Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

DEVELOPMENT OF A CENH3-BASED AND NON-VIRAL VECTOR SYSTEM

Die Dissertation wurde eingereicht
von
Stefanie F ü l ö p

München, 30. Juli 2013

Tag der mündlichen Prüfung: 14. Januar 2014

Erstgutachter: PD Dr. Josef Mautner

Zweitgutachter: Prof. Dr. Peter Becker

TABLE OF CONTENTS

I. TABLE OF FIGURES

II. TABLE OF TABLES

III. ABBREVIATIONS

α	Alpha
µg	10^{-6} gram
µl	10^{-6} liter
A	Adenosine
A.thaliana	*Arabidopsis thaliana*
aa	Amino acid
ac	Acetylation
AAV	Adeno-associated virus
Amp	Ampicillin
approx.	Approximately
APC	Anaphase promoting comlex
ATP	Adenosine triphosphate
bp	Basepair
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
C	Cytosine
°C	Degree Celsius
C.elegans	*Caenorhabditis elegans*
C-terminal	Carboxy-terminal
CAL1	Chromosome alignment defect 1
CATD	CENP-A targeting domain
CCAN	Constitutive centromere-associated network
CDE	Centromere DNA element
CENH3	Centromeric histone 3
CENP-A	Centromere Protein A (CENH3 in *H.sapiens)*
CENP-B	Centromere Protein B
CENP-C	Centromere Protein C
ChIP	Chromatin immunoprecipitation
CID	Centromere identifier (CENH3 in *D.melanogaster*)

CMV	Cytomegalovirus
CpG	Cytosine-phosphatidyl-guanosine
CsCl	Caesium chloride
D.melanogaster	*Drosophila melanogaster*
DBD	DNA binding domain
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphophates
DOC	Deoxycholic acid
DS	Dyad symmetry element
dNTP	2'-Deoxynucleosid-5'-triphosphate
DTT	Dithiothreitol
E.coli	*Escherichia coli*
e.g.	E*xempli gratia* (for example)
EBNA1	Epstein-Barr nuclear antigen 1
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
EDTA	Ethylendiamintetra acetic acid
(e)GFP	(enhanced) green fluorescent protein
et al.	*et alii* (and others)
FACT	Facilitates chromatin transcription
FCS	Foetal bovine serum
FISH	Fluorescence *in situ* hybridization
FR	Family of repeats
G	Guanine
h	Hour
H.sapiens	*Homo sapiens*
H1 – H4	Histone H1 - histone H4
H3.3	Histone H3.3
HAC	Human artificial chromosomes
HEK 293 cells	Human embryonic kidney 293 cells
HJURP	Holliday junction recognition protein
HMGA1a	High-mobility group protein 1a
hnRPN-U	Heterogenous nuclear ribonucleoprotein U
HP1α	Heterochromatic protein 1α
i.e.	*id est* (that is)
IFN	Interferon
K	Lysin
kbp	Kilobasepairs
kDa	Kilodalton
KMN network	KNL-1/Mis12 complex/Ndc80 complex
l	Liter
LacI	*Lactose repressor*
LacOperator	*Lactose operator*
LB	Luria-Bertani
MAC	Mammalian artificial chromosome
Mbp	Megabasepairs
me3	Trimethylation
min	Minute
mM	10^{-3} Mol
ng	10^{-9} gram
No.	Number
N-terminal	Amino-terminal

ORC	Origin recognition complex
Orc6	Component of the ORC complex
oriP	Epstein-Barr virus origin of latent replication
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with Tween-20
pCON	Conditional plasmid system
pCONactive	Novel version of *pCON*
pDNA	Plasmid DNA
PIPES	Piperazine-1,2-bis[2-ethanesulfonic acid]
PCR	Polymerase chain reaction
(m)RNA	(messenger) ribonucleic acid
RNAi	RNA interference
RT	Room temperature
RT-PCR	Real-time PCR
rpm	Rounds per minute
RPMI	Roswell Park Memorial Institute
s	Second
S.cerevisiae	*Saccharomyces cerevisiae*
S.pombe	*Schizosaccharomyces pombe*
SAC	Spindle assembly checkpoint
SAF-A	Scaffold attachment factor A
SAR	Scaffold-associated regions
SD	Standard deviation
SEM	Standard error mean
scTetR	*Single chain tetracycline repressor*
SDS	Sodiumdodecylsulphate
S/MAR	Scaffold/matrix attachment region
SV40	Simian virus 40
T	Thymine
TAE	Tris-acetat buffer
TetO	*Tetracycline Operator* (*TetOperator*)
TetR	*Tetracycline repressor*
TLR	Toll-like receptor
Tris	Tris(hydroxymethyl)aminomethane
U	Unit
UV	Ultraviolet
V	Volt
VLP	Virus-like particle
wt	Wildtype

ZUSAMMENFASSUNG

Neue Weiterentwicklungen von Genvektoren für die Gentherapie sind dringend notwendig. Die meisten, derzeit in der Klinik angewandten, Genvektoren basieren auf Virusgenomen. Durch Infektion mit virusähnlichen Partikeln wird ein besonders effizienter Gentransport in die Zielzelle erreicht. Der virale Hintergrund bringt jedoch negative Effekte wie z.B. immunologische Reaktionen des Patienten oder die Aktivierung von Onkogenen durch Insertionsmutagenese und somit Krebsentstehung, mit sich. Eine Möglichkeit, die Probleme zu umgehen sind extrachromosomale, autonom replizierende und nicht-virale Genvektoren. Ziel dieser Arbeit war die Entwicklung eines neuen Genvektorsystems, dessen extrachromosomale Stabilität auf dem Protein CENH3 basiert. Das Zentromer-spezifische Histon 3 CENH3 (*H.sapiens*: $CENH3^{CENP\text{-}A}$, *D.melanogaster*: $CENH3^{CID}$) ist alleine ausreichend für die Zentromerausbildung. CENH3 etabliert eine stabile, epigenetisch vererbte Markierung auf dem Chromosom sowie auf Plasmiden. Mit Hilfe von CENH3 war es möglich den passiven *Huckepack*-Mechanismus von Epstein-Barr-Virus basierten Plasmiden in einen aktiven Segregationsprozess für Plasmid-DNA Vektoren zu verändern. Durch den neuen, aktiven Segregationsmechanismus wird die Vektorstabilität in der Zielzelle erhöht. Die Grundlagenforschung liefert uns hier wertvolle Informationen, die äußerst hilfreich für die Entwicklung neuer Genvektoren ist.

ABSTRACT

New vectors for successful transgene delivery in patients are more than needed. Current gene therapeutic vectors are mostly based on virus genomes. A main reason is the very efficient administration of the vector into the target cells. The viral background of these systems also has severe side effects like *e.g.* immunological reactions of the patient or the activation of oncogenes due to insertional mutagenesis, which can result in cancer development. A possibility to overcome these problems is the establishment of extrachromosomally maintained, autonomously replicating and non-viral vectors. The goal of this work was to establish such a novel system with the help of the CENH3 protein. The centromere-specific histone 3 variant CENH3 (*H.sapiens*: $CENH3^{CENP\text{-}A}$, *D.melanogaster*: $CENH3^{CID}$) is sufficient for centromere formation. It confers a stable and epigenetically heritable mark at the centromeric region and does this also on plasmids. With the help of CENH3 I changed the passive *piggyback* segregation mechanism of Epstein-Barr virus-derived vectors into an active segregation mechanism for plasmid DNA (pDNA) vectors. I demonstrated that the new and active pDNA vector segregation mechanism prolongs pDNA vector retention in cells. The information gained from basic research might have great impact in the field of gene-therapeutic research, as this mechanism might be a helpful tool in future gene therapeutic vectors.

1 INTRODUCTION

Gene therapy is a controversial topic due to the variety of problems that come along with its application, However, it might be the future hope and cure of many diseases, ranging from cancer, to autoimmune diseases and genetic disorders. One challenge in the field is the development of safe and efficient administration vehicles for transgenes, so-called gene vectors. But after years of intense research scientists are still dealing with basic questions of this therapeutic approach like *e.g.* safety, administration and long-term stable and consistent transgene expression. Most currently used gene vectors have a viral background, mainly for the reasons of efficient administration into the target cell and stable transgene expression. Adenovirus-based vectors and vectors based on adeno-associated virus (AAV) that have a depleted integrative potential, make up roughly one quarter of applied vectors for gene therapeutic approaches[1] and offer the advantage of a non-integrating character. However, because adenovirus is a common human virus the application of adenovirus-based vectors can easily lead to immunological reactions of the patient. AAV-based vectors display low immunogenicity but this vector type has only a low transgene capacity up to a maximum of 4.7 kbp (Daya and Berns, 2008). Other vectors, like *e.g.* retro- and lentivirus-based vectors bring a big disadvantage along - integration into the host's genome. The main problem of integration is

[1] http://www.abedia.com/wiley/vectors.php (accessed on 27.05.2013)

insertional mutagenesis, which can lead to the activation of oncogenes and gene silencing (Baum *et al.*, 2006). Progress was made here by mutating the integrase-coding genes of lentiviral vectors, which leads to episomal maintenance of the virus-like particle in the cell. Insertional mutagenesis is reduced to a minimum but still not prohibited with this approach (Escors and Breckpot, 2010). Furthermore, the nature of the virus as such can become a problem and immune responses of the patient can occur with severe or even fatal consequences (Thomas *et al.*, 2003).

However, as non-viral, extrachromosomal plasmid-based gene vectors provide transgene expression only for a very limited amount of time, integrating gene vectors are still considered to be the lesser evil over short-term transgene expression (Pich *et al.*, 2008).

1.1 Non-viral, extrachromosomal and plasmid-based gene vectors

This work is dealing with the class of non-viral, extrachromosomally maintained and plasmid-based gene vectors. To avoid confusion due to the variety of synonyms, which are used in the field for this vector class, I will use the term "pDNA vector" (plasmid DNA vector) throughout this text to refer to gene vectors with the above-mentioned characteristics.

In contrast to the large group of virus-derived gene vectors, which make up 66 % of all gene vehicles in clinical applications, naked DNA and pDNA vectors represent only a small group of 18.2 %[2]. Nevertheless, the advantages offered by non-viral pDNA vectors are striking.

The lack of virus-derived material in these vectors avoids possible negative effects concerning recombination events and leads to less problems with immune responses of patients. This is why even a repeated administration of the pDNA vector raises no problems, because no specific immune responses against the gene vehicle occur (Prud'homme *et al.*, 2007). Theoretically, episomal pDNA vectors are not assumed to integrate in the host genome (Lipps *et al.*, 2003). Certainly, also non-viral gene vectors are able to integrate randomly, but methods for targeting the integration sites with the help of induced double strand breaks by zinc finger nucleases have been developed and

[2] http://www.wiley.com/legacy/wileychi/genmed/clinical/ (accessed on 30.04.2013)

help avoiding insertional mutagenesis (Moehle *et al.*, 2007). Moreover, the likelihood of mutations after genomic integration of pDNA vectors is significantly lower than the endogenous mutation probability (Prud'homme *et al.*, 2007). In addition, the non-integrating character makes expression profiles and levels for these kind of gene vectors not as ambiguous as for integrating vectors (Lipps *et al.*, 2003). Virus-based vectors take up only a certain amount of DNA depending on the choice of the virus background (Waehler *et al.*, 2007). Most viruses, like *e.g.* retroviruses, lentiviruses and AAV can harbour a maximum transgene size of only 10 kbp. Other viral vectors are capable to take up bigger DNA sequences. With Epstein-Barr virus-derived vectors *e.g.* a transgene with 123 kbp in size was successfully delivered into B-cells (White *et al.*, 2002). Not only single transgenes, but also complete gene loci are transferrable with *e.g.* S/MAR-based episomal gene vectors (Lufino *et al.*, 2007). Major problems with extrachromosomal pDNA vectors are factors concerning reliable transgene expression, efficient delivery and establishment of the pDNA vector as well as stable long-term maintenance in the target cell. For instance, only 1 - 3 % of cells establish a plasmid-based vector over time, whereas the vast moiety of pDNA is lost quickly after administration due to degradation by nucleases (Pich *et al.*, 2008). Additionally there remains the risk, that DNA is damaged or lost due to strand breakage events (Gill *et al.*, 2009). If we talk about plasmid maintenance in a cell we have to consider two key mechanisms: DNA replication and pDNA vector retention. In contrast to integrating vectors, which are replicated and retained stably in the target cell after administration, this is not the case for pDNA vectors. Proper pDNA vector replication must be guaranteed to achieve long-term maintenance in the cell; if this is not the case, plasmid DNA will be lost quickly after administration. Especially retention is not so efficient for extrachromosomal pDNA vectors and therefore they are subsequently lost from cells. Even though some vector types are present in very high copy numbers in transfected cell, these gene vehicles are lost over time and the very high initial amount of vectors in cells can cover up for increasing numbers of lost vectors only for a limited amount of time. Chromosome-like episomal vectors (*e.g.* mammalian artificial chromosomes) are retained by the same mechanism as the chromosomes of the host cell, which is a rather efficient mechanism. Last but not least autonomous pDNA vectors can be maintained in cells *via* a so-called *piggyback* mechanism. In this case the pDNA vector is attached to the host cell chromosome and gets replicated and segregated to the daughter cells

passively (Ehrhardt *et al.*, 2008). Detailed information about the different retention mechanisms of episomal, non-viral pDNA gene vectors is given in the following chapters.

For designing the best possible pDNA gene vector many different factors have to be taken into account. For instance, it was shown that the conformational state – supercoiled, open circular or linear - of pDNA is important for its stability in a cell. Stability varies in combination with the "under- or overwinding" of the supercoiled status of the DNA (Ribeiro *et al.*, 2008) and the supercoiled state is achieved after de-chromatinization of the circular pDNA vector. The idea, that maybe linearized plasmids are better maintained in cells was tested in yeast. It was demonstrated that the linearization of the transfected DNA was of no advantage and that the plasmids were lost faster from cells than in their circular state (Martin Dani and Zakian, 1983). A linearized vector in human cells showed similar transgene expression compared to the parental plasmid. The slightly better transfection efficiencies of the linearized vector were explained simply with its smaller size (Schakowski *et al.*, 2001).

Also the choice of the promoter type and its combination with enhancers can have a rather big impact on the safety of a pDNA vector. Tissue specific promoters avoid an ubiquitous expression of the transgene and the use of endogenous promoters limits transgene expression to a physiological level (Gill *et al.*, 2009).

To avoid a fast decrease in transgene expression, bacterial backbone sequences of pDNA vectors certainly deserve more attention in the future, especially because they impact on gene silencing (Gill *et al.*, 2009). The importance of this factor is reflected in the exciting progress that was made in the field of the minicircle technology during the last few years (Mayerhofer *et al.*, 2009) (see also chapter 1.1.3).

Vector administration is less efficient for naked pDNA vectors compared to virus-derived particles (Nishikawa and Huang, 2001a). The latter use the mechanisms of the virus they derive from for efficient transduction and also for their processing inside of the cell (Ehrhardt *et al.*, 2008). Non-viral vectors, however, have to encompass several obstacles in regards to cell entry and intracellular processing and are very much in danger of degradation by nucleases before even reaching the target cell (Nishikawa and Huang, 2001a).

Although clinicians are facing many problems with the currently available instruments, and autonomous pDNA vectors offer various advantages, it was not yet possible to find

a safer but equally efficient alternative to viral vectors. What are the reasons for this? Certainly the rather inefficient pDNA vector retention in cells is a big obstacle.

As pointed out before, plasmid retention, together with DNA replication, are the key factors in providing stable pDNA vector maintenance. Only if both factors function reliably with high efficacy, autonomous pDNA vectors are maintained in a cell over a prolonged period of time. Different vectors are using different strategies to maintain their genomic information in the cells over several generations and vectors can be segregated to daughter cells by either active or passive mechanisms (Figure 1.1 A-B)

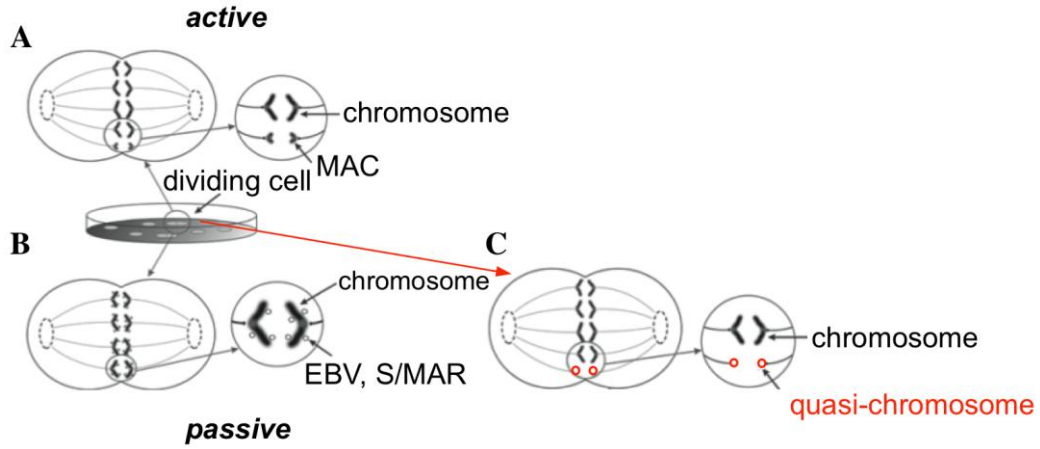


Figure 1.1 Segregation mechanisms of different vector types.

A) Active segregation mechanism of a mammalian artificial chromosome (MAC). MACs are segregated like chromosomes and are entirely integrated into the cell cycle. Other episomal vectors, like EBV or S/MAR pDNA vectors, are segregated to daughter cells *via* a passive *piggyback* mechanism **(B)**. The exact mechanism of EBV-based vector segregation is illustrated in Figure 1.5 A in more detail. **C)** shows the theoretical basis of an active "quasi-chromosome", which is a combination of a minichromosome (**A**) and an EBV-based vector (**B**).
(Lufino *et al.*, 2008)

Not all strategies are similarly effective and, therefore, it can happen that the pDNA vector is not maintained in the target cell forever, even though DNA replication and retention work properly. In particular for the design of new gene therapeutic vectors, retention mechanisms have to be optimised to ensure prolonged and stable transgene expression. It has to be mentioned, that especially an efficient pDNA vector establishment in the cells during the first few days and weeks after administration contribute to long-term stability (Leight and Sugden, 2001). It is known that a major challenge in autonomous pDNA gene vector design is the rapid loss of

extrachromosomal vectors from the target cells shortly after administration. This issue, however, has to be addressed separately and not together with retention.
In the following chapters different pDNA vector systems and their respective retention mechanisms are described in more detail.

1.1.1 S/MAR[3]-based vectors – pEPI and pEPito

S/MAR structures are usually AT[4]-rich sequences (approx. 70 %) and play important roles in cellular mechanisms like replication, establishment and maintenance of higher chromosomal structures. In 1999 an artificial pDNA vector was constructed, which replicates extrachromosomally in human cells (Piechaczek *et al.*, 1999). This pDNA vector represents one of the first non-viral gene vectors that is efficiently maintained in cells for a prolonged amount of time even in the absence of selection (Gill *et al.*, 2009).
In vivo experiments with transgenic pigs demonstrated an episomal maintenance of the *pEPI* vector in all different tissues tested and active transcription of the transgene was detectable in nine out of twelve transgenic animals (Manzini *et al.*, 2006). *pEPI* was furthermore successfully applied in an *in vivo* study focussing on the transdifferentiation of liver cells towards pancreatic β-cells in rats. Among five tested vectors, *pEPI* and a CpG-depleted vector scored best in regards to high and long-lasting expression of the three encoded reporter transgenes Pdx1, Ngn3 and MafA (Cim *et al.*, 2012).
The so-called *pEPI* vector (Figure 1.2), the prototype of S/MAR-based vectors, is 6692 bp in size and consists most importantly of a non-essential SV40[5]-origin, a resistance cassette and a S/MAR sequence derived from the 5´-region of the human interferon-γ gene (Bode *et al.*, 1992). The S/MAR sequence is essential for the extrachromosomal status of the vector and substitutes the function of the large T-antigen when it's actively transcribed (Piechaczek *et al.*, 1999). For this it is essential, that a transcription unit extends into the MAR region (Stehle *et al.*, 2003). Only if this is the case, *pEPI* is stably maintained in the nucleus attached to the chromosomes (Baiker *et al.*, 2000) *via* an interaction with the nuclear matrix protein SAF-A *in vivo* (Jenke *et al.*, 2002). It was demonstrated that *pEPI* replicates in S-phase once per cell cycle and components of the origin recognition complex (Orc1, Orc2 and Mcm3) bind the *pEPI*

[3] Scaffold/matrix attachment region
[4] Adenosine-Thymidine
[5] Simian virus 40

vector sequence-independently (Schaarschmidt *et al.*, 2004). It was furthermore demonstrated by FISH[6] and ChIP[7] experiments that *pEPI* is located predominantly at transcriptionally active chromatin in the nucleus (Stehle *et al.*, 2007).

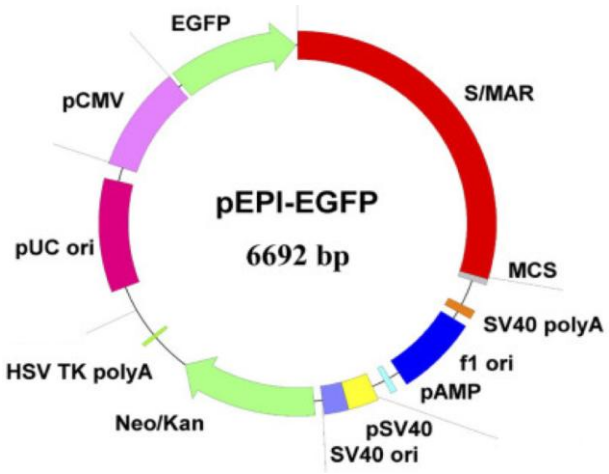


Figure 1.2 The pEPI vector.

The most important feature for maintenance of the extrachromosomal status of *pEPI* is the S/MAR element, which is responsible for the vectors attachment to the host cell chromosome. Furthermore *pEPI* comprises a non-essential SV40 origin of replication, a Neomycin/kanamycin resistance cassette and the eGFP sequence under the control of a CMV promoter.
(Stehle *et al.*, 2007)

A second generation of *pEPI* - the so-called *pEPito* vector - has been established with optimised features. The most important difference between *pEPI* and *pEPito* is the depletion of about 60 % of all present CpG[8] sites in the latter (Haase *et al.*, 2010). This brings two advantages along: first, silencing effects due to CpG-methylation are decreased and second, innate immune responses of the recipient cells, triggered by the TLR9[9] and downstream products, are reduced and therefore significantly less inflammatory cytokines are produced. Additionally, *pEPito* leads to increased transgene expression levels *in vitro* and *in vivo* due to the substitution of the CMV[10]-immediate early promoter with the human CMV enhancer/human elongation factor 1 alpha (hCMV/EF1P) promoter, which is CpG-methylation insensitive and therefore not silenced (Haase *et al.*, 2010). However, it was demonstrated in the same publication,

[6] Fluorescence in-situ hybridization
[7] Chromatin-Immunoprecipitation
[8] Cytosine-phasphatidyl-Guanine
[9] Toll-like receptor 9
[10] Cytomegalovirus

that this effect is cell type-specific. Quite recently it was demonstrated that the tissue-specific expression of a transgene from the episomal S/MAR-based *pEPito* vector was successfully established for targeting hepatocellular carcinomas (Haase *et al.*, 2013)

1.1.1.1 S/MAR-based vectors are attached to the host chromosomes via the S/MAR element

S/MAR-based vectors are segregated *via* a passive *piggyback* mechanism attached to the host's chromosomes. Via the S/MAR region the pDNA vectors are tethered to chromatin and replicated and segregated together with the chromosomes during S-phase (Figure 1.1 B). Immunoprecipitation experiments demonstrated that *e.g. pEPI* forms a specific interaction with the nuclear matrix protein hnRNP-U/SAF-A[11] *in vivo* (Jenke *et al.*, 2002). Regarding the integration of S/MAR-based vectors it has been shown that the preferential state of maintenance is extrachromosomal, although an occasional integration event cannot be excluded (Wong *et al.*, 2011).

