon attachment (Sun et al. 2012; Lee & Bell 1997). Using *Drosophila* ORC it was further elucidated that Orc1-5 form a right handed helical core with their AAA+ domains, which corresponds well to the bacterial dnaA complex that twists the origin (Clarey et al. 2006).

Although no crystallization study using human proteins has been published to date, the combination of biochemical analyses with the knowledge gained from yeast and *Drosophila* structures allows determining the organization of subunits within this complex as well. It has been elucidated that in humans, Orc2 and 3 form a very strong complex early on, to which first Orc5, then Orc4 attaches with high affinity (Ranjan & Gossen 2006; Siddiqui & Stillman 2007; Vashee et al. 2001; Kneissl et al. 2003). Orc4 recruitment is dependent on the presence of ATP or ATP γ S (a slowly hydrolyzing form

of ATP), which indicates that ATP binding, but not hydrolysis is required for this step. Thus, Orc2-5 form a stable core complex in humans to which Orc1 and Orc6 attach with low affinity (Dhar et al. 2001a; Saha et al. 1998). The core complex has been purified from human cells by the studies above, but they always found that Orc1 and 6 were only co-purified in sub-stoichiometric amounts. Orc1 attachment again requires the presence of ATP, which is in line with the DNA binding studies of ORC performed in yeast. According to the current models, the recruitment of Orc6 is less understood, but one study indicates that either Orc2-5 or Orc1-5 is able to bind Orc6 after localizing independently into the nucleus (Ghosh et al. 2011).

2.2.1 ORC6

Although the assembly and DNA binding of ORC have been extensively studied, there is surprisingly little data available on its smallest subunit, Orc6. Interestingly, this protein is unlike any other member of its family, being different in both structure and function (Duncker et al. 2009). It is a protein essential for life in all eukaryotes, it is involved in replication, but its exact role remains unclear.

Shortly after the discovery of ORC in yeast, it was established that Orc6 is not required for ORC-DNA interaction of *S. cerevisiae* ORC (Lee & Bell 1997; Li & Herskowitz 1993). Surprisingly, later it was shown that it is essential for the same step in *Drosophila* (Chesnokov et al. 2001). The same year it was also found that *in vitro* replication using the cell-free *Xenopus* replication system is functional in the absence of Orc6 (Gillespie et al. 2001), however this study only assayed the presence of ORC components via low sensitivity visualization techniques (Coomassie stains), therefore the small Orc6 protein present in sub-stoichiometric amounts might have evaded detection. Also, the *Xenopus* Orc6 homologue was not even identified at that time (Klein et al. 2002; Giordano-Coltart et al. 2005).

Orc6 was also not instantly identified in humans, due to it being less conserved than the other members of its family (Dutta & Dhar 2000; Duncker et al. 2009). So far there is no clear answer to why Orc6 is evolving faster than all the other ORC components.

Structurally, Orc6 does not contain any identified conserved domains, which are present in all eukaryotes. The sole ‘Orc6 domain’, which fills most of the human protein sequence, has no defined function, but serves as a tool to identify homologues in other species. Human Orc6, but not that of *S. cerevisiae* has two disordered regions, which are separated by a putative coiled coil motif. A similar motif in *Drosophila* Orc6 interacts with Pnut, a *Drosophila* septin protein (Chesnokov et al. 2003). This region also contains the nuclear localization signal (Ghosh et al. 2011).

Only a part of HsOrc6 has been crystallized. The region between amino acids 94 and 187 forms a globular domain containing six helices (Liu et al. 2011). The resulting structure in Figure 3B resembles the transcription factor TFIIB. This factor in concert with the TATA box binding protein is an important transcriptional regulator. Based on this, Orc6 interacting with other co-factors is likely to attach to DNA to function. The DNA binding characteristics imply an innate binding activity with no preference for human Orc6 for any sequence. Footprinting experiments confirmed that human Orc6 equally protects the entire target strand, and it even binds single stranded DNA, similarly to yeast (Lee et al. 2000).

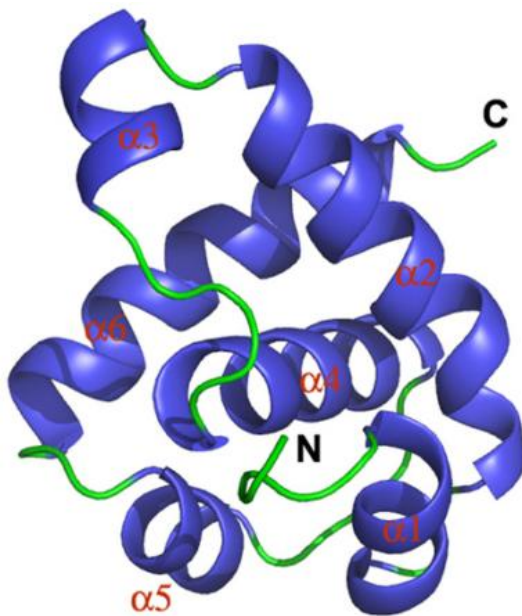
As depicted in Figure 3C, the position of Orc6 within yeast ORC is predicted to be in the vicinity of Orc2 and 3. HsOrc6 and MmOrc6 bind weakly to Orc2 and 3, however in yeast, only its interaction with Orc2 is found. It is also able to pull down a part of the Orc1 protein, but not the entire subunit (Lee & Bell 1997; Vashee et al. 2001; Sun et al. 2012; Kneissl et al. 2003). Apart from ORC, other pre-RC factors were also found to interact with Orc6, namely Cdc6 (Thomae et al. 2011); and Mcm5, Dbf4, Cdc7, Cdc45, RPA70, RPA14 in an extensive interaction screen using mouse proteins in yeast (Kneissl et al. 2003). The nature of these interactions is yet to be clarified, but it has already been observed in yeast that Orc6 tethers Cdt1 to ORC, which subsequently allows MCM complex recruitment and pre-RC formation (Chen et al. 2007; Chen & Bell 2011).

A)

```

1  mgseligr la prlglaepdm lrkaeeylrl srvcvgl sa rttetssavm cldlaaswmk
    DBD      α1      α2
61  cpldraylik lsglnketyq sclksfecll glsnigird lavqfsciea vnmaskilks
    α3      α4      α5      α6
121 yesslpqtgq vldldsrplf tsaallsack ilklkvdknk mvatsgvkka ifdrckqle
    NLS
181 kiggqv drep gdvatpprkr kkivveapak emekveemph kpqkdedltg dyewkrkil
    241 enaasaqkat ae
  
```

B)



C)

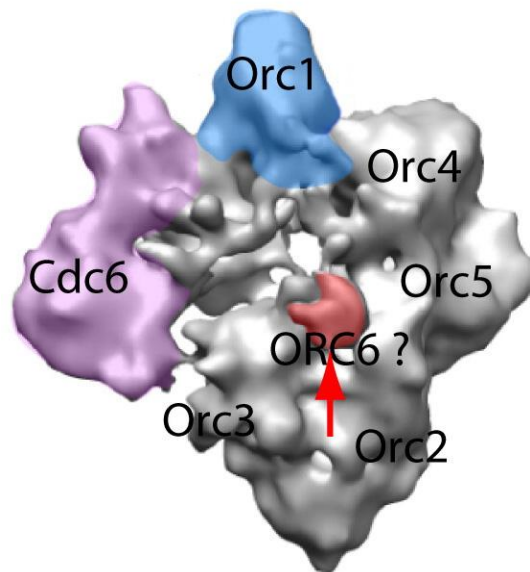


Figure 3. Overview of the human Orc6 protein. A) Amino acid sequence of HsOrc6 and main features. The area homologous to the *Drosophila* DNA binding domain (DBD) is shown in blue, with S72 and K76, the two essential amino acids for DNA binding in *Drosophila*, highlighted. The amino acids marked in red correspond to the most prominent secondary modification sites so far detected by high throughput studies (11 other sites with only a single detection are not shown). Of these sites, only T195, a known CDK1 target, has been experimentally validated (Ghosh et al. 2011). The nuclear localization signal (NLS) is depicted in orange, while the green helices labeled α 1-6 correspond to the structures seen in the crystallized middle domain of HsOrc6 shown in B) (image from: Liu et al. 2011). The three amino acids marked in green are crucial for DNA binding of HsOrc6. C) Presents the cryo-EM structure of *S. cerevisiae* ORC with the putative position of Orc6 in red (image from: Sun et al. 2012).

Going a step further, Orc6 has been postulated to interact prominently with other, non-pre-RC factors even upon its discovery. The initial Orc6 study in humans found that the protein was able to pull down several unknown factors (Dutta & Dhar 2000). Other studies later found that HsOrc6 also interacts with Cdk1 and HMGA1a (Ghosh et al. 2011; Thomae et al. 2008).

The former is an important regulator of cell cycle control, and phosphorylates both Orc2 and Orc6 in yeast. This event is mediated by Clb5 binding to an RXL motif to ScOrc6, which blocks one of the two Cdt1 interactions sites and so abolishes MCM complex loading. Mutation of this site to a constantly dephosphorylated form induces re-replication (Chen & Bell 2011; Wilmes et al. 2004). To the contrary, in HsOrc6, the Cdk1 site (T195) is within the nuclear localization signal, and microscopy studies confirmed that only the phosphorylated form of the protein is able to enter the nucleus, while the dephosphorylated form stays cytosolic (Ghosh et al. 2011). Therefore Cdk1 phosphorylation of Orc6 is crucial for replication initiation in humans.

Apart from its replication function, it is also emerging that ORC proteins are involved in processes not related to pre-RC formation (Thome et al. 2000; Scholefield et al. 2011). Specifically, a role for HsOrc6 and DmOrc6 during mitosis and cytokinesis is presented by recent studies (Prasanth et al. 2002; Huijbregts et al. 2009; Bernal & Venkitaraman 2011). These show that DmOrc6 interacts via its C terminus with Pnut, a *Drosophila* septin protein. Also, HsOrc6 localizes to the midbody during cytokinesis, and it also interacts with the kinetochore. In yeast, Orc6 depletion reduces MCM2-7 loading and disturbs S phase entry, but has no effect on mitosis - whereas in *Drosophila* and humans, aside from decreased replication, a mitotic defect is most prominent upon Orc6 depletion (Semple et al. 2006; Balasov et al. 2009).

2.3 Initiation of replication in human cells

To place the function of ORC in context, the process of DNA replication initiation has to be looked at in more detail. As ORC itself, this process is highly conserved among eukaryotes, with only minor differences between species. Most of our knowledge on it comes again from the budding yeast *S. cerevisiae*, which is supplemented by studies in higher eukaryotes.

Preparation for the upcoming replication cycle already begins at the end of mitosis. In late M phase and early G1, ORC takes the first step in replication by binding origin DNA. In *S. cerevisiae* and *S. pombe* it does not even dissociate from DNA and stays chromatin bound throughout the cell cycle (Ogawa et al. 1999; Aparicio et al. 1997). In higher eukaryotes, there is contradicting evidence as to how ORC attaches to chromatin. In frogs, initial studies found that ORC is at least partly removed from chromatin during mitosis (Romanowski et al. 1996), although other studies speculate this observation to be an artifact (Bell & Dutta 2002). It is more apparent from studies in hamster- and human cells that Orc1 dissociates from ORC during mitosis, only to be re-recruited in the next cycle (Kreitz et al. 2001; Natale et al. 2000; Li & Jin 2010; Gerhardt et al. 2006).

Regardless of ORC assembly, eukaryotic replication pathways converge on the point where ORC is successfully assembled on chromatin in the G1 phase of the cell cycle. Once this has been accomplished, it serves as a platform for pre-RC formation. It is important to note that only a subset of ORCs continue on to build pre-replication complexes, the exact selection criteria being unknown to date (Schepers & Papior 2010). The subset of ORCs that do continue in replication initiation first bind the Cdc6 protein. As seen in chapter 2.2, Cdc6 belongs to the same protein family as Orc1-5 and acts in concert with them in replication initiation. In yeast, Cdc6 is even able to interact with ORC in solution, but this is yet to be analyzed in higher eukaryotes (Sun et al. 2012).

Next, the ORC-Cdc6 complex recruits the Cdt1-MCM2-7 group. Cdt1 is a key regulator of pre-RC formation and a preventer of re-replication. It is required for the recruitment of

MCM components to origins and for promoting the formation of the heterohexameric MCM complex. In yeast it was even observed binding to MCM components and being transported to the nucleus together (Tanaka & Diffley 2002; Wu et al. 2012). MCM2-7 recruitment itself is the hallmark of pre-RC formation and the key step in replication initiation. MCM2-7 form the motor domain of the replicative helicase in eukaryotes, and are therefore crucial for replication fork progression during S phase. The six MCM components (MCM2-7) assemble into a ring-like structure and attach to chromatin in a two-step process. This ‘lock and load’ mechanism first recruits the MCM complex to ORC bound origins, and then loads it onto chromatin so that the MCM ring encircles DNA. As a review on the topic states: “*The temporal separation of helicase loading and activation is crucial for the coordination of DNA replication with cell growth and extracellular signals, the prevention of re- replication and the control of origin activity in response to replication stress*” (Remus & Diffley 2009).

During this step, the preassembled MCM rings attach to the ORC-Cdc6 complex with the aid of Cdt1 (Seki & Diffley 2000). Then, taking two hexamers head to head, a dodecameric MCM complex is loaded onto DNA in an ATP dependent manner (Sclafani et al. 2004). Since replication initiation is bidirectional at most origins, it is postulated that at least two such dodecamers are loaded up at each site (Soultanas 2012). However, since loaded MCM rings can slide on DNA, multiple such complexes are able to get established at a single origin site. MCM loading, but not ORC and Cdc6 binding is abolished in the presence of ATP γ S. Because of this, it was postulated, and recently proven, that ORC ATP hydrolysis is required for MCM loading (Machida et al. 2005; Fernández-Cid et al. 2013).

Once ORC, Cdc6, Cdt1 and the MCM complexes are loaded onto DNA, the pre-RC is complete. After passing the restriction point at the end of G1, the pre-replication complex develops into the pre-initiation complex (pre-IC), and the role of ORC becomes less crucial. Recent studies even found diminishing ORC levels after MCM loading (Tsakraklides & Bell 2010; Kundu et al. 2010; Arias & Walter 2007). The recruitment of MCM10, Cdc45 and the GINS complex, coupled to CDK and DDK (Dbf4 dependent kinase) activity prepares the origin site to load the single strand binding proteins, and

finally the polymerase-primase complexes (Sheu & Stillman 2006). Upon entering S phase, the origin fires and DNA replication initiates (Bell & Dutta 2002).

Attempts have been made at reconstructing replication complexes *in vitro*. *S. cerevisiae*, the most well-studied model organism for replication, has proven to be the most successful environment for such studies. Using purified ORC, Cdt1, Cdc6 and MCM subunit, MCM complexes load onto DNA and complete pre-RCs are assembled *in vitro* (Kawasaki et al. 2006). This method is a robust tool to study pre-RC formation in yeast (Remus et al. 2009; Mehanna & Diffley 2012; Evrin et al. 2009; Heller et al. 2011). In metazoans, an assay using *Xenopus* egg extracts was also created (Gillespie et al. 2001; Gambus et al. 2011; Waga & Zembutsu 2006). This is the only vertebrate based system to date capable of creating functional pre-RCs *in vitro* in a cell free manner. Interestingly, the *Xenopus* system is even able to function if XIORC is replaced with recombinant HsORC (Vashee et al. 2003; Giordano-Coltart et al. 2005). Although a human *in vitro* replication system using intact nuclei and cytosolic extracts also exists, this system lacks flexibility due to the need for complete nuclei (Krude 2006; Baltin et al. 2006).

2.4 The Epstein-Barr virus genome as model for human DNA replication

The processes described in the previous chapter have proven to be challenging to analyze in humans. The vastness of the human genome coupled to the amazing flexibility of origin activation and the complexness of chromatin have hindered the thorough analysis of replication initiation *in vivo* (Schepers & Papior 2010). To circumvent this problem, I took advantage of the Epstein-Barr virus latent DNA replication system. This allows observing replication initiation inside human cells, but instead of the complicated host chromosomes, the more well understood and simpler EBV genome is studied.

This virus and its 163kb circular, double stranded genome have been thoroughly studied for its pathogenicity, as well as its capability to establish a life-long latent infection (Young & Rickinson 2004; Babcock et al. 1998). To achieve this EBV uses a well-studied latent program (Takacs et al. 2010). This is centered on the function of the Epstein-Barr virus nuclear antigen 1 protein (EBNA1), which has a dual role in EBV latent replication. First, EBNA1 is able to recruit the pre-RC machinery, most notably ORC, to oriP, the latent origin of replication of the viral genome. With this EBNA1 establishes a *de novo* origin of replication at oriP (Schepers et al. 2001; Ritzi et al. 2003; Norseen et al. 2008; Wang & Sugden 2005; Lindner et al. 2008; Chaudhuri et al. 2001; Dhar et al. 2001b). Secondly, EBNA1 tethers the EBV genome to the host's chromosomes, distributing the viral genome equally among the daughter cells. This "piggyback" mechanism allows stable distribution of viral genomes which have no centromeres on their own (Middleton & Sugden 1994; Smith & Sugden 2013; Marechal et al. 1999).

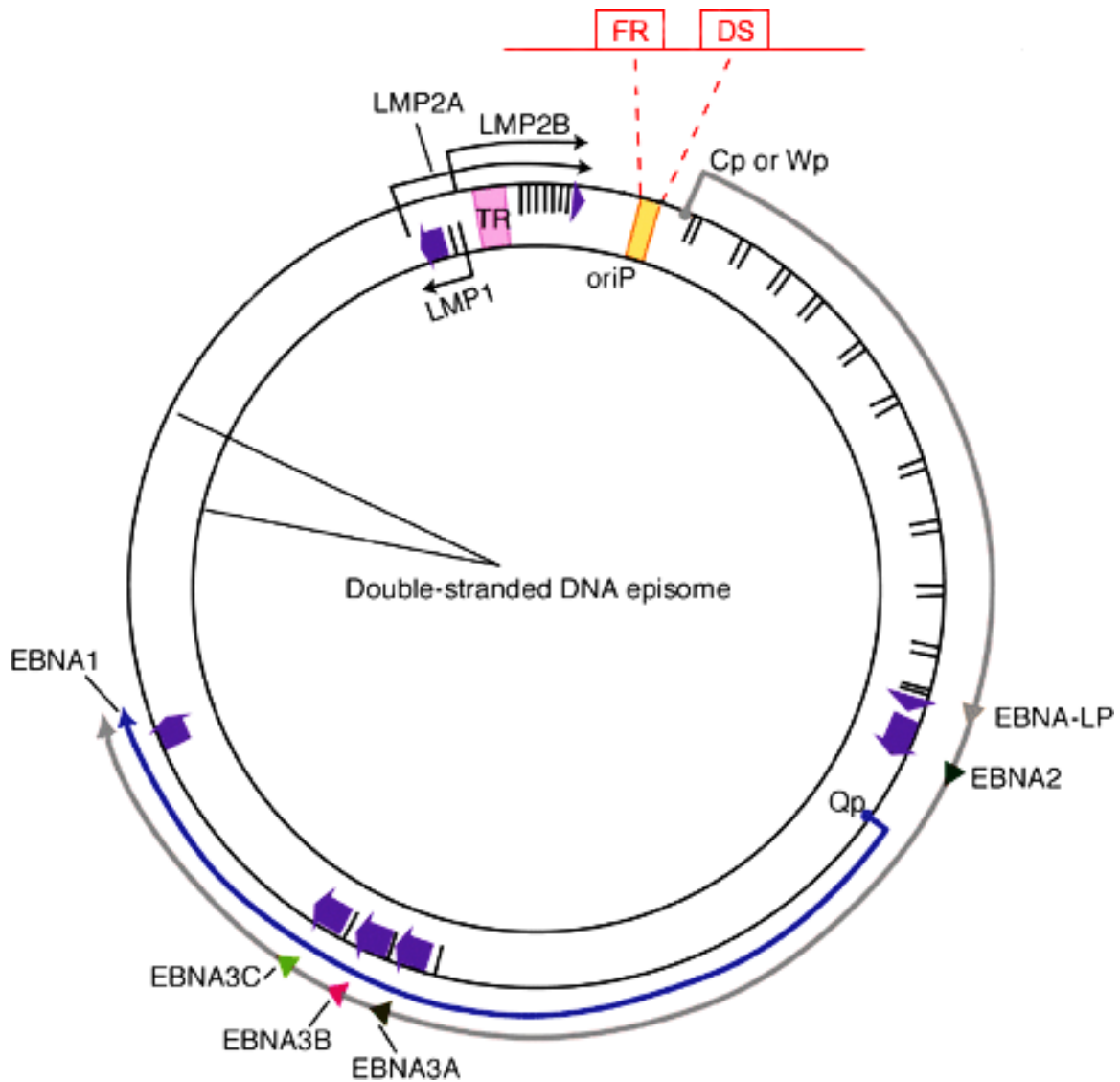


Figure 4. The Epstein-Barr virus genome. The 163kb double stranded DNA genome of the virus circularizes upon cell entry. In the latent phase, EBNA1 (blue arrow) and other latent proteins (shown in black) are expressed in the host to maintain a lifelong infection in human B cells. The latent viral origin of replication (oriP, red), responsible for the replication and retention of the latent EBV genome, consists of the dyad symmetry element (DS) and the family of repeats (FR) (image from Ravasz 2008).

Interestingly, the two duties of EBNA1 are spatially separated within oriP. The origin consists of two main regions, the family of repeats (FR) and the dyad symmetry element (DS) (Hirai & Shirakata 2001). The former is strictly responsible for recruiting EBNA1 to tether the viral genome to the host chromosomes; while the latter is used for EBNA1

dependent replication initiation, confirmed by ChIP experiments showing an accumulation of pre-RC components in G1 (Schepers et al. 2001). It is important to add that despite a prominent presence of pre-RC factors, replication initiation is distributed all over the EBV genome (Papior et al. 2012).

2.4.1 Analyzing replication competence of proteins *in vivo*: the plasmid rescue assay

The plasmid rescue assay is based on the function of oriP and EBNA1 outlined above (Thomae et al. 2011). It provides an important functional readout for this work, and is therefore presented here in detail.

To study if a protein of interest can initiate replication, it needs to be directed to the DS element of an oriP containing plasmid in a cell. If it can initiate replication, then the plasmid will be maintained through multiple generations, but if not, then the plasmid will be eventually lost from the culture. To direct proteins of interest to DS, a fusion construct is generated by attaching the protein of interest to a targeting domain, such as a single chain Tet repressor (scTetR), or a Gal4 DNA binding domain. The former specifically binds to Tet operator sites (TetO) in the genome, while the latter attaches only to GAL4 consensus sequences (Krueger et al. 2003). By replacing DS with either TetO or GAL4 sites, the fusion protein is made responsible for replication initiation, while EBNA1, losing its binding sites in DS, will only attach to FR and ensure distribution.

To carry out a plasmid rescue assay, human HEK-293 cell lines are generated first, which stably express EBNA1 and the fusion protein of interest. Also, plasmids are cloned that carry a eukaryotic and a prokaryotic selection marker, as well as either a wild-type oriP sequence or an FR-TetO construct. Next, the expression cells are transfected with either the wild-type oriP, the FR-TetO plasmid, or with water for control. After transfection the cultures are put under selection of the antibiotic to which the plasmid grants resistance. As soon as the water transfected cells die due to the toxic antibiotics, surviving cells are harvested. This is to ensure that all cells collected contain at least one copy of the transfected plasmid.

Harvested cells are then lysed using the HIRT protocol, a DNA isolation method that enriches for low molecular weight DNA, including plasmids, in the lysate. Once the DNA is isolated, it is digested with DpnI, a restriction enzyme specifically cutting non-replicated plasmids carrying the *Dam*-methylation pattern, thus removing any original plasmids introduced during transfection. Since genomic DNA will still be in abundance in the purified samples, this serves as a type of loading control. Using the same amount of DNA for subsequent steps from the individual transfections ensures that the same amount of cells was used from each sample. Therefore a uniform amount of DNA from each HIRT extract is transferred into bacteria via electroporation, after which the transformed bacteria are distributed onto agar plates. Finally, the bacterial colonies appearing on the plates are counted. This is the readout of the experiment, and it gives valuable information about the ability of the fusion proteins of interest to initiate DNA replication. If a large number of colonies are observed in the FR-TetO transfection, it means that the fusion protein efficiently recruits the replication machinery to the plasmids, and these can steadily replicate, and then get distributed via FR and EBNA1 to the next generation of cells. If a small number of colonies are observed, then the fusion protein was not able to initiate replication efficiently, and only a few plasmids were replicated and distributed upon cell division. If the fusion is completely unable to facilitate replication, then either a very small number of colonies are counted, or the cells die out along with the water control, and are considered to provide zero colonies.

Since different cell lines show slightly different growth characteristics, numbers of colonies for FR-TetO plasmids are always presented as percentage compared to wild-type oriP plasmids cultivated under similar conditions in parallel.

With the procedure established and the necessary cell lines and plasmids created, it became possible to test Orc6's ability to create functioning origins of replication *in vivo*.

3. Material and Methods

3.1 Material

3.1.1 Oligonucleotides

All the oligonucleotides listed here were synthesized by Metabion.

Name	Sequence
102 EagI_SG4_ORC6_T_forw	ATATATCGGCCGGATCGGGCGGAGGTGGCAT
103 EagI_SG4_ORC6_50_T_forw	CGGCCGTATCAGGAGGTGGTTGCCTGGACCTTGACAGCTTCCTGGATGAAG
104 ORC6_200_stop_EcoR5_T_rev	GATATCTTACACTATCTTCTTTCTCTTCCGTGGTGGAGTAGCTAC
105 ORC6_253_stop_EcoR5_T_rev	GATATCTTACTCTGCTGTAGCCTTTTGAGCACTGGCAGC
106 primer_ORC6_dNLS_forw	ACAAC TAGAGATAGTGGTTGAAGCCCCAGCAAAG
107 primer_ORC6_dNLS_rev	CAACCACTATCTCTAGTTGTTTACACAGTCGATCAAA
114 NdeI_start_SSF_EagI_forw	GAGAACATATGGACTACAAAGACGATGACGATAAATGGAGCCACCCGC AGTTCGAAAAAGGTGGTTCCGGTGGATGGAGCCACCCGCAGTTCGAAA AAGGGGGGGGCCGGCCGGT
115 NdeI_start_SSF_EagI_rev	ACCGGCCGGCCCCCCCCCTTTTTCGAACTGCGGGTGGCTCCATCCACC GGAACCACCTTTTTCGAACTGCGGGTGGCTCCATTTATCGTCATCGTCT TTGTAGTCCATATGTTCTC
132 ORC6_253_BiFC_EcoR5_T_rev	ATGATATCTTCTCTGCTGTAGCCTTTTGAG
133 ORC6_200_BiFC_EcoR5_t_rev	ATGATATCTTCACTATCTTCTTTCTCTTCCG
161 pET SSF for	TAGGATCCATGGACTACAAAGACGATGACG
162 pET SSF rev	ATCCCGGGCGGCCGGTCCCCCCCCCTTTTTCG
205 ORC6_50-125_for	AGACCAGCAGTGCAGTCATGCCCCAGACACAGCAAGTGGA
206 ORC6_125-204_for	AAAGCTATGAGTCCAGTCTTGTTGAAGCCCCAGCAAAGGA
207 ORC6_50-204_for	AGACCAGCAGTGCAGTCATGGTTGAAGCCCCAGCAAAGGA
208 ORC6_50-125_rev	TCCACTTGCTGTGTCTGGGGCATGACTGCACTGCTGGTCT
209 ORC6_125-204_rev	TCCTTTGCTGGGGCTTCAACAAGACTGGACTCATAGCTTT
210 ORC6_50-204_rev	TCCTTTGCTGGGGCTTCAACCATGACTGCACTGCTGGTCT
069-Pre-RC small oligo forw	GAGTTCCAGACTATACCCAGAGTGTCTTAATGTCTG
070-Pre-RC small oligo rev	CAGACATTAGGACACTCTGGGTATAGTCTGGAAGTCT

3.1.2 Plasmids

The plasmids used throughout this work were either created by me or obtained from the Helmholtz Centre Munich, Hämatologikum plasmid bank. Here the plasmids are stored in *E. Coli* DH5alpha cells, in LB medium supplemented with 5% glycerol and stored at -80°C. From each strain plasmid preparation was performed as detailed in chapter 3.2.13 before use.

Name	Main features
pUC19	2686bp, amp
pEPI-UPR	7211bp, amp, UPR (Schaarschmidt et al. 2004)
pJET	Clonejet kit, Thermo Scientific
p4805	pBiFC flag-orc6-1-200-yfp-1-155
p4806	pBiFC flag-orc6-50-253-yfp-1-155
p4807	pBiFC flag-orc6-50-200-yfp-1-155
p4808	pBiFC flag-orc6-S72A-K76A-yfp-1-155
p4809	pBiFC flag-orc6-T195E-yfp-1-155
p4810	pBiFC flag-orc6-dNLS-yfp-1-155
p3652	pBiFC flag-orc6-wt-yfp-1-155
p5412	pBiFC flag-orc6-d50-125-yfp-1-155
p5392	pBiFC flag-orc6-d125-203-yfp-1-155
p5393	pBiFC flag-orc6-d50-203-yfp-1-155
p3736	pBiFC ha-orc2-wt-yfp-155
p4068	pBiFC Fos
p4069	pBiFC Jun
p3725	RFP-PCNA
p3230	wt oriP Hyg
p3315	FR-4xTetO Hyg
p3243	FR only, Hyg
p5233	3230 with Gal4
p5234	3315 with Gal4

p5235	pcDNA3-zeo-Gal4:PR-SET7
p5237	pcDNA3-zeo-Gal4
p4791	pET21 Orc6-1 (AA1-200)
p4792	pET21 Orc6-2 (AA50-253)
p4793	pET21 Orc6-3 (AA50-200)
p4794	pET21 Orc6-4 (S72A K76A)
p4795	pET21 Orc6-5 (T195E)
p4796	pET21 Orc6-6 (no NLS)
p4797	pET21 Orc6-7 (wild-type, full length)
p4798	pWHE Sc-Tetr:Orc6-1 (AA1-200)
p4799	pWHE Sc-Tetr:Orc6-2 (AA50-253)
p4800	pWHE Sc-Tetr:Orc6-3 (AA50-200)
p4801	pWHE Sc-Tetr:Orc6-4 (S72A K76A)
p4802	pWHE Sc-Tetr:Orc6-5 (T195E)
p4803	pWHE Sc-Tetr:Orc6-6 (no NLS)
p4804	pWHE Sc-Tetr:Orc6-7 (wild-type, full length)

3.1.3 Antibodies

Primary antibodies

Name	Species	Dilution	Source
Orc6 3A4	rat	1:50	Kremmer, E, (Ritzi et al. 2003)
Orc1	rabbit	1:400	Santa Cruz Biotech (H-80)
Orc2	rabbit	1:1000	(Papior et al. 2012)
Orc3	rabbit	1:1000	(Schepers et al. 2001)
Orc4	mouse	1:4000	Transduction Laboratories (#83120)
Orc5	rabbit	1:4000	purified, lab stock
Cdt1	rabbit	1:0000	gift from the Knippers laboratory
Cdc6	mouse	1:400	Santa Cruz Biotech (180.2)

Material and Methods

Mcm2	rabbit	1:1000	purified, lab stock
Mcm3	rabbit	1:1000	(Ritzi et al. 2003)
Mcm5	rat	1:1	3F3, 2F3 Kremmer, E
Gal4	rabbit	1:400	Santa Cruz Biotech (sc-577)
TetR	rabbit	1:2000	gift from the Behrens laboratory (SA-1851)
Pr-Set7	rabbit	1:2000	Julien, E (Cell Signaling C18B7)
HA	rat	1:50	3F10 Kremmer, E
FLAG	mouse	1:1000	Sigma M2
EBNA1	rat	1:50	1H4 Kremmer, E (Grässer et al. 1994)

Secondary antibodies

Name	Species	Dilution	Source
anti-rat-HRP	goat	1:10000	BD-biosciences
anti-rabbit-HRP	goat	1:10000	Promega
anti-mouse-HRP	goat	1:10000	Promega
anti-rat-cy3	goat	1:200	Life Technologies
anti-mouse-cy5	goat	1:200	Life Technologies

3.1.4 Bacterial strains

Name	Source
E. coli DH5alpha	lab stock (Hanahan 1983)
E.coli DH10B (electromax)	Life Technologies
E. coli Rosetta pLysS	lab stock

3.1.5 Cell lines and culture material

The following cell lines were taken from the Helmholtz Centre Munich cell culture stocks: RAJI, HeLa, HeLa S3, HepG2, HEK-293-D, HEK-293 EBNA1⁺.