Following administration of *pEPI* in cells as naked DNA, the vectors get lost fast and only a very small fraction (approx. 0.5 - 5 %) of cells is able to maintain the vector for a longer period of time (Haase *et al.*, 2010) and this effect was similarly observed in *in vivo* experiments (Wong *et al.*, 2011). However, after successful vector establishment in the cell, the plasmid DNA is maintained for a prolonged period of time even without selection pressure *in vitro* (Haase *et al.*, 2010).

1.1.2 The Epstein-Barr virus origin of latent replication, oriP, and the conditional vector system pCON

Epstein-Barr virus is a double-stranded DNA γ-herpesvirus with 172 kb in size, which efficiently infects resting B-cells and establishes a lifelong persistent infection. Like all herpesviruses, the viral genome of EBV[12] is maintained in an extrachromosomal state (Yates *et al.*, 2000). More than 90 % of the adult population worldwide are infected with this virus and carry it in its latent form without showing any symptoms of a disease (Delecluse and Hammerschmidt, 2000). Furthermore it is known for a long time that EBV is associated with tumour development like nasopharyngeal carcinoma and Burkitt's lymphoma (Nonoyama *et al.*, 1973) as well as Hodgkin's lymphoma and

[11] Heterogeneous nuclear ribonucleoprotein U/scaffold attachment factor A
[12] Epstein-Barr virus

leukaemia (Gotlieb-Stematsky *et al.*, 1975), non-Hodgkin's lymphoma (Wutzler *et al.*, 1986) and gastrointestinal carcinomas (Shibata and Weiss, 1992). The oncogenic potential of EBV is even more pronounced in immunodeficient individuals (Delecluse and Hammerschmidt, 2000). Primary infection with EBV later in life may cause the infectious mononucleosis syndrome (Diehl *et al.*, 1968).

Epstein-Barr virus is interesting as blueprint for pDNA gene vector development as it shares several important features with potent gene vehicles in its endogenous form. First, EBV is a large virus and transgenes with sizes up to 140 kb of genomic DNA can be transferred in an EBV-derived pDNA vector (Delecluse and Hammerschmidt, 2000). Additionally these vectors are known to remain stable and extrachromosomally in the infected cell when kept under selection over long-term periods (Delecluse and Hammerschmidt, 2000; Ehrhardt *et al.*, 2008). Moreover, on EBV-derived vectors DNA replication and extrachromosomal maintenance are regulated separately from each other in *cis*. The respective elements for both functions are called the family of repeats and the dyad symmetry element. Both entities together make up the latent origin of plasmid replication *(oriP)* of EBV, which is the *cis*-component of EBV used in gene vector design of plasmid DNA vectors (Pich *et al.*, 2008) (Figure 1.3).

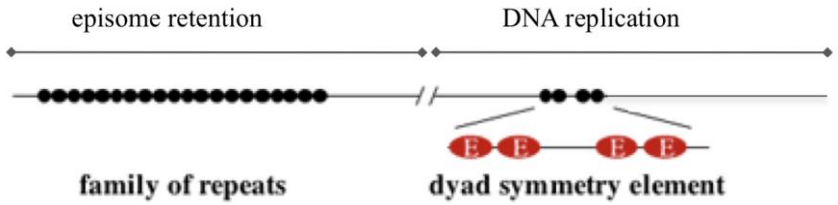


Figure 1.3 Schematic of origin of latent replication (oriP) of Epstein-Barr virus.

The latent replication origin comprises two elements: the family of repeats, FR, confers pDNA vector retention and the dyad symmetry element, DS, is responsible for DNA replication. Both elements depend on the viral transactivator EBV nuclear antigen 1 (red circles, black dots: binding sites of the transactivator at *oriP*).
(Schepers *et al.*, 2001)

The latent replication origin *oriP* is 1.7 kbp in size (Yates *et al.*, 2000). Both *oriP* modules have binding sites for the viral transactivator EBNA1[13]. The FR[14] element harbours 20 EBNA1 binding sites and is responsible for virus retention (Yates *et al.*, 2000). It was shown that seven of these binding sites are sufficient for

[13] EBV nuclear antigen 1
[14] Family of repeats

extrachromosomal maintenance (Wysokensko and Yates, 1989). The second module of *oriP* is the DS[15] element with four EBNA1 binding sites. DS supports DNA replication by recruiting the origin recognition complex, ORC, through a direct interaction between EBNA1 and the complex (Schepers *et al.*, 2001). EBNA1 is the transactivator for both functions of *oriP* and for pDNA vector retention the presence of the protein in the cell is crucial (Krysan *et al.*, 1989).

Introduction of *oriP* on a pDNA vector and transfection of this vector in EBNA1 expressing cells provides the pDNA vector with replication and retention capacity. The so-called *pCON* system, which constitutes a conditional plasmid system based on the interaction of a *scTetR*:fusion protein and *TetOperator* sites on a pDNA reporter vector, enables the dissection of the two functions of *oriP (Pich et al., 2008)*. With the help of *pCON*, DNA replication and retention can be investigated separately and the potential of endogenous (non-viral) proteins to support DNA replication, nuclear retention or both can be determined.

pCON is the first generation of gene vectors, which are extrachromosomally maintained and regulated by an allosteric switch inducible with doxycycline (Pich *et al.*, 2008). To this end fusion proteins are generated with the following composition: the protein of choice replaces the N-terminal domain of EBNA1 and two *scTetR* elements replace the DNA binding domain of EBNA1. *scTetR* is a single chain derivate of the prokaryotic DNA binding protein (*TetR*). Two *scTetR* elements form homodimers like EBNA1 and bind with very strong affinity to *TetOperator* motifs; on the pDNA reporter vectors the respective element of *oriP* (either DS, FR or both) is replaced by a number of *TetOperator* sites. *Via* the *scTetR-TetOperator* binding the protein being investigated for its replicative potential or the potential to support pDNA vector retention, is bound at the respective sites of the pDNA reporter vector (Figure 1.4). Comparable to the cellular function, also in the plasmid system EBNA1 is needed as transactivator of the potentially remaining wildtype DS or FR element on the reporter.

[15] Dyad symmetry element

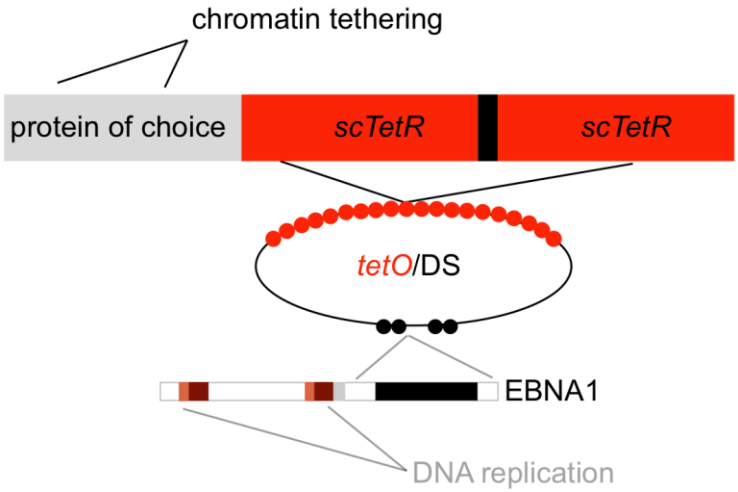


Figure 1.4 Illustration of the pCON system.

The depicted set-up illustrates the situation for testing the ability of a protein to support pDNA vector retention. The two function of *oriP* can be separated in *cis* and investigated autonomously. By replacing the family of repeats with *TetOperator* sites and expression of a protein fused to two *scTetR* elements, one is able to investigate the potential of the protein of choice to support pDNA vector retention. The same principle is applied for testing the replicative potential of a protein. In this case the FR element is maintained in its wildtype conformation and the DS element is replaced with *TetOperator* sites.
(Pich *et al.*, 2008)

First experiments with fusion proteins were performed in the Kieff lab and revealed that pDNA vector maintenance is achieved with fusion proteins of HMGA1a[16] or histone H1 and the DBD[17] of EBNA1 (Hung *et al.*, 2001). With the same experimental set-up our lab was able to confirm that the proteins HMGA1a and Orc6[18] have replicative potential, and that both, HMGA1 and HP1α[19], support pDNA vector retention (Thomae *et al.*, 2011; Thomae *et al.*, 2008) (HP1α data not published).

The usage of EBV and *wt-oriP* as pDNA gene vector is limited by the fact, that EBNA1 shows enhancement of B-cell immortalization (Humme *et al.*, 2003). Another reason is, that the virus, and even the small version of it (mini-EBV), encode for many viral genes (Ehrhardt *et al.*, 2008).

[16] High-mobility group 1a protein
[17] DNA binding domain
[18] Protein of the origin recognition complex ORC
[19] Heterochromatin protein 1α

1.1.2.1 EBV and EBV-derived vectors segregate via EBNA1-mediated attachment to the host cell chromosome

Once per cell cycle Epstein-Barr virus DNA is duplicated in proliferating B-cells using the replication machinery of the host cell. It was demonstrated that EBV recruits the human ORC complex to the dyad symmetry element of *oriP via* an EBNA1-mediated interaction (Schepers *et al.*, 2001). Virus replication happens in synchrony with host chromosome replication during S-phase. When EBV is segregated to the daughter cells after cell duplication, the virus remains attached to the host cell chromosome (Sears *et al.*, 2003). The link between *oriP* (either on the EBV genome or on a EBV-derived pDNA vector) and the chromosome is mediated by EBNA1 (Figure 1.5 B). As described before the family of repeats is responsible for virus retention, however, if FR is missing, DS is partially capable to mediate this function although to a much lesser degree (Aiyar *et al.*, 1998; Yates *et al.*, 2000). With its C-terminus, EBNA1 binds the respective binding sites at *oriP* and two linking regions (LR1 and LR2) located in EBNA1´s N-terminus mediate the attachment to the chromosome (Middleton and Sugden, 1992). It has already been shown before that R-bands of chromosomes in metaphase contain the same sequences as found in interphase chromosomes, so-called AT-rich S/MAR[20]-sequences (Saitoh and Laemmli, 1994). Only at these rare sites on the chromosome the EBNA1-mediated interaction between FR and the chromosomes is possible and pDNA vector retention is ensured (Figure 1.5 A) (Sears *et al.*, 2004). Similar to the host cells chromosomes, pDNA vectors are distributed to the daughter cells in a nearly symmetrical order (Nanbo *et al.*, 2007). The pDNA vector remains attached to the chromosome during mitosis and since this mechanism is a passive retention mode in which the pDNA vector "sits" on the chromosome, it is also referred to as *piggyback* mechanism (Calos, 1998).

Like all herpesviruses EBV is maintained in infected cells extrachromosomally (Morissette and Flamand, 2010). Also *oriP*-based gene vectors are considered to be maintained in the cells as extrachromosomal elements (Ehrhardt *et al.*, 2008). Unfortunately, stable maintenance of EBV genomes without selection is only shown for EBV-transformed cell lines, whereas *oriP*-based vectors are rapidly lost when lacking selection pressure.

[20] Scaffold/matrix-associated regions

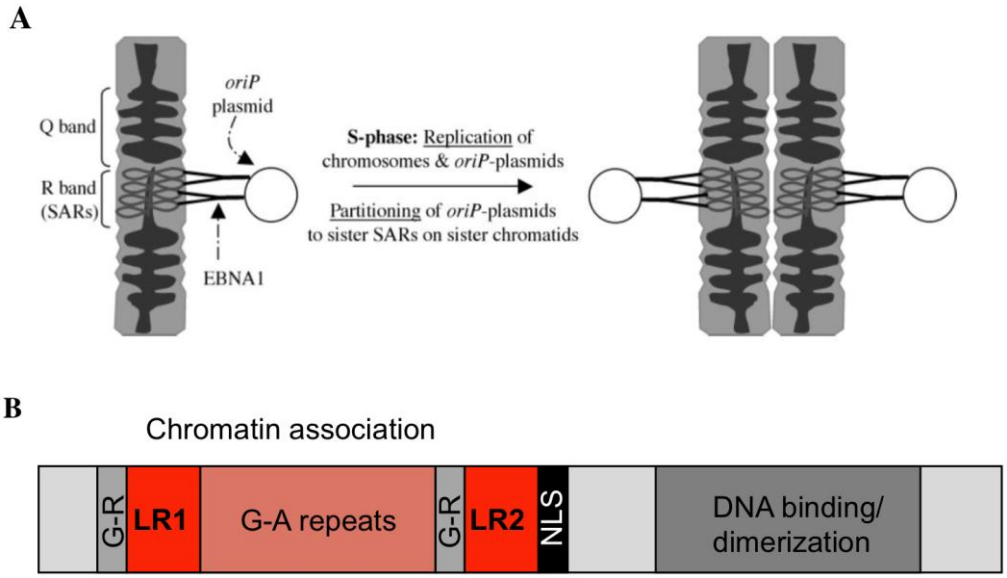


Figure 1.5 Mechanism of EBNA1-mediated, passive retention of oriP vectors.

A) *Via* EBNA1 *oriP* vectors are tethered to SAR sequences at chromosomes and are replicated like the chromosomes during S-phase and partitioned in synchrony with the host chromosomes through a passive *piggyback* mechanism. After duplication of the host cell chromosome also *oriP* vectors are doubled and segregated to both daughter cells. **B)** Scheme of EBNA1. The C-terminal part of EBNA1 is needed for DNA binding and dimerization of the protein. The N-terminal part is used for chromosome binding *via* the two linking regions LR1 and LR2 (red boxes). G-R: Glycin-Arginin repeats, G-A: Glycin-Alanin, LR: linking region, NLS: nuclear localization signal.
(Sears *et al.*, 2004).

It was found out before, that, when following a parental HEK 293 cell line and five sub-clones of this cell line over 100 generations for the former and 25 generations for the latter, respectively, the average pDNA vector number was stable in all clones. The distribution of EBV-derived plasmids in these cells was determined with FISH analysis and from the results it seems reasonable to assume that the number of pDNA vectors is predetermined and that a critical number of pDNA vectors in a cell is needed to avoid the complete loss of vectors. This critical number of pDNA vectors in the cell is referred to as "threshold level" and in HEK 293 cells on average between six and ten EBV episomes were found in investigated cells. If the number of pDNA vectors falls below the threshold, the selective pressure will become too intense and pDNA vectors are lost from the cells and if the copy number of the pDNA vector is higher than the threshold limit, pDNA vectors are not lost that easily from the cell (Nanbo *et al.*, 2007).

The establishment capacities of *wt-oriP* vectors are poor. Only 1 - 5 % of transiently transfected cell undergo an efficient establishment of the pDNA vector and can maintain the vector for a long time. After establishment, withdrawal of selection results in a loss

rate of *wt-oriP* vectors in the range of around 4 % in established cell lines per generation over time (Leight and Sugden, 2001). It was demonstrated in EBNA1 positive HeLa cells without selection pressure, that only 84 % of all EBV-derived pDNA vectors were duplicated during S-phase. The equal distribution of pDNA vectors into two daughter cells was obtained for 88 % of these co-localised (duplicated) pDNA vectors. All other pDNA vectors (12 %) segregated unevenly to only one daughter cell. 16 % of pDNA vectors carrying *wt-oriP* that were transfected into HeLa cells were unable to replicate in this study maybe due to a reduced efficacy of DS. These vectors segregated randomly and only few of them got lost from the cells (0.3 %) (Nanbo *et al.*, 2007).

1.1.3 Minicircles, mammalian/human artificial chromosomes and minichromosomes

The youngest members in the group of transgene vehicles are so-called minicircles. These molecules derive from conventional plasmid DNA (parental plasmid) but they entirely lack any bacterial backbone sequences, like antibiotic resistance genes and replication origins. In short - minicircles are circular DNA molecules depleted of all unnecessary components in regards to transgene expression (Gracey Maniar *et al.*, 2013; Kobelt *et al.*, 2013). The depletion of bacterial sequences is usually achieved by site-directed mutagenesis (Gill *et al.*, 2009). This results in a smaller size of minicircles compared to other gene therapeutic vehicles (Kobelt *et al.*, 2013) and has advantages compared to conventional plasmid-based vectors (Gill *et al.*, 2009). The depletion of sequences of bacterial origin leads to a decreased probability of DNA degradation and in turn to a prolonged stability of the pDNA vector. Additionally the problem of antibiotic resistance dissemination is circumvented by deletion of the resistance cassette, as well as of immunogenic CpG motifs (Kobelt *et al.*, 2013). It is also known, that minicircle DNA is less prone to integration and usually no concatemer formation is observed (Nehlsen *et al.*, 2006). Furthermore, by limiting DNA amounts, gene transfer rates and expression of the transgene are enhanced in minicircles compared to the parental pDNA vector (Kobelt *et al.*, 2013). For instance a 45 - 560-fold increase for serum human factor IX and alpha1-antitrypsin was observed with non-replicating minicircles compared to standard pDNA vectors (Chen *et al.*, 2003).

It was demonstrated by Kobelt *et al.*, that minicircles exhibit an enhanced transgene mRNA[21] transcription compared to parental pDNA vectors in A375 cells. Because it was excluded that this increase is a consequence of higher minicircle copy in these cells, the reason is most likely a better recruitment of the transcription machinery (Kobelt *et al.*, 2013). In the future a recently established technology for industrial-scale production established by Kobelt *et al.* will provide sufficient amounts of minicircles to perform *in-vivo* applications as well as clinical studies. They also monitored an enhanced expression of minicircle-transfected cells in six different cell lines and observed, that gene transfer rates were higher with minicircles than with controls (Kobelt *et al.*, 2013). Similar to other non-viral the problem of low transfer-rates, compared to viral vectors, remains to be solved (Nishikawa and Huang, 2001b).

Mammalian and human artificial chromosomes are rather big episomal DNA elements containing the functional elements of chromosomes, namely telomeres, centromeres and several replication origins. Artificial chromosomes behave like endogenous chromosomes in a cell, with respect to their mechanism of segregation (Larin Monaco and Moralli, 2006). The major difference between a minichromosome and a MAC/HAC[22] is size; minichromosomes cover sizes from 0.5 - 6 Mb and sizes of MACs/HACs range from 1 to 10 Mb (Lipps *et al.*, 2003). Advantages of MACs/HACs over retro- and adenoviral vectors are stable maintenance and in turn, no need for integration of the vector into the host genome. This is due to the presence of a functional centromere on the artificial chromosome itself. Transgene size in these vectors is not limited, but the fact, that gene copy numbers and the location of the gene on the MAC/HAC are not predictable is problematic. This means that transgene expression profiles and levels are not entirely controllable which resulted in only few applications of artificial chromosomes in practice (Kim *et al.*, 2011). Gene expression is checkable by construction of a HAC with 6000 *TetOperator* sites located in an alphoid DNA repeat. Targeting of *scTetR*-fusion proteins to the site leads to an inactivation of the centromere and therefore an inducible loss of the HAC (Nakano *et al.*, 2008). Recently human artificial chromosomes have been tested *in vitro* and *in vivo* in stem cell culture and mouse models and positive effects were reported for the treatment of

[21] Messenger RNA
[22] Mammalian artificial chromosomes/human artificial chromosomes

e.g. Duchenne muscular dystrophy and glioma by usage of this "pluripotent" gene vector (Kazuki *et al.*, 2011).

However, also the rather big size of MACs/HACs brings some problems along: first, only small amounts of the gene vector are available due to its complicate construction. MACs/HACs, in combination with the gene of interest, have to be co-transfected and the gene has to be inserted into the chromosome by site-specific recombination using Cre-LoxP or FLP-FRT (Lipps *et al.*, 2003). Second, due to the very large size, delivery of HACs and MACs to cells is difficult and rather laborious (Glover *et al.*, 2005; Larin Monaco and Moralli, 2006). For instance, microinjection, a technique, which is often used for the delivery of artificial chromosomes, enables only the treatment of one cell at a time and is limited to *ex vivo* application (Telenius *et al.*, 1999).

An advantage of artificial chromosomes and minichromosomes is, that they are not connected to the host chromatin. The problems that arise due to the unknown position and subsequent unknown expression of the transgene, that is characteristic for integrating vectors, is therefore of no significance in this case (Lipps *et al.*, 2003).

1.1.3.1 Segregation of minicircles, MACs/HACs and minichromosomes

Artificial chromosomes segregate *via* an active mechanisms based on the centromere on the vector. The presence of the functional centromere on MACs/HACs and minichromosomes is the reason, why they pass through mitosis like endogenous chromosomes and this also confers maintenance in the cell. During anaphase microtubules attach to the kinetochore on the HAC/MAC and the artificial chromosome are segregated symmetrically to both daughter cells (Figure 1.1 A).

The segregation mechanisms of minicircles depends on the type of vector they derive from. Minicircles are maintained in the cell extrachromosomally and are segregated *e.g.* like S/MAR-based vectors *via* the S/MAR element (Argyros *et al.*, 2011) or in case of EBV-derived vectors *via* EBNA1 attachment to the host cell chromosome (Zuo *et al.*, 2011). As discussed in chapter 1.1.1.1 and 1.1.2.1 these mechanisms are passive retention mechanisms, during which the DNA remains attached to the host cell chromosome (Figure 1.1 B).

1.1.4 The optimal extrachromosomal, plasmid-based gene vector

Summing up the positive and negative aspects of extrachromosomal and non-viral gene vectors, the following criteria seem to play important roles in the design of future vectors: *(1)* long-term guarantee of episomal retention, *(2)* high transgene size capabilities, *(3)* low immunogenicity, *(4)* small vector size and simple production, *(5)* vector administration into the target cell and *(6)* pDNA conformation.

Each of the described vectors above provides some of these factors, but only the combination of most - or even better all - of these factors will lead to the optimal gene transfer vehicle.

1.2 Centromeric DNA structures

Centromeres are defined as sites of kinetochore assembly and are absolutely essential for stable chromosome inheritance (Bergmann *et al.*, 2012). Although this function is crucial during cell division, centromeres do not contain evolutionary conserved structures in different species, *e.g.* they have no underlying DNA sequence that determines their location (Allshire and Karpen, 2008). Instead, considerable differences are found concerning centromeric structures among species (Figure 1.6) (Torras-Llort *et al.*, 2009).

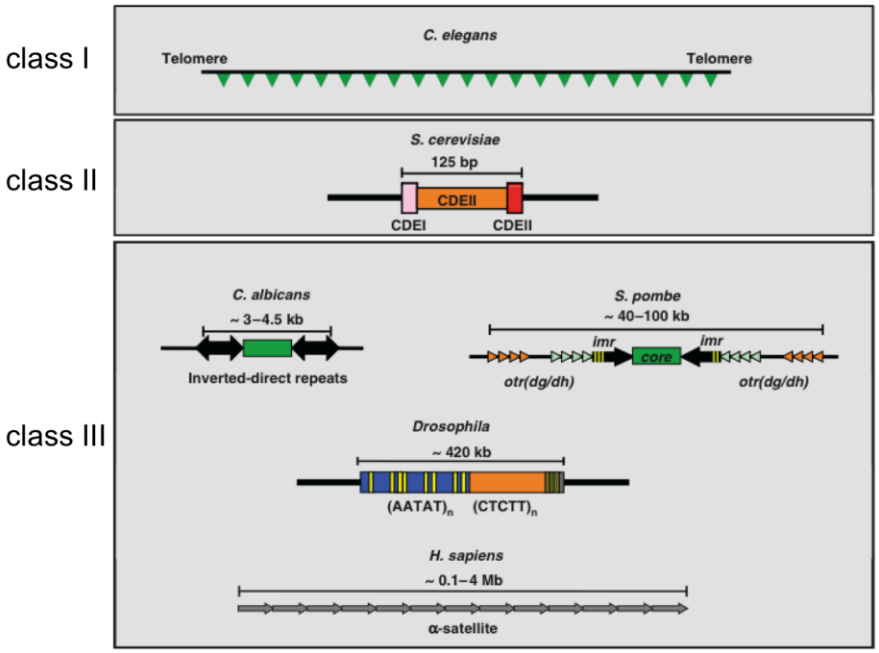


Figure 1.6 Centromeric structures among species.