All other cell lines used throughout this work were established by myself as detailed in chapter 3.2.2, namely: HEK-293 EBNA1⁺ wt Orc6⁺, HEK-293 EBNA1⁺ ΔC Orc6⁺, HEK-293 EBNA1⁺ ΔN Orc6⁺, HEK-293 EBNA1⁺ ΔCN Orc6⁺, HEK-293 EBNA1⁺ S72A K76A Orc6⁺, HEK-293 EBNA1⁺ T195E Orc6⁺, HEK-293 EBNA1⁺ ΔNLS Orc6⁺, HEK-293 EBNA1⁺ Orc6 PR-Set7⁺.

Name	Source
Culture dishes (bottles, six well, 6cm, 15cm)	Sigma-Aldrich (Nunc)
Cryotubes	Sigma-Aldrich (Nunc)
DMEM medium	Life Technologies
Fetal Calf Serum (FCS)	Life Technologies
HEPES	Life Technologies
Hygromycin	Life Technologies
Lipofectamine	Life Technologies
Neomycin (G418)	Life Technologies
Opti-MEM medium	Life Technologies
Pen-Strep	Life Technologies
RPMI 1640 medium	Life Technologies
Trypsin-EDTA	Life Technologies
Zeocine	Life Technologies

3.1.6 Software

Function	Provider
Cell line database	Filemaker Pro
DNA sequence analysis and in silico cloning	Macvector
Lab books	Google docs, MS Office
Functional clustering	DAVID, STRING
Microscopic image processing, image quantification	ImageJ
Proteomics	Scaffold 4
Sequence alignment	BLAST

3.1.7 Equipment and other material

Material not listed here was procured from Merck. Chemicals were regularly of *pro analysi* grade.

Material	Provider
Avanti J10, J25 centrifuge, L7-55 Ultracentrifuge	Beckman Coulter
Gene-Pulser II, 10ml columns	Bio-Rad
Ultrasound sonifier	Branson
PAA, SDS, Phenol, Spectra-Por 6 Dialysis membrane	Carl Roth
Bacto-Agar, Yeast extract, Trypton	Difco
Reaction tubes, Table top centrifuge 5415,	Eppendorf
Spectrophotometer, PCR machine Mastercycler personal	
Cover slip, Whatman paper, X-ray film	Hartenstein
JetStar Maxiprep kit	Genomed
Hybond ECL membrane, Protein A Sepharose	GE healthcare
Semi-Dry Blotting System	Hoefer
Strep-Tactin protein purification kit	IBA
Benchmark protein ladder, not prestained, dynabeads,	Life Technologies
Qubit fluorometer, Colloidal Blue staining kit	

Material and Methods

Leica TCS SP5 Confocal-microscope	Leica
Nucleospin Gel and PCR clean-up kit	Macherey-Nagel
Restriction enzymes and buffers, Pwo polymerase,	New England Biolabs
Electrophoresis chamber, electroporation cuvette	Peqlab
Qiaquick PCR purification kit	Qiagen
dNTPs, Tris, DNase I, RNase A, cOmplete (EDTA free)	Roche
Ampicillin, Bromophenol blue, DMSO, DTT, Glycine,	Sigma-Aldrich
HEPES, Triton X-100, Tween-20, BSA, Benzonase, MG132	
Spectra-Por dialysis membrane	Spectrum Labs
Robocycler	Stratagene
Generuler 1kb DNA ladder	Thermo Fisher Scientific
7ml Douncer	Wheaton
Vectashield, Photoprobe (S-S) biotin	Vector Labs
Axiovert 10 fluorescence microscope	Zeiss

3.2 Methods

3.2.1 Cell cultivation

All human cells used in this study were cultivated in incubators set at 37°C and 5% CO₂.

HeLa

I raised adherent HeLa cells in RPMI medium with 10% FCS and 1% Pen-Strep in 15cm culture dishes. I monitored the density of the cells and at 80% confluency I diluted the culture fivefold. To achieve this, I removed the medium, and washed the cells once in sterile PBS. Afterwards, I applied 2ml trypsin (0.05% Trypsin-EDTA) solution to the cells until they detached from the surface or for maximum 5 minutes. Next, I added 2ml medium and resuspended the cells using a pipette. From the resulting 4-5ml solution, I added 20% to a new plate and supplemented it with 20ml fresh medium.

HeLa S3

Suspension HeLa cells (HeLa S3) were cultivated in 1L roller bottles, with conditions similar to semi-adherent cells. Density was measured via using a Neubauer cell counting chamber and kept between $2-8 \times 10^5$ cells / ml.

HEK-293 and HepG2

I cultured HEK-293 and HepG2 cells and their derivatives in DMEM medium with 10% FCS and 1% Pen-Strep in 15cm culture dishes. I diluted the cultures similarly to HeLa cells. I cultured HEK-293 EBNA1⁺ cells using 220ng/μl G418, HEK-293 EBNA1⁺ Orc6⁺ cells using 300ng/ml puromycine and HEK-293 EBNA1⁺ Gal4⁺ (PR-Set7⁺) constructs with 10μg/ml zeocine.

3.2.2 Transfection of human cells

For creating HEK-293 EBNA1⁺ cells expressing Orc6 variants, Gal4, and PR-Set7, I transfected cells as follows. The day before transfection I seeded 1.5×10^5 cells into each well of a six well plate. The next day, I exchanged the medium to FCS free medium before transfection. Meanwhile, I digested 3 μ g of the expression plasmid using a restriction digest to linearize the plasmid, and supplemented it with opti-MEM medium to achieve a 50 μ l / sample volume. I added Lipofectamine transfer reagent (cat. no. 11668-027, 2 μ l/ μ g plasmid) to opti-MEM medium in a separate container in 50 μ l / sample volume and left the solutions standing on room temperature for 5 minutes. Afterwards I mixed the lipofect solution with the plasmid solution to reach 100 μ l / sample volume and lipid spheres encompassing linear plasmids were left forming for 20 minutes room temperature. Next, I added the ready solution in a careful drop-by-drop manner to the seeded cells and then left them for 4 hours in the incubator for the transfection process to complete. Following this, I added fresh medium to the cells and they were left in the incubator overnight. The next day I introduced selection pressure as presented in the previous chapter. After single-cell colonies formed on the plates under selection pressure (2-3 weeks), I picked the colonies using a piece of sterile paper soaked in trypsin and transferred them to six well plates. I tested the cultures resulting from the single cell colonies for expression by making RIPA extracts, and using these for immunoblotting.

3.2.3 Plasmid-rescue assay

I performed plasmid rescue assays using HEK-293 EBNA1⁺ cells and their derivatives. First, I transfected the cells as described in the previous chapter until the point where they were placed under selection using 80 μ g/ml Hygromycin. Hygromycin selection seemed more efficient after the first splitting of cells, and was best when cells were kept at low confluency. I kept the cells in logarithmic growth phase until the water transfected control died out under selection (approx. 2 weeks). Then, I washed full plates of the transfected cells once with TEN buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 150mM NaCl) and lysed them by adding 1.5ml TEN and 1.5ml 2x HIRT lysis buffer (1.2% SDS,

20mM Tris pH 7.5, 20mM EDTA). The resulting lysates were collected using a glass slide into a centrifuge tube and supplemented with 750µl 5M NaCl. The samples were then purified via phenol-chloroform extraction (Chomczynski & Sacchi 1987). I subjected the clean DNA samples to Dpn1 digest by adding 11µl NEB4 restriction enzyme buffer, 3µl Dpn1 (20U/µl), and 3µl RNase and incubated them for 2 hours at 37°C to remove RNA and unreplicated plasmid DNA. This was followed by an ethanol precipitation by adding 20µl 3.5M NaAc, 400µl 100% Ethanol, incubating for at least 1 hour at -20°C, centrifuging 1 hour at 4°C, 16K g, discarding the supernatant and washing once with 500µl 70% ethanol, centrifuging once more for 15 minutes at 4°C, 16K g, then discarding the supernatant, air drying and finally resuspending in 50µl dH₂O.

Then, I assayed the ready, clean DNA samples containing the rescued plasmids using the Qubit fluorometer's broad range DNA measurement kit according to the manufacturer's instructions.

From the measured samples, I introduced 100ng DNA into Electromax DH10B bacteria via electroporation with a Bio-Rad Genepulser with a resistance of 200 Ω, capacitance of 25µF, voltage set to 2.5 kV and using the manufacturer's instructions.

Transformed bacteria were plated onto agar plates with 100µg/ml ampicillin and stored overnight in an incubator at 37°C. Next day I counted the colonies on each plate.

3.2.4 Fluorescence Microscopy

To visualize the location of specific proteins in the cell, I employed fluorescence microscopy. For this, HepG2 cells were transfected similarly to HEK-293 cells as described in chapter 3.2.2, but instead of selecting the cells for 3 weeks, I carried out the following protocol after 36 hours of expression.

I washed the cells twice with PBS and fixed them with 2% Formaldehyde in PBS for 10 minutes at room temperature. Then, I carefully washed them twice with PBS, placed them on ice and gently lysed them with 0.5% Triton X-100 in PBS for 5 minutes. Blocking was performed by adding 1% BSA and 0.15% Glycine in PBS for 10 minutes at room temperature. Afterwards, I applied 80µl primary antibody in 2% BSA in PBS (rat anti-HA 1:10 for Orc2 and mouse anti-FLAG 1:300 for Orc6) to the sample and

incubated it for 1 hour at room temperature; or overnight at 4°C. After this I washed the cells twice for 5 minutes with PBS. Next, I added 80µl of the secondary antibody (anti-rat Cy3 and anti-mouse Cy5) in a 1:200 dilution using 2% BSA solved in PBS and incubated it 1 hour at room temperature. Following this, I performed DAPI staining by adding 80µl of 1:20 000 diluted 5mg/ml DAPI in PBS to the sample and incubating it for 15 minutes at room temperature. After three washes with PBS, I placed the cover slips containing the prepared cells face-down in 8µl Vectashield, and sealed them with nail polish.

3.2.5 Protein extract preparation

RIPA Extract

This protocol was used for fast, complete lysis of cells. First, I removed medium from an 80% confluent 15cm culture dish containing approximately 2×10^6 cells and washed them in PBS. After that the cells were detached from the plate by adding 2ml trypsin which I neutralized with 2ml medium once the process was complete. I then transferred the solution containing the cells to a 15ml Falcon tube and centrifuged it for 10 minutes at 1000rpm, 4°C. After removing the supernatant, I resuspended the pellet in 400µl RIPA buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% SDS, 0.5% DOC, 1% NP-40 with 1x cOmplete-EDTA free). After transferring the solution to a 1.5ml Eppendorf tube, I added 2ul Benzonase followed by incubation for 5 minutes at room temperature, to cleave long DNA stretches in the sample.

High salt extract

This protocol enriches for nuclear proteins and most factors can be renatured after extraction. Amounts in the protocol are calculated for 1×10^8 cells.

Culture plates are each first trypsinized by adding 2ml trypsin and neutralizing that with 2ml medium after the cells detached. Then, cells are centrifuged at 200g for 1 minute at room temperature. In case of HeLa S3 cells, the trypsinization is not required and the culture is centrifuged directly. After this, the pellet is washed once with PBS and once with ice-cold hypotonic buffer (10mM HEPES pH 7.9, 10mM KAc, 1.5mM MgAc2, 1mM ATP, 1mM DTT, cOmplete – EDTA free, 10µM MG132), then resuspended in 5ml

hypotonic buffer (HB). Next, the solution is placed into a 7ml douncer using the tight fit piston, and homogenized 10 times. The solution is assayed afterwards under a microscope and the percentage of nuclei/all cells is estimated, yielding typically 60-90% nuclei. The lysate is then centrifuged 10min 16k g 4°C to separate nuclei from the cytosolic fraction. After removing the cytosolic fraction, the nuclei are treated with one third volume high salt buffer (1800mM KAc, 10mM HEPES pH 7.9, 1.5mM MgAc₂, 1mM ATP, 1mM DTT, cOmplete-EDTA free), and incubated for 1 hour on ice for protein extraction. The extract is then centrifuged 80min 55 000rpm at 4°C in a TLA 100.3 rotor in thick wall tubes to remove any cell debris from solution. The soluble fraction is then dialyzed for minimum 3 hours against dialysis buffer (10mM HEPES pH 7.9, 100mM KAc, 1.5mM MgAc₂, 0.5mM DTT, cOmplete-EDTA free) using the Spectra-Por MWCO 6-8000 dialysis membrane. The dialyzed protein extract is then supplemented with glycerol to an end concentration of 10%, and finally shock-frozen in liquid nitrogen and stored at -80°C.

Co-IP extract

This extract preparation protocol offers quick cell lysis while maintaining some protein-protein interactions. In order to achieve this, I prepared the cells similarly to RIPA extracts, but instead of RIPA buffer, I used Co-IP lysis buffer (20mM Tris pH8.0, 150mM NaCl, 10% glycerol, 0.5% NP-40, cOmplete-EDTA free, 10µM MG132).

3.2.6 Western blot

I added 5x Laemmli buffer (Laemmli 1970), to a final 1x dilution to each sample prior to starting a Western blot. To denature proteins, I incubated the samples for 5 minutes in a sand bath at 95°C. After cooling on ice, they were kept at room temperature while the gel was being assembled.

To cast a gel, I first washed with deionized water and ethanol, and then assembled two glass plates separated by plastic spacers, and placed them in a film of warm, 0.8% agarose solution. I let the agarose solidify to seal the bottom of the glass container. After this, the polyacrylamide gel material was prepared. Unless otherwise stated I used 11%

polyacrylamide to buffer ratio, which is still rigid enough to be handled by hand, but also allows good separation of proteins between 30-50 kilodaltons. For an end volume of 15ml I added together 5.5ml PAA (Polyacrylamide; 33% stock, Rotiphorese gel 30), 3ml 5x TG buffer (1.875M Tris-Base, 0.5% SDS, pH 8.8), 6.35ml dH₂O, 0.03ml TEMED (Tetramethylethylenediamine), and 0.125ml 10% APS (Ammonium Peroxodisulphat). After adding together these components, I poured the gels quickly into the cast, and layered 1ml isopropanol on top to create an even surface. After the gel solidified, I decanted the isopropanol and filled the cast with a concentration gel consisting of 0.66ml PAA, 2.5ml SG buffer (0.25M Tris-Base, 0.2% SDS, pH 6.8), 1.8ml dH₂O, 0.01ml TEMED, 0.05ml APS, and a pinch of Bromophenol blue. The comb was placed into this solution and was left to solidify. The collection gel allows sample components to enter the gel together and create a sharp running front. This is divided into protein components during the electrophoresis and yields sharp, easily identifiable bands.

Once the gel was ready, it was placed into the running chamber which in turn was filled with running buffer (192mM Glycine, 25mM Tris-Base, 0.1% SDS, pH9.0 (textbooks suggest pH8.3, but with pH 9.0 the running front is better focused and the gel runs faster)). After loading the samples, I performed the electrophoresis at 300V, 35mA, until the running front reached the bottom of the gel. After this I disassembled the chamber and placed the gel into a transfer sandwich (from top to bottom: 2 pieces of whatman paper – gel – membrane (Amersham Hybond ECL from GE healthcare) – 2 pieces of whatman paper) and transferred it in a semi-dry transfer unit for 1H at 15V, 400mA using running buffer with 20% MeOH. After transfer I washed the membrane in water and PBST (PBS with 0.1% Tween-20), and blocked it for 30 minutes in 5% milk solution. I applied the primary antibody for minimum 2 hours at room temperature, or 4°C overnight with the dilution stated Chapter 3.1.3. Following incubation I washed the membrane three times in PBST and applied the secondary antibody in 2.5% milk solution for 1 hour at room temperature. Next, I washed the membrane twice in PBST, once in PBS and once in H₂O and placed it in a transparent plastic bag. For the ECL reagent, I added 10ml of solution A (200ml end volume, with 340mg (200mM) p-Coumaric acid (sigma# C9008-1G) in 500µl DMSO; 2.26g (1.25mM) luminol (Fluka# 09253) in 1ml DMSO; filled with Tris pH 8.9) and 60µl solution B (3% H₂O₂ (Sigma# 21676-3) in dH₂O) evenly on the

membrane. After 2 minutes of incubation, images were made in a dark room using CEA medical x-ray screen film blue sensitive, EC84A films. Quantification of the images was done using the ImageJ software's built in gel quantification option with the recommended settings, making sure that the exposure of the films was in the linear range, and that the scanning introduced no compression to the images.

3.2.7 Plasmid-binding assay

Plasmid-binding assays were used to study protein-DNA interactions *in vitro* as seen in Figure 7. First, I coupled biotin to the plasmids by diluting the stock photoprobe (S-S) biotin solution of 1mg/ml 1:10 in dH₂O. This I then added to 50 pmol of plasmid to achieve a 1:18 plasmid:biotin ratio and filled it to an end volume of 200µl with dH₂O. Then, I irradiated the solution with a 365nm UV lamp (4W) from a 2cm distance on ice for 30 minutes. After that, I added 400µl Tris pH9.5 (1M) and 800µl 2-butanol and mixed it with a table-top vortex vigorously. I have done a round of centrifugation for 16000g, 5 minutes at 4°C in order to separate the butanol phase with the excess biotin from the DNA-water phase. After a second 2-butanol wash I performed an ethanol precipitation (Sambrook & Russel 2001). I solved the clean, biotin-coupled DNA in 50µl dH₂O and added it to 100µl streptavidin coated magnetic beads (Dynabeads), which I prewashed once with the binding buffer from the Dynabead kit. I added 400µl binding buffer and 50µl dH₂O to the bead-DNA mix and placed the complete solution on a roller overnight at room temperature. The next day I washed it three times with 100µl washing buffer and stored it in 100µl washing buffer at 4°C.

Afterwards, I washed 10µl from the immobilized DNA twice in RB buffer (20mM HEPES pH 7.9, 5mM KAc, 1.5mM MgAc₂, 100µM DTT, 0.003% NP-40, cOmplete-EDTA free) and added 34µl RB buffer, ATP (40µM), purified proteins (if present) and 6µl of HeLa commercial nuclear extract (if not stated otherwise) to the beads and placed them on a heating block/shaker at 1200rpm for 30 minutes at room temperature. After this I washed the beads three times with 100µl RB while incubating the beads for 5 minutes on the shaker at 1200rpm at 4°C. Finally, I added 20µl laemmli buffer to the beads and continued with the Western blotting protocol.

3.2.8 Protein expression and purification

For protein expression in bacteria, the following protocol was used. I transformed ROSETTA2 DE3 pLYS S bacteria with pET vectors containing the gene of interest, by performing the bacterial transformation protocol detailed in chapter 3.2.12. For the agar plates I added chloramphenicol to an end concentration of 25µg/ml. Using a freshly picked colony from the transformation, I inoculated 3ml of LB medium with 100µg/ml ampicillin, and placed it into an incubator overnight at 37°C. The next day I used 750µl from this starter culture to inoculate 250ml of ZYM 5052 medium. 1L ZYM 5052 medium consists of 958ml ZY (1% tryptone, 0.5% yeast extract, dissolved in dH₂O), 20ml 50xM (1.25M Na₂HPO₄·7H₂O, 1.25M KH₂PO₄, 2.5M NH₄Cl, 0.25M Na₂SO₄ dissolved in 700ml dH₂O at 42°C), 20ml 50x5052 (25% glycerol, 2.5% sucrose, 10% lactose), 2ml 1M MgSO₄, 0.2ml 1000x trace elements (50mM FeCl₃, 20mM CaCl₂, 10mM MnCl₂, 10mM ZnSO₄, 2mM CoCl₂, 2mM CuCl₂, 2mM NiCl₂, 2mM Na₂MoO₄, 2mM Na₂SeO₃, 2mM H₃BO₃ in 60mM HCl) and 50µg/ml ampicillin.

I incubated the culture for 3 hours at 37°C with 200rpm shaking, and when OD₆₀₀ reached 1-2, I placed the culture to 20°C overnight. The next day I collected the cells when the culture stopped growing and centrifuged them at 3000rpm, 4°C for 10min in a JA10 rotor. Afterwards I resuspended the pellet in 20ml lysis buffer (50mM NaHPO₄, 300mM NaCl, 10mM imidazole, pH 8.0 with NaOH), added 1mg/ml lysozyme and incubated for 20 minutes on ice. Then I supplemented the solution with 10µg/ml RNase A and 5µg/ml DNase 1 and incubated it again for 10 minutes more on ice. Then I fractured the cells via sonication using an ultrasound tip for 10x10s with 40% strength on ice, followed by a centrifugation at 10000g for 1 hour at 4°C. Finally, I purified the supernatant fraction containing the soluble proteins using the Strep-tactin purification kit according to the manufacturer's instructions.

3.2.9 Immunoprecipitation

I performed immunoprecipitations on ice, or at 4°C. First, I prepared low-bind Eppendorf tubes and pipetted 5-30µg antibody along with 50µl protein extract and placed this on an

over-head rotator for 1 hour. Meanwhile I washed 20µl protein G sepharose twice with PBS. After the incubation I added the extract-antibody mix to the beads and allowed it to rotate for 1 hour more. Finally, I spun the solution down at 200g for 1min at 4°C and added 20µl 2x laemmli buffer to it.

3.2.10 Pull-down

Pull-downs were performed using the Strep-Tactin kit from IBA Lifesciences according to the manufacturer's instructions.

3.2.11 Polymerase chain reaction

PCR was performed to rapidly amplify DNA stretches for cloning (Mullis et al. 1986). For each reaction I used the Pwo DNA polymerase for high fidelity amplification, and the primers listed in chapter 3.1.1. Purification of PCR products was done by agarose gel electrophoresis as discussed in chapter 3.2.14.

3.2.12 Bacterial transformation

Transformation of *E. coli* bacteria was performed to introduce plasmids into the cells for DNA or protein production. Electroporation was only performed for plasmid rescue assays, and it is detailed there. Here, the heat-shock method is presented using competent *E. coli* cells generated according to the Inoue protocol (Sambrook & Russell 2001).

First I added 10ng plasmid DNA to a 15ml Falcon tube while I defrosted the competent cells at 37°C and then incubated them on ice for 10 minutes. Next, I added 50µl competent cell solution to the plasmids and incubated them on ice for 30 minutes. This was followed by the heat shock at 42°C for 90s, after which I put the solution back on ice for 12 minutes. Then I added 800µl LB medium to the cells and incubated them at 37°C for 45 minutes. Finally, I centrifuged the cells for 5 minutes at 3000rpm, decanted the supernatant and transferred the pellet in 200µl LB to an agar plate carrying the required

selection marker. After an overnight incubation at 37°C, I picked single cell colonies from the plate.

3.2.13 Plasmid preparation

Mini- and maxipreps were employed to isolate plasmids from *E. coli* cultures. For this, I used the JETSTAR plasmid preparation kit according to the manufacturer's instructions.

3.2.14 Agarose gel electrophoresis

To isolate DNA fragments from multiple products and to monitor plasmid integrity, I performed agarose gel electrophoresis using 0.8% agarose gels (TAE) supplemented with 1µg/ml ethidium bromide. For visualizing DNA, I used a 365nm UV lamp. For preparative work, I took advantage of a 254nm wavelength UV lamp which reduces damage to DNA. Products from the extracted gel pieces were cleaned using the Nucleospin-Extract II kit according to the manufacturer's instructions.

3.2.15 Mass spectrometry

For mass spectrometry, immunoprecipitation was performed as described in chapter 3.2.9. To reduce risk of keratin contamination I prepared samples on ice, under a fume hood and handled them with sterile gloves and sealed lab coat sleeves. For the experiment, I used 200µl of HeLa commercial nuclear extract and 100µg of antibody, the latter being covalently coupled to 100µl sepharose G beads as described elsewhere (Sambrook & Russel 2001). I loaded the samples into an electrophoresis chamber assembled in the fume hood, with gel material and glass sheets similar to those described in chapter 3.2.6 but using material specifically prepared for mass spectrometry and cleaned with ddH₂O. Also, I prepared the gel without a concentration gel on top, only the separation solution was used. I stopped the electrophoresis once the blue front reached 2cm into the separation gel. Then, I used the Colloidal Blue staining kit from Life Technologies, according to the manufacturer's instructions to stain the gel. Next, I cut out

the entire lane containing the proteins in five separate pieces, using new, sterile scalpels for each cut. I placed each gel piece into a PCR tube containing 200µl ddH₂O. The samples were then handled by the Mass Spectrometry Core facility of the Biomedical Center at the Ludwig-Maximilians-University of Munich, as described elsewhere (Schreiner et al. 2012). I then took the identified, Mascot searched spectra, and using the Scaffold proteomic software performed statistical analysis using Fisher's exact test to identify any proteins enriched over background (Zhang et al. 2006). Functional clustering was done using DAVID and STRING, with the recommended settings.

4. Aim of the Study

The aim of this study is to better understand the role of the Orc6 protein in the human cell. HsOrc6 is the least studied component of the human origin recognition complex, and its role is obscure despite numerous studies exploring DNA replication initiation. ORC is highly conserved among eukaryotes, therefore studies in model organisms provide much needed insight into the function of the complex. However, Orc6 itself is the fastest evolving subunit, which results in contradictory observations in model organisms that are often difficult to interpret in a human setting. Because of this, validation of these results in humans and further elucidation of HsOrc6 function is a much-needed addition to the current model of DNA replication initiation.

To better understand the place of HsOrc6 in ORC, first a mechanical study shall be conducted elucidating how Orc6 interacts with the pre-replication machinery. Apart from the mechanism of Orc6-Orc1-5 attachment and localization, the DNA binding of ORC and the role of Orc6 in this process shall be explored in detail.

Secondly, a functional analysis of Orc6 shall be conducted in live cells to elucidate how this factor contributes to replication initiation. There is prior evidence that Orc6 is able to recruit the replication machinery to designated sites on chromatin, and this function shall be mapped down to the domain level of this subunit.

Lastly, I shall follow up mounting evidence that Orc6 is involved in cellular processes other than DNA replication. By analyzing interaction partners of HsOrc6, I shall identify new pathways the protein is involved in to give a more complete picture of the function of Orc6 in the human cell.

5. Results

5.1 Characterization of Orc6 interaction with the replication machinery

5.1.1 Orc6 is present in abundance compared to other ORC subunits

To characterize how Orc6 interacts with the other ORC subunits, I first needed a better understanding of how ORC proteins coexist in the cell. Knowing the amounts of individual ORC components present is crucial in understanding complex formation. Therefore I first quantified the amount of Orc2 and Orc6 subunits in the cell. It has been shown by others that Orc1-6 is more abundant in transformed than normal cells (Di Paola & Zannis-Hadjopoulos 2012) and that Orc1-5 levels are higher in proliferating than in resting cells (Thome et al. 2000). Orc2, Orc4 and Orc5 levels have even been quantified in CHO cells (Wong et al. 2011). However, it was not previously known how Orc6 is expressed compared to the other subunits. To examine if Orc6 is expressed to similar levels as the other ORC components, I have quantified the level of Orc6 and Orc2 in two widely used cell lines: the HEK-293-D kidney cancer cell line, and Raji, an EBV positive B cell line. Immunoblot signal intensities of whole cell extracts and highly purified Orc2 and Orc6 proteins were employed for quantification.

First, I loaded the purified proteins onto Coomassie gels, and analyzed them for purity. Next, I assayed the concentration of these samples using the Bradford method (data not shown). Meanwhile, I also prepared RIPA extracts from HEK-293-D and Raji cells, ensuring that the cells are completely solubilized during the procedure by repeated agitation with a pipette. Subsequently, I loaded dilutions of 250x to 4000x of the purified

proteins onto a polyacrylamide gel, with three dilutions from the RIPA extracts next to them. The antibody staining after the blotting procedure (detailed in chapter 3.2.6) was done with either a purified rabbit antibody for Orc2 or the rat monoclonal antibody 3A4 for Orc6. I took special care to ensure that while films were exposed to the chemiluminescent signals of the blots, the signals stayed in the linear range of the sensitivity of the films used.

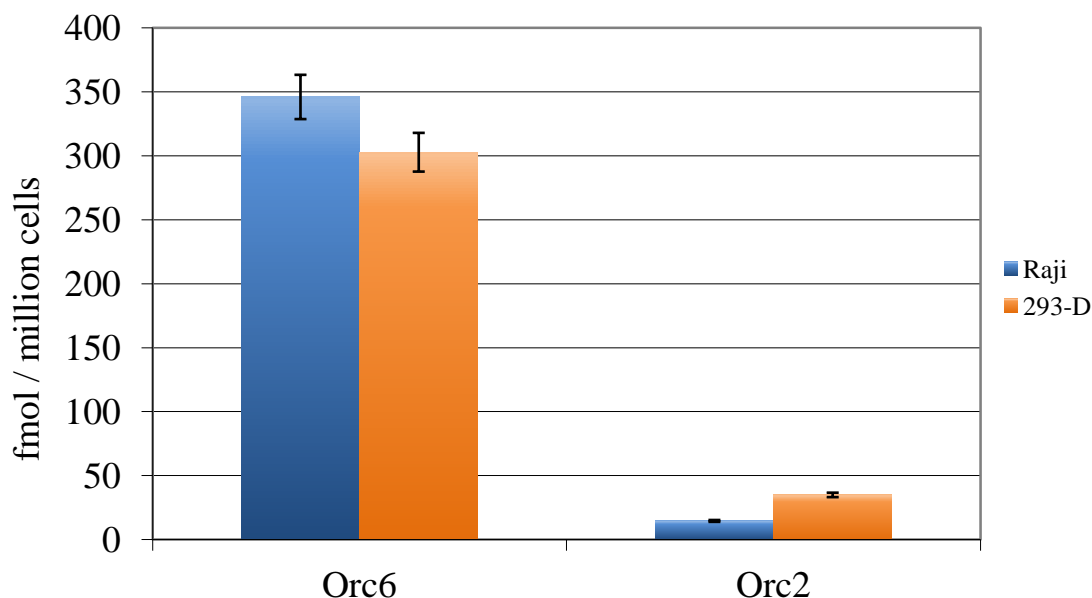


Figure 5. Amounts of Orc2 and Orc6 in HEK-293-D and Raji cells. Whole cell extracts were prepared from either Raji or HEK-293-D cells, and immunoblotting was carried out to visualize Orc6 and Orc2 signals. Protein levels were quantified by comparing these signal intensities to a dilution series of purified proteins. The average of three independent experiments are presented, the error bars show standard deviation.

I scanned the films displaying the signals of the purified proteins and the assayed cell extracts in high resolution and without image compression to ensure that the signals remain unchanged while digitalized. Finally, I analyzed the image files using ImageJ's gel quantification option. A line was fitted to the values obtained from the purified proteins, which created a curve capable of translating immunoblot signal intensities to protein concentration. The signals observed in the extract lanes were compared to the curves, while making sure that each extract signal was only compared to its unique

purified protein dilution curve, excluding inaccuracies from different exposure times and possible differences in gel handling. Immunoblotting was performed in triplicates, and from each membrane three films were developed, scanned, quantified, and averaged. Similar setups have been successfully used to quantify ORC component levels in yeast, *Xenopus*, and hamster cells (Wong et al. 2011).

Results in Figure 5 display Orc2 and Orc6 amounts and standard deviations in HEK-293-D cells in blue, and in Raji in orange. The two cell lines show surprisingly similar amounts of protein for both ORC components, with HEK-293-D cells containing about twice as much Orc2 than Raji cells. Orc6 levels are even more similar in both cell lines, and it is also apparent that Orc6 is expressed to considerably higher amounts in the assayed cells than Orc2.

Despite the abundance of Orc6 in the cell, its involvement in ORC formation is poorly understood. Therefore, to get a better understanding of this important step in replication initiation, I carried out a more detailed characterization of the binding mechanism of Orc6 to Orc1-5.

5.1.2 Orc6 interacts with Orc1-5 and Cdc6 in solution

From the pioneer experiments done in yeast, it was originally considered that ORC exists solely as a six-subunit complex in the cell. More recent work however has found that this might be a yeast specific observation despite the highly conserved subunits in all eukaryotes. The first experiments using human cells only found Orc1-5, with a separate Orc6 protein in solution (Dutta & Dhar 2000). Later it was established that purified human ORC lacked its smallest subunit, but the interaction between Orc1-5 and Orc6 was observable using human proteins expressed in insect cells (Vashee et al. 2001). The Stillman laboratory later confirmed that Orc3 and 5 are able to co-precipitate Orc6 in HEK cell nuclear extracts, but they showed no evidence of Orc6 being able to co-precipitate other ORC subunits, or Cdc6 (Siddiqui & Stillman 2007). Finally the DePamphilis laboratory was the first to show that an intact Orc1-6 complex exists in human cell extracts if they overexpress at least one of the subunits (Ghosh et al. 2011).

To continue with these experiments, I performed an Orc6 immunoprecipitation using endogenous proteins only, to show that Orc6 is able to interact with Orc1-5 in a nuclear extract with endogenous proteins, and also that the complete origin recognition complex, along with Cdc6, is assembled in solution. I performed the immunoprecipitation with established Orc6, Orc2 and Cdc6 antibodies (Ritzi et al. 2003), which I used to directly precipitate these proteins from HeLa extracts. As a control, a nonspecific IgG rat antibody was used, which was expected to provide no signal.

Results presented in Figure 6 show that Orc6 is not only able to interact with Orc1-5, but also with Cdc6. This finding is confirmed by the presented Orc2 and Cdc6 co-immunoprecipitations, which depict similar results. Taken together, these data show that the ORC-Cdc6 complex forms in solution and can be immunoprecipitated stably from human nuclear extracts. Orc3 and Orc5 are not shown in this figure since Orc2-5 has been already presented to assemble in solution and to form a stable complex. Therefore I propose that these factors are also present. The high background observable in the Orc1 input and IgG lanes are due to the fact the Orc1 is unstable in HeLa nuclear extracts, and is present in sub-stoichiometric amounts (Baltin et al. 2006). Because of this, detection of the protein requires long exposure times, which lead to an increased background.

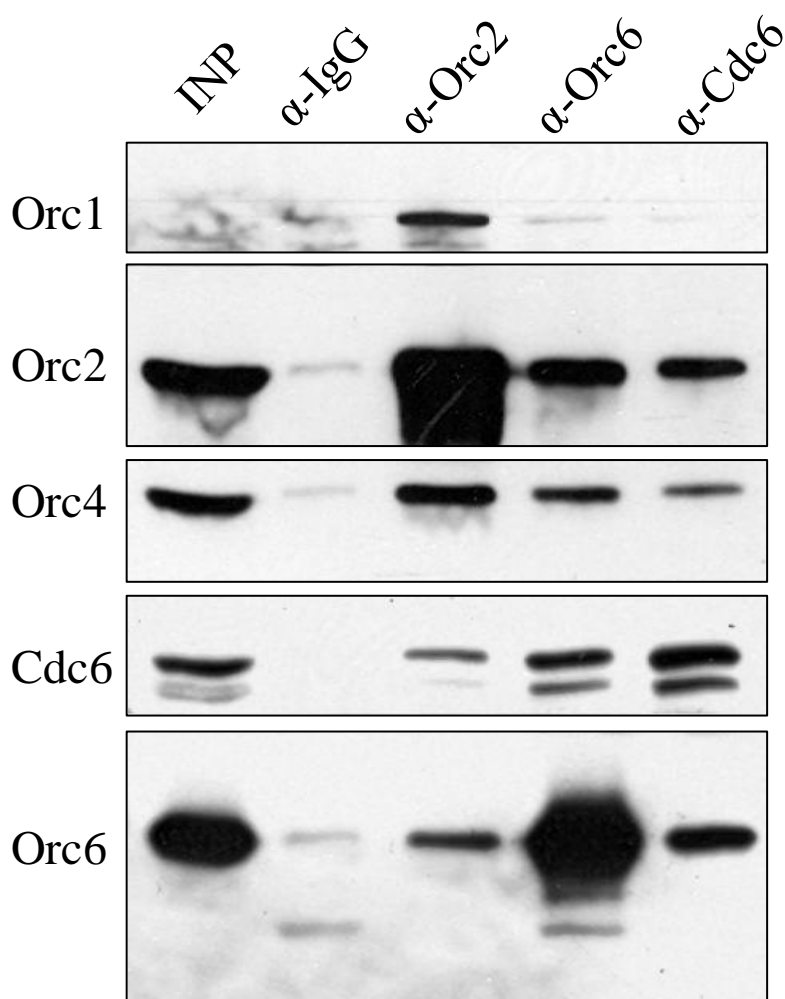


Figure 6. Co-immunoprecipitation of pre-RC subunits from nuclear extracts. The ORC-Cdc6 complex was precipitated from HeLa nuclear extracts using anti-Orc2, Orc6 and Cdc6 antibodies respectively, as indicated on the top. Immunoprecipitation with unspecific rat IgG antibody served as control. For immunoblotting, 5% of the nuclear extract used for the depletion was loaded as input. The blots were stained with the antibodies indicated on the left.

It is also notable that the various factors present in the immunoprecipitation are depleted to different levels from the extract by the precipitation (data not shown). Orc1 is highly enriched in the precipitation even compared to the residual amounts in the input lane, whereas Orc6 is still abundant in the extract if it is co-depleted with another ORC subunit. This argues for a dynamic or biochemically weak interaction between Orc6 and

the core complex, or that ORC components are present in different levels in the cells, which also underlines the difficulty of visualizing this interaction.

Although Orc6 is shown to interact with Orc1-5 and Cdc6 in this experiment, this does not contradict previous studies showing the absence of such binding. The higher sensitivity of the Orc6 antibody allows detection of minimal amounts of Orc6, which support the visualization of this protein in Orc2 and Cdc6 precipitations, but this amount might be below the detection limit of antibodies used in previous studies. Also, this antibody precipitates the smallest subunit with high efficiency, leading to a higher number of co-precipitated proteins, and a more robust signal for Orc6 interactors.

Now that the assembly of these factors in solution is better understood by showing that the full ORC-Cdc6 complex is already preassembled in solution, the next step is to analyze the DNA binding characteristics of the complex. To better understand the place of Orc6 in ORC-DNA binding, an *in-vitro* system is required which allows the study of this process.

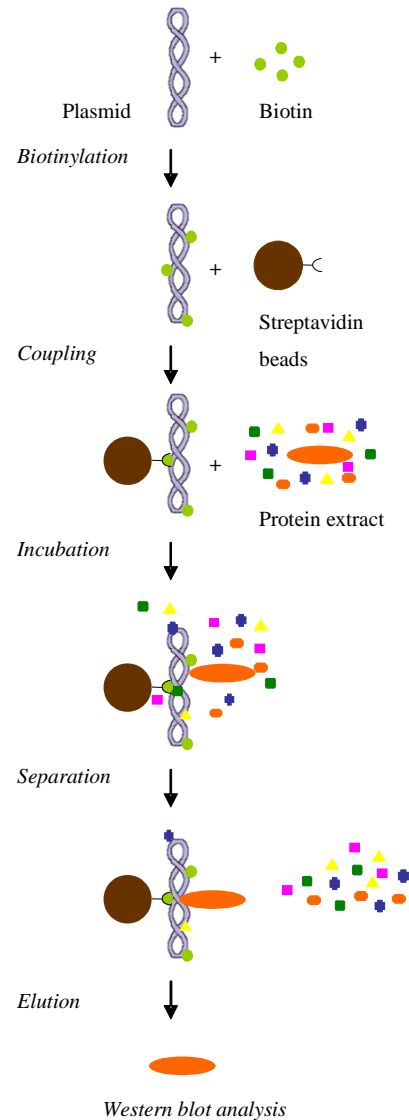
5.1.3 200bp but not 36bp of DNA is sufficient for ORC to attach

The interaction of ORC with DNA, and the role of Orc6 in it, is a focal point of replication initiation. Better understanding this step is one of the main goals of this work. In order to study it, I employed a cell-free pre-RC assembly assay. This method, termed the plasmid-binding assay, utilizes replication competent nuclear extracts, and their ability to assemble the pre-RC onto DNA in a sequence independent manner. It was originally developed to study the effect of viral factors in replication initiation (Schaarschmidt et al. 2004; Thomae et al. 2011), but has emerged as an excellent tool to elucidate the mechanism of pre-RC formation in great detail.

Figure 7. Schematics of the plasmid-binding assay. This assay allows the study of pre-RC formation on DNA in a cell free system (Thomae et al. 2011). Plasmid DNA is first coupled to biotin, and then to magnetic beads. The immobilized plasmids are then incubated with replication competent nuclear extract to allow complex formation on DNA. After incubation the beads are separated from the solution with a magnet, and washed to remove unspecific binders. Finally, DNA bound proteins are eluted and analyzed in downstream applications.

The assay requires a DNA template, usually a plasmid, which serves as the assembly site of pre-RCs. It has been established that supercoiled plasmids are better suited for *in vitro* replication assays, as the replication machinery binds with higher affinity to such plasmids (Baltin et al. 2006; Zembutsu 2006). To make use of this, I first subjected the purified plasmids to Cesium Chloride gradient centrifugation (Sambrook & Russel 2001), which enriches for supercoiled structures.

I used a small dose of UV light to incorporate reactive biotin into the prepared plasmids, which I subsequently coupled to streptavidin coated magnetic beads. Then I mixed the immobilized plasmids with a protein extract, which contained all components for pre-RC formation. It has been elucidated that nuclei prepared from HeLa cells which are then treated with 450mM KAc, release pre-RC components from chromatin creating a replication competent nuclear protein extract (Baltin et al. 2006). Once this extract has incubated with the immobilized plasmids, I separated the beads from the solution via a magnet. Then I washed the beads to remove unspecific binders, and the specific proteins were finally eluted with detergents and boiling as detailed in chapter 3.2.7. Finally I subjected the eluate to Western blot analysis to identify bound proteins.



Pre-RC components successfully bind DNA in the assay as shown in Figure 8. Here, I performed the plasmid-binding assay by using either magnetic beads without DNA, an immobilized linear construct, or a circular, supercoiled plasmid respectively. By using different templates, I was able to examine if the involved factors have a preference for a specific DNA structure. Other studies have shown that *in vitro* replication assays perform better on supercoiled plasmids. Also, it has been postulated that MCM complexes loaded during pre-RC formation may slide freely on a naked stretch of DNA, and might therefore get lost from a linear construct (Baltin et al. 2006). The results show that both the linear and the circular DNA template recruits the complete pre-RC, as seen by the presence of ORC and MCM complexes. The former is visualized via Orc4, 5 and 6, showing that not only Orc2-5, but also Orc6 is DNA bound in this experiment. By visualizing Mcm3 and Mcm5 I present evidence that the entire MCM complex is present as this is known to form a ring-like structure with all six subunits in solution, and bind as a whole to ORC. As discussed in the introduction, MCM complexes are bound to origins in two distinct steps. First, they are recruited and attached to ORC, and second, they load onto DNA encircling the double helix. The two can be distinguished by the high salt resistance of the proteins, as loaded MCMs are resistant to high salt treatment and stay DNA bound, whereas if only recruited, their signal is lost (Remus et al. 2009). My experiments established that the Mcm3 and Mcm5 signals presented in Figure 8 are not resistant to high salt, and therefore are not loaded onto chromatin but are only associated (data not shown). However, this does not hinder the assembly of ORC as this happens before MCM recruitment and shows to be complete and efficient in this *in vitro* system, and is able form the basis of subsequent experiments.

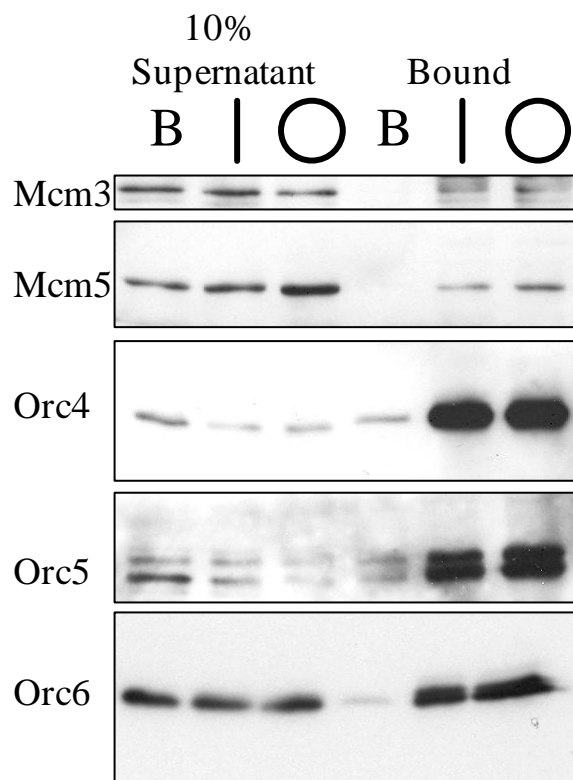


Figure 8. Recruitment of pre-RC components onto plasmid DNA *in vitro*. Plasmid-binding assay was performed by using replication competent nuclear extracts to assemble the pre-RC *in vitro*. Plasmid-bound proteins were subjected to Polyacrylamide gel electrophoresis, and visualized by immunoblotting using antibodies against the proteins indicated on the left. The first 3 lanes are supernatants collected from the assay showing 10% of the unbound proteins left after the experiment. The following three lanes show proteins bound to either magnetic beads only (B), a linear DNA fragment (vertical line) or a circular plasmid (circle).

After establishing that the plasmid-binding assay is working as intended and can be used to study ORC-DNA binding in more detail; *in vitro* DNA requirements of ORC formation were studied. First, I analyzed the minimal length of template DNA which can recruit ORC components. I excluded other pre-RC factors, such as MCMs, from further analyses as their recruitment was inefficient compared to that of ORC. Furthermore, ORC assembly and DNA binding happens before MCM loading, therefore the downstream mechanisms of MCM recruitment are not part of ORC assembly dynamics.

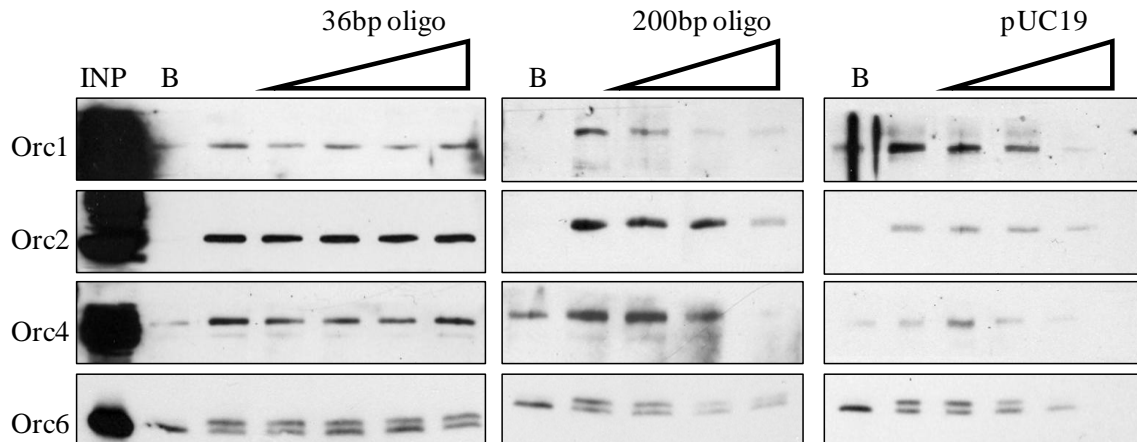


Figure 9. Effect of competitor DNA fragments on ORC-DNA binding. Plasmid-binding assays were performed with additional free DNA added to the solution before the addition of the nuclear extract. Bead-bound proteins were afterwards analyzed in Western blots by using antibodies specific for the proteins indicated on the left. The first panel shows 5% protein extract input as control. Each panel displays a control with only magnetic beads without coupled DNA (B), followed by a control assay with no competitor DNA. After this, assays loaded with increasing amounts of free competitor DNA in equimolar amounts are shown. The presence of competitor DNA is indicated on the top.

Figure 9 shows DNA length requirements of ORC formation. For this experiment I made a plasmid-binding assay (introduced in Figure 7 and tested for functionality in Figure 8) to determine if a 36bp, a 200bp, or a 3000bp long linear DNA strand is long enough to recruit ORC. I used a regular assay with immobilized pUC19 plasmids and HeLa nuclear extracts. However, before mixing the bead-bound plasmids with the extract, I added free competitor DNA to the solution as indicated in Figure 9. This way, once the extract was added, the proteins in it could either bind to the bead-bound constructs or to the free competitor DNA. Only the bead-bound plasmids were subjected to purification and their contents analyzed by Western blotting, which means if the non bound competitor DNA was able to effectively recruit ORC components, then these factors would be present on the immobilized constructs in smaller amounts, leading to a decrease in signal intensity. This is observed in the rightmost panel in Figure 9 where linearized pUC19 was used as competitor DNA in increasing amounts. The beads only control (B) displays a small signal for Orc1 and 4, and a prominent band for Orc6. The latter is of significant interest, as Orc6 has some affinity to streptavidin-coupled beads, providing background in these

experiments. However, upon DNA binding, Orc6 forms a double band on gels in an ATP, extract and time dependent manner (data not shown), which allows easy distinction of specific and unspecific signals. Since only the lower running form binds beads effectively, the upper band stands as evidence of DNA bound Orc6. With this in mind, the signals in the rightmost panel show how free competitor pUC19 DNA depletes all ORC subunits from immobilized plasmids. This however is only a control, as it was already established in Figure 8 that pre-RCs are able to assemble on plasmids of 3kb. New information is provided by the right panel, which displays the effect of a 200bp long competitor DNA piece. This has a similar effect as the pUC19 competitor, leading to the conclusion that a 200bp DNA fragment is sufficient to serve as a binding site for ORC. It is worth noting that all ORC subunits are depleted in a similar manner, despite having different amounts in the cell, as shown in Figure 5. This indicates that ORC assembles on DNA with a given stoichiometry. The result is different for the shortest oligonucleotide of 36bp displayed in the left panel of Figure 9. Here the signals are at the same level in all lanes, despite the fact that there are increasing amounts of oligonucleotides present. From this it can be concluded that a 36bp long DNA fragment is not sufficient to serve as an ORC binding template.

Since similar molar amounts of DNA strands were used in all three panels, it could be argued that the 36bp construct shows no effect because there are fewer nucleotides in the assay. This is because using an identical number of molecules in both the 200bp and the 36bp blots would lead to the presence of a lower mass (ng) of nucleotides present in the assay with the shorter template. To overcome this problem and to show that the observed depletion effect is due to the presence of DNA strands that recruit ORC and not simply because of the presence of nucleotides, an extra lane was added to the 36bp panel, where an even higher amount of DNA was used to provide the same mass of nucleic acids in the assay as in the rightmost lane of the 200bp DNA fragment. However, even this lane shows no depletion effect, therefore I concluded that not the amount of nucleotides present is actually important, but the number of possible binding sites of a critical length which are able to serve as a template for ORC. This contradicts a previous study in *Xenopus*, where it was shown that the presence of short DNA pieces increases ORC binding signals (Lebofsky et al. 2010). This however can be due to differences in

embryonic extracts used for the *Xenopus* study, which are shown to have an unusually high replication competence only observed in embryonic systems. Also, the *Xenopus* study directly measures replication of template DNAs in the presence of small DNA fragments, not the buildup of ORC or pre-RCs, which are not affected in this experiment. The size of oligonucleotides capable of recruiting ORC fall between 36 and 200 base pairs, which is in line with other studies: in yeast, a study predicted a 72bp DNA binding site for ORC-Cdc6 (Sun et al. 2012), and the 80bp DNase I footprint was observed (Bell & Stillman 1992; Speck et al. 2005). In metazoans, a *Xenopus* study using the same method as I did found that an 82bp, but not a 67bp fragment was sufficient to recruit ORC (Edwards et al. 2002).

Since the composition of DNA and the presence of particular sequence motifs has been shown to have little effect on ORC formation (Schepers & Papior 2010), and the binding site range has been here elucidated, next I studied the protein requirements of ORC binding, to obtain information on the relative binding affinities of the subunits.

5.1.4 Orc6 saturates similarly to Orc2-5 when binding to DNA

The molar amounts of Orc6 and Orc2 are different in nuclear extracts as shown in Figure 5. However, their presence in the extract and ultimately in the cell do not mean that they are equally able to be recruited to DNA, despite having shown that the ORC-Cdc6 complex forms in solution. To establish if only the complete origin recognition complex binds to DNA, or individual subunits can do so on their own, a saturation assay was performed. Increasing amounts of protein extract was added to a constant amount of DNA template. This experiment is expected to provide data on the relative binding affinity of individual ORC subunits.

Figure 10 shows Orc1, 2, 4 and 6, and Cdc6 binding and saturation on either pUC19 or the pEPI-UPR plasmids as template. In addition to the standard pUC19 plasmid for the plasmid-binding assays, I supplemented the experiment by using the pEPI-UPR construct for comparison, as it has been established that this is able to serve as template for *in vitro* replication experiments (Schaarschmidt et al. 2004). Here, the widely used pEPI plasmid backbone contains the UPR insert, which is the upstream promoter region of the human

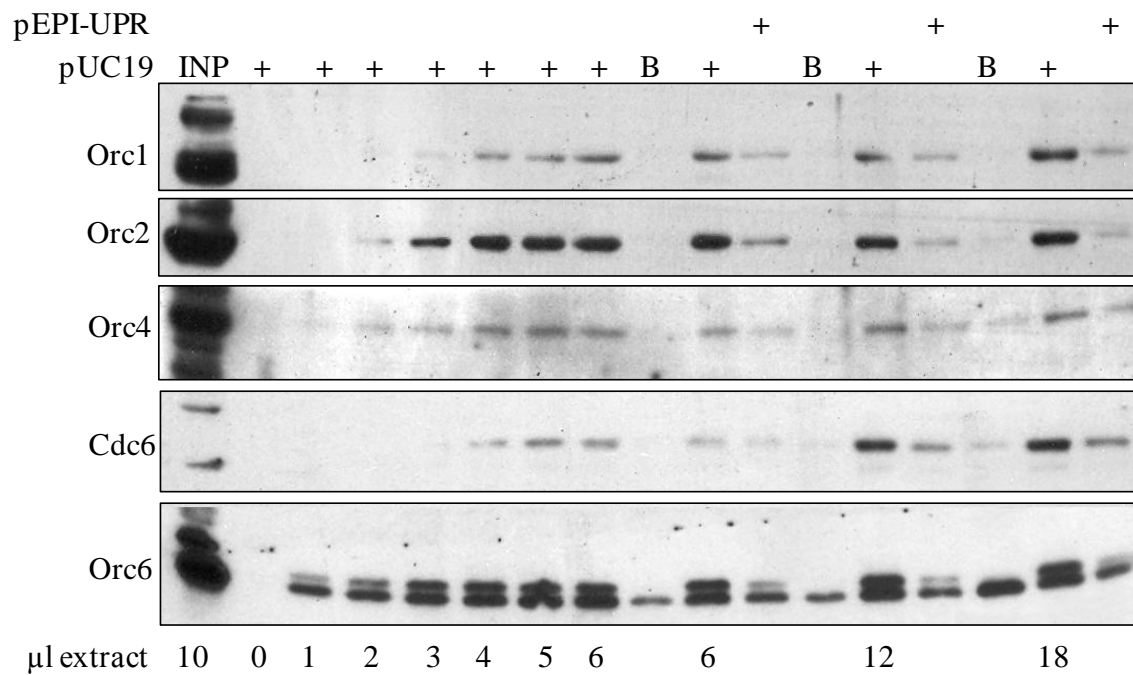
MCM4 gene, an established ORC binding site (Ladenburger et al. 2002). In Figure 10 I used equal molar amounts of the two plasmids for both biotinylation and bead coupling, and then added equal microliters of bead solution for the plasmid-binding assays. This ensured that the same molar amount of plasmids were present in all lanes, although since the pEPI-UPR plasmid is about 3 times longer than the pUC19 plasmid thus pEPI could have possibly provided more room to bind pre-RC factors per plasmid.

The amount of extract is increased from left to right, starting with zero, and gradually increasing as indicated at the bottom of Figure 10A. After 6 μ l, a beads only control and a pEPI plasmid using assay is also shown as comparison. These show similar background signals in the beads only (B) lane as in Figure 9, and surprisingly weaker signals for the pEPI plasmid as for the pUC. After these three lanes, the same experiment is shown, but with double (12 μ l) or triple (18 μ l) amounts of nuclear extract. A quantification of observed signals is presented in Figure 10B.

Results show that Western blot signals gradually increase as more extract is added for all proteins analyzed, but that Orc1 and Cdc6 show different saturation points in comparison to Orc2-Orc6. This is obvious in Figure 10B where the quantified signals are displayed in a diagram. Orc2 and Orc4, members of the core ORC, reach saturation at around 4 μ l. Also, Orc6 reaches the plateau at this amount, which is interesting as Orc6 amounts are significantly higher than core ORC component abundance in the cell. Although the extracts used are not whole cell extracts as in Figure 5 but HeLa nuclear extracts, it is still likely that Orc6 is more abundant in the nuclear extract than other ORC subunits. This leads to the conclusion that Orc6 behaves similarly to Orc2-5 when binding to DNA.

The Orc1 signal intensity on the other hand is increasing until the maximal 18 μ l, which is similar to the results of previous studies where adding extra, recombinant Orc1 to *in vitro* replication assays increased replication competence (Baltin et al. 2006).

A)



B)

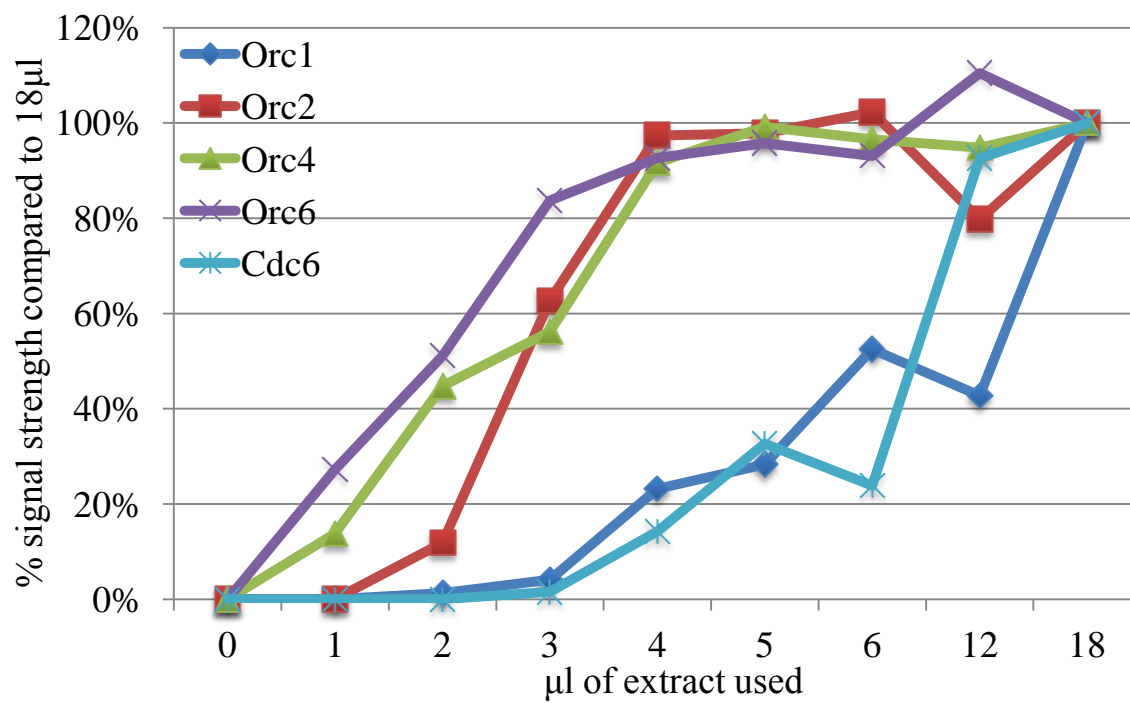


Figure 10. Binding saturation of ORC components on plasmid-DNA *in vitro*. Plasmid-binding assays were performed with constant amounts of template DNA and increasing amounts of nuclear extract to determine saturation of ORC components on DNA. Resulting Western blots with antibodies against selected pre-RC proteins are shown in A). Aside from the 3kb pUC19, the 10kb pEPI-UPR plasmid template was also tested in equimolar amounts as indicated on the top, with beads only serving (B) as control. Increasing amounts of nuclear extract is indicated at the bottom. B) presents the quantification of pUC19 signals from the blot in A), with the 18 μ l signal taken as 100% for each dataset. From Orc6 the upper, DNA specific band was quantified. The x-axis shows the amount of nuclear extract used for each data point.

Furthermore, the saturation point of Cdc6 is reached only with 12 μ l nuclear extract. This point falls between that of Orc1 and other ORC components leading to the conclusion that Orc1 is present in the most limiting amount among these factors during *in vitro* replication initiation, possibly due to either partial degradation during extract preparation, or low expression in HeLa cells.

The experiment also proved that the ORC-Cdc6 complex, despite forming in solution, can possibly assemble on DNA as well, or at least different sub-complexes of ORC-Cdc6 might differ in their DNA-binding characteristics. To further elucidate whether ORC components can bind separately to DNA, and to receive more information on how Orc6 interacts with other ORC subunits upon binding DNA, I performed plasmid-binding assays depleting specific ORC components.

5.1.5 Orc6 is able to bind DNA independently of ORC

One of the aims of this study is to better understand the mechanism of ORC-DNA binding, with respect to the function of Orc6. Looking at this mechanism in detail, this chapter explores if ORC subunits require each other to bind to DNA. To achieve this, individual subunits were depleted from the nuclear extract prior to addition to DNA. In the plasmid-binding assay.

From yeast studies it was established that Orc1-5 binds DNA without the need for Orc6 to be present (Lee & Bell 1997). However in *Drosophila*, Orc6 is crucial for ORC-DNA interaction (Chesnokov et al. 2001). Based on this knowledge, one main question in the

case of human cells is: is Orc6 required to recruit ORC to DNA? To answer this, I prepared two different depletions. First, Orc6 depletion would have to show if Orc6 is required for Orc1-5 recruitment in this assay, and second, the depletion of the core ORC (Orc2-5) should elucidate if Orc6 can bind to DNA on its own. The results of this experiment are shown in Figure 11. After an extract only control (INP), and a beads plus extract control (B), I prepared a full plasmid-binding assay that also includes DNA templates, either in the presence or absence of ATP. Following this is a depletion control using an unspecific antibody, and finally, Orc2 and Orc6 depletions in the two rightmost lanes. The image shows that Orc2, which represents the core ORC, can bind in the absence of Orc6, and also Orc6 is able to attach to DNA if the core complex is not present. Surprisingly, this feature is more similar to the results from yeast, than the ones from the more closely related insect species. Adding the fact that HsOrc6 is structurally more similar to DmOrc6 than ScOrc6 (Duncker et al. 2009), the result is even more unexpected.