Centromeric structures are classified in three main classes. The simplest version of a centromere is found in *C.elegans* where the complete genome represents the centromere (class I). In *S.cerevisiae* a 125 bp region consisting of highly conserved elements (CDEI, CDEII and CDEIII) has been identified as the area of centromere formation (class II). Higher organisms show rather large but very diverse centromere composition (class III).
(Torras-Llort *et al.*, 2009)

The most clear and conserved sequence defined centromeric structures are found in the budding yeast *S.cerevisiae*. Proper centromere function and thus genome stability in this organism depend on three elements, called centromere DNA elements (CDEI, CDE II and CDE III). Together these three elements form a 125 bp sequence (Cottarel *et al.*, 1989). Multi-cellular organisms do not exhibit such a defined centromeric region

(Wiens and Sorger, 1998) but they share the similarity, that centromeres are exclusively embedded in heterochromatin (Allshire and Karpen, 2008). It is important to note, that not only the centromeric core domain, but also the flanking pericentric heterochromatin, plays a role in the functionality of the centromere (Chan and Wong, 2012). Due to its heterochromatic state it was believed, that centromeric regions are not transcribed, but recent studies have shown that centromeres (pericentric regions and the core domain) are not only highly transcribed regions, but that transcription is also important for centromere chromatin identity (Chan and Wong, 2012).

1.2.1 Chromatin status of centromeric DNA

The diversity of centromeres in higher eukaryotes raises many questions about underlying chromatin structures and possible epigenetic regulators. The centromere specific histone H3 variant CENH3 plays a pivotal role in this context since it has been found at endogenous but also at neocentromeres, which establish at ectopic non-centromeric sites (Heun *et al.*, 2006). It has been demonstrated that CENH3 deposition is restricted to centromeric regions on chromosomes (Torras-Llort *et al.*, 2009). In most organisms the centromere specific histone CENH3 was identified in the mid 90s (Table 1.1) and are conserved mostly in their sequence (Figure 1.7 A).

Table 1.1 CENH3 variants among different species.

Organism	***CENH3 isoform***	***Original publication***
C.elegans	HCP3	(Buchwitz *et al.*, 1999)
A.thaliana	HTR12	(Copenhaver *et al.*, 1999)
S.pombe	Cnp1	(Clarke *et al.*, 1993)
S.cerevisiae	Cse4	(Stoler *et al.*, 1995)
D.melanogaster	CID	(Henikoff *et al.*, 2000)
X.laevis	XCENP-A	(Edwards and Murray, 2005)
M.musculus	CENP-A	(Rattner, 1991)
H.sapiens	CENP-A	(Palmer *et al.*, 1991)

Recently a discussion about the nomenclature of the centromere specific histone 3 variant emerged. In contrast to a before published paper that suggests to name all centromere specific histone 3 variants across species CENH3 (Talbert *et al.*, 2012) it was recommended to keep with the original nomenclature of the histones, *i.e.* CENP-A

for the human variant and CID for the *Drosophila* variant (Earnshaw *et al.*, 2013). During this thesis I made use of CENH3 of two species and to avoid confusion I use "CENH3" to refer to the centromere specific histone variant in general. If I specifically talk about one variant I use the terms "$CENH3^{CENP\text{-}A}$" (*H.sapiens*) and "$CENH3^{CID}$" (*D.melanogaster*) to discriminate between the two variants.

1.2.1.1 Chromatin status of centromeric DNA in D.melanogaster

As there is no consensus sequence known to account for centromeric regions the question remains where these functional elements establish on chromosomes. Experiments with the minichromosome *Dp1187* in *D.melanogaster* showed that the composition of the centromere-underlying DNA is primarily composed of α-satellites and, to a much lesser extent, transposons (Sun *et al.*, 1997). However, the satellite sequences are also found at multiple other positions in the genome that lack centromeric function and consequently they do not seem to be the defining element in centromere formation. Accordingly, the transposable sequences were excluded for the same reason as major determinant for centromere definition (Sun *et al.*, 1997).

As DNA sequence was excluded to be the centromere defining element, epigenetic factors like DNA methylation or centromere specific histone variants came under scrutiny and the centromere specific histone 3 variant in *D.melanogaster* was found and named centromere identifier ($CENH3^{CID}$) (Henikoff *et al.*, 2000). $CENH3^{CID}$ is a specialised histone variant present at centromeric hotspots in the fly genome (Henikoff *et al.*, 2000). The heterochromatic environment of $CENH3^{CID}$ is different in respect to conventional, heterochromatic histone modification marks and differs from both euchromatin and heterochromatin (Sullivan and Karpen, 2004). In *D.melanogaster* a specialised post-translational modification pattern might be the basis for centromeric chromatin formation. Centromeric histone modifications common for pericentric heterochromatin, H3K9 di- and trimethylation, are not found at centromeric chromatin during interphase and mitosis. Typical euchromatic reference marks (*e.g.* H3K9ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac) are not detected at centromeric sites. However, the H3K4-dimethylation mark, which usually accounts for active chromatin, is detected between CENH3 spreads but not at pericentric heterochromatin. H3K4 trimethylation is not found although it is known to be an euchromatic marker. All of these findings also hold true for human centromeres (Sullivan and Karpen, 2004).

$CENH3^{CID}$ is a 678-nucleotide long gene, which displays similarity to canonical H3 but has, like all other centromere-specific histone variants, an unique N-terminus (Henikoff *et al.*, 2000). RNAi[23]-mediated depletion of $CENH3^{CID}$ leads to compensation with H3 (Blower *et al.*, 2002). $CENH3^{CID}$ is only localised at endogenous centromeres (Henikoff *et al.*, 2000), but high overexpression also causes the induction of ectopic centromere formation (Heun *et al.*, 2006). This results in a high number of mitotic defects and chromosome aberrations, like chromosome breakage. This observation provides evidence, that induced ectopic centromeres are functional and able to assemble all kinetochore components that are necessary for microtubule attachment (Heun *et al.*, 2006). In contrast to mammalian cells, deposition of $CENH3^{CID}$ in *D.melanogaster* happens already during anaphase/metaphase (Mellone *et al.*, 2011; Schuh *et al.*, 2007). The proteins accounting for proper deposition of $CENH3^{CID}$ at centromeres are CENP-C[24] and CAL1[25]. These are two key proteins at *Drosophila* centromers, which are recruited to centromeres in early mitosis (prophase) (Dunleavy *et al.*, 2012; Mellone *et al.*, 2011). The function of CAL1 is to limit $CENH3^{CID}$ expression to avoid mitotic instability due to centromere expansion as a consequence of too high CENH3 expression (Dunleavy *et al.*, 2012). Furthermore CAL1 forms a complex with $CENH3^{CID}$ and CENP-C and both proteins are responsible for CID deposition in prophase of meiosis (Dunleavy *et al.*, 2012). Only $CENH3^{CID}$ and CENP-C are symmetrically passed on to the daughter cells, whereas this is not true for CAL1 (Mellone *et al.*, 2011). Additionally, CAL1 is present in much lower levels than $CENH3^{CID}$ and CENP-C (Dunleavy *et al.*, 2012).

1.2.1.2 Chromatin status of centromeric DNA in H.sapiens

Heterochromatin is the basis for all endogenous human centromeres (Allshire and Karpen, 2008). In fact, centromeres are located on long tandem repeats of α-satellite sequences. These AT-rich sequences are 171 bp-long monomer tandem repeats of α-satellite DNA (Choo *et al.*, 1991; Waye and Willard, 1987) and make endogenous centromeres distinguishable from neocentromeres, which do not exhibit this type of underlying DNA structure (Warburton, 2004). The DNA sequence of centromeric loci is

[23] RNA interference
[24] Centromere protein-C
[25] Chromosome alignment defect 1

not conserved (Black *et al.*, 2004) and centromeric DNA is neither sufficient nor necessary for the formation of centromeres (Torras-Llort *et al.*, 2009).

The histone composition at centromeric regions comprises the canonical histones H2A, H2B, H3 and H4, with the intercalating histone 3 variant $CENH3^{CENP\text{-}A}$ that replaces canonical H3. The epigenetic mark of centromeric DNA is definitely the accumulation of $CENH3^{CENP\text{-}A}$ (Black *et al.*, 2004). Proteins that assemble and build the kinetochore (chapter 1.2.2) are highly similar between organisms; *e.g.* CENP-A or CENP-C is always found at the respective sites (Houben and Schubert, 2003; Shen *et al.*, 2001).

Like $CENH3^{CID}$, also $CENH3^{CENP\text{-}A}$ is exclusively found at centromeric sites of human chromosomes. If overexpressed, $CENH3^{CENP\text{-}A}$ also displaces at ectopic sites (Sullivan *et al.*, 1994). The importance of $CENH3^{CENP\text{-}A}$ *in vivo* is demonstrated in homozygous knock-out mice where the depletion of $CENH3^{CENP\text{-}A}$ has lethal consequences (Howman *et al.*, 2000). The N-terminal histone tail of $CENH3^{CENP\text{-}A}$ consists of a unique sequence whereas the sequence similarity of the globular C-terminal part of $CENH3^{CENP\text{-}A}$ is more than 60 % identical to the canonical histone H3 homologue (Sullivan *et al.*, 1994) This domain is known as the histone fold domain and is found in all histones as the dimerization domain (Arents and Moudrianakis, 1995). In comparison to other histones, which share most of the sequences, $CENH3^{CENP\text{-}A}$ is poorly conserved between species at the amino acid level (De Rop *et al.*, 2012; Sarma and Reinberg, 2005). Especially those regions, which are not at all related to other histones, are supposed to determine $CENH3^{CENP\text{-}A}$´s distinct function (Palmer *et al.*, 1991). Unlike other histone 3 variants, which account for the histones with the most modified tails, $CENH3^{CENP\text{-}A}$ only has one modification on the N-terminus identified so far (Stellfox *et al.*, 2012).

Besides the non-conserved N-terminus, the major difference between H3.3 and $CENH3^{CENP\text{-}A}$ is the *CENP-A targeting domain* (*CATD*, Figure 1.7 B) which is missing in H3.3 and which targets $CENH3^{CENP\text{-}A}$ specifically to future centromeric sites. If wildtype histone 3 is supplemented with a *CATD*, the centromeric function is rescued (Black *et al.*, 2007). The *CATD* is located from loop1 to α2 helix of the histone fold domain of the protein and comprises 35 amino acids. Also in the budding yeast $CENH3^{Cse4p}$, a *CATD* homologue, was identified but $CENH3^{CID}$ does not seem to harbour this domain.

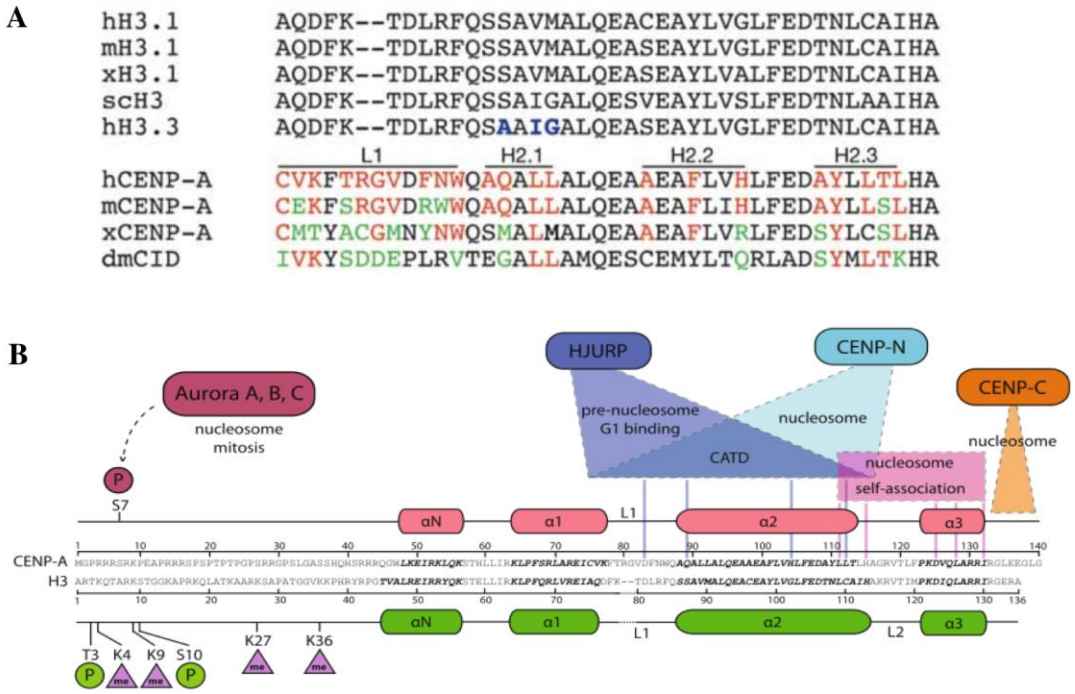


Figure 1.7 Sequence and structure alignments of histone 3 variants.

A) The sequences of loop1 (L1) and α2 helix (H2) of histone 3.1 are conserved among species (h: *H.sapiens*, m: *M.musculus*, sc: *S.cerevisiae*, x: *X.laevis*, dm: *D.melanogaster*). CENH3 sequences contain conserved motifs only in the region of L1 and H2, which represents the *CENP-A targeting domain* (*CATD*). The *CATD* is responsible for binding of CENH3^{CENP-A} to centromeric DNA and is the only difference between canonical histone 3.3 and CENH3^{CENP-A}. Deviations from CENH3 to canonical histone 3 are conserved mostly among species (red lettering) or are otherwise substituted (green lettering) (Black *et al.*, 2004). **B)** Specific differences of CENH3^{CENP-A} are shown in comparison to canonical histone 3. Additionally to the difference of the *CATD*, CENH3^{CENP-A} also has a binding site for its chaperon HJURP and kinetochore proteins like CENP-N and CENP-C. Interestingly, CENH3^{CENP-A} is much less modified by posttranslational modifications as compared to H3.
(Stellfox *et al.*, 2012)

CENH3^{CENP-A} nucleosomes build up the fundamental platform for kinetochore assembly and are the determinant tool of chromosome segregation. Like all other histones, also CENH3^{CENP-A} has a histone fold domain and it builds heteronucleosomes with histone H4. CENH3^{CENP-A} is recruited in a cell cycle-dependent manner and is tethered to the centromeric site *via* its chaperon, the Holliday junction recognition protein (HJURP) (Dunleavy *et al.*, 2009). This deposition of CENH3^{CENP-A} at centromeres takes place in late telophase and early G_1 (Jansen *et al.*, 2007) and it was shown that also all kinetochore proteins that are needed for a functional kinetochore are assembled after HJURP-mediated CENH3^{CENP-A} deposition (Barnhart *et al.*, 2011).

1.2.2 Kinetochore composition

The kinetochore is a protein structure, which is responsible for microtubule attachment to chromosomes and proper chromosome segregation (Santaguida and Musacchio, 2009; Welburn and Cheeseman, 2008). The classical kinetochore is divided into two parts – the inner kinetochore, which assembles directly on centromeric DNA and the outer kinetochore, which is loaded on top of the inner kinetochore and provides the attachment platform for microtubules during cell division (Santaguida and Musacchio, 2009) (Figure 1.8).

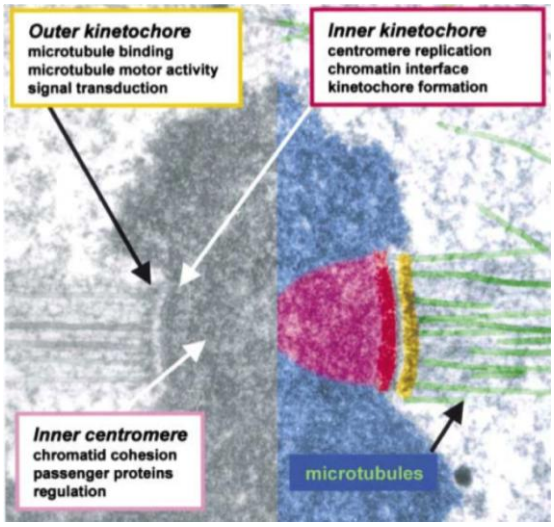


Figure 1.8 Schematic of the centromere-kinetochore-microtubule structure.

Left, grey panel: The section of a chromosome illustrates the structure of the centromere-kinetochore-microtubule arrangement during mitosis. Right, coloured panel: The different elements are pseudo-coloured: inner (magenta) and outer (yellow) kinetochore structures; inner centromere (pink) and microtubules (green). The respective functions of each element are highlighted in the coloured boxes.
(Cleveland *et al.*, 2003)

Overall more than 80 proteins are involved in kinetochore structure and function (Cheeseman and Desai, 2008). Inner kinetochores are always a prerequisite for assembly of the outer kinetochore components, like the KMN network[26] (Santaguida and Musacchio, 2009). The constitutive centromere-associated network, CCAN, is built of a subset of proteins at the inner kinetochore close to centromeric DNA and remains at this position throughout the whole cell cycle. On the other side – the outer kinetochore – the KMN complex assembles (kinetochore-microtubule-attachment complex), which is

[26] KNL-1/Mis12 complex/Ndc80 complex

bound by microtubules during mitosis and is only assembled during this specific cell cycle stage (Maresca, 2011). A complete list of kinetochore proteins in *H.sapiens* and *D.melanogaster* is given in appendix 9.3.

CENH3^{CENP-A} was identified as one of the first centromeric histones that build up the basis for kinetochore assembly and CENP-C was found to be a marker for the inner kinetochore. Furthermore, the three connected centromeric proteins CENH3^{CENP-A}, CENP-B[27] and CENP-C are only found at active centromeres (Earnshaw and Migeon, 1985). The connection between the kinetochore and its binding factors was not entirely known for a long time. Only recently the inner kinetochore marker CENP-C was found to be not only the connecting link between the centromere and the inner kinetochore *via* its C-terminal tail, but also the interface between the outer and the inner kinetochore *via* its N-terminus in *D.melanogaster* and HeLa cells (Przewloka *et al.*, 2011; Screpanti *et al.*, 2011). Ectopic kinetochore formation usually takes place near heterochromatic regions or telomeres (Olszak *et al.*, 2011). In this context, the N-terminus of CENP-C is again responsible for recruitment of the KMN network (Przewloka *et al.*, 2011).

1.2.3 Neocentromerization of chromatin

The phenomenon, that a formerly non-centromeric region of the genome acquires centromeric function, including the formation of a kinetochore, is called neocentromerization (Warburton, 2004). Due to the fact that two functional centromeres on one chromosome lead to genome instability, neocentromeres are only formed in the case of chromosomal breakage and the centromere-lacking chromosomal fragment would otherwise be lost during mitosis or meiosis (Burrack and Berman, 2012). Formation of a neocentromere on these breakage products rescues these fragments from being lost (Rocchi *et al.*, 2012). Another reason for neocentromerization is silencing of the endogenous centromere (Warburton, 2004). In this case the endogenous centromere remains unchanged, except for the fact that the site is devoid of CENH3^{CENP-A} and other CEN-proteins (Warburton *et al.*, 1997). Under these circumstances the genome is not affected in any negative way (Rocchi *et al.*, 2012).

In 1993 the first case of human neocentromere formation was reported (Voullaire *et al.*, 1993) and until today more than 90 neocentromeres were described in human genomes (Marshall *et al.*, 2008a). Usually neocentromeres are identified from patient samples as

[27] Centromere protein-B

they happen stochastically and are detectable by developmental delay or dimorphic patients (Warburton, 2004). Neocentromeres are not found directly in heterochromatic regions (Alonso *et al.*, 2010) but only in close proximity to them (Olszak *et al.*, 2011). In fact, they are located in euchromatic, highly transcribed regions of chromosomes and do not have a significant impact on protein expression (Marshall *et al.*, 2008a; Rocchi *et al.*, 2012). The structure-specific recognition protein 1 (SSRP1) and RNA polymerase II are found at enriched levels at active neocentromeres during mitosis (Chan *et al.*, 2012). SSRP1 is a subunit of the FACT[28] complex, which is believed to be involved in $CENH3^{CENP\text{-}A}$ deposition (Okada *et al.*, 2009).

In contrast to endogenous centromeres, *de novo* formed neocentromeres are not characterised by the prerequisite that repetitive sequences (α-satellites) are present at the site (Malik and Henikoff, 2000; Rocchi *et al.*, 2012) and it actually seems that neocentromeres are actively avoiding repetitive regions of the genome. Just as for endogenous centromeres, also for neocentromeres no consensus DNA sequence was identified that determines their location on the genome. It is much more likely that again an epigenetic influence directs this mechanism (Warburton, 2004).

$CENH3^{CENP\text{-}A}$ localizes to neocentromeres no matter whether α-satellite DNA is present or not (Warburton *et al.*, 1997). However, the amount of $CENH3^{CENP\text{-}A}$ at neocentromeres is reduced to one third of endogenous centromeres for yet unknown reasons (Irvine *et al.*, 2004; Marshall *et al.*, 2008a). The fact that $CENH3^{CENP\text{-}A}$ is recruited to damaged DNA pieces substantiates the suspicion, that DNA repair and neocentromere establishment are tightly connected features – in fact the latter can happen because of the first event of recruitment (Zeitlin *et al.*, 2009). It is not known if the role of the chaperone HJURP is the same at neocentromeres as for endogenous centromeres (Burrack and Berman, 2012).

Neocentromeres exhibit large differences among each other like *e.g.* size and/or the presence of H3 nucleosomes in the underlying DNA sequence (Stimpson and Sullivan, 2010). Most neocentromeres established at certain regions of the genome, with the highest indication at positions 3q, 8p, 13q, 15q and Yq (up to 16 neocentromeres are reported for these positions). At other chromosomal positions neocentromeres occur in much lower frequencies with a maximum of five cases in this group reported for chromosome 1 (Marshall *et al.*, 2008a; Rocchi *et al.*, 2012). In *D.melanogaster* the

[28] Facilitates chromatin transcription

proximity to a centromere is necessary for active neocentromere establishment (Maggert and Karpen, 2001).

1.2.3.1 Artificial neocentromeres

The centromere specific histone CENH3 is an epigenetic determinant for active centromeres and at functional neocentromeres the presence of CENH3 is a prerequisite. The fusion protein CENH3^{CID}:GFP:*LacI* has been successfully recruited to an array of integrated *LacOperators* on chromosome 3R in *D.melanogaster* (Olszak *et al.*, 2011). It was proven by immunofluorescence and ChIP analyses that a neocentromere establishes at the site similar to the endogenous centromere (Figure 1.9).

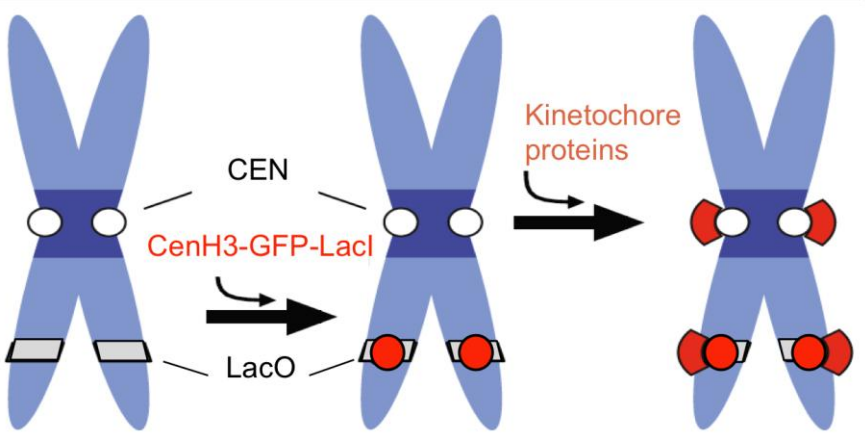


Figure 1.9 Formation of artificial neocentromeres by CENH3:GFP:LacI targeting to chromosomes.

A) 256 *LacOperator* sites (grey rectangles) integrated in the chromosomal arm of chromosome 3R in *D.melanogaster* recruit the CENH3^{CID}:GFP:*LacI* fusion protein to the site. Similarly to endogenous centromeres (white circles) neocentromeres (grey boxes) establish on the *LacOperator* arrays. These neocentromeres are able to recruit kinetochore proteins (red forms). Kinetochore assembly has been demonstrated in *D.melanogaster* and is expected to function in a similar manner in human cells.
(Mendiburo *et al.*, 2011)

The ectopic neocentromeres function like an endogenous centromere and seem to assemble similar proteins for the inner and outer kinetochore to the site (like *e.g.* POLO, Ndc80, Mad2 and CENP-C). The neocentromeres are fully functional and work like the endogenous analogue. Recently Teo *et al.* were able to prove, that even just the N-terminal domain of CENH3 fused to GFP:*LacI*:NLS is sufficient to provide targeting of the protein to integrated *LacOperator* sites in the chromosomal arms and to enable neocentromere formation in *A.thaliana* (Teo *et al.*, 2013).