Additional peculiarities can be observed in Figure 11. First, adding ATP has a distinct effect on Orc6 DNA binding, which shows a prominent, higher migrating band if the nucleotide is present, which is also observable in Figure 8, Figure 9 and Figure 10. Although posttranslational modifications of HsOrc6 are established (Dutta & Dhar 2000), and it is a known CDK target (Ghosh et al. 2011), it is still not elucidated what this modification might be. It is however dependent on the presence of ATP, and DNA binding.

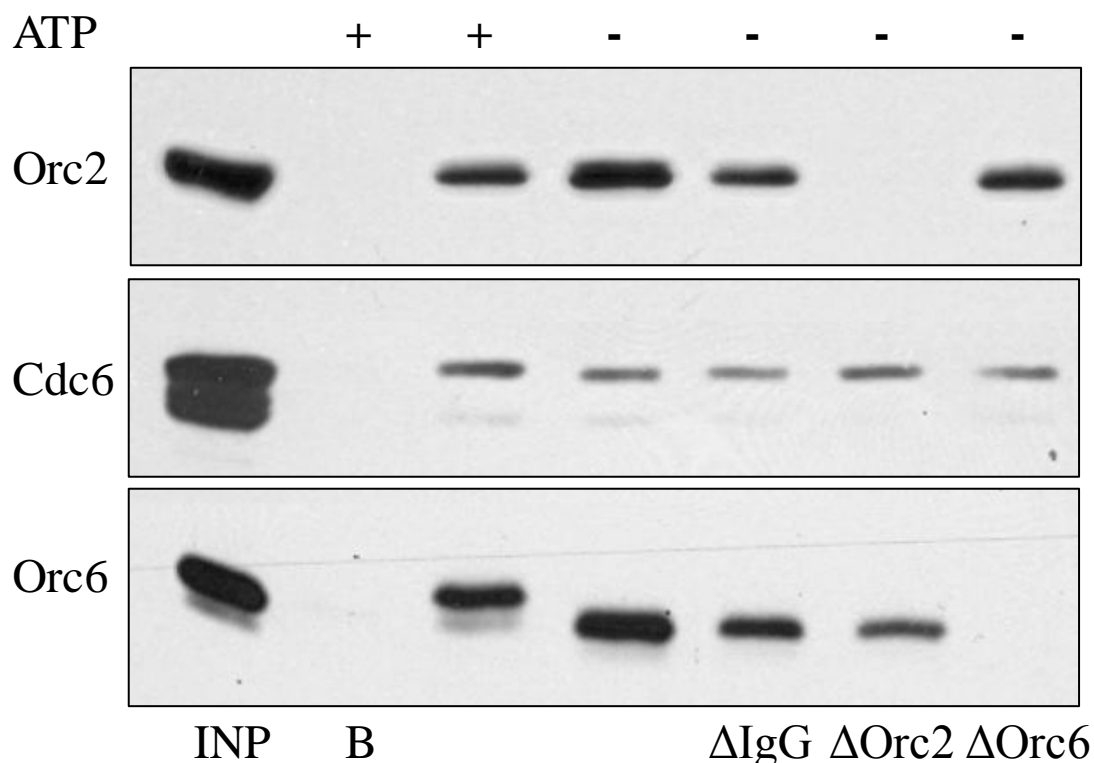


Figure 11. *In vitro* assembly of partially depleted ORC on DNA, and effects of ATP (part I). Plasmid-binding assay was performed to determine the effects of Orc2 and Orc6 depletion on ORC-Cdc6-DNA interaction. 5% extract (INP) and an assay using no DNA template (B) were loaded as control. Plasmid bound proteins in the presence or absence of ATP (indicated on the top) are shown in lanes 3 and 4. Effects of depleting proteins with an unspecific (IgG), an Orc2 or an Orc6 antibodies are shown in the last three lanes respectively.

Furthermore the IgG depletion makes all observed signals slightly weaker. It has been established that a high concentration of proteins is required for an efficient ORC formation (Thomae et al. 2011). Adding antibodies to the extract lowers the overall protein concentration, leading to a less efficient binding. However, depleting Orc2 moderately reduces Orc6 signal intensity even if compared to the IgG control. This shows that Orc6 binds both independently and in complex with ORC, but the signal difference is not prominent enough in a single experiment to be able to reliably tell the difference.

It is also unexpected that Cdc6 signals are unaffected by any ORC component depletion. This shows that either Cdc6 binds to Orc6 and Orc1-5 separately, or more likely it has an

innate ability to bind DNA, and since fully assembled ORC-Cdc6 complexes are small in number compared to free Cdc6 in extracts (see in Figure 6), the possible loss of signal due to the depletion of the full complex is negligible. Having learnt this, I concluded that Cdc6 recruitment is not an appropriate control for ORC formation, therefore the experiment was repeated using Orc4 antibodies, which stains a prominent component of core ORC.

5.1.6 ORC enhances Orc6 DNA binding

To clarify the possible Orc6 signal loss observed in Figure 11 and to see the effect of ORC component depletion on complex formation, I repeated the same experiment as in Figure 11, this time however, I included an Orc4 stain. Despite the small difference in the experimental setup, the test yielded valuable information on ORC-DNA interaction.

In Figure 12 on the left, the three supernatant lanes show the unbound proteins still available in the extract after performing the plasmid-binding assay. This shows that only a fraction of pre-RC proteins are binding to plasmid DNA from the existing pool. Comparing this result to Figure 10 it can also be concluded that upon saturation on DNA, pre-RC proteins are present in large abundance in solution.

Looking at the Orc2 and Orc6 depleted lanes, both proteins are efficiently removed from the extract, but they do not co-deplete each other. This is expected from the ORC subunit interaction results discussed in chapter 5.1.2. It is also in accordance with the current model that Orc2 efficiently co-depletes Orc4. Interestingly, the remaining Orc4 still present after Orc2 depletion does not bind to plasmid DNA but remains in the supernatant. This means that Orc4, probably with the rest of the core ORC, can only be recruited to DNA as a whole. This also indicates that despite the fact that Orc4 and other core subunits all have a winged helix domain, a hallmark of DNA binding proteins, they still require complex formation to efficiently bind to DNA (Stefanovic et al. 2003).

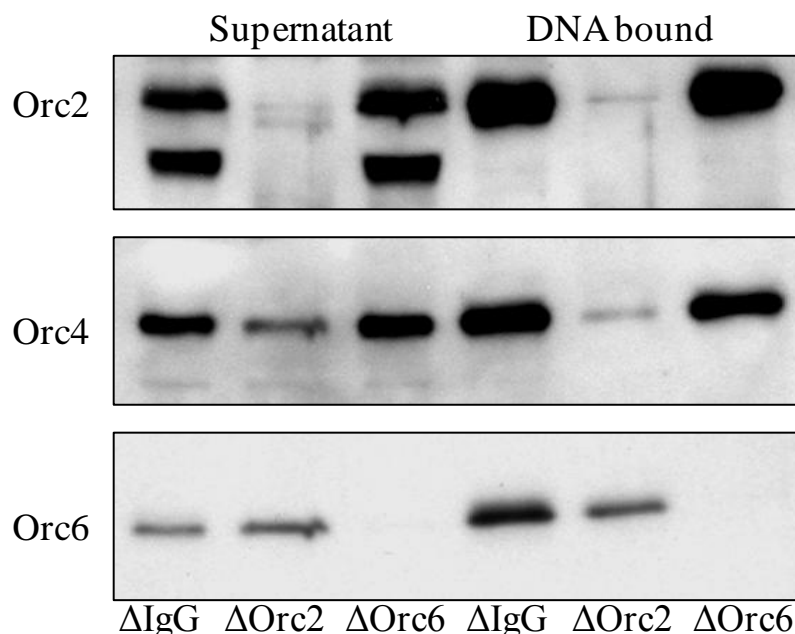


Figure 12. *In vitro* assembly of partially depleted ORC on DNA (part II). Plasmid-binding assay was performed to assess the effects of Orc2 and Orc6 depletion on ORC-DNA binding. Prior to adding the nuclear extract to the plasmids, they were depleted with either an unspecific antibody (IgG), an Orc2, or an Orc6 antibody as indicated on the bottom. 20% of unbound proteins after the assay (supernatant) and all plasmid bound proteins (DNA bound) were stained in Western blots with the antibodies indicated on the left.

It is also very important to distinguish between the role of Orc2 and Orc6 in this experiment. Whereas Orc2 is required for Orc4 DNA-binding, Orc6 is not. Therefore it can be concluded that unlike in *Drosophila*, Orc6 is not essential for the Orc1-5-DNA interaction. Orc2 on the other hand is crucial for the same process, and its presence might stabilize Orc6 DNA-binding as evidenced in lane of Orc2 depletion, where the Orc6 signal is reduced compared to the IgG depleted control.

5.1.7 Orc6 is not required for ORC DNA binding, but it enhances the process if present

Although depletion of Orc6 had no effect on ORC-DNA binding, adding excess Orc6 might do. Therefore, after depleting endogenous Orc6 from the sample, I added bacterially expressed, purified, recombinant HsOrc6 to the plasmid-binding assay. The result is presented in Figure 13.

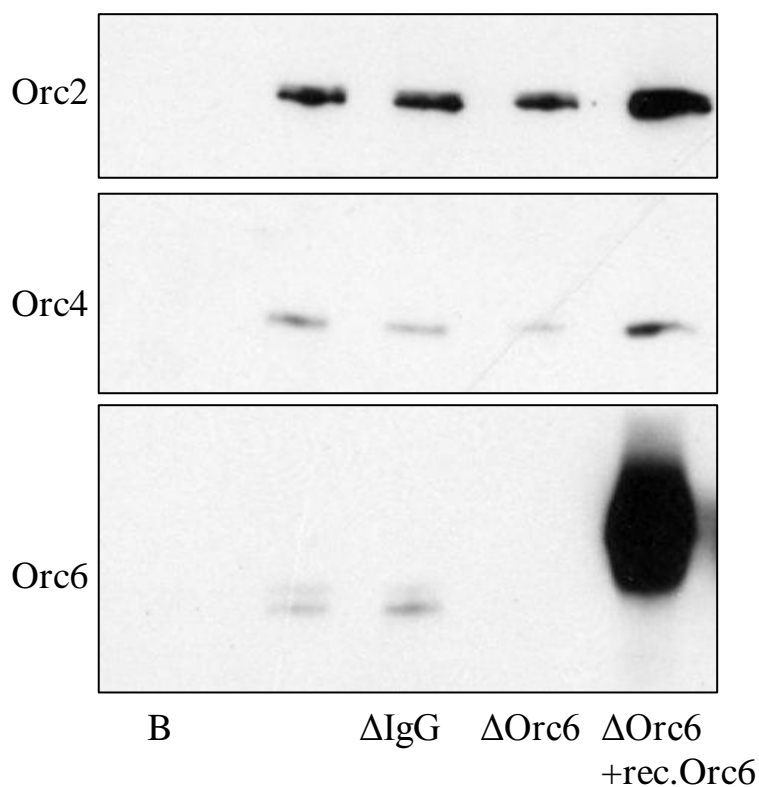


Figure 13. Effect of excess recombinant Orc6 on ORC-DNA binding. Plasmid-binding assay was performed while adding excess recombinant Orc6 to the reaction. The lane marked with “B” of the resulting immunoblot displays a control where no DNA was used, only magnetic beads. This is followed by a plasmid-binding assay with DNA as positive control. Orc6 was depleted from the last two lanes using a specific antibody. The rightmost lane contains 100ng bacterially expressed HsOrc6 which was added to the reaction before mixing it with DNA. Antibodies used for immunoblotting are indicated on the left.

After a beads only control (B) Orc6 is observed binding to plasmids as a weak doublet, similarly to previous results. Depletion of Orc6 shows no effect on the other factors, which is also seen in Figure 11 and Figure 12, but the last lane containing an excess of Orc6 induces a significant increase in Orc2 and Orc4 DNA binding. This cannot be an effect of adding extra protein, as the 100ng, concentrated, purified protein is minuscule compared to the 30µg of nuclear extract present in the assay. Also, in preliminary experiments, I used BSA to block the beads prior to adding the nuclear extract to reduce background binding. Although this method did not reliably reduce Orc6-background signals, and so was subsequently abandoned, it also led to the conclusion that adding extra protein has no effect on ORC-DNA binding (data not shown). Because of this I concluded that specifically the addition of Orc6 leads to the increased binding of ORC components.

Although the effect is present, the mechanism behind Orc6 enhancement of ORC-DNA binding is not yet fully understood. It is speculated that at least in yeast, Orc6 assists in the stabilization of ORC by binding Orc2 and a specific segment of Orc1 (Sun et al. 2012). What domains within Orc6 are active in this process, and how this occurs *in vivo* is yet to be explored.

5.1.8 Orc6 binds via its C terminus to ORC *in vitro*

Having elucidated that Orc6 interacts with ORC and Cdc6 in solution (Figure 6), and that the protein is not required for Orc1-5-DNA binding (Figure 11) but enhances the process if present (Figure 13), it is still unclear which domains of Orc6 contribute to replication initiation. Previous studies trying to dissect or crystallize Orc6 found no known domains, contrary to other ORC subunits (see chapter 2.2). Although Orc6 seems to have an evolutionary conserved body in most species, even that has only a modest similarity between distant eukaryotes.

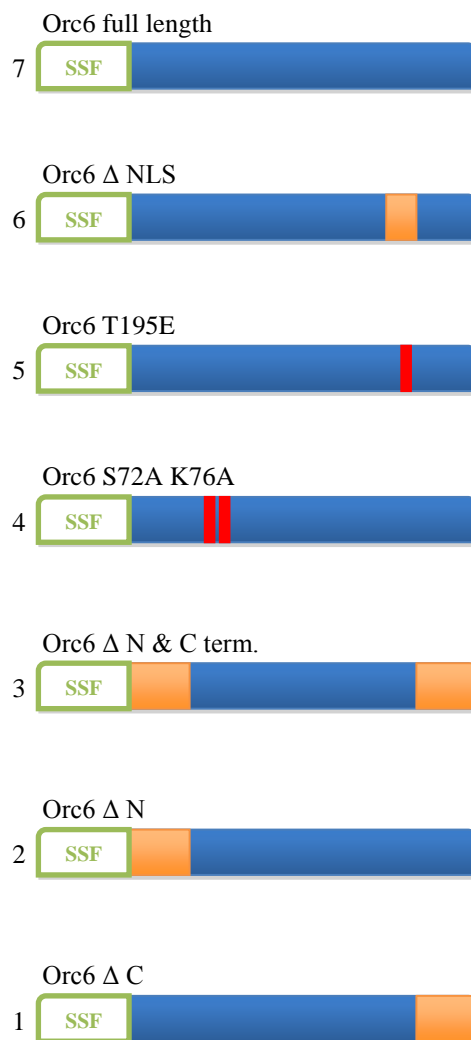


Figure 14. Bacterial expression of recombinant Orc6 variants. The depicted proteins from 1-7 were expressed and purified from bacteria. All variants have a Strep-Strep-Flag tag on the N terminus for purification and detection purposes. Deletions in the variants are indicated by orange marks. The red bars in the S72A K76A and the T195E variant represent point mutations which are indicated in their corresponding labels. The full length construct is the tagged, wild-type Orc6, which was used as control.

To characterize Orc1-5 interacting domains of Orc6, I generated a set of variants as displayed in Figure 14. I N-terminally tagged the 252 aminoacid long, full length, wild-type Orc6 protein with a strep-strep-flag tag for purification and detection purposes. Then I truncated the full-length protein by 50 amino acids from either the C terminus, or the N terminus, leading to two new variants with unique truncations. The third variant contains

both of these truncations, so it is missing both its first and last 50 amino acids, leading to a missing N and C terminus. The next one corresponds to a *Drosophila* homologue in which it was shown that two highly conserved amino acids, S72 and K76, are crucial for DNA binding of DmOrc6. By mutating both the serine at position 72 and the lysine at position 76 to alanines, the DNA binding ability was lost (Balasov et al. 2007). Both amino acids are conserved in the human homologue, so I created this variant to provide a potentially DNA binding defective protein carrying the S72A, K76A point mutations. An additional variant contains a single point mutation compared to the wild-type protein: the threonine at position 195 is changed into a glutamate. The purpose of this is to mask a predicted CDK phosphorylation site on the protein, and mimic the phosphorylated state. Finally, the last derivative is lacking the predicted nuclear localization signal of Orc6, which also contains the predicted CDK target site. Therefore this mutant, unlike the wild-type protein, is expected to remain mainly cytosolic *in vivo*, and to show a constant, unphosphorylated phenotype, as seen in a previous study (Ghosh et al. 2011).

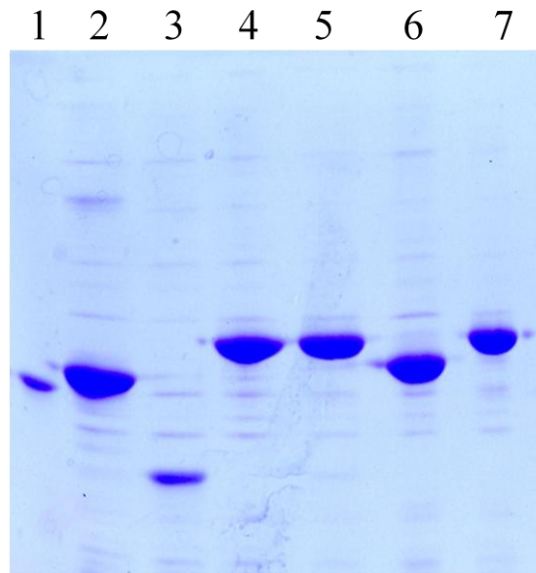


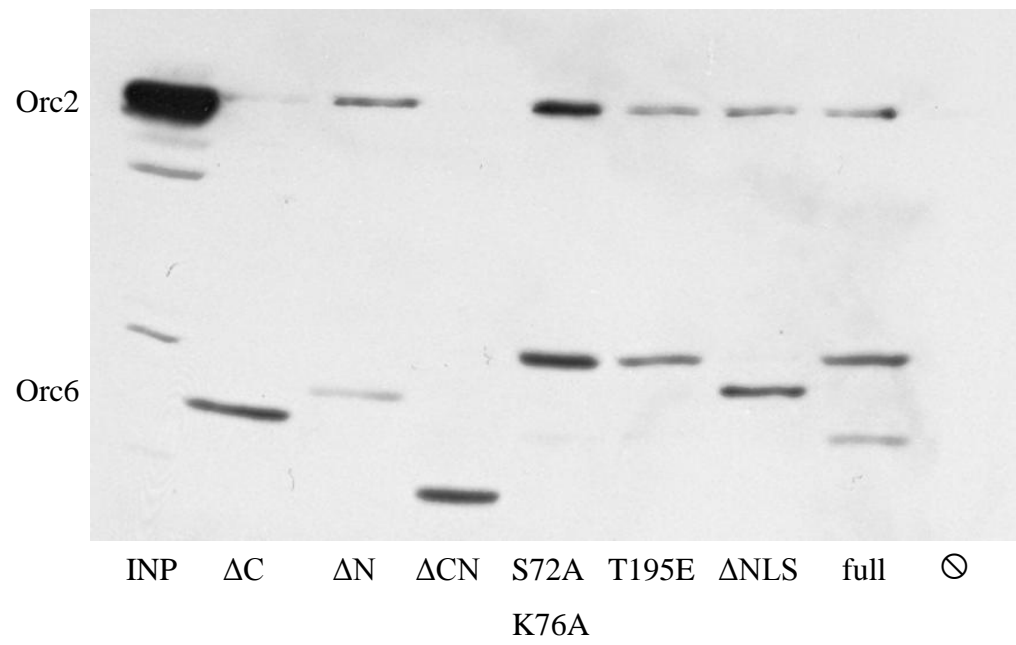
Figure 15. Polyacrylamide gel with purified, bacterially expressed Orc6 variants. Orc6 variants depicted in Figure 14 were expressed and purified from bacteria, and loaded onto a polyacrylamide gel respectively. After electrophoresis, the gel was stained with the Colloidal Blue staining kit. The number of each variant is shown on the top, and correspond to the numbering presented in Figure 14.

I first cloned all seven versions into pET21 bacterial expression vectors, and then expressed them in BL21 Rosetta cells. Following purification from the bacteria culture using their Strep tags, I analyzed the purified proteins via Polyacrylamide gel electrophoresis to assess correct product size and purity.

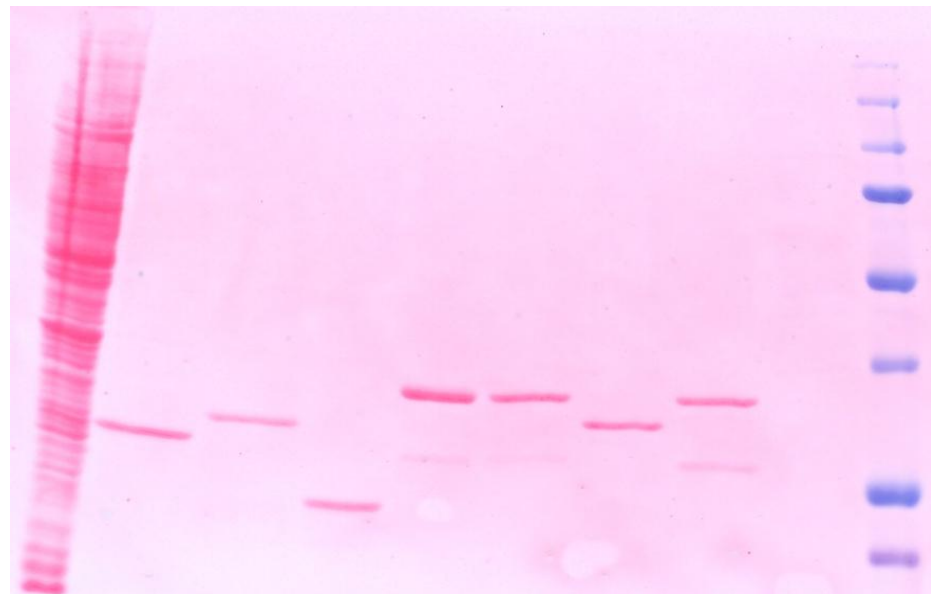
The purified proteins were then used in pull-downs, and tested for interaction with other ORC subunits. To do this, I coupled each purified protein onto a column and subsequently added Orc6 depleted nuclear protein extract to it. This allowed binding of factors in the extract to the immobilized Orc6, which after washing were eluted and finally subjected to Western blotting.

Figure 16. Interaction of different Orc6 protein variants with Orc1-5 *in vitro*. Pull-down experiments were performed by coupling Orc6 derivatives to columns and adding Orc6 depleted nuclear extracts to them. A) Eluates from the 8 separate pull-down experiments were fractionated and fraction 2 loaded onto a Western blot gel, which was stained with an Orc2 specific antibody. The Orc6 derivative used for each pull-down is indicated on the bottom. The leftmost lane contains 10µl of nuclear extract as input control. The last lane is a negative control, where no Orc6 was added. Apart from to the Orc2 signals labeled on the left, the Orc6 variants are also visible due to the high amount of eluted protein present on the membrane, and the minor cross reactivity of the anti-Orc2 antibody. B) Ponceau-red stain showing proteins on the membrane. The Orc6 variants are visible even in the absence of an antibody, due to their high amounts in the eluates. A protein size marker is included on the right.

A)



B)



Results presented in Figure 16 show which Orc6 variants were able to pull down Orc2, as representative for Orc1-5. The last lane features only buffer, no added Orc6 protein, and is therefore lacking an Orc6 signal, but also a signal corresponding to Orc2, showing that Orc2 has no significant background binding. The input lane shows the correct position of Orc2, serving as a positive control. Using these two endpoints as reference, I concluded that Orc2 is able to bind all Orc6 variants which have an intact C terminus. Orc6 Δ C and Orc6 Δ N Δ C however, which lack this domain show reduced Orc2 binding activity, which can only be observed with extended exposure times (data not shown). The weak signal next to the input is more likely to be a spillover than an actual signal, confirmed by the consistent results with the double truncated construct and replications of this experiment. Orc6 S72A K76A contains a surprisingly stronger Orc2 signal, but this is most likely due to the higher amount of recombinant protein present in the sample. Although the effort was made to use identical amounts of protein in each lane, the different eluates used contain varying amounts of Orc6. I also have to note that the blot was not stained for Orc6, as the epitope is located in the C terminus of Orc6, and would therefore not give a signal with the variants lacking this domain. Instead, the signals seen are from cross-reaction of the Orc2 antibody with the excess amounts of Orc6 in the sample. The amount of recombinant Orc6 was so high in the elution fraction in this experiment that it quickly bleached the signal when using an Orc6 specific antibody, and gave a very prominent band even with a Ponceau-red stain (Figure 16B). This is however in line with the results from Figure 6, which conclude that the biochemical interaction between Orc6 and Orc1-5 is weak. The result does not in itself mean that Orc6 directly interacts with Orc2. Since an extract was used, it cannot be excluded that the interaction detected is independent, and occurs via other ORC components or auxiliary factors. Taken together, this experiment demonstrates how Orc6 interacts via its C terminus with core ORC *in vitro*. To confirm these results *in vivo*, other methods were employed.

5.1.9 Orc6 needs a nuclear localization signal or its C terminus to enter the nucleus

In vitro experiments allow more direct control over the experimental setup, as they allow the study of various systems of an organism independently. This however leads to results, which do not hold true in a live organism, as the complexity of an entire cell is much greater than that of the individual parts. Therefore, *in vitro* studies and their results need to be validated *in vivo*. While studying the *in vitro* association of Orc6 to Orc1-5 I came to the conclusion that the smallest subunit requires its C terminal 50 amino acids to interact with Orc1-5. To validate this result *in vivo*, I carried out immunofluorescence experiments, which visualize the localization of the proteins of interest inside a complete cell.

For this experiment, new, eukaryotic Orc6 expression vectors were required, as contrary to the previous experiments where expression was in bacteria, here the goal was to express and study the different Orc6 proteins in human cells. Therefore, I transferred the 7 already established gene variants introduced in Figure 14 into the bimolecular fluorescence complementation vector backbone (pBiFC), which allows eukaryotic expression of these proteins. Each construct received an N terminal HA tag for detection purposes, and a C terminal YFP tag, which codes either for an N-terminal or C-terminal part of the fluorescent YFP protein. The separation of YFP into two domains can be used to conduct BiFC assays, which allow the detection of close interactions between two proteins (Kodama & Hu 2012). Although wild-type Orc6 has been shown to be able to provide a BiFC signal with Orc4, Orc5 and Cdc6 (Thomae et al. 2011) no such signal was observed with Orc2 (data not shown). In this experimental setup however, I attempted immunofluorescence experiments, which identify the constructs via the use of specific antibodies.

Apart from the seven previously generated versions of Orc6, I established three new ones to gain a better understanding of the role of the central area of the Orc6 protein. The first of the new variants contains a deletion from amino acids 50-125. This 75 amino acid deletion starts just after the 1-50 deletion tested in ΔN variant. The next construct follows this by having a deletion from 125-203. As the ΔN variant already has a deletion from

203-252, the Δ 125-203 construct completes the set of deletion mutants by omitting the only fragment of the protein, which all other variants have. The last mutant is a control for Δ 50-125 and Δ 125-203, as it is lacking the entire area from 50-203. This mutant basically consists of a C and N terminus only, with all other areas removed in between. A cartoon depicting these variants is shown in Figure 16.

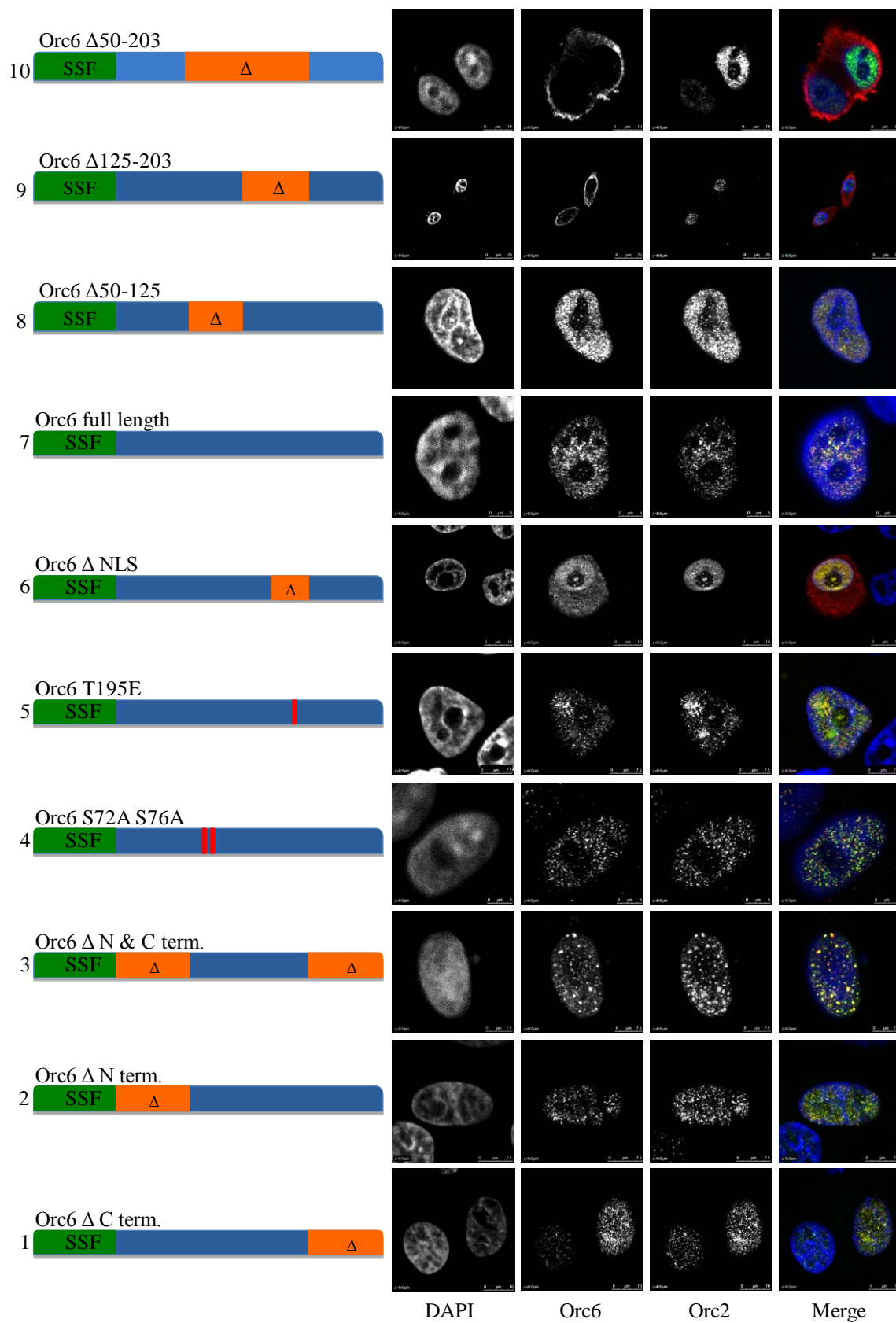
I carried out the immunofluorescence experiment by using HepG2 cells, which is an established cell line showing primary cell-like morphology, and is often used for microscopy experiments (Knowles et al. 1980). I transfected HepG2 cells with a pBiFC plasmid carrying Flag:Orc2:YFP155-253, and each of the 10 different HA:Orc6:YFP1-155 variants in separate wells. After 18-24 hours of expression, I cross-linked the cells with formaldehyde, and stained them with anti-Flag and anti-HA antibodies coupled to Cy3 and Cy5 dyes to visualize the location of the expressed Orc2 and Orc6 proteins.