1.3 Aim – Establishment of a CENH3-dependent, active pDNA vector segregation mechanism

Gene therapy is a field with great potential for the treatment of various diseases like cancer, genetic disorders or autoimmune diseases. The major goal in gene vector design is to improve transgene vehicles with regard to genome integrity and immunological issues of the patient and making them more efficient concerning retention and transgene expression. Viral vectors, which are currently used in most clinical applications, integrate into the host genome and permit a stable expression of the transgene. However, integration leads to genotoxic effects (*e.g.* the activation of oncogenes) and viral particles are always a potential source of immunological reactions by the patient. To avoid these side effects much effort is put on the development of non-viral and extrachromosomal pDNA gene vectors.

Epstein-Barr virus derived vectors are one sub-group of plasmid-based, non-viral gene vectors. The essential tool of EBV-based vector is the use of *oriP*, a bipartite structure responsible for DNA replication and pDNA vector retention of the virus. The conditional plasmid system *pCON*[29] was established and studies by Pich *et al*. led to the first generation of gene vectors in an *oriP*-based context (Pich *et al.*, 2008). During earlier work in our laboratory the roles of HP1α and HMGA1a in regards to pDNA vector retention were investigated (Deutsch, M., unpublished) (Thomae *et al.*, 2008). Both proteins are considered to be able to substitute the function of retention of *oriP*. Like *oriP* and *pEPI* vectors, they use the passive *piggyback* mechanism to distribute the pDNA vector to daughter cells.

Based on these observations, the aim of this work was to establish an entirely new generation of non-viral gene vectors, which use an active segregation mechanism. By using CENH3 as segregation factor, I aimed to create a new pDNA vector family that behaves like a "quasi-chromosome" (in this respect like mammalian artificial chromosomes) with is own CENH3-induced centromere (Figure 1.10).

[29] Conditional plasmid system

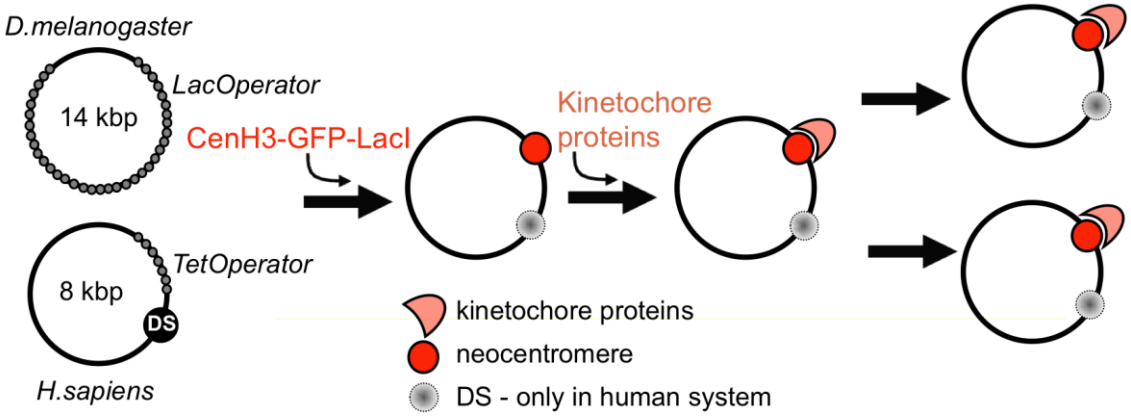


Figure 1.10 CENH3 induced formation of neocentromeres on pDNA vectors.

Overexpression of the centromere specific histone variant CENH3^{CID} in cells leads to the ectopic deposition of the protein and the formation of neocentromeres (Heun *et al.*, 2006). We transferred this system from the chromosomal level to the pDNA vector. With this set-up it is possible to test the functional relevance of CENH3 incorporation by targeting of the protein to usually non-centromeric sites. For experiments in *D.melanogaster* I used the *LacO-LacI* targeting system, for experiments in *H.sapiens* I used the *TetO-scTetR* targeting system. Due to the formation of neocentromeres and transforming the pDNA vector into "quasi-chromosomes", pDNA vectors are maintained in the cell over a prolonged period of time by an active segregation mechanism that is independent of the host chromosomes as shown before in Figure 1.1. The formation of neocentromeres is epigenetically conserved and maintains also in the daughter pDNA vectors after replication and segregation.

2 MATERIALS

2.1 Materials

2.1.1 Chemicals, antibiotics and buffers

Table 2.1 shows all chemicals, antibiotics, buffers and other substances that were used during this work.

Table 2.1 Substances used in this work.

Substance	***Distributor***
Agarose	Invitrogen GmbH, Karlsruhe
Albumin, bovine	Sigma-Aldrich Chemie, München
Ampicillin sodium salt	Carl Roth GmbH und Co KG, Karlsruhe
ATX Poncau S red staining solution	Fluka Analytical, Steinheim
Bio-Rad Protein Assay	Bio-Rad, München
BSA, purified, 100x	New England Biolabs, Schwalbach
Caesium chloride	AppliChem GmbH, Darmstadt
Calciumchlorid ($CaCl_2$)	Merck-Eurolab GmbH, Darmstadt
Chloroform	Merck-Eurolab GmbH, Darmstadt
dCTP [α-32P]	Hartmann Analytic GmbH, Braunschweig
DNA loading dye, 6x	New England Biolabs, Schwalbach
Dimethylsulphoxide (DMSO)	Carl Roth GmbH und Co KG, Karlsruhe
Disodiumhydroxyphosphat	AppliChem GmbH, Darmstadt

Dithiothreitol (DTT)	Sigma-Aldrich Chemie, München
Deooxycholic acid (DOC)	Sigma-Aldrich Chemie, München
Drosophila Schneider S2 medium	PAN Biotech GmbH, Aidenbach
Dulbecco's Eagle modified medium	Gibco Life Technologies GmbH, Darmstadt
Ethylendiamintetraacetic acid	Carl Roth GmbH und Co KG, Karlsruhe
Ethanol	Merck-Eurolab GmbH, Darmstadt
Eosin	Merck-Eurolab GmbH, Darmstadt
G418/Geneticin	Carl Roth GmbH und Co KG, Karlsruhe
GeneRuler™ 1 kb DNA ladder	MBI Fermentas, St. Leon-Rot
Glycerin	AppliChem GmbH, Darmstadt
Glycin	Carl Roth GmbH und Co KG, Karlsruhe
Hydrochloric acid	Merck-Eurolab GmbH; Darmstadt
Hygromycin	PAA-Laboratories, Wien, Austria
Isoamyl alcohol	Merck-Eurolab GmbH; Darmstadt
Isopropanol (2-Propanol puriss.)	Merck-Eurolab GmbH, Darmstadt
Lipofectamine 2000	Invitrogen GmbH, Karlsruhe
Manganese chloride ($MnCl_2$)	Fluka Analytical, Steinheim
Methanol	Merck-Eurolab GmbH; Darmstadt
Milk powder	Carl Roth GmbH und Co KG, Karlsruhe
NP-40 (Igepal CA-630)	Sigma-Aldrich Chemie, München
Ortho-phosphoric acid	Merck-Eurolab GmbH; Darmstadt
PBS Dulbecco, pH 7.2	Biochrom AG, Berlin
Piperazine-N,N'-bis(ethanesulfonic	ICN Biomedicals Inc., Aurora, Ohio, USA
Polyacrylamide	Carl Roth GmbH und Co KG, Karlsruhe
Potassium hydroxide	AppliChem GmbH, Darmstadt
Puromycin	AppliChem GmbH, Darmstadt
RPMI-1640 cell medium	Gibco Life Technologies GmbH, Darmstadt
Select agar	Invitrogen GmbH, Karlsruhe
Sodium acetate	Merck-Eurolab GmbH, Darmstadt
Sodium dodecylsulfate (SDS)	Serva Electrophoresis GmbH; Heidelberg
Sodium chloride	Merck-Eurolab GmbH, Darmstadt
Sodium citrate	Merck-Eurolab GmbH, Darmstadt
Sodium hydroxide	Carl Roth GmbH und Co KG, Karlsruhe
‚N',N',N',N'-Tetramethylendiamine	Carl Roth GmbH und Co KG, Karlsruhe
Triton-X100	Sigma-Aldrich Chemie, München
Trypsin-EDTA	Gibco Life Technologies GmbH, Darmstadt
Tween-20	AppliChem GmbH, Darmstadt
Tryptone enzymatic digest from	Carl Roth GmbH und Co KG, Karlsruhe
Yeast extract	Gibco Life Technologies GmbH, Darmstadt

2.1.2 Devices

The following table (Table 2.2) files all devices used during this work.

Table 2.2 Devices used in this work.

Device	***Distributor***
Analysis balance sealtec (< 120 g)	Bayrische Waagenwerkstätte, Utting a.A.
Balance kern 470 (< 2000 g)	Kern, Albstadt
Bio photometer	Eppendorf, Hamburg
Biorupter UCD-200	Diagenode, Liège, Belgium
Centrifuge 2K15	Sigma, Ostenrode am Harz
Centrifuge Rotina 38R	Hettich, Tuttlingern
Developer machine CP100	AGFA, Köln
Electroporation cuvettes 1mm	Peqlab Biotechnologie GmbH, Erlangen
Electroporation device Gene	Bio-Rad Laboratories, Richmond CA, USA
Film cassettes	FujiFilm, Kleve
Gel documentation system	peqlab GmbH, Erlangen
Gene Amp PCR system 2400	Perkin Elmer, USA
High-speed centrifuge Avanti J-	Beckman-Coulter, München
Hotplate/magnetic stirrer RH	IKA Labortechnik, Staufen
Illustra Microspin G-50 columns	GE Healthcare, München
Incubator shaker innova 4400	New Brunswick Scientific GmbH, Nüttingen
Incubator shaker innova 44	New Brunswick Scientific GmbH, Nüttingen
Incubator kelvitron-t, 26.5 °C	Heraeus, Hanau
Incubator Napco L5410, 37 °C,	UniEquip, Martinsried
Intensifier screen	FujiFilm, Kleve
Light-Cycler™	Roche Diagnostics GmbH, Mannheim
Microcon filter device	Millipore, Schwalbach
Milli-RO 60 PLUS water filter	Millipore, Schwalbach
NanoDrop® ND-1000	ThermoScientific, USA
Optimax x-ray film processor	Protec GmbH & Co KG, Oberstenfeld
Phosphoimager FLA 5100	FujiFilm, Kleve
Qubit fluorometer	Invitrogen, Darmstadt
SemiDry blotting system	Hoefer Scientific Instruments, USA
SevenEasy InLab413 pH-meter	Mettler Toledo, Gießen
Spectrometer DU 640	Beckmann, Heidelberg
Syringe filter	Asahi Techno Glass Co, Singapur
Table centrifuge 5415R	Eppendorf, Hamburg
Thermomixer comfort	Eppendorf, Hamburg
Thermomixer AccuBlock Digital	Labnet International Inc., USA

Ultracentrifuge Optima L-70	Beckman-Coulter, München
UV transilluminator TS40 254 nm	Herolab Inc., USA
Vortex Genie 2	Scientific Industries Inc., USA
Waterbath Acoline 100 Lauda	Dr.R.Wobser GMBH & Co KG, Königshofen

2.1.3 Kits

Table 2.3 shows all kits that were used during this work.

Table 2.3 Kits used in this work.

Kit	***Distributor***
JETstar Maxi-prep kit	Genomed GmbH, Löhne
NucleoSpin Gel and PCR Clean-	Macherey-Nagel, Düren
CloneJET PCR Cloning kit	MBI Fermentas, St. Leon-Rot
Random primed DNA labelling	Roche, Penzberg

2.1.4 Enzymes

All utilised restriction enzymes were purchased from New England Biolabs, Schwalbach and Fermentas, Thermo Scientific, USA. More enzymes are listed in Table 2.4.

Table 2.4 Enzymes used in this work.

Enzyme	***Distributor***
T4 DNA ligase	Affymetrix/usb products, UK
Pwo polymerase	Peqlab, Erlangen
RNase, DNase-free	Roche, Penzberg
Proteinase K	Carl Roth GmbH und Co KG, Karlsruhe
Antarctic phosphatase	New England Biolabs, Schwalbach
Taq polymerase	kindly provided by C.-E. Mayer

2.1.5 Plasmids

Plasmids listed in Table 2.5 were used in this thesis. Numbers indicate the Helmholtz Zentrum München (HMGU) database ID. *-labelled plasmids stem from the HMGU AGV plasmid collection and °-labelled plasmids were provided by other research groups. All other plasmids were cloned during this work.

Table 2.5 Plasmids used in this work.

HMGU ID	*Characteristics*	*Creator*
3230*	*wt-oriP*	(Gerhardt *et al.*, 2006)
3231*	DS/20 *TetOperator*	WH/AT[30], HMGU
3293.5*	DS/40 *TetOperator*	WH, HMGU
3223.9*	Puromycin-cassette	WH, HMGU
4201*	*scTetR*: CENH3^{CENP-A} (mini-CMV - 53 + 74)	M. Deutsch, HMGU
4202*	*scTetR*: CENH3^{CENP-A} (E-Cad - 178 + 92)	M. Deutsch, HMGU
4203*	*scTetR*: CENH3^{CENP-A} (mini E-Cad - 21 + 92)	M. Deutsch, HMGU
4890	*scTetR*:H3.3 (E-Cad -178 + 92)	S. Fülöp, M. Amman
4892	*scTetR*: CENH3^{CENP-A}-*CATD*^{mut} (E-Cad – 178 + 92)	S. Fülöp, M. Amman
5403°	256 x *LacOperator*	P. Heun, MPI Freiburg

2.1.6 Primer sequences

Following tables (Table 2.6 - Table 2.8) list primers used during this work. Numbers before the primer name indicate for the primer ID in the Schepers group primer collection.

Table 2.6 Primers used for qPCR.

Primer ID	*Sequence (5´ > 3´)*
118 G418 fw	ATGATTGAACAAGATGGATTG
119 G418 rev	TCAGAAGAACTCGTCAAGAAG

Table 2.7 Primers used for cloning.

Primer ID	*Sequence (5´ > 3´)*
033 5´CENP-A fw	AGTCGGCCGGCCATGGGCCCGCGCCGCCGG
034 3´CENP-A rev	ATGCGCTAGCTCAGCCGAGTCCCTCCTC
043 H3.3 fw	CGACGGCCGGCCATGGCCCGAACCAAGCAGACT
044 H3.3 rev	CCCGCTAGCTTAAGCTCTCTCTCCCCGTAT
151 HJURP s	AAAACGGCCGCTATGCTGGGTACGCTGCGC
152 HJURP as	AATTCGGCCGCTACACACTTTTAGTTTCCAATTTTTCTAG

Table 2.8 Primers used for sequencing.

Primer ID	*Sequence (5´ > 3´)*
255 Amp 3´fw	GGGATCATGTAACTCGCCTTG

[30] Wolfgang Hammerschmidt/Andreas Thomae

256 Amp 5´rev	CGACACGGAAATGTTGAATACTC
317 Amp follow-up fw2	CGGGTTGGACTCAAGACGATAG
321 Amp follow-up fw3	TGATCATCTCAGTGCAACTAAAGG

2.1.7 Commercial bacteria strains

- DH5α for cloning (preparation see chapter 3.2.3)

F-; lacI-; recA1; endA1; hsdR17; (lacZYA-argF), U169, F80dlacZ M15; supE44; thi-1; gyrA96; relA1 (Hanahan, 1983).

- Electromaxx DH10B for electroporation in plasmid rescue assays (Invitrogen, Karlsruhe)

F-; mcrA; (mrr-hsdRMS-mcrBC), 80dlacZM15; lacX74; deoR; recA1; endA1; araD139; (ara, leu)7697; galU; galK; -; rpsL; nupG

2.1.8 Commercial cell lines

- HEK 293 EBNA1

human embryonic kidney cells; stably integrated EBNA1

- *Drosophila* Schneider S2 (kindly provided by P. Heun)

late stage (20-24h) *D.melanogaster* embryonic cells

2.1.9 Antibodies

The α-CENH3^{CENP-A} antibody was stored at -20 °C, α-*TetR* antibody was stored at 4 °C. Detailed information about the application of primary antibodies is shown in Table 2.9.

Table 2.9 Primary antibodies used in this work.

Specificity	*Origin*	*Application*	*Dilution*	*Distributor*
α-CENH3^{CENP-A}	rabbit	Western Blot	1:2500	LifeSpan (Biozol) Eching
α-*TetR*	rabbit	Western Blot	1:2000	C. Berens, Erlangen

The secondary α-rabbit antibody was purchased from Jackson Immuno Research Labs and applied in Western Blot in a 1:10000 dilution in 2.5 % milk in PBS-T.

2.1.10Blotting membranes

Amersham Hybond ECL for Western Blot (GE Healthcare, München)

Nylon membrane for Southern Blot (Millipore, Bedford, UK)

3 METHODS

3.1 Cell biological methods

3.1.1 Cultivation of cells

HEK 293-EBNA1-*scTetR*:CENP-A and HEK 293-EBNA1-*scTetR* cells were cultivated at 37 °C and 5 % CO_2 in DMEM and RPMI-1640 medium, respectively. Schneider S2 cells were cultivated at 26.5 °C in PAN Schneider's *Drosophila* medium and did not require CO_2 supply. All cell culture media were supplemented with 10 % FCS[31] and 1 % Penicillin/Streptomycin as standard practice. The respective antibiotic concentrations for selection are listed in Table 3.1.

Table 3.1 Concentration of antibiotics for selection of cells.

Antibiotic	*HEK 293*	*Drosophila Schneider S2*
Puromycin	300 – 700 ng/ml	2 µg/ml
Neomycin/G418	220 ng/ml	1 mg/ml
Hygromycin	150 µg/ml	100 µg/ml

3.1.2 Cell number determination

Cells were counted with the help of a Neubauer cell counting chamber. To differentiate between dead and living cells the samples were mixed with eosin in a 1:1 ratio. Eosin stains all dead cells and cell fragments whereas living cells do not take up the dye. The dilution must be considered in the calculation of the total cell count. Cells in the middle squares of two chambers were counted and according to the following equation the number of cells/ml was determined:

$$[(n/N)]*d*2*10^4 = \text{cells/ml}$$

n: number of counted cells in all counted squares
N: number of counted squares
d: dilution factor of the cell suspension

[31] Foetal calf serum

3.1.3 Cryopreservation

Cells were frozen down in a suspension of 90 % FCS mixed with 10 % DMSO[32] to -80 °C in a 2 ml CryoTube (Nunc) and slowly cooled in a "Nalgen Nunc Cryo 1 °C Freezing Container" (Nunc) with a cooling rate of -1 °C/minute. After a few days cells were transferred to the liquid nitrogen tank (-196 °C) for long-term storage.

For thawing cells were warmed to 37 °C, washed with 10 ml fresh medium to get rid off DMSO and plated on 15 cm cell culture dishes in 20 ml of fresh medium.

3.1.4 Transfection of HEK 293 cells with Lipofectamine 2000 (Invitrogen)

Transfections were performed according to manufacturer's instructions. 1.5 x 10^5 cells were seeded per well of a 6-well plate and transfected 24 h later. 1 µg of DNA and 2 µl of Lipofectamine (µl Lipofectamine = 2 x µg DNA) were diluted separately with Opti-MEM medium to a final volume of 50 µl each and incubated for five minutes. Subsequently both solutions were mixed and incubated for 20 minutes at room temperature. The resulting 100 µl of transfection solution were applied to the cells in a drop-wise manner. Prior to this procedure FCS was removed from the cell culture medium to avoid interference with the complex formation reaction. 24 hours post transfection cells were transferred to a 15 cm cell culture dish and the medium was changed to serum-supplemented medium containing the according selection antibiotic (Table 3.1).

3.1.5 Establishment of stable cell lines

A plasmid encoding for the *scTetR*-fused protein of choice (CENH3 and H3.3) and one with a puromycin cassette were co-transfected in HEK 293 EBNA1 cells. Both plasmids were linearised prior to transfection to promote integration into the host genome. To increase the efficiency of selection, the *scTetR*-fusion protein was used in a higher concentration compared to the puromycin cassette plasmid (ratio 10:1). Transfections were performed as described in 3.1.4. After a selection period of two to three weeks single colonies were picked with autoclaved pieces of thin Whatman paper soaked in trypsin. The cells sticking on the paper pieces were transferred to 6-wells and cultivated

[32] Dimethylsulfoxide

as described in 3.1.1. To confirm the expression of the fusion protein, Western Blot analysis was performed (see chapter 3.3.3). The following cell lines were established during this thesis:

- HEK 293 EBNA1-*scTetR*:CENH3^{CENP-A}

three cell lines under the control of different promoters

1. E-Cadherin promoter (- 178 + 92)
2. mini-E-Cadherin promoter (- 21 + 92)
3. mini-CMV promoter (- 53 + 74)

HEK 293 EBNA1-*scTetR*:H3.3

1. E-Cadherin promoter (- 178 + 92)

3.2 Molecular biological methods

General lab techniques like agarose gel electrophoresis, phenol-chloroform extraction and DNA precipitation were performed according to standard protocols as described in Sambrook *et al.* (Sambrook and Russell, 2001).

3.2.1 Bacterial culture

Culturing of *Escherichia coli* (*E.coli*) was performed in LB-broth (Luria-Bertani medium). For cloning purposes cells were grown as single colonies on LB-plates. After transformation cells were cultivated with the respective antibiotic (ampicillin 100 µg/ml, carbenicillin 50 µg/ml, kanamycin 40 µg/ml). For long-term storage overnight cultures of bacteria were mixed with 25 % glycerol and stored at -80 °C.

LB-broth:

1 % NaCl, 1 % Tryptone enzymatic digest from Casein, 0.5 % yeast extract

+ 1.5 % select agar (for plates only)

3.2.2 Plasmid-DNA amplification and purification from bacteria

400 ml of LB-broth were supplemented with the respective antibiotic and inoculated with *E.coli* carrying the plasmid of choice. After overnight incubation at 37 °C and 200 rpm on orbital shakers, cells were harvested by centrifugation at 4500 rpm for 15 minutes at 4 °C. The pellet was washed once with PBS[33] and plasmid DNA was isolated

[33] Phosphate-buffered saline

following the instructions of the JETstar Maxi-prep kit (Genomed). The DNA content of the sample was measured with the NanoDrop® at 260 nm.

3.2.3 Preparation of chemically competent cells

Preparation of competent cells was performed as described previously in Sambrook *et al.* Protocol 24: The Inoue Method for Preparation and Transformation of Competent *E.coli*: "Ultra-competent" cells (Sambrook and Russell, 2001). Previously prepared competent bacteria were used to inoculate 25 ml of a new starting culture. Cells were incubated on an orbital shaker for six to eight hours at 37 °C and 250 – 300 rpm. For overnight incubation a 250 ml culture was inoculated with 4 ml of the starting culture, OD^{34}_{600} = 0,05, and kept on the shaker at 18 °C and 200 rpm. OD_{600} measurement was performed continuously until an OD_{600} of 0.55 was reached. The culture vessel was kept in an ice-water bath for 10 minutes and cells were harvested by centrifugation at 2500 x g for 10 minutes at 4 °C. Medium was discarded and cells were resuspended in ice-cold Inoue transformation buffer. Cells were centrifuged as described before and resuspended in 20 ml of ice-cold Inoue transformation buffer. After addition of 1.5 ml of DMSO cells were kept at 4 °C for 10 minutes. After snap-freezing the aliquots in liquid nitrogen bacteria were stored at -80 °C until use.

PIPES:

0.5 M PIPES pH 6.7 (with 5M KOH)

Inoue transformation buffer (sterile):

55 mM $MnCl_2 \bullet 4\ H_20$, 15 mM $CaCl_2 \bullet 2H_20$, 250 mM KCl, 10 mM PIPES (0.5 M, pH 6.7)

3.2.4 Transformation of chemically competent cells

- *Heat-shock transformation*

E.coli DH5α were thawed on ice and 100 µl of bacteria were mixed with DNA. The mixture was incubated 30 minutes on ice, transferred to 42 °C for 90 seconds and again put on ice for five more minutes. Samples were diluted in 3 ml of LB-broth and put on the orbital shaker for 45 minutes at 37 °C. Overnight incubation was done on selective agar plates at 37 °C.

- *Electroporation*

[34] Optical density at 600 nm

Electroporation was performed with the gene pulser II electroporator (Biorad) and Electromaxx DH10B bacteria (Invitrogen). All buffers and cuvettes were cooled throughout the experiment. 100 µl of bacteria were diluted 1:6 with ice-cold water and 100 µl of this suspension were mixed with 50 µl DNA sample (150 ng and 350 ng). Electroporation was performed in 1 mm cuvettes at 25 µF and 2.5 kV. After electroporation cells were rescued in 3 ml of LB-broth on a shaker at 37 °C for 45 minutes. Afterwards cells were spun down, plated on selective agar plates and incubated overnight at 37 °C.

3.2.5 Primerdesign- and synthesis

Primers were designed using the MacVector software. All oligonucleotides and primers, used in this work, were synthesised at Metabion (Martinsried, Germany). All relevant primer sequences for this work are found in chapter 2.1.6.

3.2.6 Polymerase-chain reaction (PCR)

Polymerase chain reaction is used for amplification of specific DNA fragments (Mullis *et al.*, 1986). For standard reactions self-made *Taq* polymerase or *Pwo* polymerase were used and the annealing temperature of the reaction was set accordingly to the respective primer melting temperature. PCR products were purified with the Machery-Nagel PCR purification kit after the amplification was performed and samples were checked on 1 % agarose gels.

3.2.7 Restriction digest

With the help of specific endonucleases, DNA palindrome sequences were cut and samples can be investigated regarding their genomic integrity. Gel analysis was done to verify the reaction efficiency and correct band formation. Restriction digests were performed at 37 °C for one to two hours and the enzyme was heat inactivated at 65 °C for 20 minutes if the sample was not purified in an agarose gel.

Restriction digest (V = 20 µl)

500 ng DNA
2 µl 10x buffer (NEB Buffer 1 - 4 according to enzyme)
0,2 µl BSA

x µl enzyme (depending on U/µl, restriction site and enzyme)
add H_20 to a final volume of 20 µl

3.2.8 Dephosphorylation

To avoid re-ligation of a DNA fragment after single-cut digestion a dephosphorylation of the vector was performed. Phosphate molecules are removed from the overhangs and prevent re-ligation. Prior to the phosphorylation reaction the sample was desalted by centrifugation through a G50-Sepharose column. Dephosphorylation of the vector backbone was performed at 37 °C for one hour.

Dephosphorylation reaction with Antarctic Phosphatase (V = 50 µl)
30 µl digested vector backbone (desalted)
5 µl 10x buffer
3 µl Antarctic Phosphatase
add H_20 to a final volume of 50 µl

3.2.9 Fill-in reaction

To create blunt-ends for ligation a fill-in reaction with *Klenow* enzyme was performed. After the *Klenow*-reaction the DNA was purified from an agarose gel and applied in a ligation reaction. The reaction was performed for 30 minutes at 37 °C and stopped by heat-inactivation of the enzyme at 75 °C for 20 minutes.

Fill-in reaction with Klenow enzyme (V = 100 µl)
30 µl purified digested DNA
10 µl NEB buffer 2
5 µl Klenow enzyme (5 U/µl)
5 µl dNTPs (10 mM)
add H_20 to a final volume of 100 µl

3.2.10Ligation

Ligases build up phosphodiester bonds between 3´-OH ends and a neighbouring 5´-phosphate group in an ATP-dependent manner. Vector and insert are cut with enzymes that create compatible overhangs. Because of high error rates ligations have to be tightly controlled. To check for re-ligation of the vector, which happens if the vector is cut with a single enzyme (see dephosphorylation, chapter 3.2.8), the reaction was performed

without insert. The second control is checking the restriction efficiency of the digest. To do so insert and ligase were left out in the reaction. Each ligation was transformed into competent bacteria. The reaction was performed for two to three hours at room temperature or overnight at 16 °C.

Ligation reaction with T4 ligase (V=15 µl)

vector backbone:insert 1:1, 1:4 and 1:10 (ratio in pmol)

1,5 µl 10x Rx buffer

0,5 µl T4 ligase

1 µl ATP 100 mM

add H_20 to a final volume of 15 µl

3.2.11Isolation and purification of DNA/PCR products from agarose gels

DNA was purified from gels according to manufacturer's instructions with the NucleoSpin Gel and PCR[35] Clean-up kit. DNA was eluted twice with 15 µl of elution buffer.

3.2.12Southern Blot

Southern Blot was performed as described previously (Church and Gilbert, 1984; Southern, 1975). 10-15 µg of genomic DNA were digested with restriction enzymes and separated on a 0.8 % agarose gel. DNA integrity was checked with ethidiumbromide after the run. Depurination of the sample was achieved by washing the gel for 20 minutes in 0,25 M HCl and was followed by washing the gel in denaturation buffer twice for 10 minutes. The samples were transferred from the gel to a nylon membrane *via* capillary forces for three to four hours and afterwards washed in 2x SSC for 10 minutes. The membrane was pre-hybridised with Church buffer for 1 h at 65 °C. Subsequently the radioactive labelled probe was added and incubated overnight at 60 °C. The following day the membrane was washed three times for 10 minutes with washing buffer. The membrane was wrapped up in plastic wrap, a film was put on the membrane and the cassette was stored at -80 °C. Duration of exposition depended on the intensity of signals. Alternatively, a phosphoimager screen can be used instead of a film and the cassette must then be stored at room temperature.

[35] Polymerase chain reaction

1x TAE buffer:
40 mM Tris-HCl (pH 8.0), 5 mM NaAc, 1 mM EDTA
Denaturation buffer:
1,5 M NaCl, 0,5 M NaOH
2x SSC:
300 mM NaCl, 30 mM sodium citrate
Church buffer (pH 7.2):
Solution 1: 7 % SDS, 400 mM Na_2HPO_4, 100 mM Ortho-Phosphore acid, 1 mM EDTA pH 7.2
heat up to dissolve ingredients
Solution 2: 1 % BSA in H_20
mix solution 1 and 2 and fill up to 1 l with H_20
Washing buffer:
0,2 x SSC, 1 % SDS

3.2.12.1 Radioactive labelling of DNA probes for Southern blotting

Labelling of 25-50 µg of DNA probe was performed according to manufacturer's instructions with the "Random primed DNA labelling kit" (Roche). Removal of not incorporated labelled nucleotides was achieved by centrifugation of the sample through Sephadex-G50 columns. Prior to hybridization the probe was denatured (5 minutes at 95 °C, 5 minutes on ice).

3.3 Protein biological methods

3.3.1 Preparation of whole cell protein lysates

Cells were harvested with trypsin, washed twice with ice-cold PBS and centrifuged at 1000 rpm for 8 minutes at 4 °C. The cell pellet was resuspended in 60-100 µl RIPA buffer, according to the pellet size. Lysed cells were incubated on ice for 30 minutes and centrifuged at full speed for 30 minutes at 4 °C. The protein containing supernatant was kept on -20 °C. Protein concentration was determined using the Bradford method (3.1.3.2).

RIPA buffer:
1 mM EDTA, 150 mM NaCl, 0,1 % SDS, 0.5 % DOC, 1 % NP-40, 50 mM Tris pH 8.0

3.3.2 Determination of protein concentration

The protein concentration of RIPA extracts was determined according to the protocol of Bradford (Bradford, 1976). Protein assay solution was diluted 1:5 with PBS and 1 ml was mixed with 1 µl of the RIPA-extract in an appropriate cuvette. The mixture was incubated for 15 minutes in the dark and measured with the Eppendorf Biophotometer.

3.3.3 Western Blot

Proteins for Western blotting were run according to the standard procedure on SDS-polyacrylamid gels (Laemmli, 1970). Samples were blotted on a nitrocellulose membrane following the protocol for semi-dry blotting (Towbin *et al.*, 1979) vertical to the running direction of the gel. The proteins were stuck to the matrix due to hydrophobic interactions and intercalated SDS was washed out with methanol-containing blotting buffer. Furthermore methanol enhances the binding of proteins to the membrane. To avoid the formation of air bubbles between the gel and the membrane, a glass pipette was rolled over the sandwich with little pressure. Blotting was performed with a semi-dry chamber for 1h at 15 V, 400 mA and 150 W. Proteins of interest were detected with specific primary antibodies, which in turn were detected with peroxidase-coupled secondary antibodies with ECL[36]-solution.

Running buffer:
192 mM Glycin, 24 mM Tris, 3.4 mM SDS, pH 7.4
2x Concentration gel buffer:
0.25 M Tris, 7 mM SDS, > pH 6.8 (HCl)
5x Separation gel buffer:
1.86 M Tris, 17 mM SDS, > pH 8.8 (HCl)
5x Laemmli buffer:
100 mM Tris pH 6.8, 25 % SDS, 50 % β-mercaptoethanol, 0.5 % Bromphenolblue
ECL solution:
solution 1: 0,1 M Tris pH 8.8, 200 mM p-choumaric acid, 1.25 mM luminol
solution 2: 3 % H_2O_2
For usage 1 ml of solution 1 and 3 µl of solution 2 were mixed and applied on the membrane
PBS-Tween20 (PBS-T):
1x PBS, 0,1 % Tween-20
5 % and 2.5 % milk-blocking solution:

[36] Enhanced chemiluminescence

5 g or 2.5 g milk powder in 100 ml 1xPBS-T

1x Running buffer;

192 mM Glycin, 24 mM Tris, 3.4 mM SDS, pH 7.4

1x Blotting buffer:

1x Running buffer + 20 % methanol

To check the blotting efficiency, the membrane was stained with Ponceau S solution. Ponceau S is a red coloured azoic dye that stains all proteins on nitrocellulose or PVDF membranes by binding to the positive charged amino groups of proteins. The staining is reversible and the dye can easily be removed from the membrane by washing with water.

The membrane was blocked with 5 % milk-blocking solution for 1h at room temperature to saturate all unspecific binding sites. The membrane was washed three times with PBS-T for 5 minutes and afterwards the primary antibody, diluted in 2.5 % milk-blocking solution, was added overnight at 4 °C or 1h at room temperature. The membrane was washed three times for 5 minutes with PBS-T. The peroxidase-coupled secondary antibody (1:10.000 in 2.5 % milk-blocking solution) was added and incubated for 45 minutes at room temperature and antibodies were visualised by incubation with ECL- solution.

3.4 Plasmid rescue assay

3.4.1 Human cells

1.5×10^5 cells were transfected with 1 µg of the respective pDNA reporter vector (DS/20 *TetOperator*, DS/40 *TetOperator*, *wt-oriP*) following the instructions for Lipofectamine 2000 (chapter 3.1.1.4). Hygromycin selection (150 µg/ml) was applied for three weeks. Low molecular weight extrachromosomal DNA was isolated at different time points by HIRT extraction as described before (Hirt, 1966). For harvesting, cells were washed with TEN buffer and afterwards resuspended in 1.5 ml TEN buffer and 1.5 ml 2x HIRT buffer. Proteins and chromatin were precipitated with 750 µl of 5 M NaCl (end-conc 1.25 M) and samples were stored at 4 °C overnight. After centrifugation for 1h at 15.000 rpm and 4 °C, DNA was purified from the supernatant by phenol-chloroform extraction and 1 µg of low-molecular-weight DNA was digested with *DpnI* and RNase. Digest of the extracted DNA with *DpnI* allows

eliminating the initially transfected plasmids, which retain the *dam* methylation pattern acquired during propagation in *E. coli*. 350 ng of digested DNA were electroporated into Electromaxx DH10B competent cells (Invitrogen) and after overnight incubation at 37 °C ampicillin-resistant colonies were counted.

TEN buffer:
10 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl
2x HIRT buffer:
1,2 % SDS, 20 mM Tris-HCl pH 7.5, 20 mM EDTA

3.4.2 Schneider S2 cells

1.4 µg *LacOperator* reporter vector were transfected into S2 cells expressing CENH3^{CID}:GFP:*LacI* and GFP:*LacI* protein. Cells were kept under selection with 1 mg/ml G418 throughout the experiment. All further steps were performed as described for the human system in 3.1.4.1. Cells were lysed in 0.5 ml TEN buffer and 0.5 ml HIRT buffer (2x) and 150 ng of DNA were electroporated into Electromaxx DH10B bacteria (Invitrogen).

3.5 BrdU-incorporation assay

CENH3^{CID}:GFP:*LacI* or GFP:*LacI* cells transfected with the *LacOperator* reporter vector. Selection with G418 (1 mg/ml) was applied for two weeks before cells were pulsed with 10 mM 5-bromodeoxyuridine (BrdU) for 48 hours. With the help of a CsCl gradient centrifugation cells with incorporated BrdU and cells without BrdU were separated due to their differences in buoyant density of replicated DNA (the gradient is adjusted to 1.403 refractive index). From 1.5 x 10^7 cells DNA was isolated and applied on the gradient after *Bam*HI digest. Gradients were spun in a Beckman SW41 rotor at 48.000 rpm for 48 h. The refractive index of each fraction (250 µl) was determined and the samples were analysed by either quantitative PCR using PCR primer pairs recognizing the backbone of the plasmid (Table 2.7) or measuring the A_{260} index with the NanoDrop for the genomic DNA content in the sample.

4 RESULTS I – D.MELANOGASTER

pDNA vectors (*i.e.* plasmid-derived, non-viral and extrachromosomally maintained vectors) of different origin employ various mechanisms for their retention in target cells. They either integrate into the host genome and remain there for the lifetime of the cell, or are maintained in an extrachromosomal state. Examples for non-integrating vectors are *e.g.* the *pEPI* vector and EBV-derived gene vectors (Ehrhardt *et al.*, 2008). The extrachromosomal EBV-derived wt-*oriP* vector is dependent on two components: *oriP* in *cis* and its transactivator EBNA1 in *trans*. In the absence of EBNA1 wt-*oriP* vectors as well as EBV genomes are lost rapidly from infected cells (Humme *et al.*, 2003).

I explored a novel, artificial and CENH3-dependent mechanism of pDNA vector retention and investigated its efficiency. The idea was, to turn the pDNA vector into a "quasi-chromosome" by establishing a pivotal cellular feature for chromosome segregation also on the pDNA vector – a (neo)centromere. It was the first aim of this project to evaluate, whether or not a neocentromere on the pDNA vector is sufficient to prolong pDNA vector maintenance in the cell and in this way guarantee transgene expression and long-term stability of the pDNA vector.

Previous experiments showed, that through artificial targeting of $CENH3^{CID}$, functional kinetochores were established at ectopic sites in *D.melanogaster* (Heun *et al.*, 2006).

These results were further validated with targeting experiments of a CENH3^{CID}:GFP:*LacI* fusion protein to *LacOperator* arrays integrated into chromosome 3R of *D.melanogaster* (Olszak *et al.*, 2011). Due to the *LacI-LacOperator* interaction, formation of a neocentromere was specifically induced at the *LacOperator* sites on the chromosome.

In respect to previous studies it seemed reasonable to test our hypothesis of inducible neocentromere formation first on pDNA vectors in a *D.melanogaster* model (Heun *et al.*, 2006). Follow-up experiments were performed in human cells and results from these experiments are found in the Results II section (chapter 5).

4.1 Rescue of autonomous pDNA vectors from cells

Autonomous pDNA vector maintenance in a proliferating cell is dependent on two mechanisms: DNA replication and pDNA vector retention (Pich *et al.*, 2008). The overall goal of this work was to develop a mechanism of pDNA vector retention, that is independent of host cell chromosome attachment and *trans*-acting factors like viral proteins (*e.g.* EBNA1 in the case of *oriP*-derived vectors). With the knowledge that was gained in the past (Heun *et al.*, 2006; Pich *et al.*, 2008) I wanted to test, if pDNA vector autonomy in the retention mechanism is obtained *via* the induction of a neocentromere on the pDNA vector itself. To investigate whether the centromere specific protein CENH3 has the ability to induce neocentromeres on pDNA vectors and might enhance pDNA vector retention, I used plasmid rescue assays. This assay is one possible way to detect exclusively extrachromosomal and self-propagating vectors in a cell. Figure 4.1 illustrates the experimental set-up.

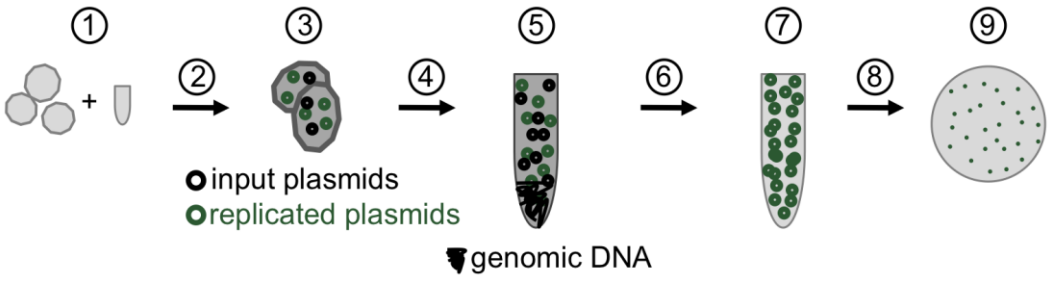


Figure 4.1 Plasmid rescue assay – experimental procedure.

Cells are transfected with a pDNA reporter vector (1) and kept under selection (2) until certain harvesting times. During selection two sets of pDNA vectors are present in the cells (3): first, plasmids derived from bacterial propagation, which are *dam*-methylation positive and digestible

by *DpnI* (input plasmids) and second, plasmids, which have been actively replicated in the cell after transfection (replicated plasmids). These plasmids do not carry the *dam*-methylation profile and are therefore not digested by *DpnI*. Cells are harvested following the HIRT protocol (4) (Hirt, 1966). During this step, plasmid DNA is enriched in the supernatant of the sample and bulk chromatin is precipitated with proteins (5). After phenol/chloroform extraction (6) the sample is digested with RNase and *DpnI* and the resulting solution contains only plasmids, which stem from propagation in human cells (7). Purified DNA is electroporated into competent bacteria and plated on Amp^+ agar plates (8). Following overnight incubation at 37 °C the number of outgrown bacterial colonies are counted (9).

Drosophila Schneider S2 cells, stably expressing either $CENH3^{CID}$:GFP:*LacI* or GFP:*LacI*, were transfected with the pDNA reporter vector encoding for two antibiotic resistance genes and 256 *LacOperator* sites (Figure 4.2 B). The strong binding of the *LacI*-fusion protein to the *LacOperator* sites on the pDNA vector enables the site-specific targeting of $CENH3^{CID}$. Cells were selected with G418 (1 mg/ml) and harvested at indicated times during a period of one month (Figure 4.2 A).

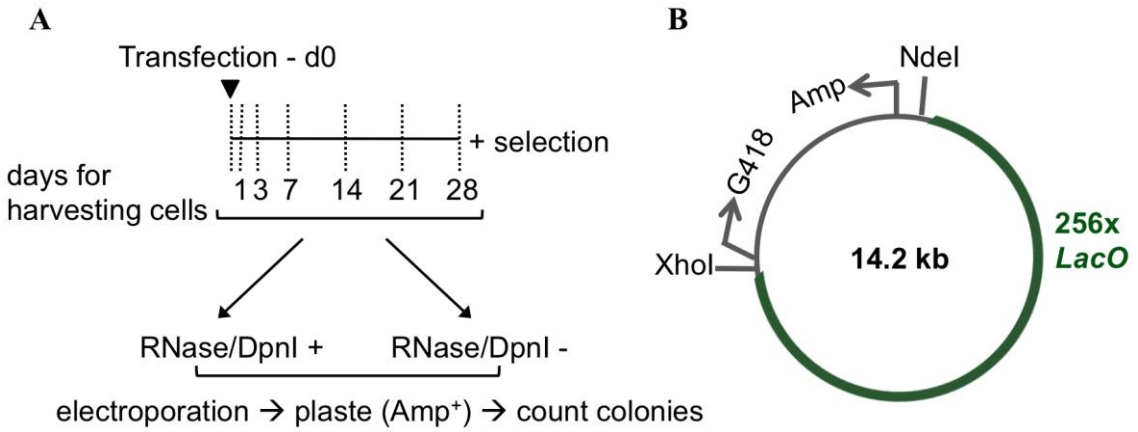


Figure 4.2 Plasmid rescue assay in D.melanogaster – experimental set-up.

A) Schneider S2 cells (stably expressing $CENH3^{CID}$:GFP:*LacI* or GFP:*LacI*) were transfected with a pDNA reporter vector and selected under high selective pressure (1 mg/ml G418). Bulk chromatin was precipitated with a high salt concentration and the plasmid-enriched supernatant was phenol/chloroform extracted. DNA was digested with *DpnI* (for elimination of *dam*-methylated input plasmids derived from generation in bacteria) and RNase (for enhanced purity of DNA samples), 150 ng of DNA were electroporated into DH10B competent bacteria and plated on Amp^+ agar plates. The read-out of the assay is the number of bacteria colonies growing after overnight incubation at 37 °C. **B)** The 14.2 kb pDNA reporter vector used in the assay encodes for an ampicillin-resistance cassette used for selection in bacteria and a G418 resistance cassette for selection in Schneider S2 cells. 256 *LacOperator* sites, which are bound by the $CENH3^{CID}$:GFP:*LacI* and GFP:*LacI* fusion proteins expressed in the transfected cells, occupy the main part of the plasmid.

During selection two sets of pDNA vectors were present in the cells: first, input vectors derived from bacterial propagation and used for transfection in the beginning of the

experiment. They carry the *Dpn*I-sensitive *dam*-methylation profile. Second, vectors are present, which have replicated in the cell only after transfection. These vectors have lost the bacterial methylation pattern and are distinguishable from input vector with a *Dpn*I digest. Cells were lysed with 0.6 % SDS and proteins plus all chromatinised DNA were precipitated with 1.25 M NaCl. The sample was incubated overnight at 4 °C before most of the genomic DNA was removed by centrifugation. After phenol/chloroform extraction of the supernatant, the sample was highly enriched in plasmid DNA, even if the main part of DNA in the sample was still genomic DNA. Ethanol precipitation of DNA was followed by a *Dpn*I digest in combination with RNase treatment. After re-precipitation, 150 ng of purified DNA were electroporated in Electromaxx DH10B competent bacteria (Invitrogen) and plated on ampicillin agar plates. Colonies were counted after overnight incubation at 37 °C and represented the read-out of the experiment (Figure 4.2 A).

4.2 Targeting of CENH3 to the pDNA vector confers autonomous and prolonged pDNA vector retention in cells

The stability of extrachromosomal pDNA vector retention in a cell was analysed with plasmid rescue assays and quantified by the number of colonies growing on Amp^+ agar plates after electroporation. The number of outgrowing colonies from plasmid rescue assays in cells expressing the $CENH3^{CID}$:GFP:*LacI* fusion protein (Figure 4.3 A, green bars) and cells expressing GFP:*LacI* (Figure 4.3 A, grey bars) differed strongly after two weeks under selection. Results revealed prolonged pDNA vector retention in $CENH3^{CID}$:GFP:*LacI* expressing cells compared to cells lacking $CENH3^{CID}$:GFP:*LacI* (Figure 4.3 A, day 21 and 28, green bars). The only difference between the two samples was the presence and absence of targeted $CENH3^{CID}$. Therefore, it is likely that the difference in pDNA vector numbers was caused by the $CENH3^{CID}$:GFP:*LacI* -induced neocentromere formation on the vectors, which enabled them to segregate in a host cell independent mechanism. pDNA vectors in cells lacking $CENH3^{CID}$:GFP:*LacI* did not establish neocentromeres and were lost from cells more rapidly.

The high number of pDNA vectors in $CENH3^{CID}$:GFP:*LacI* negative cells in the initial phase of the experiment (until day 14) was due to a delay in selection response of Schneider S2 cells to G418. Selection with G418 in these cells was becoming effective

only after two weeks (Figure 4.4). Until this point was reached, pDNA vectors were replicated. After passing that time frame pDNA vectors were lost rapidly from cells if no CENH3^{CID}:GFP:*LacI* protein was expressed due to the fact that the pDNA vectors carrying the resistance cassette were lost (Figure 4.3 A, day 21, grey bars).

Control experiments confirmed, that plasmid rescue assay experiments were started with an initial transfection efficiency being comparable between both cell lines. I checked this by performing plasmid rescue assays without the *Dpn*I digest (*Dpn*I-experiments in Figure 4.3 B – C). The background of the experiment was determined by performing plasmid rescue assays on cells missing the initial transfection of the pDNA reporter vector. The background here is considered as negligible (experiments with non-transfected cells in Figure 4.3 B – C).