Although the dyes are excited by different lasers and their fluorescence also occurs at distinct wavelengths (Mujumdar et al. 1993), I confirmed independence of the signals by identifying clear green and red signals in the cell lines studied. This ensures that the co-localization observed is not an effect of detecting the same dye in both channels.

Results shown in Figure 17 present cartoons of each construct used and representative microscopy images obtained. A DAPI stain is shown for each image to visualize the chromatin. After the antibody signals, the final column shows a merge of the previous three images, displaying DAPI in blue, Orc6 in red, and Orc2 in green. Co-localizing signals in the merge image appear yellow.

Figure 17. Cellular localization and Orc2 interaction of different Orc6 variants *in vivo*. Each of the presented HepG2 cells were transfected with FLAG:Orc2 carrying expression plasmids, and plasmids encoding the Orc6 variant depicted on the left of each row. After allowing the expression of the fusion proteins, the cells were prepared for fluorescence microscopy by fixing them onto microscopy cover slips. Subsequently they were treated with anti-FLAG and anti-HA antibodies, and finally DAPI to visualize chromatin. Resulting images are displayed showing the color channel of the signal indicated at the bottom. The last image in each lane contains a merge of the previous three, depicting DAPI in blue, Orc6 in red, and Orc2 in green. These images have been digitally enhanced for printing. For the raw images, see **Figure 26**.

Results



The results show that all Orc6 variants generally localize to the nucleus if they contain the nuclear localization signal. This is similar to the results of other studies conducted in yeast and HeLa cells (Semple et al. 2006; Ghosh et al. 2011). Orc6 is absent from the nucleolus, but otherwise shows an even, granular distribution in the nucleus without a clear preference for any chromatin environment. Similarly, Orc2 shows a granular nuclear pattern in all transfected cells, showing that mutant Orc6 does not alter its spatial organization. Endogenous Orc6 is still present in all cells however; therefore this could complement any effect the mutant Orc6 might have on Orc2 function.

Compared to Figure 16 it is surprising that both variants lacking the C terminus show a wild-type signal, and generally co-localize with Orc2.

The effect of the nuclear localization signal is even more striking. It is expected that deletion of the NLS results in the complete loss of the protein from the nucleus, but the results show more than that. Variant 6, containing the 11 amino acid deletion of the predicted consensus NLS is present in the cytosol as expected, in contrast to the nuclear wild-type protein. However, some of these proteins are still able to enter the nucleus and shows a near complete co-localization with Orc2, evidenced by the lack of red signal in the nucleus and the sole presence of yellow in the merge image. This means that despite losing its NLS, Orc6 can still enter the nucleus, most likely via forming a complex with ORC already in the cytosol, and traversing the nuclear membrane as part of a complex, not requiring a separate NLS. Although the formation of ORC prior to nuclear entry is not directly supported by this evidence, it is a surprising coincidence that an even larger deletion, as in the case of the $\Delta 125-203$ and the $\Delta 50-203$, will lead to a near-complete absence from the nucleus, showing prominent cytosolic signals.

In conclusion this experiment provided unexpected results, as C terminus lacking Orc6 variants showed similar localization as the wild-type protein, and their signals largely overlapped with those of Orc2. This is contradictory to the *in vitro* results presented in Figure 16 as there a clear loss of interaction is observed upon deletion of the C terminus. This means that *in vivo*, ORC assembly is different than *in vitro*. To further elucidate this interaction *in vivo*, and to characterize Orc6 function in the cell, a different experimental approach was required.

5.2 Functional analysis of Orc6 at origins of replication

5.2.1 Orc6 variants are able to create origins of replication

The previous experiments provided data on the mechanism of Orc6-Orc1-5 interaction, but no functional read out. The method used in this chapter addresses this with results that show if Orc6 is able to create functional origins of replication in living cells.

The plasmid rescue assay introduced in chapter 2.4.1 allows *in vivo* analysis of replication competence of proteins. To carry out the experiment I inserted the seven variants introduced in Figure 14 into the pWHE vector, which adds a single chain Tet repressor domain (scTetR) to the N terminus of all constructs and allows eukaryotic expression of these fusion proteins (Krueger et al. 2003). In addition, I prepared the empty vector expressing the scTetR tag only as control. Using these vectors, I transfected HEK-293-D cells expressing EBNA1 and created cell lines which each stably express EBNA1 and one of the Orc6 variants. Next, I also introduced a wild-type oriP plasmid, an FR-TetO plasmid, or water to each of these cell lines. After carrying out the plasmid rescue assay I collected the results and summarized them in Figure 18.

The goal of this experiment was to study the function of Orc6. Figure 18 shows the percentage of colonies counted with an FR-TetO plasmid, relative to the number of colonies obtained with a wild-type oriP plasmid in the same cell line. The higher the percentage, the more colonies were counted in the assay relative to oriP. This number depends on the replication efficiency of the plasmids, which in turn depends on the ability of the Orc6 variants to create functional origins of replication. Each experiment was performed in triplicates using newly thawed aliquots of each cell line. Figure 18 displays the average number of colonies obtained in each setup, and their respective standard deviations.

Both wild-type Orc6 and all variants, which I tested, provide around 2-10% as much colonies as EBNA1 with DS. The scTetR-tag only is strikingly different from these, as it

was not able to provide any colonies at all. All other variants yielded colonies in all experiments, showing that all Orc6 variants tested are able to create new origins of replication, including the full-length protein. This provides additional knowledge to previous experiments, where our lab has shown that full length Orc6 is able to generate new origins of replication (Thomae et al. 2011). An independent, unpublished thesis work made by Vishal Agrawal from the Gossen lab also confirmed these results. Albeit Orc6 is only about 5-10% as efficient as EBNA1 in facilitating replication, this activity is still sufficient to retain plasmids in human cells for up to a month when under selection, but the protein is likely to be able to carry out this function for even longer intervals.

The results are in alignment with the microscopy data presented in Figure 17, which also underline that despite the mutations, Orc6 variants are able to co-localize with Orc2 and possibly function at origins. That result is now complemented with the functional readout of this experiment, which concludes that the variants are not only able to co-localize, but are also functionally active and are able to support replication initiation to about 5-10% of EBNA1.

The plasmid rescue assay however is able to say more than that. Not only are these variants functionally active, but they are also able to establish an origin of replication, which is completely absent if they are not specifically directed to this site, as evidenced by the complete plasmid loss observed with the tag only. Surprisingly, all variants show a more or less similar replication activity, which is about a tenth of EBNA1.

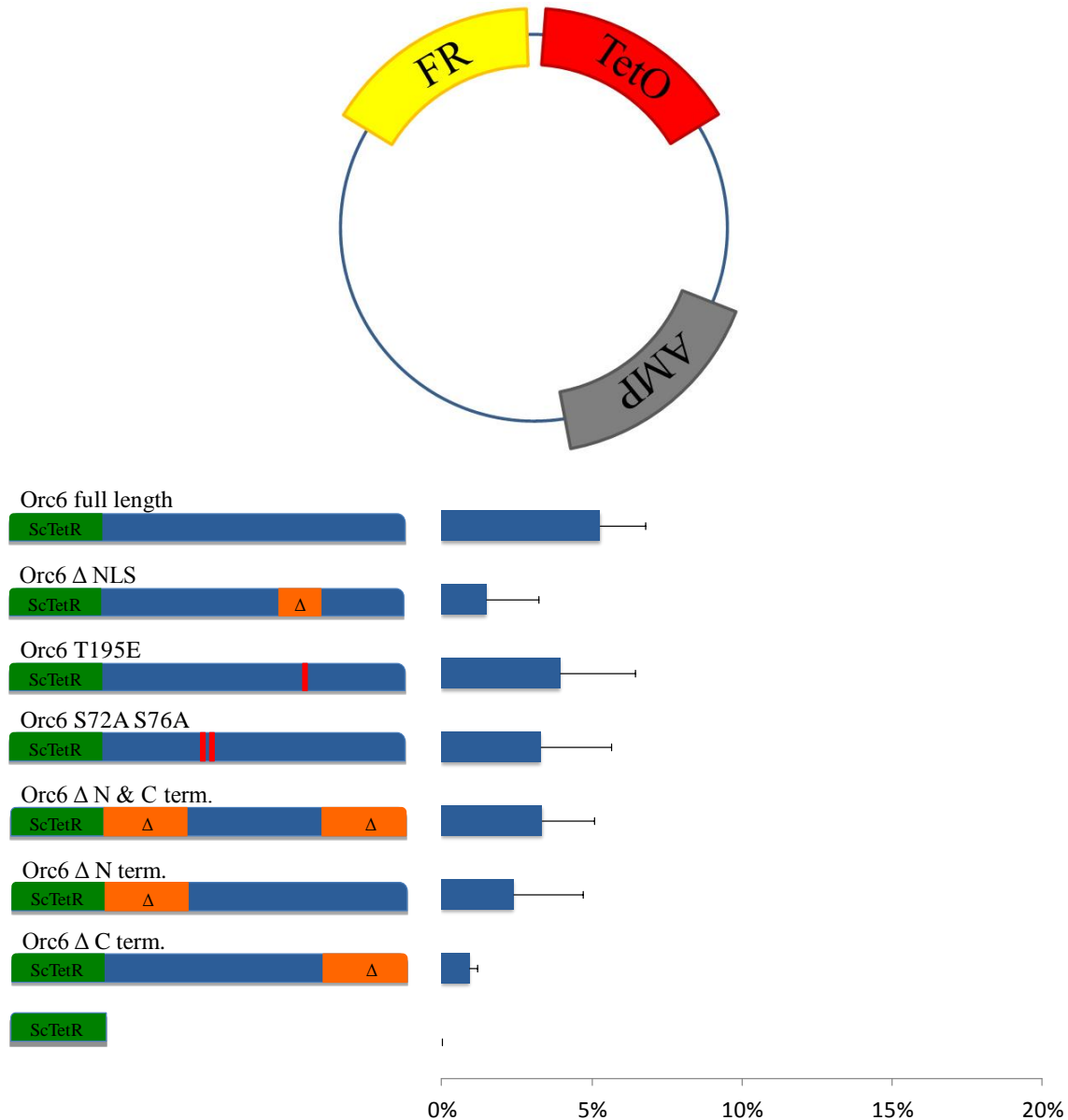


Figure 18. Ability of Orc6 variants to create functional origins of replication. The protein constructs displayed on the left were expressed in HEK-293-D EBNA1 cells. Subsequently, these cell lines were transfected with the FR-TetO plasmid depicted at the top. The bars show the results observed upon performing a plasmid rescue assay with this setup in the individual cell lines. The percentage displayed is the ratio of colonies observed with the FR-TetO plasmid versus a wild-type oriP plasmid in the same cell line. N=3, error bars show standard deviation.

5.2.2 PR-Set7 is able to create origins of replication

Seeing the positive results with Orc6 and its variants in creating functional origins of replication, I designed another experiment to test if this effect can be enhanced by also directing a histone methyltransferase to Orc6 dependent origins. Since Orc6 is only about 5-10% as successful in replicating plasmids as EBNA1, a more favorable chromatin environment might be needed to boost this effect and lead to a more robust replication of plasmids in human cells, without the need for viral factors. To this end, I targeted PR-Set7 to origins established by Orc6, and tested in plasmid rescue assays for plasmid maintenance efficiency.

It has been established that PR-Set7 plays a role in origin activation by methylating the H4K20 residue locally on chromatin. This modification is required for proper S-phase replication initiation (Falbo & Shen 2009; Beck et al. 2012). H4K20 levels increase during G1, and diminish through S phase. This is linked to the cell-cycle regulated expression pattern of PR-Set7. The degradation of PR-Set7 is induced by the binding of it to PCNA in S phase (Tardat et al. 2010).

To carry out the experiment, first I fused PR-Set7 to the Gal4 DNA binding domain (Gal4-DBD). This tag specifically targets the construct to GAL4 operator sites on DNA. I have also modified the reporter plasmids to carry GAL4 operator in the vicinity of the TetO sites (see Figure 17). Four new HEK-293-D cell lines were also created: one that expresses EBNA1, TetR:Orc6 and Gal4:PR-Set7, a second one with EBNA1, TetR:Orc6, Gal4-DBD, the third with EBNA1 and Gal4:PR-Set7 and a final cell line with EBNA1 and Gal4-DBD. I transfected these four cell lines with either the FR-TetO plasmid, or the FR-GAL4-TetO reporter plasmid, and assayed them in plasmid rescue experiments.

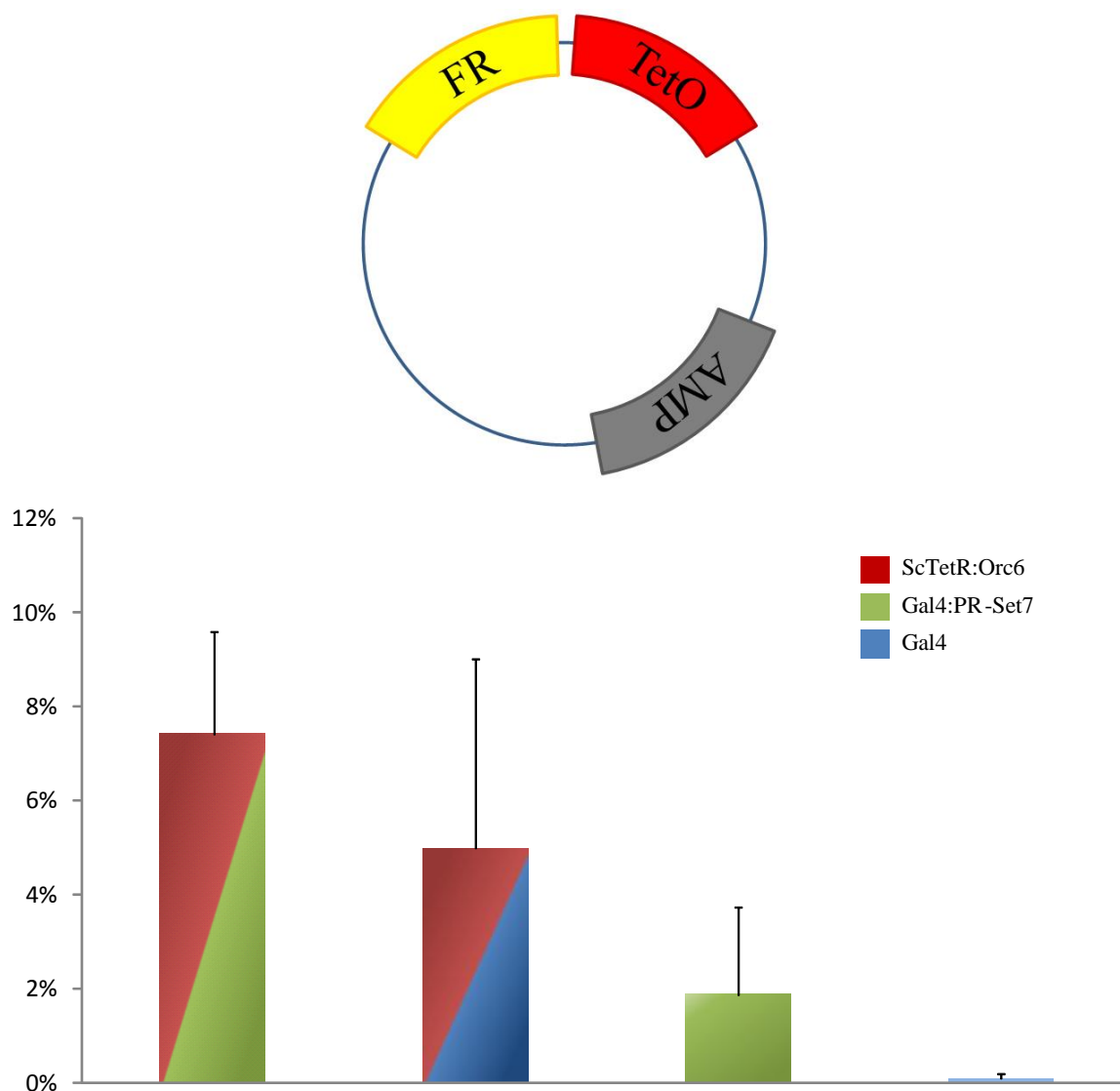


Figure 19. Effect of PR-Set7 expression on Orc6 dependent plasmids. Four cell lines were generated from HEK-293-D EBNA1 cells by expressing combinations of scTetR:Orc6, Gal4:PR-Set7 or Gal4-DBD as displayed by the color-coded bars. These cell lines were transfected with the FR-TetO plasmid depicted at the top. Plasmid rescue assay was performed with this setup, resulting in the rescue efficiencies presented for each cell line. The percentage is a ratio of the FR-TetO plasmids rescued compared to wild-type oriP plasmids rescued in the same cell line. N=3, error bars show SEM.

Results presented in Figure 19 show data obtained with an FR-TetO reporter, while Figure 20 depicts the same experiment using the FR-GAL4-TetO plasmid. As expected from the previous experiment, FR-TetO plasmids in cells expressing TetR:Orc6 show between 5-10% rescue compared to the wild-type oriP plasmid. This was not significantly changed by the ectopic expression of Gal4:PR-Set7 or Gal4-DBD, as these constructs are not targeted to the plasmid, and therefore are likely not involved in the rescue process, leading to similar results in the first two bars of Figure 19.

When expressing, but not directing PR-Set7 to the plasmids, there is a slight increase of plasmids rescued compared to Gal4-DBD only, as evidenced by comparing the last two bars of Figure 19. This suggests that overexpressing PR-Set7 might influence the replication competence of reporter plasmids, even when not specifically directed to origins. The plasmid rescue experiments using the FR-GAL4-TetO reporter plasmid on the other hand showed different results. Targeting PR-Set7 to plasmids yielded a high rescue level (15%), but I observed no rescue with the tag only. This result suggests that PR-Set7 shows no combined effect when both it and Orc6 are targeted to the plasmid, as in this case, rescue is only 2-3%, staying in the range of that of the two proteins individually. This means that the presence of PR-Set7 does not increase Orc6 mediated plasmid rescue. Also, the histone methyltransferase induces replication activity on the reporter plasmid in this assay. This is in line with previous experiments where PR-Set7 induced replication activity when targeted to a GAL4 operator cassette introduced into a chromosomal location (Tardat et al. 2010; Beck et al. 2012).

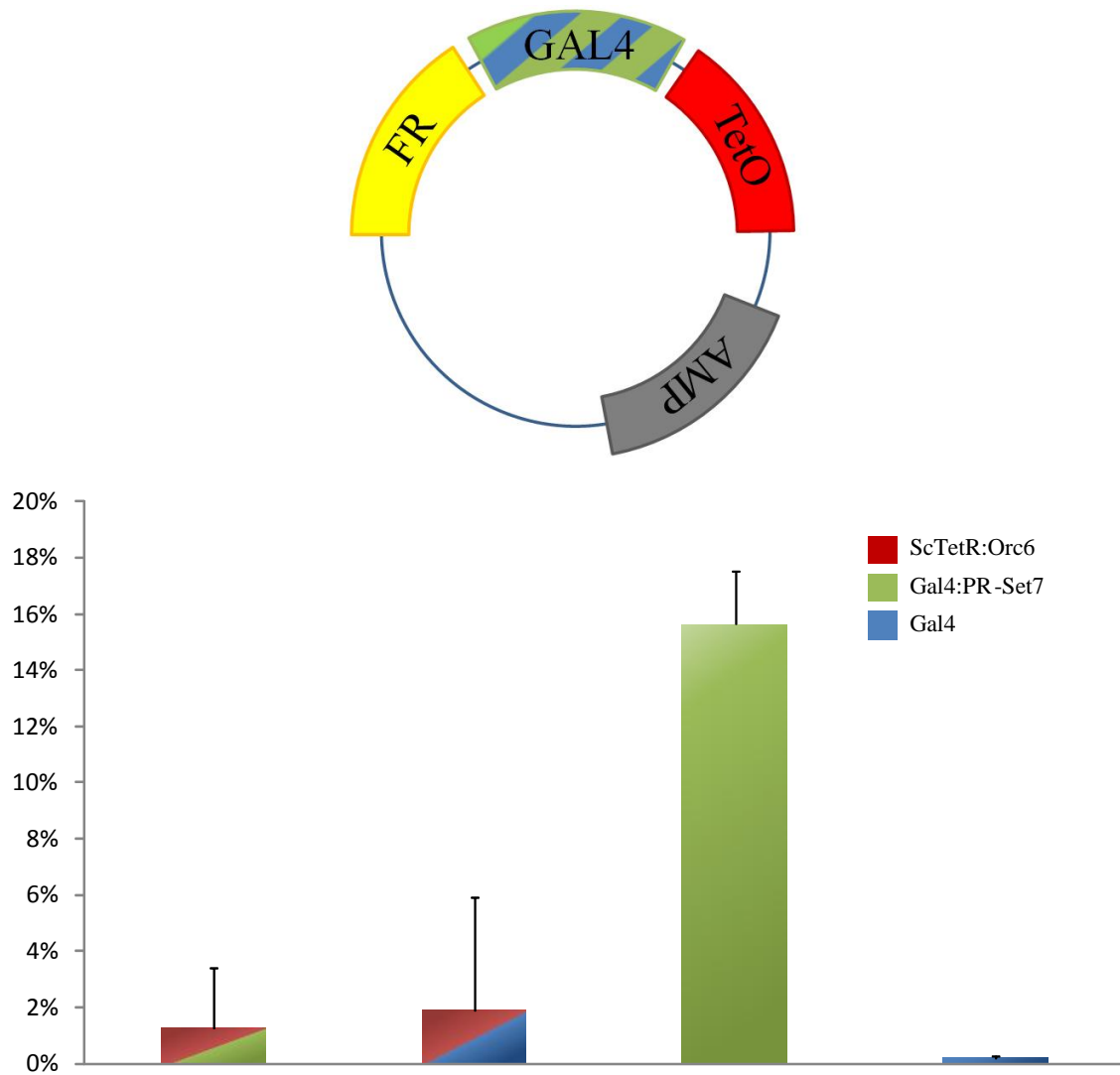


Figure 20. Effect of targeting PR-Set7 to Orc6 dependent plasmids. Cell lines expressing HEK-293-D EBNA1, and combinations of scTetR:Orc6, Gal4:PR-Set7 or Gal4-DBD as displayed by the color-coded bars, were used for this plasmid rescue assay. Contrary to the previous experiment, here the FR-GAL4-TetO plasmid depicted at the top was used for the assay, allowing Gal4 mediated targeting of fusion proteins to the origin site. The resulting plasmid rescue levels are presented as a ratio of FR-GAL4-TetO plasmids rescued compared to wild-type oriP plasmids rescued in the same cell line. N=3, error bars show SEM.

5.2.3 PR-Set7 enhances existing replication origins

The previous experiment indicates that targeted PR-Set7 induces replication. However, the impact of induced H4K20 methylation on an already existing replication origin is unclear. The experiments using the scTetR:Orc6 dependent origin did not show an increase in replication competence, therefore I wanted to determine if PR-Set7 has an impact on the more robust EBNA1 mediated replication initiation at the oriP origin. To test this I generated an FR-GAL4-DS plasmid, which, since it contains both the FR and the DS element of oriP, is only dependent on EBNA1 for efficient maintenance. Then I transfected the four cell lines from the previous experiment with this plasmid, and a wild-type oriP plasmid as control.

The experiment showed that PR-Set7 enhances EBNA1 dependent replication very efficiently, but only if TetR:Orc6 is not present in the cell. The tag itself has no detectable effect on EBNA1 function, whereas the histone methyltransferase can increase plasmid replication efficiency as much as tenfold if directed to EBNA1 mediated origins. Similarly to the previous experiment, if both PR-Set7 and TetR:Orc6 are expressed in the cell, plasmid maintenance efficiency is not increased, but reduced. Here, taking the wild-type oriP plasmid as 100% in the same cell line, the FR-GAL4-DS plasmid could only deliver approximately half as many colonies in the rescue assay.

In conclusion these experiments show information on the function of Orc6 and PR-Set7. Both can create new origins of replication, Orc6 is even able to do so if expressed as various truncated variants. Orc6 targeting to plasmids coupled with EBNA1 mediated plasmid retention forms a system which can stably maintain FR-TetO containing plasmids in human cells, independently of PR-Set7 mediated chromatin modification. PR-Set7, although being able to induce replication on its own, does not act in concert with Orc6 in the current experimental setup.

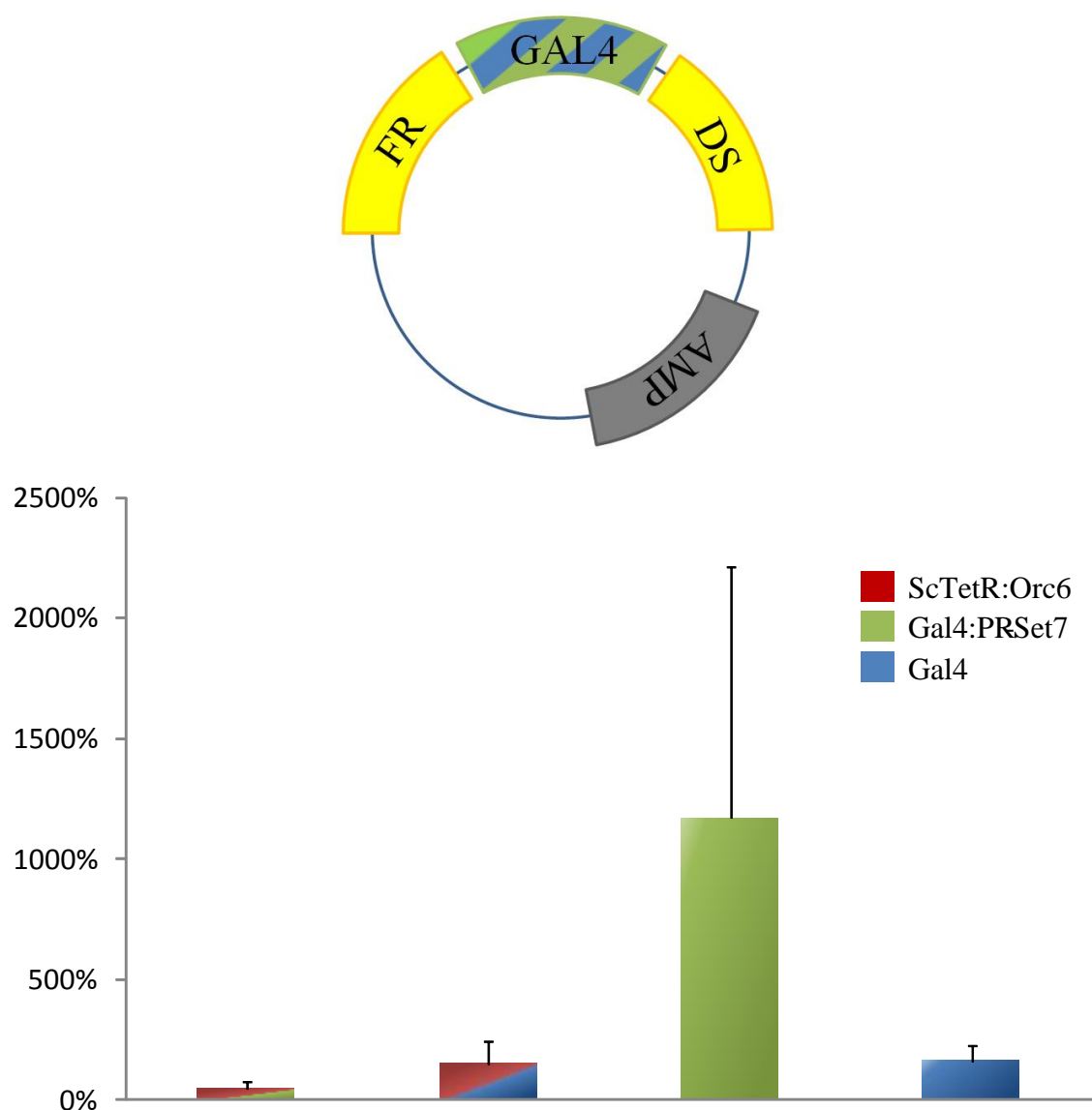


Figure 21. Effect of targeting PR-Set7 to EBNA1 dependent plasmids. Similarly to the previous experiment, here the same cell lines expressing EBNA1, and combinations of scTetR:Orc6, Gal4:PR-Set7 or Gal4-DBD as displayed by the color-coded bars, were used for plasmid rescue assays. This time the FR-GAL4-DS plasmid was tested, which is a solely EBNA1 dependent plasmid harboring GAL4 operator sites. The resulting plasmid rescue levels are presented as a ratio of FR-GAL4-DS plasmids rescued compared to wild-type oriP plasmids rescued in the same cell line. N=3, error bars show SEM.

5.3 Orc6 interacts with various other factors in the cell

After characterizing the mechanics of Orc6 interaction with Orc1-5 and elucidating the function of the protein with the plasmid rescue experiments, I took a new approach to even better understand Orc6 in the human cell. Using immunoprecipitation coupled to mass spectrometry, I mapped the interaction network of HsOrc6. The idea behind this was to get information on the protein's role outside replication initiation.

So far, there have been no fully successful attempts in establishing the function of Orc6 throughout the cell cycle. Orc6 has been seen already upon its discovery in humans to interact with unknown, non pre-RC proteins (Dutta & Dhar 2000). Further experiments showed that Orc6 depletion causes decreased replication, but also mitotic defects, multinuclear cells and multipolar spindles (Prasanth et al. 2002). The same effects were observed in *Drosophila* (Balasov et al. 2009), and avian cells (Bernal & Venkitaraman 2011). The function behind these phenotypes is not yet fully elucidated, although it has been established that DmOrc6 interacts with septins and is involved in cytokinesis; and that ScOrc6 works together with Rrb1, the depletion of which also causing mitotic defects (Chesnokov et al. 2003; Huijbregts et al. 2009; Killian et al. 2004).

To create the interaction network I immunoprecipitated Orc6 from nuclear extracts, using an Orc6 specific antibody covalently coupled to sepharose beads. As negative control, a non specific IgG antibody was employed. The positive control was an Orc2 immunoprecipitation, the results of which were already published by another group (Shen et al. 2010). I loaded the precipitated proteins onto specially prepared, low contamination polyacrylamide gels and stained them with Coomassie dye after electrophoresis. Gel slices were analyzed by mass spectrometry (LC-MS/MS at the Zentrum für Proteinanalytik). After receiving the obtained spectra, I identified the proteins present using the Scaffold 4 software. I have made three replicates of each experiment, and statistically evaluated the results to rule out background and non-specific hits. I subjected the final list to DAVID, a functional clustering program, to identify proteins of the same

pathway or those involved in the same cellular mechanism. To complement this, I also conducted a study using the STRING protein interaction prediction tool.

First, to test this experimental setup, I prepared an Orc2 immunoprecipitation and a corresponding IgG control, and analyzed it via mass spectrometry. This experiment was already conducted and published, and showed that LRWD1 interacts with ORC (Shen et al. 2010).

Accession number	Gene name	IgG	Orc2
gi 5453830	ORC2	3	53
gi 32483367	ORC3	2	52
gi 32454746	ORC4	0	7
gi 4505525	ORC5	0	32
gi 23097240	LRWD1	1	49

Table 1. Control experiment to identify known interactors of Orc2 via mass spectrometry. The table shows the number of total spectra obtained via mass spectrometry of an Orc2 immunoprecipitation versus IgG as control to test the validity of the experimental setup. The list of protein GI accession numbers and gene names only show members of the origin recognition complex, which were identified by this analysis. Although LRWD1 (OrcA) is not strictly a member of ORC, it was included as it was originally discovered using the same experimental setup in another study (Shen et al. 2010). Other factors identified are not shown, as they were also not reported in the Shen study, and therefore cannot be used to confirm the validity of this setup.