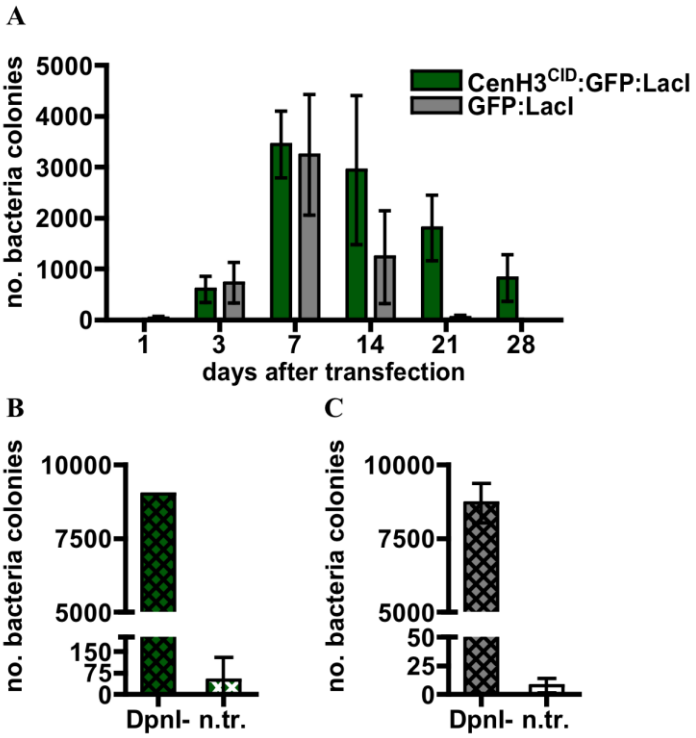


Figure 4.3 Targeting of CENH3^{CID}:GFP:LacI to pDNA vectors improved their retention in Schneider S2 cells.

A) Cells expressing CENH3^{CID}:GFP:*LacI* (green bars) and cells that express GFP:*LacI* (grey bars) were transfected with a pDNA reporter vector (Figure 4.2 B) and analysed in plasmid rescue assays (n = 3, +/- SEM). Experiments without *Dpn*I digest (left columns in **B)** and **C)** show the initial transfection efficiency of the two cell lines. Experiments of non-transfected (n.tr.) cells show the background colony number of the experiment. Control experiments without *Dpn*I digest and without transfection of the pDNA reporter vectors (n.tr.) are shown in **B)** for CENH3^{CID}:GFP:*LacI* cells (n = 3, +/- SD) and **C)** for GFP:*LacI* cells (n = 3, +/- SD).

4.2.1 Selection features of Schneider S2 cells with G418

The results that I obtained in plasmid rescue assays with CENH3^{CID}:GFP:*LacI* and GFP:*LacI* cells were quite similar between the two cell lines during the first week of the experiment. I expected a faster loss of pDNA vectors in CENH3^{CID}:GFP:*LacI* negative cells, but also after two weeks of selection still many pDNA vectors were rescued from these cells. A possible explanation for this observation is the growth behaviour of Schneider S2 cells under selection with G418. Therefore I wanted to find out, how Schneider S2 cell proliferation is influenced by selection pressure with G418 (1 mg/ml). Additionally proliferation assays of transfected and non-transfected cells with and without selection enable the determination of cell generation numbers and thus the calculation of plasmid loss rates.

I seeded Schneider S2 cells (not transfected and transfected) with a density of 1 x 10^6 cells/ml in a total volume of 2.4 ml in 6-well plates and counted the cells repeatedly about every second day during one month. Cell numbers were determined by counting cell aliquots in a Neubauer cell counting chamber (chapter 3.1.2).

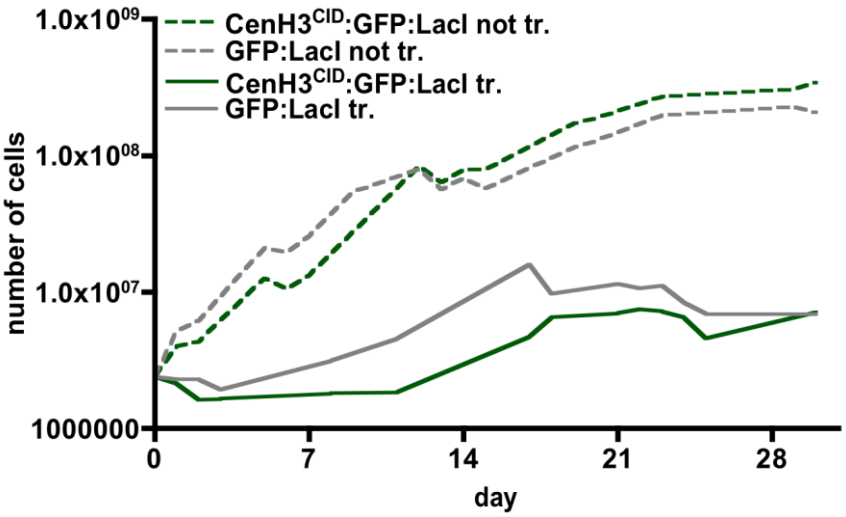


Figure 4.4 Proliferation curves of Schneider S2 cells with and without selection pressure.

CENH3^{CID}:GFP:*LacI* and GFP:*LacI* cells were seeded with a density of 1 x 10^6 cell/ml in a total volume of 2.4 ml. One set of cells was transfected with a pDNA reporter vector, whereas the other half of cells remained non-transfected. Non-transfected cells were cultivated in antibiotic free Schneider S2 cell medium and transfected cells were kept under selection with 1 mg/ml G418. Cells were counted in a Neubauer cell counting chamber and proliferation curves were determined (n = 1) tr.: transfected, n.tr.: not transfected.

The proliferation curves demonstrated, that proliferation in cells that were transfected and kept under selective pressure with G418 was slowed down strongly in comparison to non-transfected cells, which grew in medium lacking G418 (Figure 4.4). Cells under selection showed an increase in total cell numbers during the first two weeks of selection. The first evidence that selection hit in was observed at days 17/18. Afterwards cells under selective pressure were not able to recover and died whereas not transfected cells in non-selective medium proliferated constantly.

Proliferation rates of individual cell lines are required to determine pDNA vector loss rates per generation. As the proliferation did not follow a linear curve, I was not able to determine the pDNA vector loss rates for the respective cell lines.

4.3 Genetic integrity and conformational state of transfected pDNA vectors

In regards to therapeutic pDNA vector design, genetic stability (*i.e.* no alterations in the DNA sequence) of the pDNA vector and its conformational state in the cell are very important aspects. From a clinical perspective, reliable transgene expression and genetic stability of the pDNA vector must be guaranteed throughout the complete medical procedure. While resting in a cell, the pDNA vector might recombine, lose important (regulatory) elements or take up foreign sequences. This can lead to problematic consequences especially in regards to transgene expression. If the pDNA vector is genetically unstable, the expression of the therapeutic protein from the pDNA vector might become irregular with potential impact on the therapeutic activity. It is also possible that the transgene is lost from the cell completely, if the genetic integrity of the pDNA vector is not ensured.

To check genetic integrity of the pDNA vector I performed restriction digest analysis of colonies taken from rescued pDNA vectors (chapter 4.3.1). To elucidate the conformation in which the pDNA reporter vector was maintained in the cell, I performed Southern Blot analysis on DNA of transfected cells (chapter 4.3.2).

4.3.1 Genetic integrity of the pDNA vector is preserved over time

Single bacteria colonies from plates of plasmid rescue assays were taken and DNA was isolated. To confirm the genetic integrity of the rescued pDNA vector an analytical

restriction enzyme analysis was performed with the restriction enzymes *Xho*I and *Nde*I. Both of these enzymes are single cutters on the input pDNA vector (Figure 4.5). The predicted fragment sizes after the digest are shown in Table 4.1.

Table 4.1 Fragment prediction of restriction digest of the LacOperator pDNA reporter vector with XhoI/NdeI.

Predicted fragment size (kbp)	***Annotation***
4.1	Amp- and G418 resistance cassettes
10.1	*LacOperator* repeats

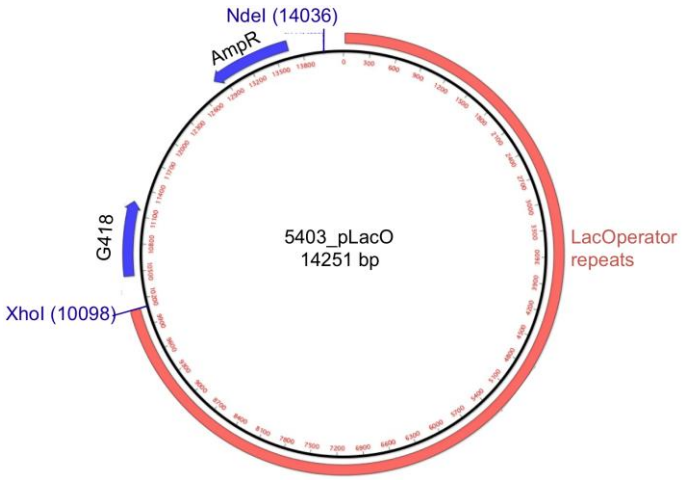


Figure 4.5 Vector map of the LacOperator pDNA reporter vector displaying the cut sites of XhoI and NdeI.

Both restrictions enzymes cut the pDNA vector at one site in close proximity to the *LacOperator* repeats. The predicted fragment sizes and annotations of the vector after the digest are given in Table 4.1 and Figure 4.6 illustrates the analytical restriction digest.

For the analysis I picked three clones from agar plates at day 7 and day 21 and performed restriction digests. Additionally I picked one clone from plates deriving from experiments with CENH3^{CID}:GFP:*LacI* negative cells at days 7 and 21 (Figure 4.6).
On the agarose gel the input pDNA vector was loaded in undigested and digested form as reference. The analysis demonstrated, that the genetic integrity of the *LacOperator* pDNA vector was maintained throughout the experiment. One has to mention that the *LacOperator* fragment was detected at approx. 8 kb instead of 10.1 kb also in the input pDNA vector, indicating that some of the 256 *LacOperator* sites have been lost, most likely due to the high recombination activity of highly repetitive sequences during the propagation in bacteria. Nevertheless, the input pDNA vector restriction pattern was confirmed in all rescued pDNA vectors except for one (Figure 4.6, clone

CENH3^{CID}:GFP:*LacI*, 21.1). This clone displayed a smaller band for the *LacOperator* fragment at approx. 5 kbp, indicating that even more of the *LacOperator* repeat sequences have been deleted due to recombination. More important is the fact, that the pDNA vector backbone, where a potential transgene is located, remained stable in all investigated pDNA vectors (detected at 4.1 kbp).

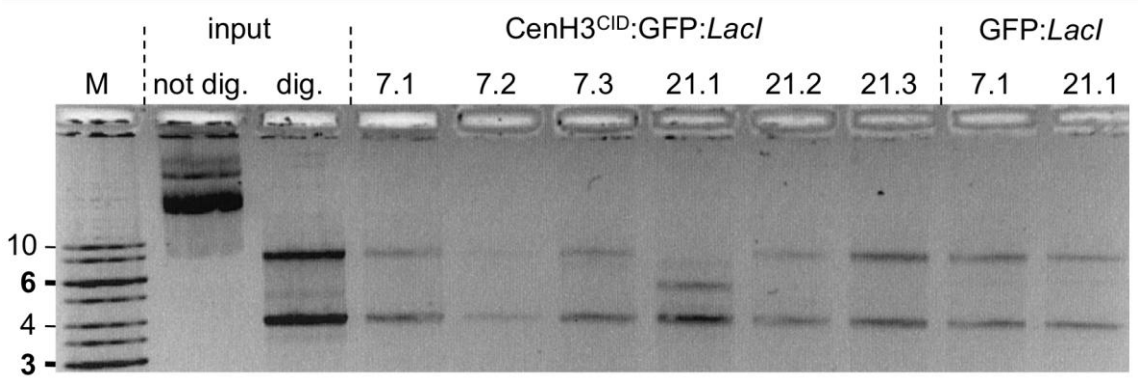


Figure 4.6 Restriction digest of rescued bacteria colonies confirmed the genetic integrity of the pDNA vector.

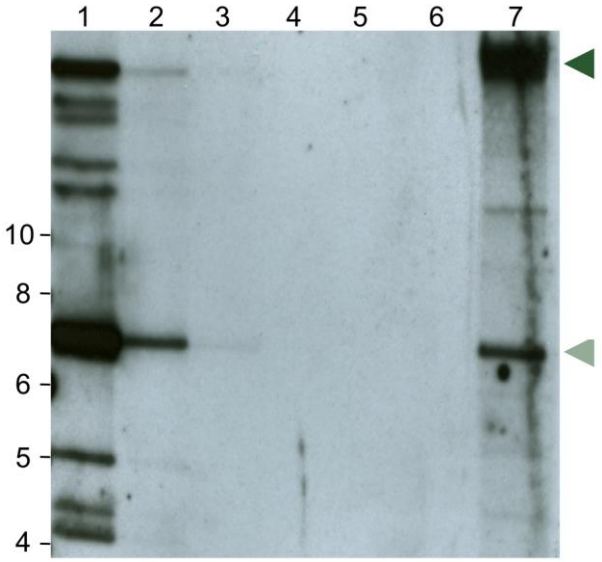
Clones picked from plates of plasmid rescue assays, and the transfected input pDNA vectors were analysed in a restriction digest with *Xho*I and *Nde*I and separated on an agarose gel. Three clones were analysed deriving from experiments with CENH3^{CID}:GFP:*LacI* positive cells after one and three weeks and one clone was analysed for clones deriving from experiments with GFP:*LacI* cells. The digest led to a fragmentation of the pDNA vector into the backbone (4.1 kb) and the fragment with the *LacOperator* sites (10.1 kb).

4.3.2 pDNA vectors are maintained in cells in a concatemer conformation

Plasmid DNA can be maintained in different states of DNA conformations in a cell – *e.g.* as concatemers, which are single circular DNA strands attached to each other, or in linearised form. To find out, which conformation was the preferred state of pDNA vectors in my system, I performed Southern Blot analysis with the radioactive probe directed against the backbone of the pDNA vector. To eliminate background signals at the predicted height of the pDNA vector, genomic DNA was digested with *Fse*I, a non-cutter enzyme on the pDNA vector sequence (Figure 4.7, lane 7). Transfected cells were lysed 16 days post transfection with 0.1 % SDS and DNA was isolated, digested with RNase and ProteinaseK and purified with phenol/chloroform extraction. Samples were loaded on a 0.7 % agarose gel and separated over night (70 V, 150 mA).

For quantification purposes I loaded different concentrations of input pDNA vector in the first five lanes (Figure 4.7, lanes 1 - 5). Amounts lower than 100 pg were not detectable in the analysis.

In my experiments the preferred status of pDNA vectors was the concatemeric form. In Southern Blot analysis the concatemer structure is visible as an additional higher migrating signal at the upper end of the gel (Figure 4.7, dark green arrow). The bright green arrow in Figure 4.7 indicates the band for monomeric, supercoiled plasmid DNA.



Lane	*Sample*
1	10 ng input plasmid
2	1 ng input plasmid
3	100 pg input plasmid
4	10 pg input plasmid
5	1 pg input plasmid
6	empty lane
7	CENH3^{CID}:GFP:*LacI*

Figure 4.7 Southern Blot analysis illustrates the formation of concatemers 16 days post transfection.

Undigested input pDNA vector was loaded in five different concentrations for quantification purposes (lanes 1 - 5). DNA amounts lower than 100 pg were not detected on the blot. Transfected CENH3^{CID}:GFP:*LacI* cells were lysed with 0.1 % SDS after 16 days under selection. Prior to loading, the sample was digested with *Fse*I, a non-cutter enzyme on the plasmid, to get rid off high background signal of genomic DNA (lane 7). Dark green arrow: concatemeric DNA molecule, bright green arrow: monomeric, supercoiled plasmid DNA.

Replication of (EBV-derived) pDNA vectors follows the theta replication principle, a bidirectional replication model (Gahn and Schildkraut, 1989). I wanted to know if the CENH3^{CID}:GFP:*LacI* expression in cells influences DNA replication of the *LacOperator* pDNA vector and performed Meselson-Stahl experiments, which are discussed in the following chapter.

4.4 pDNA vector replication is independent of CENH3^{CID}

Replication and retention are the two main factors regulating pDNA vector maintenance. To confirm whether or not pDNA vector replication is integrated into the cell cycle and to find out if CENH3^{CID}:GFP:*LacI* overexpression plays a role in DNA replication of the pDNA vector I performed Meselson-Stahl density transfer experiments. The Meselson-Stahl experiment is based on the weight differences that occur, when BrdU is incorporated instead of thymidine into the newly synthezised DNA strand during replication. One round of semi-conservative replication in the presence of BrdU increases the buoyant density of DNA strand from "light-light" (1.70 g/cm^3) to "heavy-light" (hemisubstituted, 1.73 g/cm^3). After separation of the samples on a CsCl-gradient, DNAs are found in different density fractions of the gradient according to their respective weight, which correlates with the incorporation of BrdU. If the DNA is not actively replicated, BrdU incorporation does not occur.

In order to investigate CENH3^{CID}'s influence on pDNA vector replication CENH3^{CID}:GFP:*LacI* cells and GFP:*LacI* cells were transfected with a *LacOperator* pDNA reporter vector (Figure 4.2 B) and selected for two weeks with G418 (1 mg/ml). Cells were pulsed with 10 mM BrdU for one generation (48 hours) to allow one complete substitution of thymidine with BrdU. Total DNA was isolated, digested with *Bam*HI and separated on a CsCl gradient. The CsCl solution had a starting refractive index of 1.403 (corresponding 1.74 g/cm^3). After centrifugation for 48 hours at 48.000 rpm fractions of 250 µl were collected. The refraction indices were measured for every second fraction to determine the CsCl density. Refraction indices were measured for three gradients, indicating that the different gradients are comparable (Figure 4.8 A).

Chromatin DNA of the individual samples was measured photometrically (Figure 4.8 B), and pDNA vector DNA was quantified by quantitative PCR using primers for the pDNA vector backbone (Figure 4.8 C). DNA peaks were observed for both BrdU-pulsed cell lines (with and without CENH3^{CID}:GFP:*LacI*) in the identical fractions of the gradient (fractions 15-17, CsCl density: 1.4025 - 1.3990), whereas DNA from non-labelled cells was found in higher fractions (fractions 17-21, CsCl density: 1.4000 – 1.3965) (Figure 4.8 B-C). This experiment demonstrated, that the pDNA vector replicated once-per-cell-cycle and that CENH3^{CID}:GFP:*LacI* was not required for pDNA vector replication. This also explains the result of plasmid rescue assays, in which pDNA vector DNA accumulated also in CENH3^{CID}:GFP:*LacI* negative cells until

day 7 (Figure 4.3 A). pDNA vectors segregated CENH3^{CID}:GFP:*LacI*-dependently became dominant after two weeks of selection.

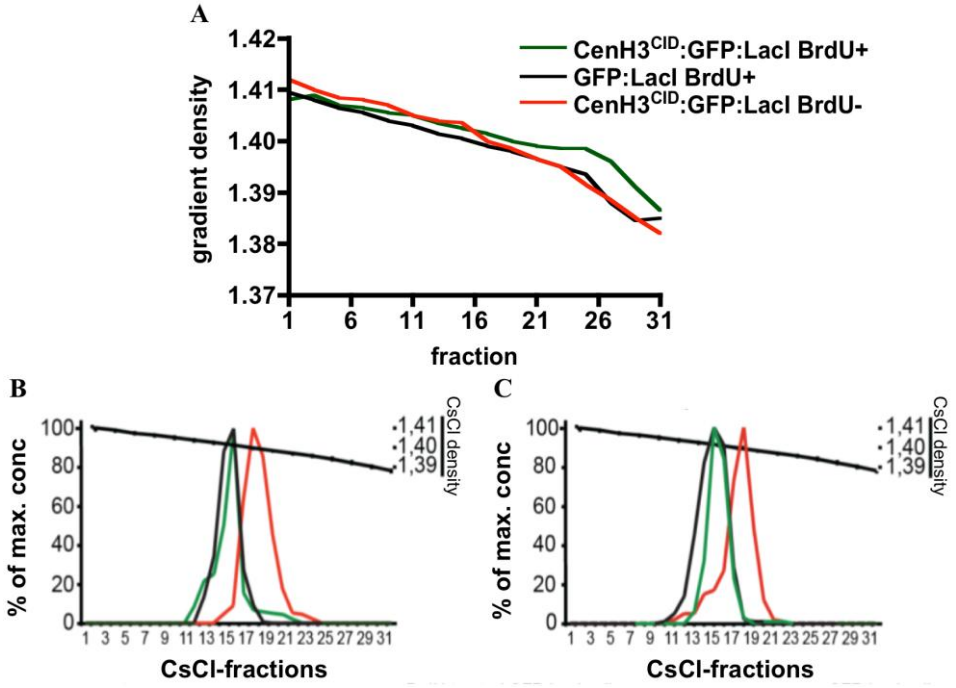


Figure 4.8 pDNA vector replication occurs independently of CENH3^{CID}:GFP:LacI overexpression in the cell.

Transfected cells were selected for two weeks and pulsed with BrdU for 48 hours. BrdU was incorporated in DNA during active replication and due to its higher molecular weight DNA with incorporated BrdU was heavier than DNA without BrdU. In CsCl gradients the density of the fractions increased from bottom to top and DNA with incorporated BrdU was found in lower fractions of the gradient than BrdU-free DNA. **A)** CsCl gradient refraction indices. **B)** Photometric bulk chromatin measurement at A_{260} and **C)** pDNA vector DNA measurement with RT-PCR using primers for the pDNA vector backbone. Green line: CENH3^{CID}:GFP:*LacI* cells BrdU+, black line: GFP:*LacI* cells BrdU+, red line: CENH3^{CID}:GFP:*LacI* cells BrdU-.

4.5 Extrachromosomal pDNA vectors establish a functional spindle apparatus

During mitosis chromosomes segregate following a strict time-schedule, guarded by several checkpoint control mechanisms. After duplication of the genome during S-phase, the chromosomes separate during mitosis. Mitosis is divided into five sub-phases: prophase, prometaphase, metaphase, anaphase and telophase. In prophase, mitotic spindles grow starting from the centrosomes, which finally attach to the centromeres of chromosomes. In prometaphase the chromosomes start to assemble at

the metaphase plate. This process is finished in metaphase and during anaphase the spindle tubules start pulling the chromosomes to the spindle poles.

To illustrate the behaviour of pDNA vectors in the cells and to follow them during mitosis I transfected cells with the pDNA reporter vector. In principle the induction of a neocentromere on the pDNA vector leads to a segregation mechanism as it is illustrated in Figure 4.9 A. pDNA vectors behave like "quasi-chromosomes" and their segregation is integrated into the endogenous mechanism of chromosome segregation. Immunofluorescence is a powerful method to follow up this development in the cell and members of Patrick Heun's group at the MPI Freiburg performed immunofluorescence microscopy of the transfected cells at different states during mitosis (Figure 4.9 B-C).

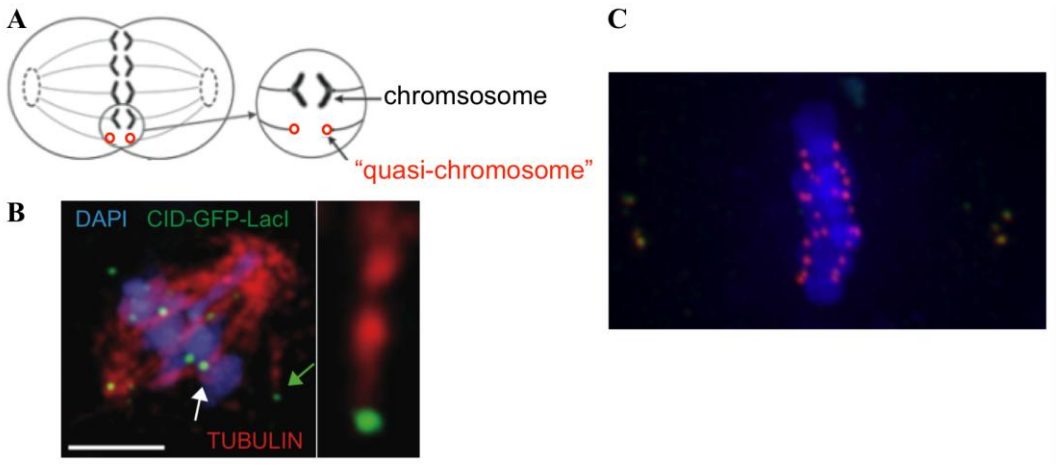


Figure 4.9 Immunofluorescence staining of transfected cells in different cell cycle stages.