My results fully confirmed the ones by the Prasanth lab. Total spectra of LRWD1, along with the core ORC subunits were enriched in the immunoprecipitation, and nearly absent in the IgG control. As I did this control experiment to test the efficiency of the method, it was not prepared in triplicates and therefore no statistical analysis was possible, but a robust difference is still present when comparing the immunoprecipitation to the control in Table 1. Although other factors were also enriched in this immunoprecipitation (data not shown), Orc6 was not among them, confirming the weak interaction between Orc6 and Orc1-5.

After confirming that the experimental setup is functional, I also carried out the Orc6 immunoprecipitations, and subjected these to mass spectrometry.

Scaffold 4 identified 1563 different proteins in the samples. A protein was considered identified if it had at least two unique peptides not present in any other protein, each of

the peptides was identified with at least 95% accuracy from the obtained spectra, and the protein itself was more than 99% accurately identified from these peptides. These rather stringent conditions ensure a low false positive rate, but also filter out many specific hits, since if a single unique peptide of a protein is present in the sample, then that protein is detected, but excluded from these results.

From this list of proteins I collected the ones with spectrum counts significantly higher in the Orc6 immunoprecipitation samples than in the IgG control. Since the statistical method used here greatly influences the outcome of the experiment, I compared two different analysis methods.

First, to get an overview I calculated a simple fold change: if a protein had at least twice as many total spectra in the Orc6 immunoprecipitation than in the IgG, and this was reproducible in all three replicates, then it was considered significant. To rule out infinite enrichments and divisions by zero, all values of zero spectra were replaced with 0.5.

The strength of this analysis is its simplicity and the robustness of the members included. This method however does have weaknesses. The first of these is the inclusion of proteins with very few spectrum counts, since if 1-2 spectra were detected for the protein in the immunoprecipitation, and none in the IgG control, the protein would show up on this list, despite being on the detection limit in both samples. This is the case for factors like U1C or SFR2, which were barely detected in the immunoprecipitation, but because of the empty control, they are considered significant. The second limitation is excluding proteins, which have an unspecific binding to the beads or the control antibody. If such a factor provides a number of spectra in the IgG control, then it is more difficult to achieve twice as many in the Orc6 immunoprecipitations, so the protein might be excluded despite being an important interactor. Finally, experimental variance can also negatively affect the outcome of the analysis. If a protein does not meet the criteria in a single experiment, then it will be excluded, even if it provided robust results in the other two replicates.

Accession number	Gene name	IgG-1	IgG-2	IgG-3	Orc6-1	Orc6-2	Orc6-3
gi 12056971	ANAPC1	0,5	0,5	0,5	2	1	6
gi 15055539	RPS2	0,5	1	0,5	3	2	3
gi 1840467	MSH2	3	1	0,5	8	2	5
gi 22122445	CDC73	0,5	0,5	1	2	3	2
gi 34582345	DDX52	0,5	1	0,5	1	5	3
gi 37196760	EPPK1	2	0,5	1	4	1	7
gi 40807443	PRC1	0,5	2	1	1	5	2
gi 42476299	GIGYF2	0,5	2	1	2	7	9
gi 4506619	RPL24	0,5	0,5	2	1	2	4
gi 4507127	U1C	0,5	0,5	1	1	1	3
gi 50659095	DDX21	0,5	1	0,5	1	2	1
gi 52630326	CHD3	0,5	0,5	0,5	6	3	7
gi 55666362	RPRD2	0,5	1	0,5	1	9	6
gi 56676371	CPSF1	0,5	0,5	0,5	1	2	1
gi 56757608	RFC1	1	1	1	3	4	4
gi 57209883	PHF8	0,5	4	1	2	13	4
gi 58331218	ARS2	3	1	1	6	2	3
gi 6754472	KIF23	0,5	1	0,5	2	3	2
gi 119609849	SFRS2	1	1	1	2	2	2
gi 68533103	NUP153	0,5	0,5	0,5	1	1	1
gi 71361682	NUMA1	3	7	10	10	18	22
gi 4506691	RPS16	1	0,5	3	3	2	6
gi 7657427	ORC6	0,5	0,5	1	6	21	21
gi 7706501	WWBP11	0,5	0,5	0,5	1	2	2
gi 8922712	SEP11	1	3	0,5	3	7	1

Table 2. Factors identified by mass spectrometry with at least twice as many total spectra in all three Orc6 immunoprecipitations as in the control. Protein GI accession numbers and gene names are presented for proteins which were least twofold enriched in all three replicates in the Orc6 immunoprecipitation over the IgG immunoprecipitation. The numbers indicate the amount of total spectra obtained for each protein in the respective sample. Values of 0 are exchanged to 0.5 to allow fold change calculations. Rows marked in **bold** show factors, which are also present in Table 3.

Taking this into account I only used this protein list to get a quick overview of the proteins detected, and to check internal controls. This method yielded 25 hits, shown in Table 2, as significantly enriched proteins in the Orc6 immunoprecipitation over IgG. Orc6 as a positive control is present on the list with a high enrichment over background, and 15%, 39% and 25% sequence coverage in the three precipitations respectively.

However, other ORC subunits were not detected in the analysis, and also no pre-RC or pre-IC hits emerged. Since the Orc6-Orc1-5 interaction is weak and barely detectable with sensitive immunoblots, this is still acceptable, especially as the list shows various other factors which are enriched over the control. Also cell-cycle specific interaction might be weaker in the asynchronous extracts used.

To use a more powerful statistical method to determine Orc6 interactors, Fisher's exact test was also applied to the dataset with a p value cutoff at 0.05. As shown in Table 3, this yielded 61 proteins, one of the most significant hits being Orc6 itself, with a p value below 10^{-4} (Scaffold does not calculate exact p values below 10^{-4}). No other ORC components were found to be significant even with this method. This larger protein pool, while containing more than twice as many hits as the fold change analysis, contains false positives as well, while only 15 proteins are present in both lists (these are marked in bold in Table 2). For example, three different keratins were included using Fisher's exact test, although these show only a single spike in one of the samples, and do not provide reproducibly higher signals in all three experiments. As keratins are known as the most often appearing sample contaminations in mass spectrometry, despite being significant, they are likely to be false positives (Hodge et al. 2013). However for the integrity of the analysis and to use an unbiased approach, these are also included in the subsequent analyses.

After obtaining this final list of possible interactors, the next step is to turn this data into information, and address the question 'which other processes is Orc6 involved in besides replication initiation'. To achieve this, I used further tools of bioinformatics to characterize the functions of the proteins found, and to identify pathways and clusters from the list of hits.

Results

Accession number	Gene name	IgG-1	IgG-2	IgG-3	Orc6-1	Orc6-2	Orc6-3	P value
gi 158420731	CHD3	0	0	0	6	3	7	0,0001
gi 33350932	DYHC1	10	4	4	21	7	40	0,0001
gi 4506713	RPS27A	11	0	0	7	0	49	0,0001
gi 7657427	ORC6	0	0	1	6	21	21	0,0001
gi 119608214	SPTA	11	4	3	19	3	27	0,00017
gi 119573938	RPRD2	0	1	0	1	9	6	0,00018
gi 119595227	NUMA1	3	7	10	10	18	22	0,00038
gi 1195531	KRT16	0	2	0	16	2	0	0,00048
gi 156766047	GIGYF2	0	2	1	2	7	9	0,00098
gi 119613715	EPRS	1	2	3	7	3	14	0,001
gi 119610995	PRP8	2	4	1	8	2	15	0,0015
gi 12056971	ANAPC1	0	0	0	2	1	6	0,0023
gi 58530840	DSP	0	1	1	3	0	11	0,0026
gi 116063573	FLNA	9	9	4	18	5	24	0,0028
gi 62243332	MAD1	1	8	3	1	27	3	0,0039
gi 119593672	POLDIP3	3	23	7	13	34	14	0,0043
gi 32698700	PHF8	0	4	1	2	13	4	0,0043
gi 33636719	TIM44	0	0	0	0	7	1	0,0045
gi 15215421	VAR5	1	2	0	6	2	7	0,0047
gi 4758012	CLTC	7	1	3	10	1	17	0,0065
gi 13654237	PRKDC	20	12	14	31	6	39	0,0074
gi 12803709	KRT14	0	0	0	8	0	0	0,0089
gi 14150141	PDCD2L	0	0	0	0	2	5	0,0089
gi 21626466	MATR3	7	17	6	14	31	9	0,0092
gi 119604485	DNMT1	2	3	5	10	3	12	0,011
gi 119573397	GEF2	0	1	0	0	9	0	0,012
gi 1840467	MSH2	3	1	0	8	2	5	0,012
gi 34582345	DDX52	0	1	0	1	5	3	0,012
gi 14719392	CFL2	11	0	0	9	0	18	0,013
gi 54607053	GCN1	8	3	7	13	1	22	0,014
gi 11935049	KRT1	6	13	15	26	13	18	0,016
gi 119604500	RAVER1	5	23	13	13	36	17	0,016
gi 101943240	TF3C-1	0	0	0	0	0	6	0,017
gi 25777671	PPP1R10	0	0	0	0	4	2	0,017
gi 27754056	TUBB6	0	0	0	0	0	13	0,017
gi 119605748	TRAP1	3	3	3	4	2	16	0,019
gi 37196760	EPPK1	2	0	1	4	1	7	0,021
gi 15055539	RPS2	0	1	0	3	2	3	0,022
gi 119585644	PBR1	1	0	0	1	2	5	0,022
gi 4557365	BLM	0	0	1	0	4	4	0,022
gi 61743954	AHNAK1	0	0	1	0	1	5	0,022

Results

gi 119588370	ZNF289	0	5	0	2	13	0	0,025
gi 4506787	IQGAP1	18	19	5	34	7	24	0,026
gi 119568637	HDAC2	2	2	2	8	0	7	0,029
gi 12803479	HNRNPUL1	0	0	4	5	1	7	0,029
gi 119589485	TUBB4	21	26	0	30	17	24	0,032
gi 119594342	DDB1	3	2	3	13	1	5	0,032
gi 112382250	SPTB	11	1	9	15	4	18	0,033
gi 119613328	RFC1	1	1	1	3	4	4	0,033
gi 119568930	MDN1	0	0	0	1	0	4	0,034
gi 7706501	WWBP11	0	0	0	1	2	2	0,034
gi 12052826	RAB10	1	1	0	1	1	7	0,037
gi 119571569	MYO18A	0	4	1	1	5	8	0,038
gi 12017959	CDC73	0	0	1	2	3	2	0,039
gi 13786127	CDC42EP4	1	9	7	7	14	10	0,039
gi 20143967	KIF23	0	1	0	2	3	2	0,039
gi 119614476	ANLN	6	10	5	8	10	18	0,042
gi 126507451	CSE1L-2	1	1	2	5	1	6	0,045
gi 40217847	SNRNP200	8	5	3	11	6	12	0,047
gi 4503509	EIF3L	2	2	2	7	2	6	0,047
gi 119574685	CCAR1	3	6	2	10	5	7	0,05

Table 3. Factors identified by mass spectrometry from Orc6 and control immunoprecipitations with p values lower than 0.05 using Fisher's exact test. GI protein accession numbers and gene names are presented for factors, which were significantly enriched in the Orc6 immunoprecipitation over an IgG control. The number of total spectra in each sample is presented for each protein, followed by the calculated p value.

5.3.1 Functional clustering of Orc6 interacting proteins

The proteins identified in the previous chapter contain a number of interesting hits, but a list of 61 different proteins is hard to grasp. To overcome this, I used two different grouping methods to characterize the functions and cellular processes affected by the factors identified from the immunoprecipitation.

First, I grouped the significant hits from the Orc6 immunoprecipitation according to gene ontology (GO) terms. GO terms are a set of popular, standardized keywords, describing gene product properties (Ashburner et al. 2000). For example in the GO Biological Process database, Orc6 is associated with two keywords: “DNA metabolic process” and “DNA replication”.

For the grouping I used the online accessible DAVID (Database for Annotation, Visualization and Integrated Discovery) toolkit. DAVID is able to handle a gene list as input, collect the GO terms associated with the list members, and statistically analyze which terms are enriched compared to the human proteome as background using a modified Fisher’s exact test. This is done by the functional classification tool, which concluded that 46 out of 61 hits from the immunoprecipitation are associated with Biological Process GO terms, totaling to 63 statistically significant BP GO terms ($p < 0.05$), using the standard settings for the program. Although these keywords help to better understand functions of the Orc6 interactors, they still build up to a large, complicated pool of keywords. To further refine this, DAVID also offers a functional clustering tool, which groups similar GO terms together into clusters, giving an overview of the biological processes Orc6 interactors are involved in.

This analysis created 6 significant groups, shown in Table 4. The most prominent of these is the term “mitotic cell cycle” which shows by far the highest enrichment score. The score is the average $-\log_{10}$ of the p values of the cluster members, meaning that the average p value is $10^{-4.17}$ or 6.76×10^{-5} for the GO terms in the cluster of “mitotic cell cycle”. So the keywords similar to “mitotic cell cycle” are highly overrepresented in the list of immunoprecipitation hits, compared to the human proteome as background. The cluster itself contains keywords related to mitosis and cell cycle control. Both of these processes are expected for Orc6 interacting proteins, despite no other pre-RC components

were entered into the analysis. This result underlines the importance of the so far poorly understood mitotic function of Orc6, which was first observed in human cells by the Prasanth lab more than a decade ago, but still remains uncharacterized (Prasanth et al. 2002). Proteins like the Bloom syndrome protein (BLM), Mitotic arrest deficient-like 1 (MAD1), Nuclear mitotic apparatus protein 1 (NUMA1), Anaphase promoting complex subunit 1 (ANAPC1) are among the most prominent interactors of Orc6, while even mitosis related scaffolding proteins like Septins (SEPTA, SEPTB) and Anillin (ANLN) are present in the list.

Cluster with most significant gene ontology term	Enrichment score
mitotic cell cycle	4.17
macromolecular complex assembly	2.19
negative regulation of organelle organization	1.78
cellular protein localization	1.72
RNA splicing, via transesterification reactions	1.55
DNA metabolic process	1.32

Table 4. Functional annotation clusters of Orc6 immunoprecipitation hits. This table contains biological process GO term clusters significantly enriched among Orc6 interaction partners compared to the human proteome as background. Each of the six clusters contains GO terms similar to the name of the cluster. The enrichment score is the combined (-lg) p value of the respective cluster's individual GO terms.

The other expected group, the “DNA metabolic process” cluster is still considered significant for this study, with an enrichment score of 1.32 ($p=0.047$). This contains keywords connected to DNA repair, and cellular stress response. Orc6 being an efficient DNA binder, part of the pre-RC, and a known CDK target fits well into this group, and may have a yet to be pinpointed role in cell cycle control upon DNA damage.

The rest of the keywords in Table 4 are less meaningful in a biological way. “Macromolecular complex assembly” is attributed to proteins, which form larger complexes. As this is a common feature for many proteins from tubulins to ORC itself, this keyword cluster has little value interpreting the functions of Orc6 interactors. Similarly, “negative regulation of organelle organization” denotes proteins, which are involved in the disassembly of any structure in the cell, like protein complexes, cellular components or the cytoskeleton. Again, this general term can mean different things

depending on the context, and therefore has little value. “Cellular protein localization” is a keyword cluster containing GO terms of protein import, transport, targeting and localization, and other similar keywords. This can mean transcription factors, DNA damage cascades or signaling pathways alike, anything which involves the localization of proteins. As the term does not identify the type of pathway these proteins are involved in, this biological process is yet again a too general attribute of Orc6 interacting proteins.

Lastly however, “RNA splicing” is an unexpected, interesting cluster. It contains keywords, which are connected to transcription and the processing of mRNA. Transcription related proteins are present in the immunoprecipitation hits, but so far Orc6 has not been connected to transcription. Looking at the interacting gene names hnRNP-UL1, CCAR1, WBP11, PRPF8 and SNRNP200 (snRNP-U5) are all transcription related gene products. They are however involved in very different processes, and although Orc6 has been shown to be involved in ribosome biogenesis (Killian et al. 2004), transcription related functions of this protein are not known from other studies to date.

Apart from the DAVID analysis, I also used an independent method to integrate Orc6 and its newly found putative interactors into cellular pathways. For this I used the STRING program, which offers to map known interactions between proteins and to predict new ones (Franceschini et al. 2013). It uses the genomic context of genes, published high throughput experiments, co-expression data and also mines the text of scientific publications to gather knowledge on protein interactions. Using this data, it takes a list of input proteins, and assembles them into networks. Factors taken from Table 3 are assembled into the network presented in Figure 22.

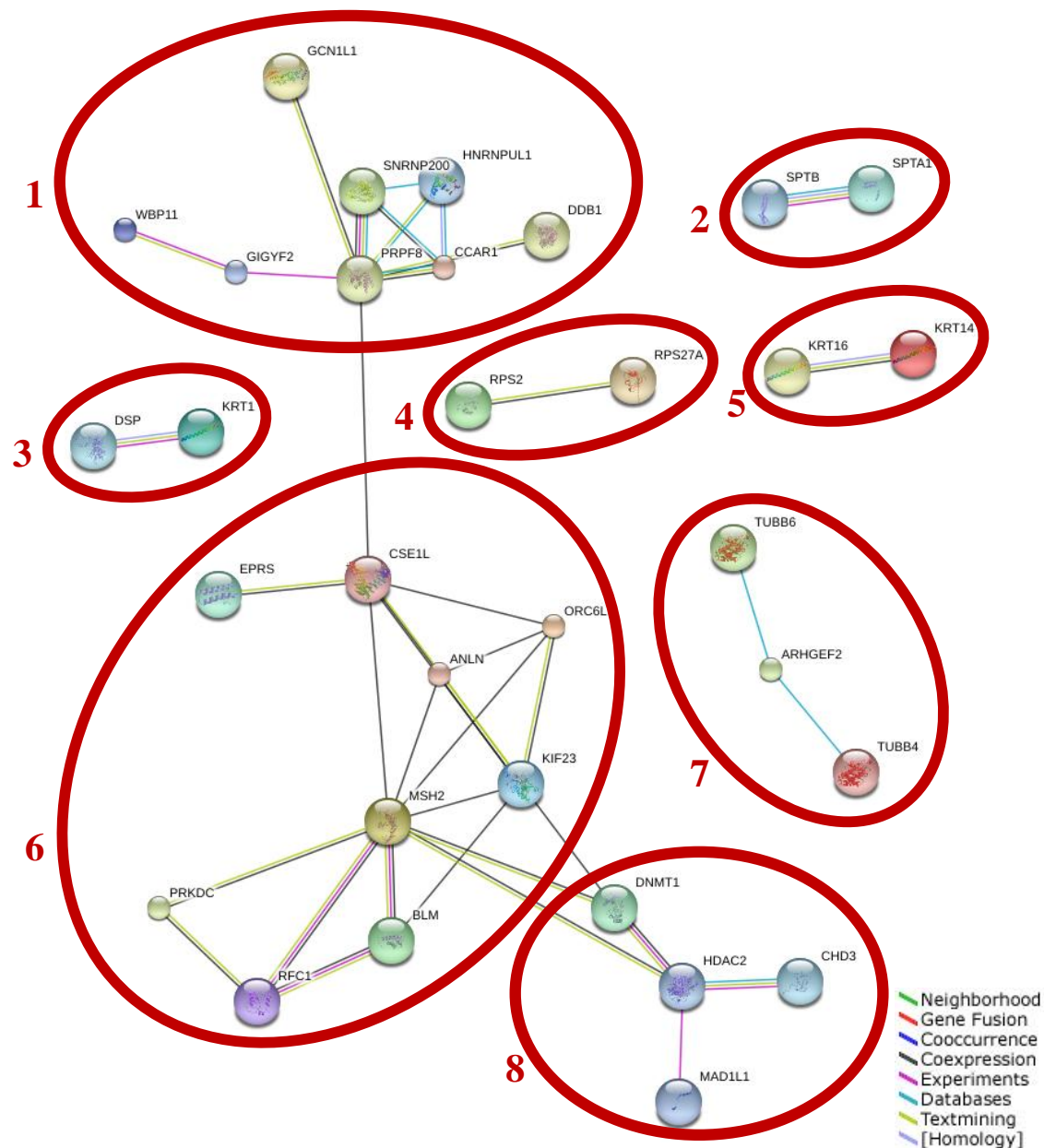


Figure 22. STRING network analysis of Orc6 interaction candidates. Interaction network generated from the list of Orc6 immunoprecipitation hits by the STRING-DB algorithm using its standard settings (Franceschini et al. 2013). Out of the 61 input factors 32 are displayed which had at least one connection to another factor. The links between the proteins are color coded as shown in the bottom right. The 8 clusters marked in red indicate proteins involved in similar processes.

Out of the 61 members, STRING found 32 that had a connection to at least one other factor. The interaction map displays one mayor cluster with 21 factors (groups 1, 6 and

8), one with 3 (group 7), and 4 other clusters with pairs of proteins in each one (groups 2-5). The main cluster can be divided into three subgroups as presented in Figure 22.

Analyzing these 8 groups in detail, group 1 contains transcription factors, similarly to the “RNA splicing” cluster identified by DAVID. All five proteins from the “RNA splicing” cluster are present in group 1, with GCN1, GIGYF2 and DDB1 added to them. GCN1, based on its Uniprot identifier (Q92616) is factor involved in transcription control, and has been most extensively studied in yeast. Based on this function, it fits to other members in this group. GIGYF2 on the other hand is a poorly characterized protein, with no function known to date. It was shown in two separate studies to be involved in either insulin-like growth factor signaling (Giovannone et al. 2003) or Parkinson’s disease (Lautier et al. 2008). However, the former study only presents evidence on GIGYF1, a related, but different factor, and only in yeast. And the Lautier paper was challenged by a follow up study, which found no connection to Parkinson’s. The reason why it is included in the group is its published interaction with PRPF8 and WBP11 which was conducted during a study of yeast mRNA processing pathways (Ash et al. 2010). A further connection to WBP11 is found in a paper detailing GYF domain function, using GIGYF2 as an example (Kofler et al. 2005). Based on the little evidence available on this protein, it may be involved in mRNA processing, and therefore might fit to the group of splicing factors. Finally, DDB1, although predicted to be part of the mRNA processing group, is in reality a factor involved in DNA repair, which is why it was included in the “DNA metabolic process” cluster in the DAVID analysis. The reason why it is connected to PRPF8 is that these two proteins were found to be co-expressed in a *Drosophila* mitosis study, and in another paper on macrophage transcription regulation (Somma et al. 2008; Patino et al. 2006). More importantly, DDB1 is known to attach to p21, and p21 levels are shown to be affected directly by Orc6 activity in human colon cancer cell lines (Abbas et al. 2008; Gavin et al. 2008).

Group 2 contains two members of the septin family. These proteins are required for cytokinesis by providing the scaffolding required for chromosome segregation and cell division. Their depletion leads to mitotic defects, similarly to those of Orc6 (Spiliotis et al. 2005), also a study from *Drosophila* indicates that Orc6 interacts with Pnut, a member of the septin family. Similarly, group 3 and 5 also contain structural proteins. Although

the role of septins fits well into the profile of Orc6, desmoplakin (DSP) and the three keratins should be evaluated with caution. Structural proteins are often present in large numbers in cellular extracts, and can also be present as contamination in the samples as discussed in the previous chapter. Since I found no other evidence in literature of Orc6 interacting with keratins, groups 3 and 5 will not be discussed in more detail.

Group 4 on the other hand fits well to a proposed role of Orc6 in ribosome biogenesis (Killian et al. 2004). Both RPS2 and RPS27A are part of the 40S ribosomal subunit (Kenmochi et al. 1998), and provide the long sought link between ORC and ribosome biogenesis (Thomae et al. 2008).

Looking at group 7, the last small group identified by STRING, leads again to mitosis. The two tubulin genes (TUBB4 and TUBB6) and ARHGEF2, a GTPase, all take part in microtubule formation (Ren et al. 1998). Experimental evidence confirming the interaction of Orc6 with microtubules comes from microscopy experiments, which showed that Orc6 localizes to the midbody, which is itself made up of bundles of microtubules (Skop et al. 2004). ANLN and KIF23, although present in group 6, are also related to this process: the former is a scaffolding protein required for cleavage furrow formation (Oegema et al. 2000), while the latter is a plus-end directed microtubule-dependent motor protein expressed in mitosis (Nislow et al. 1992).

Finally looking at groups 6 and 8, they essentially form one larger cluster, which I divided into two small subgroups. The reason behind the splitting is that group 8 contains a well-characterized assembly of proteins, whereas group 6 is more heterogeneous. In group 8, DNMT1 and HDAC2 are both widely studied epigenetic factors, involved in transcriptional regulation, DNA damage response and DNA replication (Jin & Robertson 2013; Krämer 2009). The two factors have also been published to interact with each other in several independent studies from both man and mouse (Rountree et al. 2000), HDAC2 was also shown to bind to MAD1, an important mitotic regulator (Yoon et al. 2004; Rottmann et al. 2005), and CHD3, a member of the Mi2/NuRD chromatin remodeling complex (Hakimi et al. 2002; Kunert et al. 2009).

Group 8, in contrast to group 6, contains a more disperse pool of DNA replication and repair related proteins, which do not associate to a single pathway. Most of the links between the factors within are marked with black color, indicating similar coexpression

patterns in different organisms, but no direct functional link. However, looking at the individual members, BLM, RFC1 and ORC6 are all involved in different aspects of DNA replication. The direct interaction proposed by this analysis would hint at a cooperation of these proteins in replication control. Repair proteins are represented by MSH2, a central member of the cluster, which is also a post-replicative DNA mismatch repair protein and oncogene (Iaccarino et al. 1998). Secondly, the product of the PRKDC gene is a crucial component of double-strand brake repair. Lastly, there are two hits which are functionally unrelated to the other members of the group, and are likely to be experimental artifacts: CSE1L is a member of the nuclear importing complex, and EPRS is an aminoacyl tRNA synthetase.

In summary, the results in Table 4 and Figure 22 introduce novel interactors of Orc6 and connect this protein to various cellular processes other than DNA replication. From the interaction network Orc6 appears as a prominent mitotic factor, likely to be involved in multiple aspects of this process, and is also linked to ribosome biogenesis and DNA repair.

6. Discussion

6.1 Orc6 is an abundant ORC component

Before any analysis of Orc6, my priority was to ensure that this factor is readily detectable and present in the studied cell types. For this, I quantitated Orc6 levels in the cell lines used, which allowed subsequent experiments by showing a clear presence of this protein in my samples (Figure 5).

Looking at the data in this figure, it is apparent that Orc6 is about ten times more abundant in the cell than Orc2. Given that ORC itself contains exactly one of each subunit (Sun et al. 2012), Orc6 shows no sign of degradation as opposed to Orc1 (Ritzi et al. 2003), and Orc6 saturates on DNA even more quickly than Orc2 (Figure 10), it would be expected that Orc2 and Orc6 are expressed to similar levels. Since Orc6 is about tenfold more abundant than Orc2, this either means that large amounts of Orc6 are needed for the Orc6-Orc1-5 interaction, or that Orc6 might have additional functions for which the other molecules are required. The former scenario is less probable, as literary evidence from knock-down of Orc6 in HeLa and human colon cancer cells states that a five-fold reduction of Orc6 levels still supports replication, but leads to polyploidy and multinucleated cells (Prasanth et al. 2002; Gavin et al. 2008). This means that pre-RC formation and replication initiation is still possible with decreased Orc6 levels. However, the less well understood mitotic functions become defective if protein levels are reduced. Based on this, the excess Orc6 seen in Figure 5 is indeed required for the suspected non-pre-RC related functions of this factor.

To further support this, using knowledge from literature it can even be estimated how many Orc6 molecules are actually required for replication initiation. The following calculation is not accurate, but it aims to estimate the number of ORC molecules in a human cell according to the current models:

There are approximately 30.000 replication forks in a human cell (Nasheuer et al. 2002; Aladjem 2007) these would require half as many functional pre-RCs, due to the bidirectional nature of human DNA replication. Furthermore, the number of assembled

pre-RCs are assumed to be significantly higher than the actual origins that fire every cell cycle, with approximately 10-20% pre-RCs used in each S phase (Papior et al. 2012). This would mean that according to current models, at least 75.000 ORC molecules are expected to be present in every cell at pre-RC sites alone, not including protein turnover or other cellular processes.

After this theoretical calculation, looking at the experimental data in Figure 5 again, Raji cells contain 14.5495 femtomole Orc2 molecules per million cells. Calculating from this would lead to 8730 Orc2 molecules in a single Raji cell, while for HEK-293-D cells this number is 20929 according to the quantification results. Of course the accuracy of quantified immunoblots is not expected to be this high, so these results should not be treated as hard numbers. Instead, the magnitudes of these numbers are more meaningful: there are around 1×10^4 and 2×10^4 Orc2 molecules in Raji and HEK cells respectively, which is in the ballpark of the 7×10^4 theoretical estimate, albeit nearly a magnitude lower. This is either due to the inaccuracy of the model, which is still lacking quantitative results despite concentrated efforts (Schepers & Papior 2010), or the experimental setup. Possible fallacies in the latter include the inaccurate measurement of the purified proteins, cell-counting deviations, and extract preparation and sample handling errors. Still, the same measurement yielded around 2×10^5 Orc6 molecules per cell (207634 in Raji and 181703 in HEK-293-D), which is in range of the predicted amount and possibly more.

These results fit well to a similar quantification study, which found 7.5×10^4 Orc2 molecules per cell in the hamster CHO cell line, and also showed that Orc4 and 5 are expressed to similar levels (Wong et al. 2011). The obtained data also confirm that Orc6 is an abundant ORC component in the cell, which enables it to take part in multiple cellular processes. This also serves as the first step in characterizing Orc6, as its abundance allows all subsequent analyses.

6.2 Orc6 has two distinct ways of attaching to Orc1-5

The excess Orc6 discussed in the previous chapter may have multiple functions. As I outlined in the aims of the project, I started the characterization of Orc6 by looking at the mechanism of its attachment to Orc1-5, and its DNA binding ability.

I have found that since Orc6 is able to bind a DNA segment as small as 200bp (Figure 9), without the aid of other ORC components (Figure 11 and Figure 12), it might firstly act as a protein that helps direct pre-RC factors to DNA. In other words, according to the experimental evidence presented in this work, Orc6 might be able to create binding sites for ORC. This is supported by Figure 18, which shows that Orc6 alone is able to create functional origins of replication if targeted to specific sites on chromatin. This is possibly achieved by recruiting Orc1-5, or other pre-RC factors to itself to allow replication initiation. This suggests that the abundance of Orc6 might be required to bind sites for future pre-RC assembly. Since Orc6 is more abundant than Orc1-5 and it is able to bind DNA without the aid of other ORC components, it might bind several possible origins, which then compete for Orc1-5. This mechanism would be similar to that of ORC itself, which attaches to DNA at several sites, but only some are able to recruit the complete pre-RC in each cell cycle (Papior et al. 2012).

Secondly, Orc6 is able to form a complex with Orc1-5 and Cdc6 already before DNA binding (Figure 6), and binds as the holocomplex to DNA. This means that the Orc1-5-Orc6 interaction can occur both on DNA and in solution. The two mechanisms converge on the point where ORC is assembled on DNA.

Figure 11 and Figure 12 approach the two-way binding mechanism from a different perspective. These figures provide evidence on Orc6 forming a complex with Orc1-5 on DNA or in solution. Here, the two experiments show that the absence of Orc1-5 lessens Orc6 binding, as the Orc6 signal is weaker in the Orc2 depleted lane than in the IgG depleted lane. Secondly, adding excessive Orc6, but not BSA to the plasmid-binding assay enhances Orc2-DNA binding. So the binding of Orc2 and Orc6 to DNA mutually

enhance each other, while the two proteins are still able to bind DNA in the absence of another to a lesser extent. Given that I also show the complete ORC-Cdc6 complex in solution in Figure 6, a convenient explanation for all these experiments would be that Orc6 is able to bind Orc1-5 both on DNA and in solution. This model extends our knowledge on the so far poorly understood Orc6-Orc1-5 interaction and is also supported by a another study published while this project was in progress (Ghosh et al. 2011).