A) Model of active pDNA vector segregation. pDNA vectors with functional centromeres and kinetochores behave like chromosomes. The functional kinetochore is recognised by microtubules and after attachment the pDNA vectors are pulled to the spindle poles and symmetrically segregated to the daughter cells. Immunofluorescence microscopy of transfected cells in prometaphase **(B)** and metaphase **(C).** Blue: DAPI, green: $CENH3^{CID}$:GFP:*LacI* pDNA vectors, red: tubulin.
(Lufino *et al.*, 2008; Mendiburo *et al.*, 2011)

I transfected Schneider S2 cells with the same pDNA reporter vector as previously used in plasmid rescue assays. After nine days under selection prometaphase cells were used to stain DNA (DAPI), the transfected pDNA vector ($CENH3^{CID}$:GFP:*LacI*; green) and microtubules (tubulin, red). Subsequently cells were analysed with a fluorescence microscope. In Figure 4.9 B the green arrow and the enlargement illustrate the attachment of microtubules to the neocentromeres on pDNA vectors. The white arrow indicates an endogenous microtubule attachment site. Similar to the chromosomes also

pDNA vectors with neocentromeres are recognised by the microtubules and subsequently pulled to the spindle poles depicted in Figure 4.9 C (metaphase spread). In contrary to the chromosomes, which were still assembled at the metaphase plate during this cell cycle stage, at least a part of the pDNA vectors was pulled to the spindle poles earlier. It is conceivable that pDNA vector segregation was not fully integrated into the mitotic checkpoints or that segregation was happening much faster for pDNA vectors possibly due to their smaller size.

4.6 Epigenetic inheritance of the centromeric mark

Another question that arose was, how stable the neocentromeric mark is maintained on the pDNA vector over several cell generations. To this end, members of the group of Patrick Heun at the MPI in Freiburg performed experiments, in which they microscopically checked cells under the microscope for $CENH3^{CID}$/CENP-C positive foci after the transient co-transfection of the *LacOperator* pDNA reporter vector and a $CENH3^{CID}$:GFP:*LacI* expression plasmid. To complete the information gained with my experiments I included this result for the matter of a better understanding of the process of CENH3-mediated neocentromere formation (Figure 4.10) (Mendiburo *et al.*, 2011).

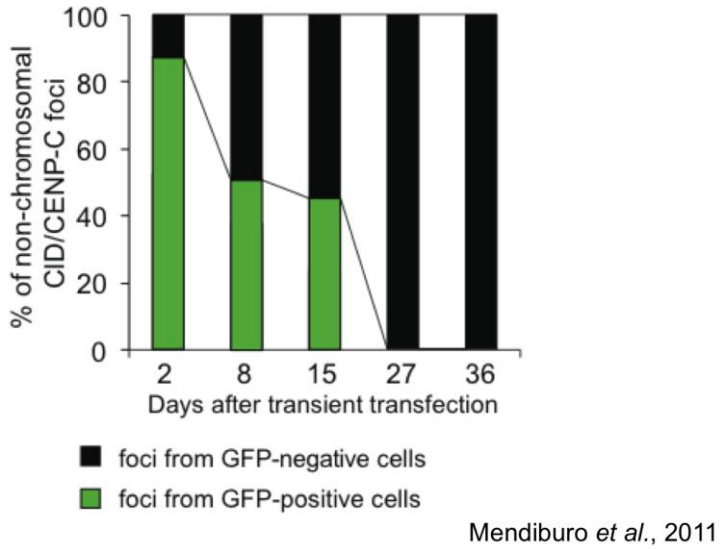


Figure 4.10 Initial targeting of CENH3:GFP:LacI for the nucleation of centromere function.

Cells were transiently co-transfected with the *LacOperator* pDNA reporter vector and a $CENH3^{CID}$:GFP:*LacI* construct. The percentage of extrachromosomal $CENH3^{CID}$/CENP-C foci was determined in mitotic cells with a fluorescent microscope. Black: GFP-negative cells containing $CENH3^{CID}$/CENP-C foci, green: GFP-positive cells (at least one GFP-positive focus) containing $CENH3^{CID}$/CENP-C foci.
(Mendiburo *et al.*, 2011)

After an initial high number of GFP-positive cells (green bars) with positive signals for CENH3^{CID} and CENP-C (representing extrachromosomal pDNA vectors with functional neocentromeres) the fraction of these cells subsequently decreased to about 50 % after one week and was finally lost after one month. Nonetheless, it was still possible to detect CENH3^{CID}/CENP-C positive foci in these GFP-negative cells (black bars). This means, that even after the loss of the *LacI*-targeting system, centromere formation and kinetochore assembly worked efficiently pointing towards epigenetically inherited marks recognised by CENH3^{CID} at centromeres (Mendiburo *et al.*, 2011).

4.7 Summary – Results I

The main question of my thesis was, if induced neocentromere formation on a pDNA vectors would turn the pDNA vector into a "quasi-chromosome" and enable an active segregation mechanism similar to the one of chromosomes. In *D.melanogaster* it was demonstrated that targeting of CENH3^{CID}:GFP:*LacI* indeed led to the induction of neocentromeres on the pDNA vectors. Our collaborators provided impressive evidence that the pDNA vectors were recognised by microtubules during mitosis and segregated actively to the daughter cells. Induced neocentromere formation on pDNA vectors increased the stability of the pDNA vector in cells after transfection, compared to pDNA vectors lacking neocentromeres.

I observed a pronounced difference in pDNA vector retention efficacy between the two systems (+/- CENH3^{CID}:GFP:*LacI* targeting), which is due to the differences in segregation mechanisms of the pDNA vectors. Furthermore I was able to demonstrate that *(1)* pDNA vectors remain their genetic integrity throughout selection, *(2)* that the preferred conformational state of the pDNA vector in the cell is concatemeric and *(3)* that DNA replication is independent of CENH3^{CID}:GFP:*LacI*-induced mechanism of pDNA vector segregation. Additionally, our collaborators demonstrated that the centromeric mark is epigenetically inherited. With these encouraging results in hands I tested the functionality of this principle also in the human system using HEK 293 cells.

5 RESULTS II – H.SAPIENS

Our studies in *D.melanogaster* provided proof of principle that artificially targeted CENH3^{CID}:GFP:*LacI* allows the formation of functional neocentromeres to support the active segregation of pDNA vectors. For all my experiments in human cells I used human embryonic kidney cells (HEK 293 cells). In the human system I also tested longer time spans of pDNA vector retention and additionally I characterised the system in regards to dose dependencies in *cis* and *trans*. Also an *all-in-one* pDNA vector system was established and I investigated the role of the CENH3^{CENP-A}´s targeting domain in neocentromere formation.

5.1 CENH3^{CENP-A} improves pDNA vector retention in HEK 293 cells

To investigate pDNA vector retention in the mammalian cell culture system I started with plasmid rescue assays as described before (Figure 5.1 A). For my experiments in human cells I used the well-established *pCON* system, relying on the interaction of the *TetOperator* sites on the reporter vectors and the *scTetR*:CENH3^{CENP-A} protein to investigate the function of pDNA vector retention of an *oriP*-based pDNA vector. A detailed introduction and illustration about the *pCON* system is given in chapter 1.1.2.

On the pDNA reporter vector 40 *TetOperator* binding sites replaced the functional domain of vector retention of *oriP*, the family of repeats (Figure 5.1 B). The wildtype dyad

symmetry element was present on the pDNA reporter vector to guarantee DNA replication. HEK 293 cells stably expressed EBNA1 to support DS function. In addition these cells stably expressed the *scTetR:*CENH3^{CENP-A} fusion protein, which was targeted to the *TetOperator* sites on the pDNA reporter vector. With this set-up I investigated, if *scTetR:*CENH3^{CENP-A} substitutes the function of FR by induction of a neocentromere on the pDNA vector.

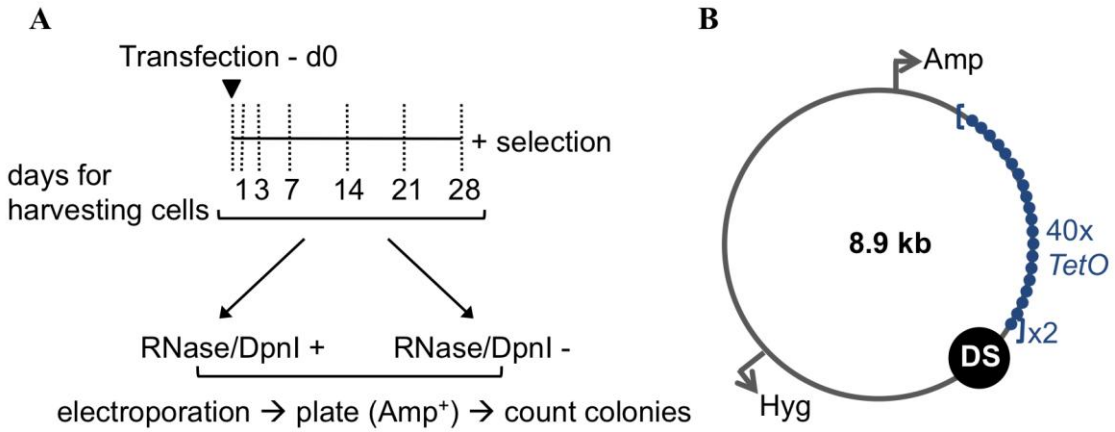


Figure 5.1 Experimental set-up of the plasmid rescue assay and map of the 40 TetOperator reporter vector.

A) HEK 293 EBNA1-*scTetR*:CENH3^{CENP-A} cells were transfected with the pDNA reporter vector and selected with 150 µg/ml hygromycin until indicated times. After lysis, bulk chromatin was precipitated with 1.25 M NaCl and the plasmid-enriched supernatant was phenol/chloroform extracted. DNA was digested with *Dpn*I, to eliminate *dam*-methylated input pDNA vectors derived from generation in bacteria and RNase for enhanced purity of DNA samples. 350 ng of low molecular weight DNA were electroporated into competent DH10B bacteria and plated on Amp$^+$ agar plates. The result of the assay was obtained after overnight incubation at 37 °C by counting outgrowing colonies on agar plates. **B)** The pDNA reporter vector contained an ampicillin resistance cassette for selection in bacteria and a hygromycin resistance cassette for selection in HEK 293 cells. DS was present on the pDNA vector in its wildtype status and 40 *TetOperator* sites were introduced instead of FR to test for pDNA vector retention efficiency with the *scTetR*:CENH3^{CENP-A} fusion protein.

Plasmid rescue experiments were performed with cells expressing *scTetR:*CENH3^{CENP-A} at more than five-fold increased levels compared to endogenous CENH3^{CENP-A} expression (Figure 5.3, lane 6) and cells lacking *scTetR:*CENH3^{CENP-A} expression. After transfection with the pDNA reporter vector, cells were selected with hygromycin (150 µg/ml) and harvested at indicated times (Figure 5.1 A).

The result of the experiment is depicted in Figure 5.2 A. Similar to analyses performed with *D.melanogaster*, the number of rescued pDNA vectors during the first three days after transfection was at comparable levels in both cell lines (Figure 5.2 A, day 1 and day 3). After seven days of selection *scTetR:*CENH3^{CENP-A} expressing cells retained the reporter

much more efficient than *scTetR*:CENH3^{CENP-A} lacking cells. These cells have lost almost all pDNA vectors at that time (Figure 5.2 A, day 7 and day 14). Cells carrying the selection resistance cassette survived under selection pressure, whereas cells that did not maintain the pDNA vector encoding for hygromycin resistance died shortly after day 3 under selection.

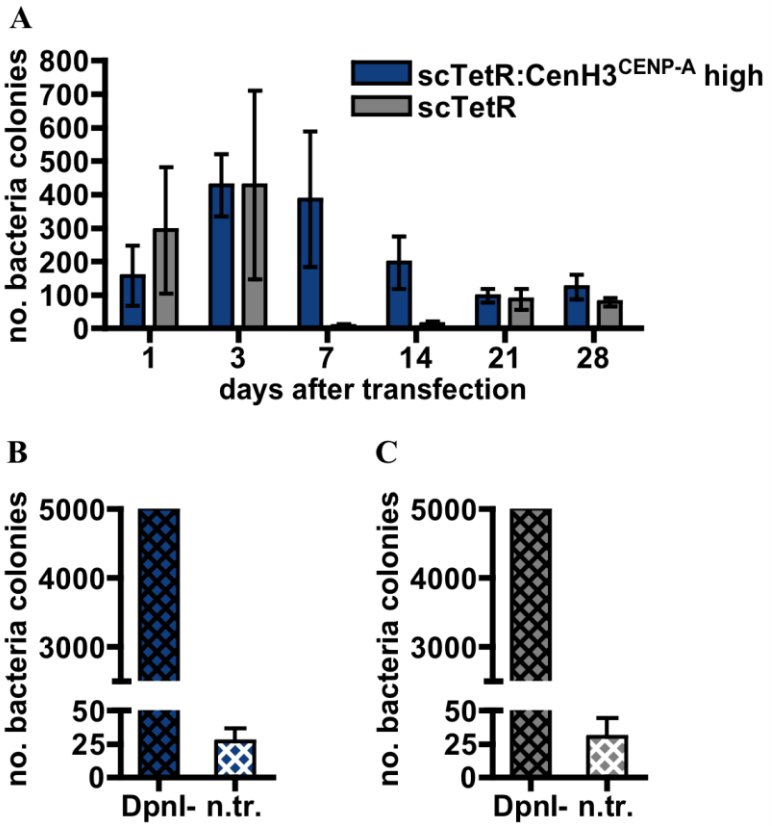


Figure 5.2 Plasmid rescue assays in HEK 293 cells displayed an enhanced pDNA vector retention after scTetR:CENH3^{CENP-A} expression in the cells.

A) Cells, which expressed *scTetR*:CENH3^{CENP-A} at high levels (blue bars) and cells that expressed only *scTetR* (grey bars) were transfected with 40 *TetOperator* pDNA reporter vectors (Figure 5.1 B) and plasmid rescue assays were performed (*scTetR*:CENH3^{CENP-A} cells: n = 4, +/- SEM, *scTetR* cells: n = 5, +/- SEM). Plasmid rescue assay experiments without *Dpn*I digest indicate that the initial transfection efficiency of the two cell lines was at similar levels in the beginning of the experiment. Plasmid rescue assay experiments of not transfected (n.tr.) cells show the background of the experiment. Control experiments are shown in **B)** for *scTetR*:CENH3^{CENP-A} cells (n = 3, +/- SD) and **C)** for *scTetR* cells (n = 3, +/- SD).

Until day 14 after transfection the rescue rates were as expected: *scTetR*: CENH3^{CENP-A} expressing cells maintained the pDNA vectors with the highest rescue rate between three and seven days post transfection. After a subsequent small loss of pDNA vectors the number of rescued pDNA vectors remained stable at a certain level after three weeks post transfection. The situation was different for *scTetR*:CENH3^{CENP-A} lacking cells, which lost pDNA vectors quite fast after three days post transfection. Most of these cells died due to

selective pressure, but unexpectedly some cells recovered and after cultivation for three weeks they were again able to give rise to clones in plasmid rescue assays. After three weeks under selection an average number of 96 colonies was detected from *scTetR*:CENH3^{CENP-A} positive cells and 86 colonies were found in assays from *scTetR* cells. At day 28 plasmid rescue assays led to 123 colonies from *scTetR*:CENH3^{CENP-A} expressing cells and 77 colonies from *scTetR* cells (Figure 5.2 A). The background level of colonies in the experiment was on average 27 colonies for *scTetR*:CENH3^{CENP-A} cells and 30 colonies for *scTetR* cells (Figure 5.2 B-C).

From the obtained data I conclude, that *(1)* the background (colonies from n.tr. cells) had a higher impact in the human system, as the overall number of colonies in these assays were much lower and *(2)* after pDNA vector establishment in the cell the copy number of pDNA vectors in cells expressing *scTetR*:CENH3^{CENP-A} drops to a certain level, where it seems to remain constant (Figure 5.2, day 21 and day 28). To investigate this hypothesis in more detail I performed long-term plasmid rescue assays spanning a period of several months. Results from these experiments are presented in chapter 5.3.

To solve the question, why colonies are growing in *scTetR* cells after three and four weeks one has to also perform long-term experiments with this cell line to see if the number of colonies decreases after a prolonged period of time. Additionally restriction enzyme analysis has to be done on the rescued clones to check if really the input pDNA vectors are rescued from these cells.

5.2 Exploring the potential of the system

The *pCON* system allows the independent investigation of DNA replication and pDNA vector retention. Within the system it is quite easy to modify certain parameters like *scTetR*:CENH3 expression or the number of targeted binding sites on the pDNA reporter vector. A higher expression level of the *scTetR*-fusion protein might influence the efficacy of the system in a positive or negative direction. Similar effects are possible by variation of the number of *TetOperator* binding sites on the pDNA reporter vector.

To test, if the expression level of *scTetR*:CENH3^{CENP-A} had any impact on the outcome of the experiment I additionally used a cell line expressing *scTetR*:CENH3^{CENP-A} at a lower level than in the short-term experiment (dose-dependency in *trans*). Further, I was interested if the number of *TetOperator* binding sites on the pDNA reporter vector influences the efficacy of the system. The pDNA reporter vector used in the first experiment comprised 40 *TetOperator* sites. Here, I additionally used a pDNA reporter

vector with only 20 binding sites, which reflects the number of EBNA1 binding sites in the FR element of *oriP*. A comparison of the two pDNA reporter vectors would elucidate a possible dose-dependency in *cis*.

5.2.1 Dose-dependencies in trans and cis

5.2.1.1 Cell lines and pDNA reporter vectors

To determine if the expression level of *scTetR*:CENH3^{CENP-A} had an impact on pDNA vector retention I established cell lines varying in the levels of *scTetR*:CENH3^{CENP-A} expression. To this end I used three different expression constructs for stable *scTetR*:CENH3^{CENP-A} cell lines. The expression constructs varied in the promoter type controlling *scTetR*:CENH3^{CENP-A} expression: a mini-CMV promoter (-53 +74) for high expression, an E-Cadherin promoter (-178 +92) for medium expression and a mini-E-Cadherin promoter (-21 +92) for low expression of the artificial protein. The aim was to establish a cell line with a *scTetR*:CENH3^{CENP-A} expression level comparable to the endogenous CENH3^{CENP-A} expression, since it has been shown in *D.melanogaster* that too high expression of CENH3^{CID} led to mitotic defects in cells (Heun *et al.*, 2006). Contrary to this a too low expression might result in no effect on *scTetR*:CENH3^{CENP-A} recruitment to the *TetOperator* pDNA reporter vectors.

Western Blot analysis demonstrated that *scTetR*:CENH3^{CENP-A} expression levels with all promoters tested were high compared to endogenous CENH3^{CENP-A} and so I used the mini-E-Cadherin promoter for further generation of cell lines. To this end I co-transfection a linearised expression plasmid, encoding for *scTetR*:CENH3^{CENP-A} and a puromycin resistance cassette in HEK 293 EBNA1 cells. Cells were kept under selection with puromycin (300 ng/ml) for two to three weeks, before single clones were isolated and analysed for their expression profile by Western Blot analysis with a CENH3^{CENP-A}-specific antibody. With the mini-E-Cadherin expression cassette I generated five different cell clones and their individual expression profiles are illustrated in the Western Blot in Figure 5.3.

Endogenous CENH3^{CENP-A} migrated at 17 kDa (Figure 5.3, grey arrow head) and the *scTetR*:CENH3^{CENP-A} fusion protein was detected at 65 kDa (Figure 5.3, black arrow head). The short exposure illustrates the expression differences between the five cell lines (Figure 5.3, top panel) and the long exposure allows the comparison of expression levels of *scTetR*:CENH3^{CENP-A} and the endogenous CENH3^{CENP-A} protein (Figure 5.3, middle

panel). Not transfected HEK 293 EBNA1 cells (lane 1) were used as negative control and tubulin served as loading control (Figure 5.3, lower panel).

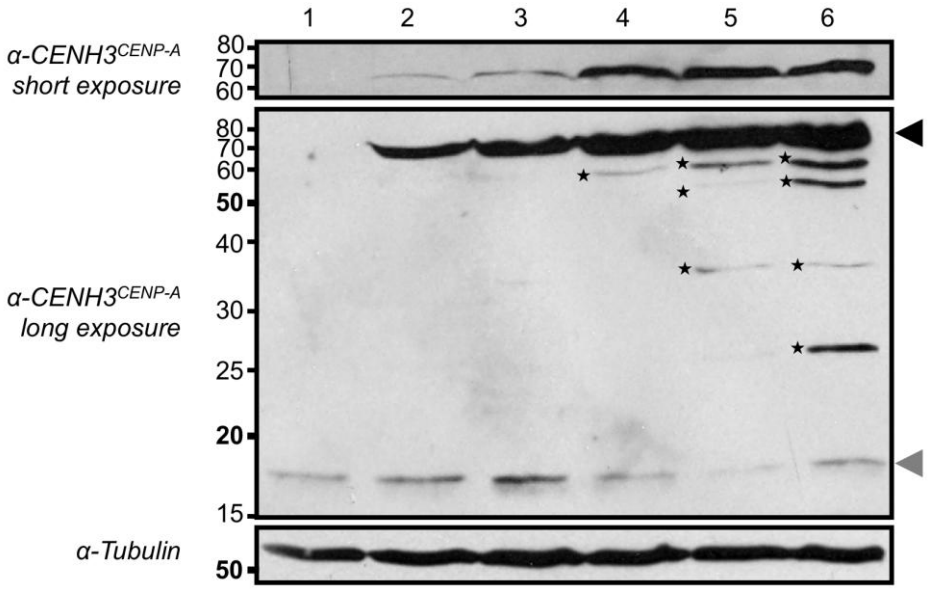


Figure 5.3 Western Blot of scTetR:CENH3^{CENP-A} expression levels in stable cell lines.

Stable expression of *scTetR*:CENH3^{CENP-A} in HEK 293 cells was achieved by co-transfection of linearised expression plasmids encoding for *scTetR*:CENH3^{CENP-A} and a puromycin resistance cassette. After two to three weeks of puromycin selection (300 ng/ml) single clones were analysed for the expression of *scTetR*:CENH3^{CENP-A} using a CENH3^{CENP-A}-specific antibody. Endogenous CENH3^{CENP-A} was detected at 17 kDa and *scTetR*:CENH3^{CENP-A} was detected at 65 kDa. Short exposure illustrates the expression differences between the five cell lines (top panel) and the long exposure compares the expression levels of *scTetR*:CENH3^{CENP-A} and endogenous CENH3^{CENP-A}; lane 1: non-transfected HEK 293 EBNA1 cells (negative control), lane 2: *scTetR*:CENH3^{CENP-A} clone 1, lane 3: *scTetR*:CENH3^{CENP-A} low, lane 4: *scTetR*:CENH3^{CENP-A} clone 3, lane 5: *scTetR*:CENH3^{CENP-A} clone 4, lane 6: *scTetR*:CENH3^{CENP-A} high, tubulin served as loading control. Asterisks indicate unspecific bands. Black arrow head: *scTetR*:CENH3^{CENP-A}, grey arrow head: endogenous CENH3^{CENP-A}.

Additionally to the level of *scTetR*:CENH3^{CENP-A} expression, the number of *TetOperator* sites on the pDNA reporter vector might have an impact on the effectiveness of the *pCON* system. The number of *TetOperator* sites could influence *e.g.* the binding efficiency between *scTetR*:CENH3^{CENP-A} and the *TetOperator* repeats on the pDNA vector. It has been published before that an increase in the number of *TetOperator* sites on the pDNA reporter vector leads to higher copy numbers in a *scTetR*:HMGA1a cell background (Thomae *et al.*, 2008). To test if this dose-dependency also holds true for the *scTetR*:CENH3^{CENP-A} fusion protein and pDNA vector retention I compared three different pDNA reporter vectors in the experiment: a wildtype *oriP* vector and two pDNA reporter vectors with differing amounts of *TetOperator* sites (Figure 5.4).