6.3 Orc6 has domains that control cellular localization and ORC interaction

In order to better understand the mechanisms of Orc1-5 attachment and DNA binding of Orc6 outlined in the previous chapter, I also conducted domain mapping analyses, especially because other publications on this topic reported contradicting results to date. Although there are three studies published exploring Orc6 domains, none of them provided the full picture. First, a study was conducted in *Drosophila* using Orc6-S72A, Orc6-K76A and Orc6 Δ C (Balasov et al. 2007). In another publication, a systematic domain mapping was performed using HsOrc6 with regards to cellular localization, along with the study of the predicted phosphorylation site at T195 (Ghosh et al. 2011). The final publication in this respect was the crystallization of the middle of Orc6, which identified Q129, R137 and K168 crucial for DNA binding (Liu et al. 2011).

These studies provide valuable information, but an incomplete picture of Orc6 domains. DmOrc6 for example proved to function differently than HsOrc6. DmOrc6 is indispensable for ORC-DNA binding in fruit flies, a feature distinct from any other species observed. The S72A and K76A mutations, although proven to be required for DNA binding of Orc6 in *drosophila*, had no such effect in human proteins (see Figure 17 and data not shown). Also, the C terminal deletion mutant in *Drosophila* presented no

change in binding to Orc1-5, contrary to Figure 16. A possible explanation for the different behavior is the large evolutionary distance between HsOrc6 and DmOrc6. As seen in Figure 23, HsOrc6 is most closely related with its homologues in primates, and other vertebrates. Insects however, are very distantly related to this group, and are even further away than the Orc6 proteins of plants, despite both insects and primates being metazoans. Therefore, the large evolutionary distance between DmOrc6 and HsOrc6 can account for the observed difference in the effect of mutants.

Contrary to the *Drosophila* study (Balasov et al. 2007), my results confirmed, and further developed the findings of the two 2011 papers on this topic (Liu et al. 2011; Ghosh et al. 2011). Namely I confirmed the activity of the consensus nuclear localization signal of Orc6 studied by the Ghosh laboratory (Figure 17), and provided evidence that S72 and K76 do not affect DNA binding in humans as stated in the Liu paper (Figure 18 and data not shown). Further, I identified a new domain of Orc6: the Orc1-5 interaction domain in the C terminus (Figure 16). This was achieved by first creating 7 different Orc6 derivatives (detailed in Figure 14) and testing their nuclear localization (Figure 17), interaction with Orc1-5 *in vitro* and *in vivo* (Figure 16 and Figure 17, respectively) and ability to create origins of replication (Figure 18).

Concluding these experiments, I have established that Orc6 requires its C terminus to interact with Orc2 *in vitro*. Surprisingly, C terminal deletion does not alter the nuclear localization of Orc6, it even co-localizes with Orc2 *in vivo* despite the absence of interaction *in vitro*. A possible explanation for this observation is that although Orc6 requires its C terminus to interact with Orc1-5 directly, an indirect interaction through another domain might still be possible *in vivo* via additional factors present at the pre-RC. These hypothetical proteins which help recruit Orc6 to the pre-RC *in vivo* are likely to be missing from the *in vitro* origin recognition complex, as even ORC-Cdc6 complexes are challenging to assemble in solution (see Figure 6).

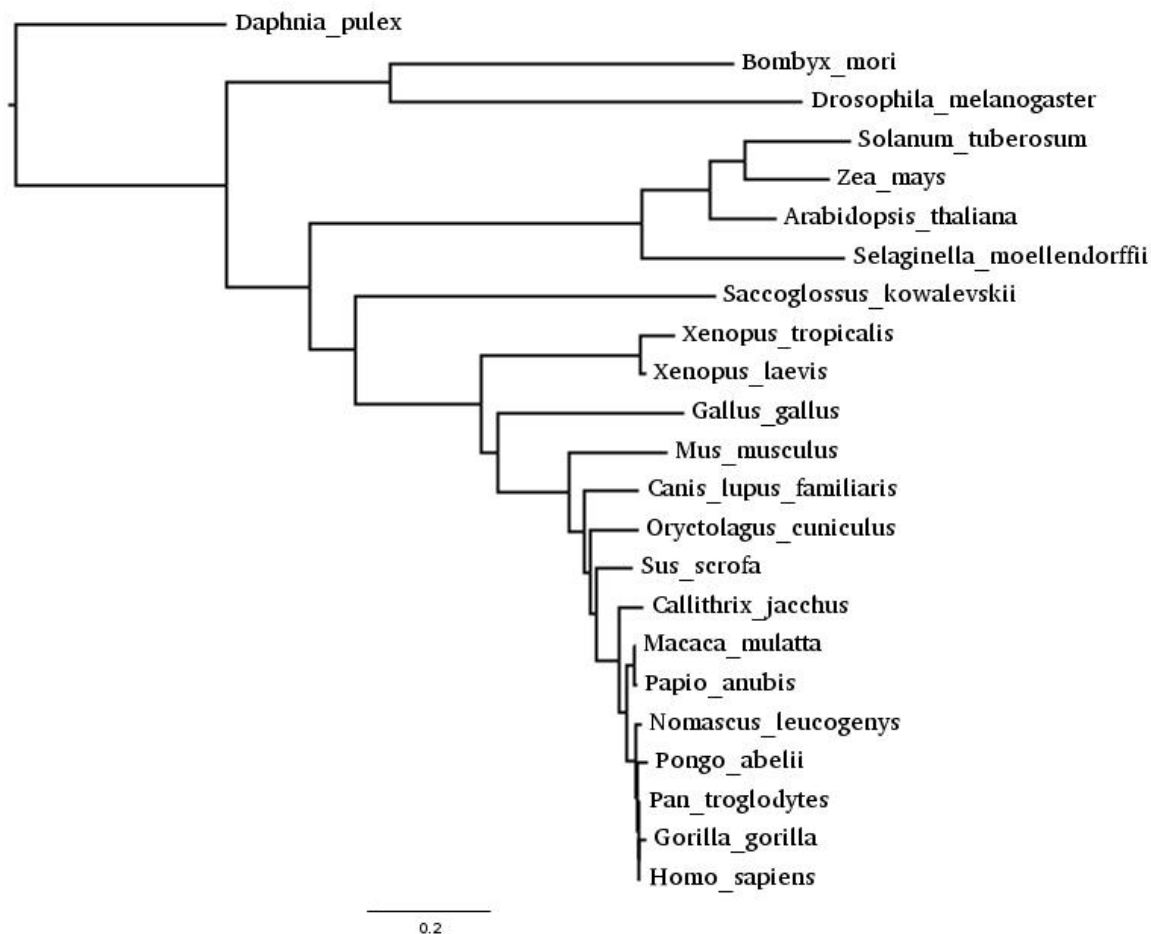


Figure 23. Evolutionary conservation of Orc6. Homologues of the HsOrc6 protein sequence were identified by BLAST, and plotted using the built-in distance tree option, with the recommended settings for both programs (McGinnis & Madden 2004). The branch lengths correspond to the sequence divergences of each node from their hypothetical common ancestor using the fast minimal evolution algorithm (Desper & Gascuel 2004). The tree is rooted in the middle of the longest branch. For this image, model organisms and species representing large taxonomic groups found were chosen out of the total 112 HsOrc6 homologues found.

Looking at the microscopy experiments in more detail, evidence can be found for this second proposed pre-RC binding site of Orc6 being at AA125-181. In Figure 17 the AA125-203 deleted construct is present largely in the cytosol, similarly to the AA50-203 deleted variant. Surprisingly, removing only the nuclear localization signal (AA181-200) leads to an only partially cytosolic localization, with the remaining nuclear Orc6 co-localizing with Orc2. This means that missing AA125-181 changes cellular localization of Orc6 from partially cytosolic to dominantly cytosolic. Since I reason in the previous chapter that the Δ NLS mutant is still able to enter the nucleus in complex with other factors, the combined deletion of the NLS and AA125-181 might block the binding of other factors to Orc6, leading to the observed dominantly cytosolic localization. The area AA125-181 is a prime candidate for pre-RC interaction, as it also contains the $\alpha 3$ - $\alpha 6$ helices and all three essential amino acids required for origin binding of HsOrc6 (Liu et al. 2011).

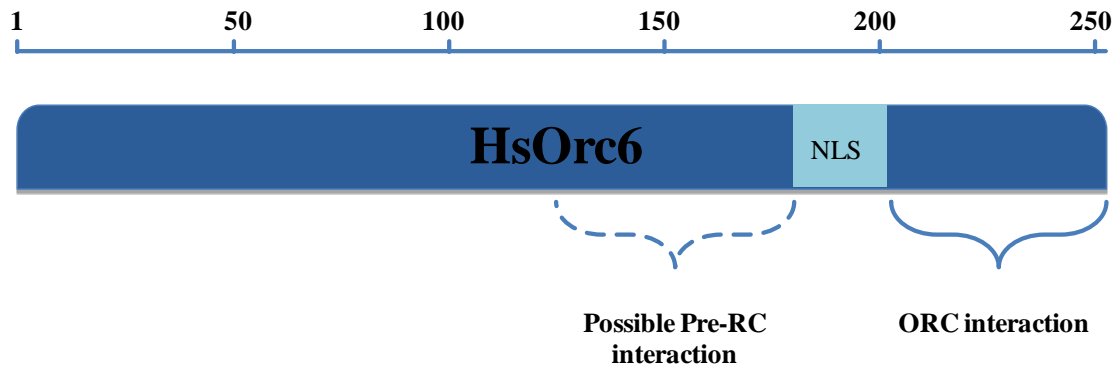


Figure 24. Model of HsOrc6 interaction and localization domains. Model depicting the results of Orc6 domain mapping experiments with the scale on top representing amino acids in the protein. *In vitro*, the C terminus (AA203-252) is required for Orc6-ORC interaction. *In vivo*, deletion of the C terminus has no effect on Orc6 localization, therefore another domain must be also responsible for Orc6-pre-RC interaction. N terminus deletion also has no G1 localization phenotype. Deletion of the consensus NLS sequence leads to partially cytosolic presence, while deleting AA125-203 or AA50-203 shows a prominently cytosolic localization. Based on these results, the second Orc6 targeting domain might be present in the AA125-181 region, which has already been shown to be important for DNA binding (Liu et al. 2011).

Another explanation could be that despite AA181-200 having a consensus NLS signal, a second NLS is present at AA125-181, or the observed signals might simply be overexpression artifacts. However, a previous study already tested the former possibility, and found no change in localization if a small deletion was present anywhere in the AA125-181 region, but also found a similar, only partially cytosolic localization phenotype if the consensus NLS sequence was removed (Ghosh et al. 2011; Kalderon et al. 1984).

In conclusion, my results show that the C terminus of Orc6 is required for direct interaction with ORC, and I also found evidence of a second pre-RC binding site, which might be located in the region AA125-181. For nuclear entry, Orc6 uses its own NLS, but in the absence of this, it is still able to localize specifically to Orc2 in the nucleus via other means, possibly by its additional pre-RC interaction domain. These experiments, together with the previous chapter, conclude the mechanical analysis Orc6-ORC and Orc6-DNA interactions, completing the first goal of this study.

6.4 Derivatives of Orc6 are sufficient to create origins of replication

Moving on from the first goal of this project which deepened our knowledge on the mechanical interactions between Orc6 and ORC, and the impact of this interaction on DNA binding, I continue by addressing the second set of problems outlined in the aim of this study. Here, I summarize my results on Orc6 function in the living cell.

Although in a previous paper our group has already shown that Orc6 is able to create functioning origins of replication (Thomae et al. 2011), this study further elucidates that even parts of the protein are sufficient for this process. In fact, all Orc6 derivatives tested were able to create weak, but functional origins of replication, as seen in Figure 18. This is possibly the consequence of Orc6 having two separate pre-RC interaction domains, as discussed in the previous chapter. All the Orc6 variants tested had at least one of these

intact (either the C terminus or AA125-181), so they should interact with pre-RC factors despite mutations or deletions in other domains. Since targeting was achieved by the fused scTetR tag, no DNA binding ability was necessary for the Orc6 part itself. Therefore it is no surprise that all seven variants tested provided similar amounts of rescued plasmids. Although minor differences are present between the constructs, these are not robust enough to be conclusive. 2-3% differences are not justified from the experimental setup; therefore I consider all constructs equally effective in inducing replication. However, I do consider the highly reproducible 0 colony result of the tag only construct to be markedly different from the 1-2% replication, as here the difference means either not replicating at all, versus inducing replication. Therefore the negative control is clearly surpassed by all constructs tested.

It could also be argued that any protein with a certain minimum length can support replication. This however was disproven by our lab previously when we showed that certain variants of HMGA1a are able to support replication, but not others (Thomae et al. 2011).

Despite the Orc6 constructs being on average around 20 times less efficient in inducing replication than EBNA1, I still conclude that Orc6 is sufficient in inducing replication on plasmids, and even derivatives of the protein are able to do so, as evidenced by often hundreds of plasmids rescued after each assay. Therefore, linking parts of HsOrc6 to scTetR is a possible way to create self-replicating plasmids for use in gene therapy.

Attempts at boosting replication efficiency by co-targeting PR-Set7 to the initiation sites have not been successful. PR-Set7 alone, and Orc6 alone are able to induce replication on plasmids, but no combined effect is observed. A possible explanation for this is that oriP is bent and twisted during the initiation process, and then melted during pre-RC formation (Bashaw & Yates 2001). In the presence of one targeted protein, this is apparently still possible, but if two tagged fusion proteins are targeted to the site, steric hindrance might disallow efficient replication initiation. Still, Figure 20 and Figure 21 show that although PR-Set7 is not able to boost Orc6 replication competence, it is alone able to induce replication initiation, and can boost EBNA1 dependent replication initiation if tagged Orc6 is not present in the assay.

In conclusion, this study found that Orc6, and its derivatives, can function in the cell by creating origins of replication. These origins are independent of the newly found PR-Set7 mediated replication initiation. These results shed more light on Orc6 function and fulfill the second goal of this project.

6.5 Orc6 is a versatile cell cycle regulation factor

After discussing the mechanical interaction between Orc6 and members of its complex, and studying how the protein works in vivo, I set out to meet the final goal of this project and map the interaction network of the protein to find new pathways it is involved in.

The fact that HsOrc6 is involved processes outside replication initiation was already proposed upon its discovery (Dutta & Dhar 2000). Shortly afterwards it was shown by microscopy that depletion of Orc6 from human cells causes defects in G1 but has even bigger effects in mitosis (Prasanth et al. 2002). The results presented in this thesis further underscore the importance of these findings, and characterize Orc6 not as a dominant factor in pre-RC formation, but a more versatile nuclear component also involved in cell cycle regulation.

By employing immunoprecipitation coupled to mass spectrometry, I was able to identify a large number of Orc6 interactors, the final list of which are shown in Table 3. From the list, pre-RC factors are mostly missing. This surprising result is probably due to the weak interaction between Orc6 and Orc1-5. Supporting this, doing the same experiment using Orc2 I was able to clearly identify most ORC components (Table 1), and even LRWD1 (OrcA), a pre-RC related factor originally indentified by the same approach in another lab (Shen et al. 2010), but not Orc6.

As the expected hits were not present in the Orc6 hit list, but many other factors were, I employed grouping algorithms to identify cellular processes and pathways that are connected to these factors. Both functional clustering (Table 4) and STRING network

analysis (Figure 22) concluded that interacting factors are mostly related to cell cycle control. Visualizing this using DAVID's built-in KEGG pathway finder, yet again the cell cycle related proteins were identified as the only and highly enriched group of identified proteins.

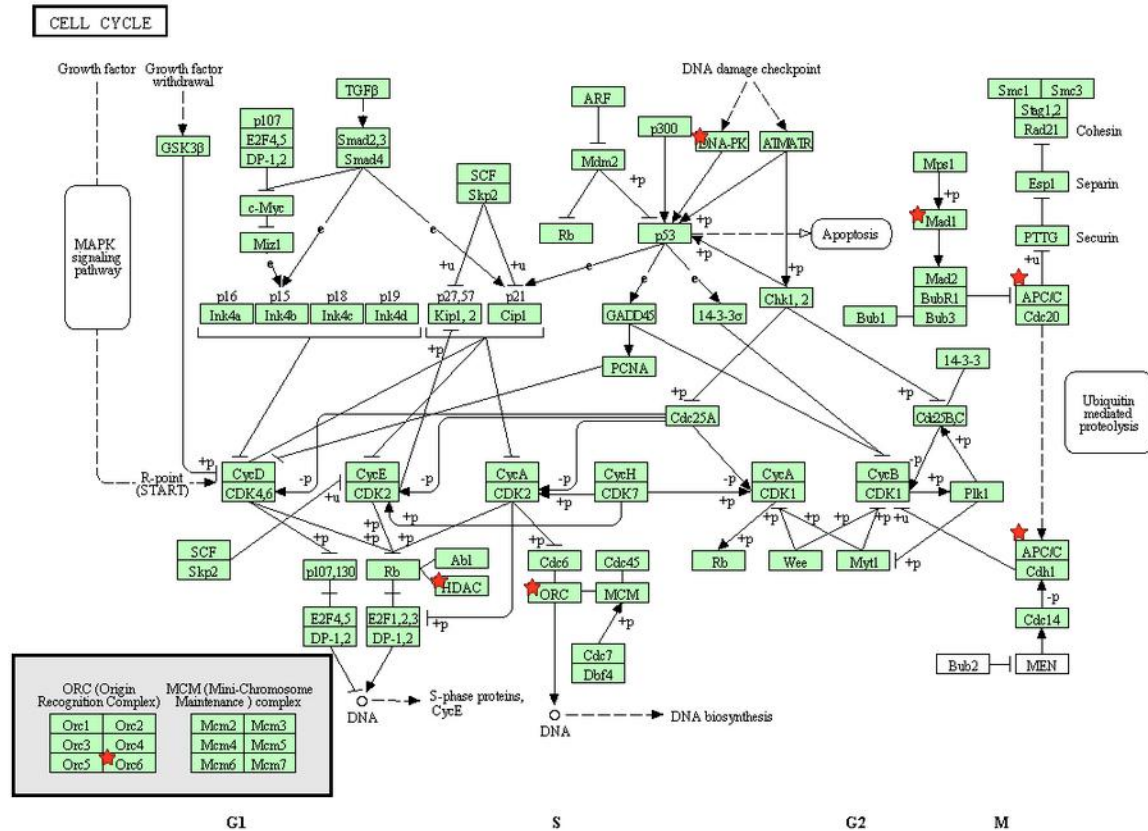


Figure 25. KEGG pathway map with significant hits from the Orc6 immunoprecipitation. Using the list of identified and statistically significant proteins of the Orc6 mass spectrometry analysis, the DAVID functional annotation tool identified one group of proteins from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. The cell cycle protein group contains crucial pathways for cell proliferation control. The factors marked by a red star indicate the proteins directly present among Orc6 interactors. This pathway was identified by DAVID's standard settings with a p value of 0.0029.

The pathway map of Figure 25 I included here in the discussion as it not only shows the relevant hits and the connecting process, but also displays the pathways Orc6 is linked to in cell cycle control. In early G1 (until early S phase) it is involved pre-RC formation although connecting factors like other ORC components or Cdt1 are not marked in the image. Next, detailed under S phase, are the damage checkpoint pathways, also crucial

elements of cell cycle control and G1/S transition. Here, Orc6 interacts with the product of the PRKDC gene (shown as DNA-PK), which is a crucial factor for double strand break repair, and also recombination. Although not present in the image, the bloom syndrome protein (BLM) is another interacting partner of P53 and Orc6, and is also involved in managing recombination, hinting at a role of Orc6 in DNA repair.

After S phase, numerous mitotic factors are present in the interaction list, establishing mitosis as the most prominent cellular process identified among Orc6 interactors. Although only ANAPC1 (APC/C) and MAD1 are present in the image, septins (SEPTA, SEPTB), anillin (ANLN), NUMA1, and CHD3 are all present on the list, and are mostly involved in chromosome segregation and cytokinesis.

Apart from cell cycle control proteins some other identified interactor groups provide hints at more distinct functions. Epigenetic factors for example are among the most interesting candidates on the list. There are 3 histone modifying enzymes found: CHD3, HDAC2, and PHF8. Not only are these all epigenetic factors and involved in cell cycle control (CHD3 in cytokinesis (Hakimi et al. 2002), HDAC2 in ORC expression (Wang et al. 2001), and PHF8 replication initiation and G1-S transition (Liu et al. 2010)) but they are all involved in X-linked mental retardation, one of the most common causes of mental retardation in males (Hakimi et al. 2003; Ropers & Hamel 2005; Loenarz et al. 2010). As Orc6 itself is linked to developmental diseases (namely the Meier-Gorlin syndrome (de Munnik et al. 2012)) it would be interesting to pursue Orc6's function in human development.

PHF8 itself is especially interesting for this study, as it demethylates H4K20 residues, the same residues which are methylated by PR-Set7 (Beck et al. 2012), and which have been shown to affect replication initiation (see chapter 5.2.2, 5.2.3, and (Liu et al. 2010)). Also, PHF8, along with HDAC2 has been shown to regulate E2F mediated transcription (Jung et al. 2012; Liu et al. 2010), which pathway is responsible for the expression of ORC components (Diaz-Trivino et al. 2005). Therefore this interaction might foretell a feedback loop regulating Orc6 expression during the cell cycle.

Looking at the patterns emerging from Orc6's interaction network I conclude that instead of considering Orc6 as a pre-RC factor, it should be regarded instead as a multifunctional protein involved in various aspects of cell cycle regulation.

In summary, I provide experimental data and a model to elucidate Orc6-Orc1-5 interaction by demonstrating that complex assembly might occur in solution and on DNA, and is dependent on the C terminus of Orc6. I also explore Orc6 function and conclude that the protein itself, and truncated constructs derived from it are able to create functioning origins of replication. Finally I place Orc6 into a new context by demonstrating a rich interaction network of this multifunctional cellular factor.

7. Summary

Orc6 is a crucial component of the replication initiation machinery in eukaryotes. Its study helps us to better understand one of the most basic cell functions: DNA replication. The first step in replication initiation is the assembly and DNA binding of the origin recognition complex (ORC). Although members of ORC are highly conserved and well studied, Orc6 evolved faster than the other members of the complex, has a different structure and is less well understood than Orc1-5. The low number of studies on Orc6 in human cells and the low predictability of results obtained from model systems necessitates an in depth analysis of HsOrc6.

The aim of this study is to better understand the role of Orc6 in the human cell by exploring the mechanism of Orc6-Orc1-5 interaction and Orc6-DNA binding, studying the function of Orc6 in live cells, and mapping its interaction network to find new pathways and functions of the protein.

The study concludes that Orc6 interacts via its C terminus with Orc1-5 *in vitro*. It might also have a second pre-RC interaction domain, probably close to the nuclear localization signal. Orc6 is not required for Orc1-5-DNA binding, but is able to enhance the process *in vitro*. It is also able to bind DNA in the absence of Orc1-5 with high affinity.

In vivo experiments present evidence on the ability of Orc6 to recruit the replication machinery to itself on DNA. Although this ability is less efficient than that of EBNA1, a viral factor, it is sufficiently strong to support autonomous replication of plasmids in human cell lines for at least four weeks. Truncated variants of the protein containing at least one of the predicted Orc1-5 interaction sites are sufficient to allow replication. This finding can be used to design autonomous-replicating plasmids for use in gene therapy.

Mapping the interaction network of Orc6 leads to the conclusion that HsOrc6 is only weakly attached to pre-RC factors, and is likely to have an important role in mitosis and possibly be involved in other pathways.

8. Zusammenfassung

Orc6 ist eine entscheidende Komponente der Initiationsmaschinerie der eukaryotischen Replikation. Molekulare Studien zur Funktion dieser Faktoren tragen nicht nur zum grundlegenden Wissen über Zellfunktionen bei, sondern auch zum Verständnis der Krankheiten, die durch Replikationsdefekte hervorgerufen werden.

Der erste Schritt der Replikationsinitiation ist die Rekrutierung des Origin Recognition Complex (ORC) an die DNA. Obwohl die unterschiedlichen ORC Komponenten hochkonserviert und gut erforscht sind, evolviert Orc6 schneller als die anderen ORC-Mitglieder. Orc6 besitzt im Vergleich zu Orc1-5 eine andere, weniger verstandene Struktur und ist als einziges ORC Protein nicht Mitglied der AAA+ Familie.

Die geringe Zahl an Studien über Orc6 in menschlichen Zellen und die geringe Berechenbarkeit der Ergebnisse in den Modellsystemen macht eine eingehende Analyse des HsOrc6 notwendig.

Ziel dieser Arbeit ist es, die Rolle von Orc6 auf molekularer Ebene besser zu verstehen. Dazu studierte ich den Mechanismus der Orc6-Orc1-5 Interaktion, die DNA-Bindung von Orc6 in humanen Zellen, und generierte ein Protein-Interaktionsnetzwerk von Orc6.

Als Ergebnis dieser Untersuchung konnte ich zeigen, dass Orc6 über seinen C-Terminus mit Orc1-5 interagiert. Es ist wahrscheinlich, dass eine zweite Interaktionsdomäne existiert, die vermutlich neben dem Kernlokalisierungssignal liegt. Orc6 wird zur Bindung von Orc1-5 an die DNA nicht benötigt, kann aber diese Bindung zumindest *in vitro* stabilisieren. Orc6 kann unabhängig von Orc1-5 mit hoher Affinität an DNA binden.

In vivo Experimente deuten darauf hin, dass Orc6 die Rekrutierung von ORC an potentiellen Origins unterstützt. Das gerichtete Binden von Orc6 an ein Plasmid unterstützt die autonome Replikation über mindestens vier Wochen, auch wenn die Effizienz geringer ist, vergleichbar zu EBNA1. Deletionsmutanten von Orc6, die eines der beiden Interaktionsdomänen besitzen, sind ausreichend, um Replikation zu

induzieren. Dieses Ergebnis kann dafür benutzt werden, autonomreplizierende Plasmide zur Anwendung in Genetherapie zu generieren.

Aus dem Interaktionsnetzwerkes von Orc6 geht hervor, dass HsOrc6 nur schwach an pre-RC Faktoren assoziiert ist, die Interaktion mit mitotischen Faktoren ist sehr viel ausgeprägter, was auf eine wichtige Rolle Orc6 in der Mitose hindeutet.

9. References

- Abbas, Sivaprasad, Terai, Amador, Pagano & Dutta, 2008. PCNA-dependent regulation of p21 ubiquitylation and degradation via the CRL4Cdt2 ubiquitin ligase complex. *Genes & Development*, 22(18), pp.2496–506.
- Aladjem, 2007. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nature Reviews. Genetics*, 8(8), pp.588–600.
- Aparicio, Weinstein & Bell, 1997. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell*, 91(1), pp.59–69.
- Arias & Walter, 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes & Development*, 21(5), pp.497–518.
- Ash, Faelber, Kosslick, Albert, Roske, Kofler, Schuemann, Krause & Freund, 2010. Conserved beta-hairpin recognition by the GYF domains of Smy2 and GIGYF2 in mRNA surveillance and vesicular transport complexes. *Structure*, 18(8), pp.944–54.
- Ashburner et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1), pp.25–9.
- Babcock, Decker, Volk & Thorley-Lawson, 1998. EBV persistence in memory B cells in vivo. *Immunity*, 9(3), pp.395–404.
- Balasov, Huijbregts & Chesnokov, 2009. Functional analysis of an Orc6 mutant in *Drosophila*. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(26), pp.10672–7.
- Balasov, Huijbregts & Chesnokov, 2007. Role of the Orc6 protein in origin recognition complex-dependent DNA binding and replication in *Drosophila melanogaster*. *Molecular And Cellular Biology*, 27(8), pp.3143–53.

- Baltin, Leist, Odronitz, Wollscheid, Baack, Kapitza, Schaarschmidt & Knippers, 2006. DNA replication in protein extracts from human cells requires ORC and Mcm proteins. *The Journal Of Biological Chemistry*, 281(18), pp.12428–35.
- Bashaw & Yates, 2001. Replication from oriP of Epstein-Barr virus requires exact spacing of two bound dimers of EBNA1 which bend DNA. *Journal Of Virology*, 75(22), pp.10603–11.
- Beck, Burton, Oda, Ziegler-Birling, Torres-Padilla & Reinberg, 2012. The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes & Development*, 26(23), pp.2580–9.
- Bell & Dutta, 2002. DNA replication in eukaryotic cells. *Annual Review Of Biochemistry*, 71, pp.333–74.
- Bell, Kobayashi & Stillman, 1993. Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science*, 262(5141), pp.1844–9.
- Bell, Mitchell, Leber, Kobayashi & Stillman, 1995. The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell*, 83(4), pp.563–8.
- Bell & Stillman, 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, 357(6374), pp.128–34.
- Bernal & Venkitaraman, 2011. A vertebrate N-end rule degron reveals that Orc6 is required in mitosis for daughter cell abscission. *The Journal Of Cell Biology*, 192(6), pp.969–78.
- Cayrou et al., 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Research*, 21(9), pp.1438–49.

- Chaudhuri, Xu, Todorov, Dutta & Yates, 2001. Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 98(18), pp.10085–9.
- Chen & Bell, 2011. CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes & Development*, 25(4), pp.363–372.
- Chen, de Vries & Bell, 2007. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes & Development*, 21(22), pp.2897–907.
- Chesnokov, Chesnokova & Botchan, 2003. A cytokinetic function of Drosophila ORC6 protein resides in a domain distinct from its replication activity. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 100(16), pp.9150–5.
- Chesnokov, Remus & Botchan, 2001. Functional analysis of mutant and wild-type Drosophila origin recognition complex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 98(21), p.11997.
- Chomczynski & Sacchi, 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162(1), pp.156–9.
- Clarey, Erzberger, Grob, Leschziner, Berger, Nogales & Botchan, 2006. Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nature Structural & Molecular Biology*, 13(8), pp.684–90.
- Clyne & Kelly, 1995. Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *The EMBO Journal*, 14(24), pp.6348–57.

- Dai, Chuang & Kelly, 2005. DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(2), pp.337–42.
- DePamphilis, 2006. *DNA replication and human disease*, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.
- Desper & Gascuel, 2004. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Molecular Biology And Evolution*, 21(3), pp.587–98.
- Dhar, Delmolino & Dutta, 2001a. Architecture of the human origin recognition complex. *The Journal Of Biological Chemistry*, 276(31), pp.29067–71.
- Dhar, Yoshida, Machida, Khaira, Chaudhuri, Wohlschlegel, Leffak, Yates & Dutta, 2001b. Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell*, 106(3), pp.287–96.
- Diaz-Trivino, del Mar Castellano, de la Paz Sanchez, Ramirez-Parra, Desvoves & Gutierrez, 2005. The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differently expressed in proliferating and endoreplicating cells. *Nucleic Acids Research*, 33(17), pp.5404–14.
- Dueber, Corn, Bell & Berger, 2007. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science*, 317(5842), pp.1210–3.
- Duncker, Chesnokov & McConkey, 2009. The origin recognition complex protein family. *Genome Biology*, 10(3), p.214.
- Dutta & Dhar, 2000. Identification and characterization of the human ORC6 homolog. *The Journal Of Biological Chemistry*, 275(45), pp.34983–8.

- Edwards, Tutter, Cvetic, Gilbert, Prokhorova & Walter, 2002. MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *The Journal Of Biological Chemistry*, 277(36), pp.33049–57.
- Evrin, Clarke, Zech, Lurz, Sun, Uhle, Li, Stillman & Speck, 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(48), pp.20240–5.
- Falbo & Shen, 2009. Histone modifications during DNA replication. *Molecules And Cells*, 28(3), pp.149–154.
- Fernández-Cid et al., 2013. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Molecular Cell*, 50(4), pp.577–88.
- Foss, McNally, Laurenson & Rine, 1993. Origin recognition complex (ORC) in transcriptional silencing and DNA replication in *S. cerevisiae*. *Science*, 262(5141), pp.1838–44.
- Franceschini et al., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(Database issue), pp.D808–15.
- Gambus, Khoudoli, Jones & Blow, 2011. MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *The Journal Of Biological Chemistry*, 286(13), pp.11855–64.
- Gaudier, Schuwirth, Westcott & Wigley, 2007. Structural basis of DNA replication origin recognition by an ORC protein. *Science*, 317(5842), pp.1213–6.
- Gavin, Song, Wang, Xi & Ju, 2008. Reduction of Orc6 expression sensitizes human colon cancer cells to 5-fluorouracil and cisplatin. *PloS One*, 3(12), p.e4054.

- Gerhardt, Jafar, Spindler, Ott & Schepers, 2006. Identification of new human origins of DNA replication by an origin-trapping assay. *Molecular And Cellular Biology*, 26(20), pp.7731–46.
- Ghosh, Vassilev, Zhang, Zhao & DePamphilis, 2011. Assembly of the human origin recognition complex occurs through independent nuclear localization of its components. *The Journal Of Biological Chemistry*, 286(27), pp.23831–41.
- Gilbert, 2004. In search of the holy replicator. *Nature Reviews. Molecular Cell Biology*, 5(10), pp.848–55.
- Gillespie, Li & Blow, 2001. Reconstitution of licensed replication origins on *Xenopus* sperm nuclei using purified proteins. *BMC Biochemistry*, 2, p.15.
- Giordano-Coltart, Ying, Gautier & Hurwitz, 2005. Studies of the properties of human origin recognition complex and its Walker A motif mutants. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(1), pp.69–74.
- Giovannone, Lee, Laviola, Giorgino, Cleveland & Smith, 2003. Two novel proteins that are linked to insulin-like growth factor (IGF-I) receptors by the Grb10 adapter and modulate IGF-I signaling. *The Journal Of Biological Chemistry*, 278(34), pp.31564–73.
- Grässer et al., 1994. Monoclonal antibodies directed against the Epstein-Barr virus-encoded nuclear antigen 1 (EBNA1): immunohistologic detection of EBNA1 in the malignant cells of Hodgkin's disease. *Blood*, 84(11), pp.3792–8.
- Hakimi, Bochar, Schmiesing, Dong, Barak, Speicher, Yokomori & Shiekhattar, 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*, 418(6901), pp.994–8.

- Hakimi, Dong, Lane, Speicher & Shiekhattar, 2003. A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *The Journal Of Biological Chemistry*, 278(9), pp.7234–9.
- Hanahan, 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal Of Molecular Biology*, 166(4), pp.557–80.
- Heinzel, Krysan, Tran & Calos, 1991. Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Molecular And Cellular Biology*, 11(4), pp.2263–72.
- Heller, Kang, Lam, Chen, Chan & Bell, 2011. Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell*, 146(1), pp.80–91.
- Herman, Wegrzyn & Wegrzyn, 1994. Combined effect of stringent or relaxed response, temperature and rom function on the replication of pUC plasmids in *Escherichia coli*. *Acta Biochimica Polonica*, 41(2), pp.122–4.
- Hirai & Shirakata, 2001. Replication licensing of the EBV oriP minichromosome. *Current Topics In Microbiology And Immunology*, 258, pp.13–33.
- Hodge, Have, Hutton & Lamond, 2013. Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *Journal Of Proteomics*, 88, pp.92–103.
- Huijbregts, Svitin, Stinnett, Renfrow & Chesnokov, 2009. *Drosophila* Orc6 facilitates GTPase activity and filament formation of the septin complex. *Molecular Biology Of The Cell*, 20(1), pp.270–81.
- Iaccarino, Marra, Palombo & Jiricny, 1998. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. *The EMBO Journal*, 17(9), pp.2677–86.

- Jacob, Brenner & Cuzin, 1963. On the Regulation of DNA Replication in Bacteria. *Cold Spring Harbor Symposia On Quantitative Biology*, 28, pp.329–348.
- Jin & Robertson, 2013. DNA methyltransferases, DNA damage repair, and cancer. *Advances In Experimental Medicine And Biology*, 754, pp.3–29.
- Jung et al., 2012. HDAC2 overexpression confers oncogenic potential to human lung cancer cells by deregulating expression of apoptosis and cell cycle proteins. *Journal Of Cellular Biochemistry*, 113(6), pp.2167–77.
- Kaguni, 2011. Replication initiation at the Escherichia coli chromosomal origin. *Current Opinion In Chemical Biology*, 15(5), pp.606–13.
- Kalderon, Roberts, Richardson & Smith, 1984. A short amino acid sequence able to specify nuclear location. *Cell*, 39(3 Pt 2), pp.499–509.
- Kawasaki, Kim, Kojima, Seki & Sugino, 2006. Reconstitution of Saccharomyces cerevisiae prereplicative complex assembly in vitro. *Genes To Cells*, 11(7), pp.745–56.
- Kelman & Kelman, 2003. Archaea: an archetype for replication initiation studies? *Molecular Microbiology*, 48(3), pp.605–15.
- Kelman & Kelman, 2004. Multiple origins of replication in archaea. *Trends In Microbiology*, 12(9), pp.399–401.
- Kenmochi, Kawaguchi, Rozen, Davis, Goodman, Hudson, Tanaka & Page, 1998. A map of 75 human ribosomal protein genes. *Genome Research*, 8(5), pp.509–23.
- Killian, Le Meur, Sesboué, Bourguignon, Bougeard, Gautherot, Bastard, Frébourg & Flaman, 2004. Inactivation of the RRB1-Pescadillo pathway involved in ribosome biogenesis induces chromosomal instability. *Oncogene*, 23(53), pp.8597–602.

- Klein, Strausberg, Wagner, Pontius, Clifton & Richardson, 2002. Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative. *Developmental Dynamics*, 225(4), pp.384–91.
- Klemm, Austin & Bell, 1997. Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell*, 88(4), pp.493–502.
- Kneissl, Pütter, Szalay & Grummt, 2003. Interaction and assembly of murine pre-replicative complex proteins in yeast and mouse cells. *Journal Of Molecular Biology*, 327(1), pp.111–28.
- Knowles, Howe & Aden, 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science*, 209(4455), pp.497–9.
- Kodama & Hu, 2012. Bimolecular fluorescence complementation (BiFC): a 5-year update and future perspectives. *BioTechniques*, 53(5), pp.285–98.
- Kofler, Motzny & Freund, 2005. GYF domain proteomics reveals interaction sites in known and novel target proteins. *Molecular & Cellular Proteomics : MCP*, 4(11), pp.1797–811.
- Kornberg & Baker, 2005. *Dna Replication*, University Science Books.
- Krämer, 2009. HDAC2: a critical factor in health and disease. *Trends In Pharmacological Sciences*, 30(12), pp.647–55.
- Kreitz, Ritzi, Baack & Knippers, 2001. The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells. *The Journal Of Biological Chemistry*, 276(9), pp.6337–42.
- Krude, 2006. Initiation of chromosomal DNA replication in mammalian cell-free systems. *Cell Cycle*, 5(18), pp.2115–22.

- Krueger, Berens, Schmidt, Schnappinger & Hillen, 2003. Single-chain Tet transregulators. *Nucleic Acids Research*, 31(12), pp.3050–6.
- Krysan, Smith & Calos, 1993. Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. *Molecular And Cellular Biology*, 13(5), pp.2688–96.
- Kundu et al., 2010. Deregulated Cdc6 inhibits DNA replication and suppresses Cdc7-mediated phosphorylation of Mcm2-7 complex. *Nucleic Acids Research*, pp.1–10.
- Kunert, Wagner, Murawska, Klinker, Kremmer & Brehm, 2009. dMec: a novel Mi-2 chromatin remodelling complex involved in transcriptional repression. *The EMBO Journal*, 28(5), pp.533–44.
- Kunkel, 2004. DNA replication fidelity. *The Journal Of Biological Chemistry*, 279(17), pp.16895–8.
- Ladenburger, Keller & Knippers, 2002. Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. *Molecular And Cellular Biology*, 22(4), pp.1036–48.
- Laemmli, 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), pp.680–685.
- Lautier, Goldwurm, Dürr, Giovannone, Tsiaras, Pezzoli, Brice & Smith, 2008. Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease. *American Journal Of Human Genetics*, 82(4), pp.822–33.
- Lebofsky, van Oijen & Walter, 2010. DNA is a co-factor for its own replication in *Xenopus* egg extracts. *Nucleic Acids Research*, pp.1–11.
- Lee & Bell, 1997. Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Molecular And Cellular Biology*, 17(12), pp.7159–68.

- Lee, Makhov, Klemm, Bell & Griffith, 2000. Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. *The EMBO Journal*, 19(17), pp.4774–82.
- Leonard & Grimwade, 2010. Regulating DnaA complex assembly: it is time to fill the gaps. *Current Opinion In Microbiology*, 13(6), pp.766–72.
- Li & Herskowitz, 1993. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science*, 262(5141), pp.1870–4.
- Li & Jin, 2010. DNA replication licensing control and rereplication prevention. *Protein & Cell*, 1(3), pp.227–36.
- Lindner, Zeller, Schepers & Sugden, 2008. The affinity of EBNA1 for its origin of DNA synthesis is a determinant of the origin's replicative efficiency. *Journal Of Virology*, 82(12), pp.5693–702.
- Liu et al., 2010. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature*, 466(7305), pp.508–12.
- Liu, Balasov, Wang, Wu, Chesnokov & Liu, 2011. Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 108(18), pp.7373–8.
- Loenarz, Ge, Coleman, Rose, Cooper, Klose, Ratcliffe & Schofield, 2010. PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. *Human Molecular Genetics*, 19(2), pp.217–22.
- Machida, Hamlin & Dutta, 2005. Right place, right time, and only once: replication initiation in metazoans. *Cell*, 123(1), pp.13–24.
- MacNeill, 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *The Biochemical Journal*, 425(3), pp.489–500.

- MacNeill, 2001. Understanding the enzymology of archaeal DNA replication: progress in form and function. *Molecular Microbiology*, 40(3), pp.520–9.
- Marahrens & Stillman, 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science*, 255(5046), pp.817–23.
- Marechal, Dehee, Chikhi-Brachet, Piolot, Coppey-Moisan & Nicolas, 1999. Mapping EBNA-1 domains involved in binding to metaphase chromosomes. *Journal Of Virology*, 73(5), pp.4385–92.
- McGinnis & Madden, 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32(Web Server issue), pp.W20–5.
- Mehanna & Diffley, 2012. Pre-replicative complex assembly with purified proteins. *Methods*, 57(2), pp.222–6.
- Middleton & Sugden, 1994. Retention of plasmid DNA in mammalian cells is enhanced by binding of the Epstein-Barr virus replication protein EBNA1. *Journal Of Virology*, 68(6), pp.4067–71.
- Mott & Berger, 2007. DNA replication initiation: mechanisms and regulation in bacteria. *Nature Reviews. Microbiology*, 5(5), pp.343–54.
- Mujumdar, Ernst, Mujumdar, Lewis & Waggoner, 1993. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjugate Chemistry*, 4(2), pp.105–11.
- Mullis, Faloona, Scharf, Saiki, Horn & Erlich, 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia On Quantitative Biology*, 51 Pt 1, pp.263–73.
- De Munnik et al., 2012. Meier-Gorlin syndrome: growth and secondary sexual development of a microcephalic primordial dwarfism disorder. *American Journal Of Medical Genetics. Part A*, 158A(11), pp.2733–42.

- Myllykallio & Forterre, 2000. Mapping of a chromosome replication origin in an archaeon: response. *Trends In Microbiology*, 8(12), pp.537–9.
- Nasheuer, Smith, Bauerschmidt, Grosse & Weissart, 2002. Initiation of eukaryotic DNA replication: regulation and mechanisms. *Progress In Nucleic Acid Research And Molecular Biology*, 72, pp.41–94.
- Natale, Li, Sun & DePamphilis, 2000. Selective instability of Orc1 protein accounts for the absence of functional origin recognition complexes during the M-G(1) transition in mammals. *The EMBO Journal*, 19(11), pp.2728–38.
- Nislow, Lombillo, Kuriyama & McIntosh, 1992. A plus-end-directed motor enzyme that moves antiparallel microtubules in vitro localizes to the interzone of mitotic spindles. *Nature*, 359(6395), pp.543–7.
- Norseen, Thomae, Sridharan, Aiyar, Schepers & Lieberman, 2008. RNA-dependent recruitment of the origin recognition complex. *The EMBO Journal*, 27(22), pp.3024–3035.
- Oegema, Savoian, Mitchison & Field, 2000. Functional analysis of a human homologue of the *Drosophila* actin binding protein anillin suggests a role in cytokinesis. *The Journal Of Cell Biology*, 150(3), pp.539–52.
- Ogawa, Takahashi & Masukata, 1999. Association of fission yeast Orp1 and Mcm6 proteins with chromosomal replication origins. *Molecular And Cellular Biology*, 19(10), pp.7228–36.
- Di Paola & Zannis-Hadjopoulos, 2012. Comparative analysis of pre-replication complex proteins in transformed and normal cells. *Journal Of Cellular Biochemistry*, 113(4), pp.1333–47.

- Papior, Arteaga-Salas, Günther, Grundhoff & Schepers, 2012. Open chromatin structures regulate the efficiencies of pre-RC formation and replication initiation in Epstein-Barr virus. *The Journal Of Cell Biology*, 198(4), pp.509–28.
- Patino, Kang, Matoba, Mian, Gochuico & Hwang, 2006. Atherosclerotic plaque macrophage transcriptional regulators are expressed in blood and modulated by tristetraproline. *Circulation Research*, 98(10), pp.1282–9.
- Prasanth, Prasanth & Stillman, 2002. Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. *Science*, 297(5583), pp.1026–31.
- Ranjan & Gossen, 2006. A structural role for ATP in the formation and stability of the human origin recognition complex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103(13), pp.4864–9.
- Rao, Marahrens & Stillman, 1994. Functional conservation of multiple elements in yeast chromosomal replicators. *Molecular And Cellular Biology*, 14(11), pp.7643–51.
- Ravasz, 2008. *Methylation patterns and in vivo protein-DNA interactions between the C promoter and the latent replication origin of the Epstein-Barr virus genome*. Eötvös Loránd University.
- Remus, Beall & Botchan, 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *The EMBO Journal*, 23(4), pp.897–907.
- Remus, Beuron, Tolun, Griffith & Morris, 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell*, 139(4), pp.719–30.
- Remus & Diffley, 2009. Eukaryotic DNA replication control: lock and load, then fire. *Current Opinion In Cell Biology*, 21(6), pp.771–7.

- Ren, Li, Zheng & Busch, 1998. Cloning and characterization of GEF-H1, a microtubule-associated guanine nucleotide exchange factor for Rac and Rho GTPases. *The Journal Of Biological Chemistry*, 273(52), pp.34954–60.
- Ritzi, Tillack, Gerhardt, Ott, Humme, Kremmer, Hammerschmidt & Schepers, 2003. Complex protein-DNA dynamics at the latent origin of DNA replication of Epstein-Barr virus. *Journal Of Cell Science*, 116(Pt 19), pp.3971–84.
- Romanowski, Madine, Rowles, Blow & Laskey, 1996. The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Current Biology*, 6(11), pp.1416–25.
- Ropers & Hamel, 2005. X-linked mental retardation. *Nature Reviews. Genetics*, 6(1), pp.46–57.
- Rottmann et al., 2005. Mad1 function in cell proliferation and transcriptional repression is antagonized by cyclin E/CDK2. *The Journal Of Biological Chemistry*, 280(16), pp.15489–92.
- Rountree, Bachman & Baylin, 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genetics*, 25(3), pp.269–77.
- Saha, Chen, Thome, Lawlis, Hou, Hendricks, Parvin & Dutta, 1998. Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. *Molecular And Cellular Biology*, 18(5), pp.2758–67.
- Sambrook & Russel, 2001. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.
- Schaarschmidt, Baltin, Stehle, Lipps & Knippers, 2004. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *The EMBO Journal*, 23(1), pp.191–201.

- Schepers & Papior, 2010. Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Research*, 18(1), pp.63–77.
- Schepers, Ritzi, Bousset, Kremmer, Yates, Harwood, Diffley & Hammerschmidt, 2001. Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *The EMBO Journal*, 20(16), pp.4588–602.
- Scholefield, Veening & Murray, 2011. DnaA and ORC: more than DNA replication initiators. *Trends In Cell Biology*, 21(3), pp.188–94.
- Schreiner, Westerburg, Forné, Imhof, Neupert & Mokranjac, 2012. Role of the AAA protease Yme1 in folding of proteins in the intermembrane space of mitochondria. *Molecular Biology Of The Cell*, 23(22), pp.4335–46.
- Sclafani, Fletcher & Chen, 2004. Two heads are better than one: regulation of DNA replication by hexameric helicases. *Genes & Development*, 18(17), pp.2039–45.
- Seki & Diffley, 2000. Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 97(26), pp.14115–20.
- Semple, Da-Silva, Jervis, Ah-Kee, Al-Attar, Kummer, Heikkila, Pasero & Duncker, 2006. An essential role for Orc6 in DNA replication through maintenance of pre-replicative complexes. *The EMBO Journal*, 25(21), pp.5150–8.
- Shen, Sathyan, Geng, Zheng, Chakraborty, Freeman, Wang, Prasanth & Prasanth, 2010. A WD-repeat protein stabilizes ORC binding to chromatin. *Molecular Cell*, 40(1), pp.99–111.
- Sheu & Stillman, 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Molecular Cell*, 24(1), pp.101–13.

- Siddiqui & Stillman, 2007. ATP-dependent assembly of the human origin recognition complex. *The Journal Of Biological Chemistry*, 282(44), pp.32370–83.
- Skarstad, Boye & Fanning, 2003. Circles in the sand. *EMBO Reports*, 4(7), pp.661–5.
- Skop, Liu, Yates, Meyer & Heald, 2004. Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science*, 305(5680), pp.61–6.
- Smith & Sugden, 2013. Potential Cellular Functions of Epstein-Barr Nuclear Antigen 1 (EBNA1) of Epstein-Barr Virus. *Viruses*, 5(1), pp.226–40.
- Somma et al., 2008. Identification of Drosophila mitotic genes by combining co-expression analysis and RNA interference. *PLoS Genetics*, 4(7), p.e1000126.
- Soultanas, 2012. Loading mechanisms of ring helicases at replication origins. *Molecular Microbiology*, 84(1), pp.6–16.
- Speck, Chen, Li & Stillman, 2005. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nature Structural & Molecular Biology*, 12(11), pp.965–71.
- Spiliotis, Kinoshita & Nelson, 2005. A mitotic septin scaffold required for Mammalian chromosome congression and segregation. *Science*, 307(5716), pp.1781–5.
- Stefanovic, Stanojcic, Vindigni, Ochem & Falaschi, 2003. In vitro protein-DNA interactions at the human lamin B2 replication origin. *The Journal Of Biological Chemistry*, 278(44), pp.42737–43.
- Stinchcomb, Struhl & Davis, 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature*, 282(5734), pp.39–43.
- Sun, Kawakami, Zech, Speck, Stillman & Li, 2012. Cdc6-Induced Conformational Changes in ORC Bound to Origin DNA Revealed by Cryo-Electron Microscopy. *Structure*, 20(3), pp.534–44.

- Tada, Kundu & Enomoto, 2008. Insight into initiator-DNA interactions: a lesson from the archaeal ORC. *BioEssays*, 30(3), pp.208–11.
- Takacs, Banati, Koroknai, Segesdi, Salamon, Wolf, Niller & Minarovits, 2010. Epigenetic regulation of latent Epstein-Barr virus promoters. *Biochimica Et Biophysica Acta*, 1799(3-4), pp.228–35.
- Takisawa, Mimura & Kubota, 2000. Eukaryotic DNA replication: from pre-replication complex to initiation complex. *Current Opinion In Cell Biology*, 12(6), pp.690–6.
- Tanaka & Diffley, 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nature Cell Biology*, 4(3), pp.198–207.
- Tardat, Brustel, Kirsh & Lefevbre, 2010. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nature Cell*, 12(11), pp.1086–1093.
- Thomae, 2008. *Rolle des Chromatin-Proteins HMGA1a für die Definition von Replikationsursprüngen*. Ludwig Maximilian University.
- Thomae, Baltin, Pich, Deutsch, Ravasz, Zeller, Gossen, Hammerschmidt & Schepers, 2011. Different roles of the human Orc6 protein in the replication initiation process. *Cellular And Molecular Life Sciences*, 68(22), pp.3741–56.
- Thomae, Pich, Brocher, Spindler, Berens, Hock, Hammerschmidt & Schepers, 2008. Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 105(5), pp.1692–7.
- Thome, Dhar, Quintana, Delmolino, Shahsafaei & Dutta, 2000. Subsets of human origin recognition complex (ORC) subunits are expressed in non-proliferating cells and associate with non-ORC proteins. *The Journal Of Biological Chemistry*, 275(45), pp.35233–41.

- Tsakraklides & Bell, 2010. Dynamics of pre-replicative complex assembly. *The Journal Of Biological Chemistry*, 285(13), pp.9437–43.
- Tye, 2000. Insights into DNA replication from the third domain of life. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 97(6), pp.2399–401.
- Vashee, Cvetic, Lu, Simancek, Kelly & Walter, 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes & Development*, 17(15), pp.1894–908.
- Vashee, Simancek, Challberg & Kelly, 2001. Assembly of the human origin recognition complex. *The Journal Of Biological Chemistry*, 276(28), pp.26666–73.
- Waga & Zembutsu, 2006. Dynamics of DNA binding of replication initiation proteins during de novo formation of pre-replicative complexes in *Xenopus* egg extracts. *The Journal Of Biological Chemistry*, 281(16), pp.10926–34.
- Wang, Fu, Mani, Wadler, Senderowicz & Pestell, 2001. Histone acetylation and the cell-cycle in cancer. *Frontiers In Bioscience*, 6, pp.D610–29.
- Wang & Sugden, 2005. Origins of bidirectional replication of Epstein-Barr virus: models for understanding mammalian origins of DNA synthesis. *Journal Of Cellular Biochemistry*, 94(2), pp.247–56.
- Wilmes, Archambault, Austin, Jacobson, Bell & Cross, 2004. Interaction of the S-phase cyclin Clb5 with an “RXL” docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes & Development*, 18(9), pp.981–91.
- Wong, Winter, Zaika, Cao, Oguz, Koomen, Hamlin & Alexandrow, 2011. Cdc45 limits replicon usage from a low density of preRCs in mammalian cells. *PloS One*, 6(3), p.e17533.

- Wu, Wang & Liang, 2012. Cdt1p, through its interaction with Mcm6p, is required for the formation, nuclear accumulation and chromatin loading of the MCM complex. *Journal Of Cell Science*, 125(Pt 1), pp.209–19.
- Yoon, Baek, Jeong, Shin, Ha, Jeon, Hwang, Chun & Lee, 2004. WD repeat-containing mitotic checkpoint proteins act as transcriptional repressors during interphase. *FEBS Letters*, 575(1-3), pp.23–9.
- Young & Rickinson, 2004. Epstein-Barr virus: 40 years on. *Nature Reviews. Cancer*, 4(10), pp.757–68.
- Zembutsu, 2006. De novo assembly of genuine replication forks on an immobilized circular plasmid in *Xenopus* egg extracts. *Nucleic Acids Research*, 34(13), p.e91.
- Zhang, VerBerkmoes, Langston, Uberbacher, Hettich & Samatova, 2006. Detecting differential and correlated protein expression in label-free shotgun proteomics. *Journal Of Proteome Research*, 5(11), pp.2909–18.

10. Appendix

10.1 Raw microscopy images

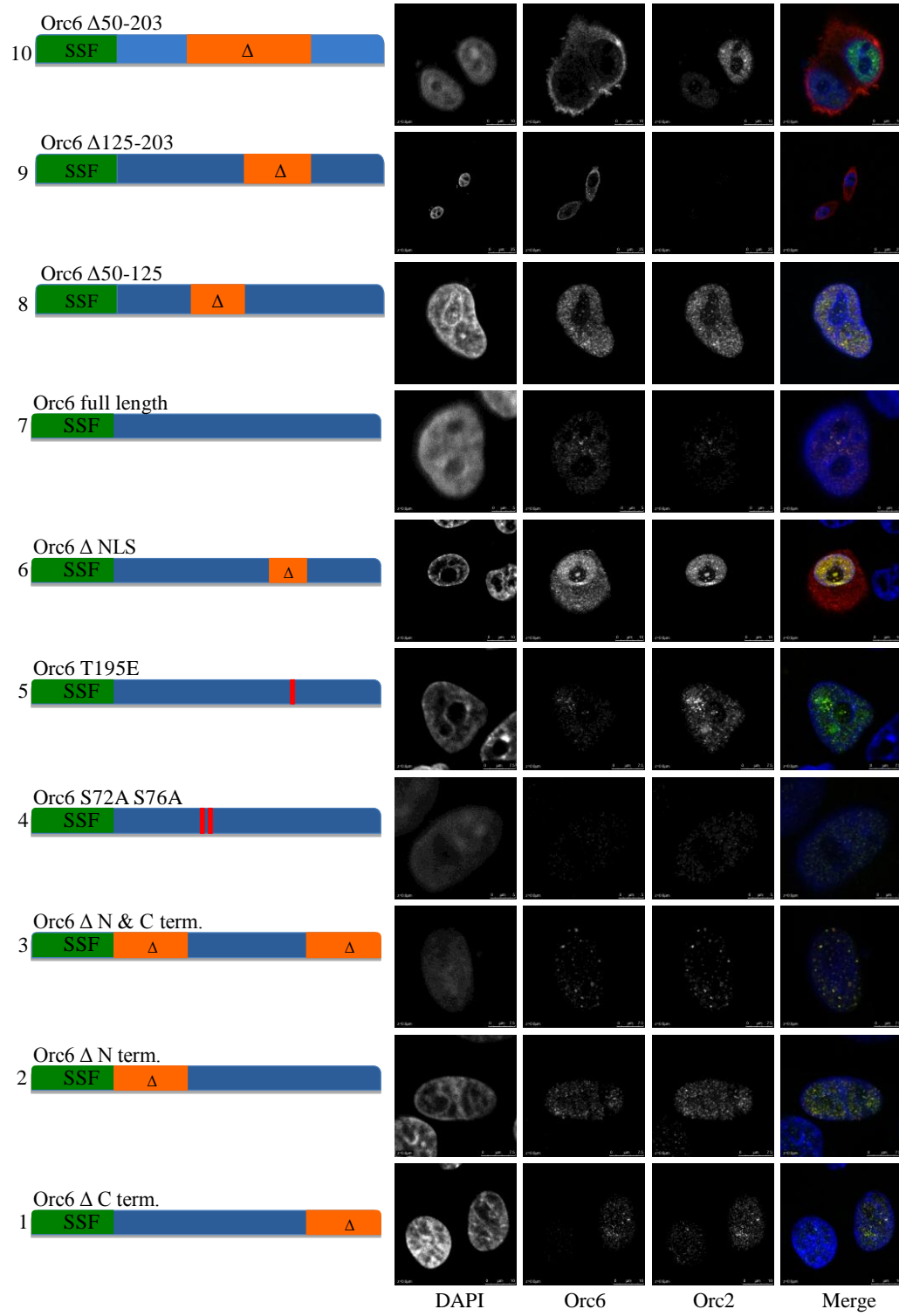


Figure 26. Raw images obtained from fluorescence microscopy. This figure contains the unedited microscopy images that supplement the ones presented in **Figure 17**. As several of these signals are faint, they are difficult to present in a printed format or on dimmed displays. To overcome this, **Figure 17** presents these images with edited color levels as a way to provide better visibility. As some of the color balancing applied there is not a linear transformation, here I included the raw images to show that the actual data obtained indeed supports the conclusions made, and that image manipulation does not introduce significant artifacts in this experiment. The DAPI column represents the blue color channel from the raw fluorescence microscopy image, Orc6 shows red, and Orc2 the green color channel for each cell line analyzed. A merged, color image is presented at the end of each row. The cartoons on the left display the fusion proteins of Orc6 expressed in each cell line. This experiment is detailed in chapter 5.1.9.

10.2 Glossary

A	Adenine
AA	Amino acid
APC	Anaphase promoting complex
APS	Ammonium peroxodisulphat
ATP	Adenosine-5'-triphosphate
Bp	(DNA) base pair
BP	Biological process
BSA	Bovine serum albumin
C	Cytosine
Cdc6	Cell Division Cycle 6
CpG	Methyl-Cytosine-Guanine dinucleotide
DAPI	4',6-diamidino-2-phenylindole
DDK	Dbf4 dependent kinase
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOC	Na-Deoxicholate
DS	Dyad symmetry element
DTT	Dithiothreitol
EBNA1	Epstein-Barr viral nuclear antigen I
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal calf serum
FR	Family of repeats
G	Guanine
GI	GenInfo identifier
GO	Gene ontology
HB	Hypotonic buffer
IP	Immunoprecipitation

Appendix

KAc	Potassium acetate
MCM	Minichromosome maintenance complex
NLS	Nuclear localization signal
OBP	Origin binding protein
ORC	Origin recognition complex
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pre-RC	pre-replication complex
R	Adenine or Guanine (purine)
RT	Room temperature
SDS	Sodium dodecyl sulfate
T	Thymine
TEMED	Tetramethylethylenediamine
Tris	Trisaminomethane
W	Thymine or Adenine (weak bonds)
scTetR	Single chain Tet repressor domain
Y	Thymine or Cytosine (pyrimidine)
YFP	Yellow fluorescent protein

10.3 Publications

Thomae AW, Baltin J, Pich D, Deutsch MJ, **Ravasz M**, Zeller K, Gossen M, Hammerschmidt W, Schepers A. 2011. Different roles of the human Orc6 protein in the replication initiation process. *Cell Mol Life Sci.* 2011 Nov;68(22):3741-56.

Salamon D, Banati F, Koroknai A, **Ravasz M**, Szenthe K, Bathori Z, Bakos A, Niller HH, Wolf H, Minarovits J. 2009. Binding of CCCTC-binding factor *in vivo* to the region located between Rep* and the C promoter of Epstein-Barr virus is unaffected by CpG methylation and does not correlate with Cp activity. *J Gen Virol.* 2009 May;90(Pt 5):1183-9.

10.4 Acknowledgements

I take this opportunity to show my gratitude towards those who have helped me with this project.

First and foremost I would like to thank Dr Aloys Schepers for the opportunity to work in this intriguing field of research and for the immense support provided for me during this project. I also really appreciate the excellent discussions and guidance which I have received under his supervision that allowed me to greatly increase my knowledge and experience in research.

I would also like to show my gratitude to Dr Axel Imhof, Dr Andreas Thomae, and Prof. Dirk Eick for being part of my thesis advisory committee and providing me feedback and useful input on my project. I would specifically like to thank Dr Imhof and his lab for lending their expertise in mass spectrometry, Dr Thomae for teaching me several biochemical techniques, and Prof Eick for his keen insight and support as doctoral supervisor.

A big thanks goes out to all the members of our group at the AGV: Manuel, Nina, Vroni, Krisz, Peer, Stefanie, Julia, Jens, Stoffl, Lara – I consider myself very lucky to have worked together with all you folks. I might even going to miss you.

I am deeply grateful to my girlfriend and soon-to-be wife, Zsuzsi, for bearing with me all this time. I would not have had a chance to succeed without you.

To my parents I thank all the support I have received throughout the years that allowed me to pursue a career in research. It is hard to express in words how grateful I am.

10.5 Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit mit dem Titel

„The role of Orc6 in the human cell“

von mir selbständig und ohne unerlaubte Hilfsmittel angefertigt wurde und ich mich dabei nur der ausdrückliche bezeichneten Quellen und Hilfsmittel bedient habe. Diese Arbeit wurde weder in der jetzigen noch in einer abgewandelten Form einer anderen Prüfungskommission vorgelegt.

München, im September 2013

Máté Ravasz