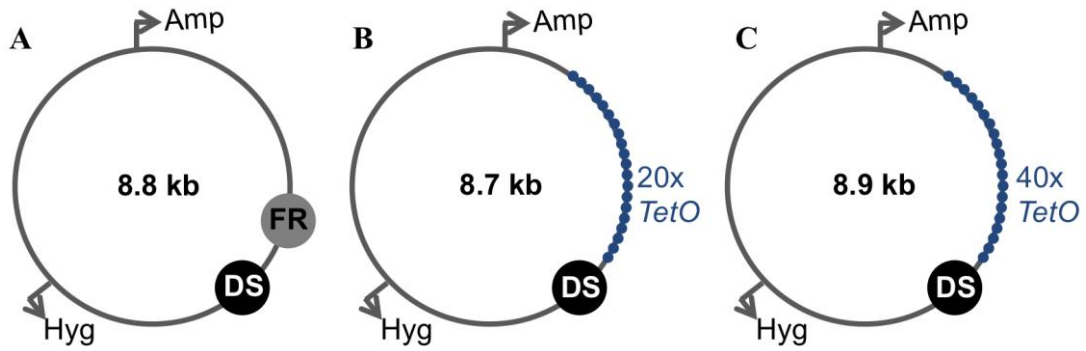


Figure 5.4 pDNA reporter vectors for plasmid rescue assays in HEK 293 cells.

All three reporter vectors contained an ampicillin resistance gene for propagation in bacteria and a hygromycin resistance cassette for selection in human cells. **A)** wt-*oriP* vector. The control vector contained both elements of *oriP* in wildtype status. DS and FR function *via* the viral protein EBNA1, which is stably expressed in all used cell lines. **B)** 20 *TetOperator* pDNA reporter vector. The number of *TetOperator* sites represents the number of EBNA1 binding sites of the wildtype FR element. The DS element is kept in its wildtype configuration and functions *via* EBNA1 interaction. **C)** 40 *TetOperator* pDNA reporter vector. The reporter pDNA vector contained 40 *TetOperator* sites. The DS element is present on the pDNA vector to support DNA replication *via* EBNA1.

5.2.2 No dose-dependencies were detected in the pCONactive model

From the established cell clones I chose two cell lines to perform plasmid rescue assays, namely *scTetR*:CENH3^{CENP-A} low (Figure 5.3, lane 3) and *scTetR*:CENH3^{CENP-A} high (Figure 5.3, lane 6). Both cell lines exceeded the endogenous CENH3^{CENP-A} expression level but differed strongly in regards to *scTetR*:CENH3^{CENP-A} expression. I performed plasmid rescue assays as described before with both pDNA reporter vectors and both cell lines (Figure 5.5).

No substantial difference was detected in the amount of rescued reporters from both investigated cell lines (Figure 5.5, A - B). In the case of the 20 *TetOperator* pDNA reporter vector similar numbers of pDNA vectors were rescued from both cell lines after three weeks and more pDNA vectors were rescued from *scTetR*:CENH3^{CENP-A} low expressing cells after four weeks under selection. However, also the SEM[37] value was much higher in this case than for the *scTetR*:CENH3^{CENP-A} high expressing cell line. The 40 *TetOperator* pDNA reporter vector behaved differently: after three weeks more pDNA vectors were rescued from *scTetR*:CENH3^{CENP-A} high expressing cells and after four weeks comparable amounts of pDNA vectors were rescued from both cell lines. According to the amount of rescued colonies I can state that the number of rescued pDNA vectors was generally lower with cells that have obtained the 40 *TetOperator* pDNA reporter vector. I observed the tendency, that 40 *TetOperator* reporters were lost faster from cells than 20 *TetOperator*

[37] Standard error mean

pDNA reporter vectors, especially in *scTetR*:CENH3^{CENP-A} low expressing cells (Figure 5.5 C - D).

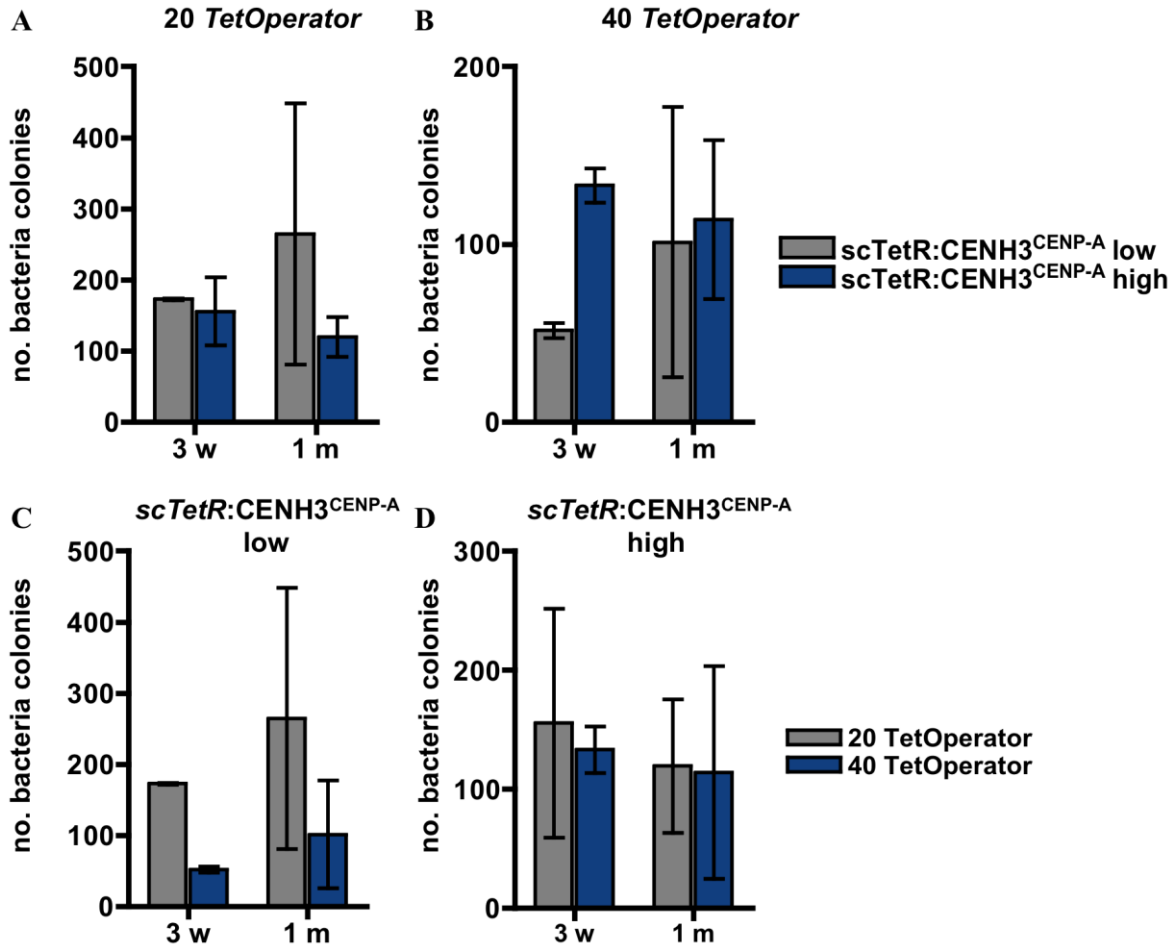


Figure 5.5 scTetR:CENH3^{CENP-A} expression levels in cells as well as the number of TetOperator sites on the pDNA vectors have no impact on pDNA vector retention.

Plasmid rescue assays were performed with *scTetR*:CENH3^{CENP-A} low expressing cells (grey bars) and *scTetR*:CENH3^{CENP-A} high expressing cells (blue bars). Experiments were performed with two different pDNA reporter vectors after three weeks (3 w) and one month (1 m) under selection with 150 µg/ml hygromycin. The 20 *TetOperator* pDNA reporter vector and the 40 *TetOperator* pDNA reporter vector were transfected in *scTetR*:CENH3^{CENP-A} low and high expressing cells. *scTetR*:CENH3^{CENP-A} low: n = 3, +/- SEM, *scTetR*:CENH3^{CENP-A} high: n = 4, +/- SEM.

I conclude from the experiment, that *scTetR*:CENH3^{CENP-A} expression levels have no impact and the number of *TetOperator* sites might have little impact on pDNA vector retention in cells.

5.3 pDNA vectors are maintained in cells over a period of five months

Next I wanted to investigate the extrachromosomal stability of the gene vector over a prolonged period of time and to do so I set up a long-term plasmid rescue assay. In addition I aimed to determine the requirement of a constitutive antibiotic selection. To this end I performed parallel experiments with cells under sustained selection and cells from which the selection media was withdrawn after the initial selection period of three weeks. The comparison of these two sets of cells in plasmid rescue assays allowed the investigation of the epigenetic inheritance of the neocentromeric mark and confirmed the results obtained in the *Drosophila* system (Figure 4.10) with a second experimental approach.

With the experimental set-up illustrated in Figure 5.6 I was able to determine the overall pDNA vector retention for a high number of cell generations. HEK 293 cells are fast dividing cells and during the experiment I monitored the pDNA vector retention efficiency in more than 100 cell generations during five months.

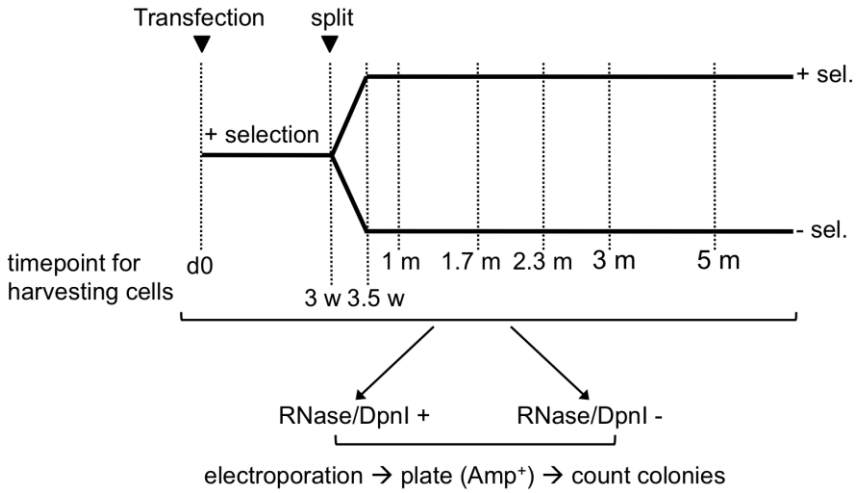


Figure 5.6 Long-term experimental set-up of the plasmid rescue assay with and without selection pressure.

HEK 293 cells (*scTetR*:CENH3$^{CENP\text{-}A}$ high and low expressing) were transfected with three pDNA reporter vectors (Figure 5.4). Cells were divided into two fractions after selection for three weeks with 150 µg/ml hygromycin. One set of cells was continuously kept under the same selection conditions as before until the end of the experiment; the other fraction was cultivated without selection pressure after an initial selection period of three weeks. Bulk chromatin was precipitated with a high salt concentration (1.25 M NaCl) and the pDNA vector-enriched supernatant was phenol/chloroform extracted. DNA was digested with *Dpn*I and RNase and 350 ng of DNA were electroporated into competent DH10B bacteria and plated on Amp$^+$ agar plates. The result of the assay was obtained after overnight incubation at 37 °C by counting outgrowing colonies on agar plates.

As different expression levels of *scTetR*:CENH3^{CENP-A} might not have an impact in short-term experiments, but possibly in a long-term set-up, I again tested high and low *scTetR*:CENH3^{CENP-A} expression levels and both pDNA reporter vectors in plasmid rescue assays. I used the wt-*oriP* vector (Figure 5.4 A) as control in the experiment.
With my experiments I demonstrated that apparently a different mechanism of pDNA vector establishment for *oriP* and *TetOperator* pDNA reporter vectors led to major differences in the amount of rescued pDNA vectors. After three weeks wt-*oriP* vectors were present in 27-fold higher numbers than both *TetOperator* pDNA reporter vectors (Figure 5.7). Again, I confirmed that the level of *scTetR*:CENH3^{CENP-A} expression as well as the number of *TetOperator* sites on the pDNA reporter vectors have no impact on pDNA vector retention also during a longer time span. Not only during the beginning of the experiment but also later I observed differences between the wt-*oriP* vector and the *TetOperator* pDNA vectors. For instance the very high initial copy number of wt-*oriP* vectors was lost quickly whereas the *TetOperator* reporters remained the level of pDNA vectors relatively stable over time and independently of selective pressure. Summing up the findings that were obtained from these experiments I conclude that:

(1) Much less *TetOperator* pDNA reporter vectors were rescued from cells during the first month after transfection compared to wt-*oriP* vectors. A possible reason for this is, that the copy number of the *TetOperator* pDNA reporter vectors in cells is lower than for the wt-*oriP* vectors. However, the establishment efficiency of both types of reporters might be similar. Additionally the retention mechanism of *TetOperator* pDNA vectors should be altered in comparison to the wt-*oriP* vectors.

(2) wt-*oriP* vectors display a faster loss rate as the *TetOperator* pDNA reporter vectors in the presence and absence of selection. The number of rescued pDNA vectors is reduced to less than 10 % of the initial wt-*oriP* vector count within only a few days. Selection pressure is required for stable extrachromosomal maintenance of wt-*oriP* vectors. Cells without selection lose the wt-*oriP* vectors more quickly (93 %) than cells under selection (73 %) within a few days (Figure 5.8 C). This very rapid loss of *wt-oriP* vectors especially in the presence of selection is unusual. Normally the vector should have established in the cell at that time and display a loss rate of only 3 -5 % per generation afterwards (Leight and Sugden, 2001).

(3) scTetR:CENH3^{CENP-A}-dependent pDNA vectors were maintained throughout the complete duration of the experiment. After five months with and without selection pressure similar numbers of pDNA vectors were rescued as in the beginning of the experiment.

(4) pDNA vector retention is not dependent on selection pressure in the active segregation mechanism. This suggests in accordance with our previous finding from *D.melanogaster*, that CENH3^{CENP-A} establishes an epigenetic mark on the pDNA vector itself, which is able to be retained over many generations (Mendiburo *et al.*, 2011).

(5) To summarize these findings, the copy number of *TetOperator* pDNA reporter vectors in the cells is lower than the number of *wt-oriP* vectors. However, the retention of established reporter vectors is much more constant for the *TetOperator* pDNA vectors and enables a much longer stable maintenance of the vectors in the transfected cells.

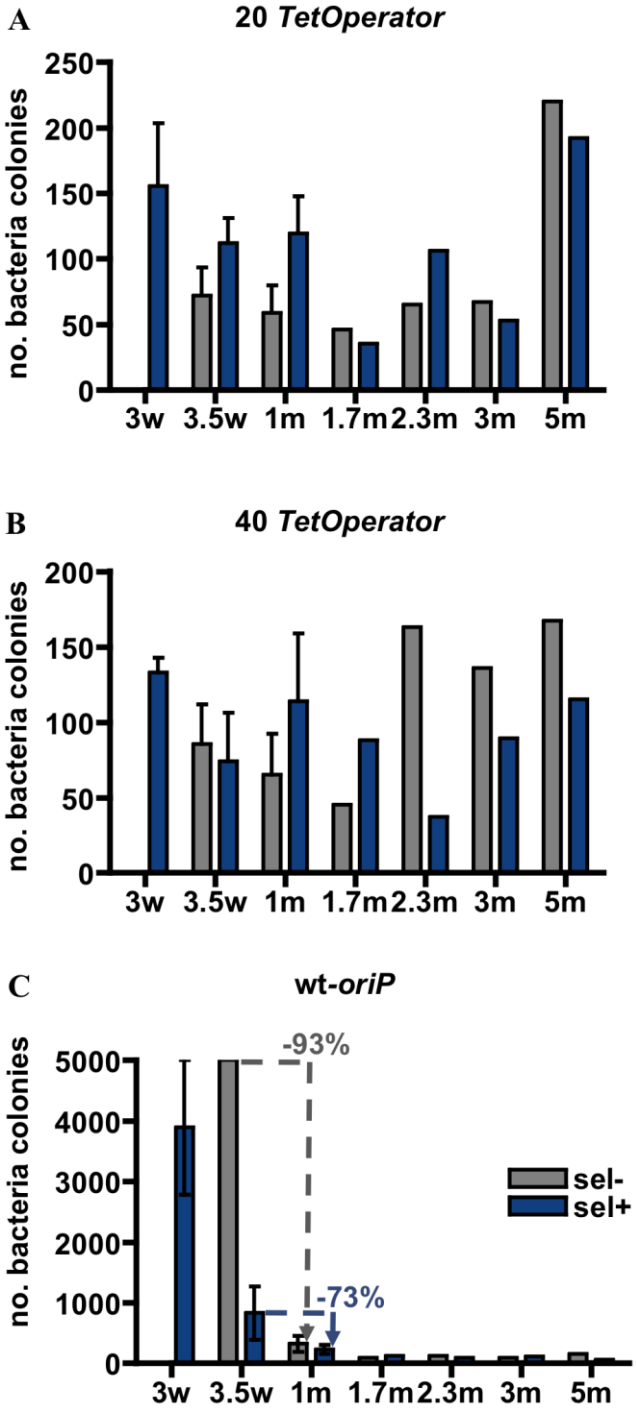


Figure 5.7 Long-term plasmid rescue assay set-up with and without selection pressure.

Plasmid rescue assays were performed as described before with three different pDNA reporter vectors over a period of five months in *scTetR*:CENH3$^{CENP\text{-}A}$ high expressing cells. Transfected cells were selected for three weeks in the presence of hygromycin and then divided into half. Half of the cells were cultivated without selection (grey bars) and the other half was kept under antibiotic selection with 150 µg/ml hygromycin (blue bars). Time points 3 w, 3.5 w, 1 m: n = 4 +/- SEM, all other time points: n = 1, sel+: selection pressure applied, sel-: no selection pressure applied.

5.4 Histone H3 variants and mutants

Different variants of histone H3 exist, namely histone H3.1, H3.2, H3.3, H3T (H3.4), H3.5, H3.X, H3.Y and the centromere specific variant CENH3^{CENP-A} (Kurumizaka, 2012). Histone H3.3 is most related to CENH3^{CENP-A}. The only difference between the histones H3.3 and CENH3^{CENP-A} is the CENP-A targeting domain (*CATD*), which is present only in the centromere specific CENH3 variant (Figure 5.8 B-C). The *CATD* domain spans the α2-helix and loop1 within the histone fold domain of CENH3^{CENP-A} and is absolutely necessary for protein targeting (Black *et al.*, 2007).

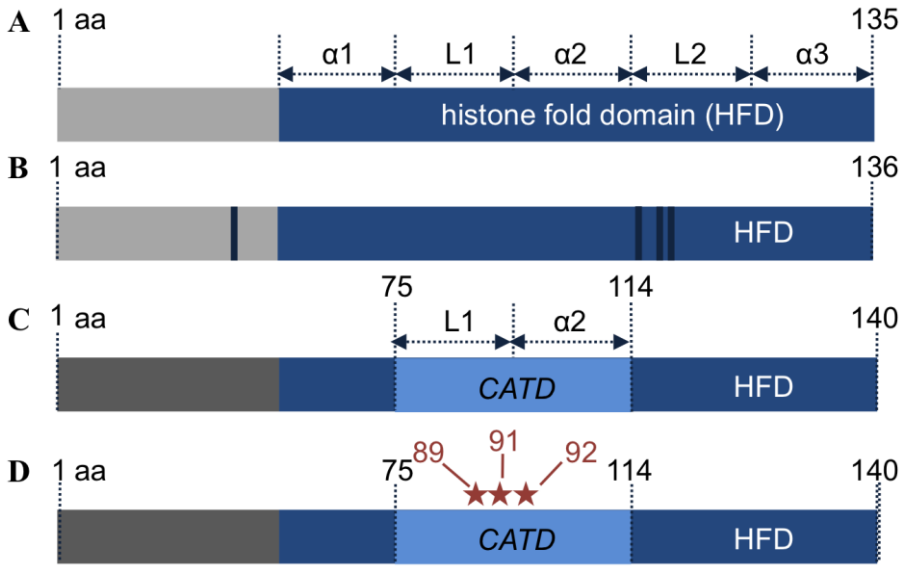


Figure 5.8 Histone 3 variants and mutants used in this work.

Different variants and mutants of histone 3 are shown. **A)** canonical histone 3.1. **B)** histone H3.3. The black bars indicate the positions of H3.3 specific modification compared to H3.1 modifications. **C)** CENH3^{CENP-A} wildtype. **D)** CENH3^{CENP-A}*CATD*^{mut}. The N-terminal parts of CENH3^{CENP-A}, canonical H3.1 and H3.3 differ strongly whereas the C-terminal parts of the proteins are highly conserved. HFD: histone fold domain, aa: amino acid, L1: loop1, L2: loop2, α1: α-helix 1, α2: α-helix 2, α3: α-helix 3, *CATD*: CENP-A targeting domain, ★ point mutation
(Sarma and Reinberg, 2005)

The data presented in the previous chapters demonstrated the ability of CENH3^{CENP-A} to support pDNA vector segregation most likely by neocentromere formation on the pDNA vector. In fact, the actual formation of a neocentromere on the pDNA vector was so far only shown in *D.melanogaster* by immunofluorescence experiments (Figure 4.9). To analyse, whether this function is exclusively dependent on the *CATD* of CENH3^{CENP-A} or other histones might also support pDNA vector retention, I performed plasmid rescue experiments with CENH3^{CENP-A} containing a non-functional *CATD* domain and also with the histone variant H3.3.

5.4.1 The integrity of CEN3$^{CENP\text{-}A}$´s CATD domain is essential for pDNA vector retention

The *CATD* domain - in interaction with CENH3$^{CENP\text{-}A}$ ´s chaperone HJURP - is absolutely necessary for binding of the protein to chromatin (Black *et al.*, 2007) and represents the only structural difference between histone CENH3$^{CENP\text{-}A}$ and H3.3 (Barnhart *et al.*, 2011). I investigated the effect of three amino acid exchanges in the *CATD* domain of CENH3$^{CENP\text{-}A}$ on pDNA vector retention by performing plasmid rescue assays with cells expressing the mutant CENH3$^{CENP\text{-}A}$ coupled to *scTetR*. This loss-of-function *CATD* mutant was generated with three mutations at the amino acid positions 89, 91 and 92 (Figure 5.8, D) (GENEART AG, Regensburg). The wildtype DNA sequence **CAG** GCC **CTA TTG** was changed to **AGC** GCC **GTA ATG**, which results in a change from amino acids **Q**A**LL** to **S**A**VM** (Black *et al.*, 2007). The mutations represent the respective sequences for histones H3.1 and H3.2 and lead to a mistargeting of CENH3$^{CENP\text{-}A}$. The mutant protein served as negative control in my assay.

In my experiments I used HEK 293 EBNA1 cells in which I co-transfected a 40 *TetOperator* pDNA reporter vector and either a plasmid for the transient expression of the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} fusion protein or a *scTetR*:CENH3$^{CENP\text{-}A}$ expression plasmid as reference. After three weeks of selection with 150 µg/ml hygromycin I performed plasmid rescue assays with these cells as described before.

To test the transient expression level of the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} and the *scTetR*:CENH3$^{CENP\text{-}A}$ proteins I performed Western Blot analysis 48 hours post transfection using a *TetR*-specific antibody (Figure 5.9 B). Cell lysates were obtained from the same cells as used in plasmid rescue assays 48 hours post transfection. *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} and *scTetR*:CENH3$^{CENP\text{-}A}$ were detected at 65 kDa. The *scTetR* element alone was detected at 48 kDa.

I obtained the highest level of protein expression with the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} construct (Figure 5.9, lane 3). Transient expression of *scTetR*:CENH3$^{CENP\text{-}A}$ (Figure 5.9, lane 2) resulted in much lower protein expression compared to the expression of this construct in stable cell lines (Figure 5.9, lane 1). The Western Blot against tubulin demonstrated, that much less extract was loaded in from the latter sample compared to other protein samples (Figure 5.9 B, lower panel).

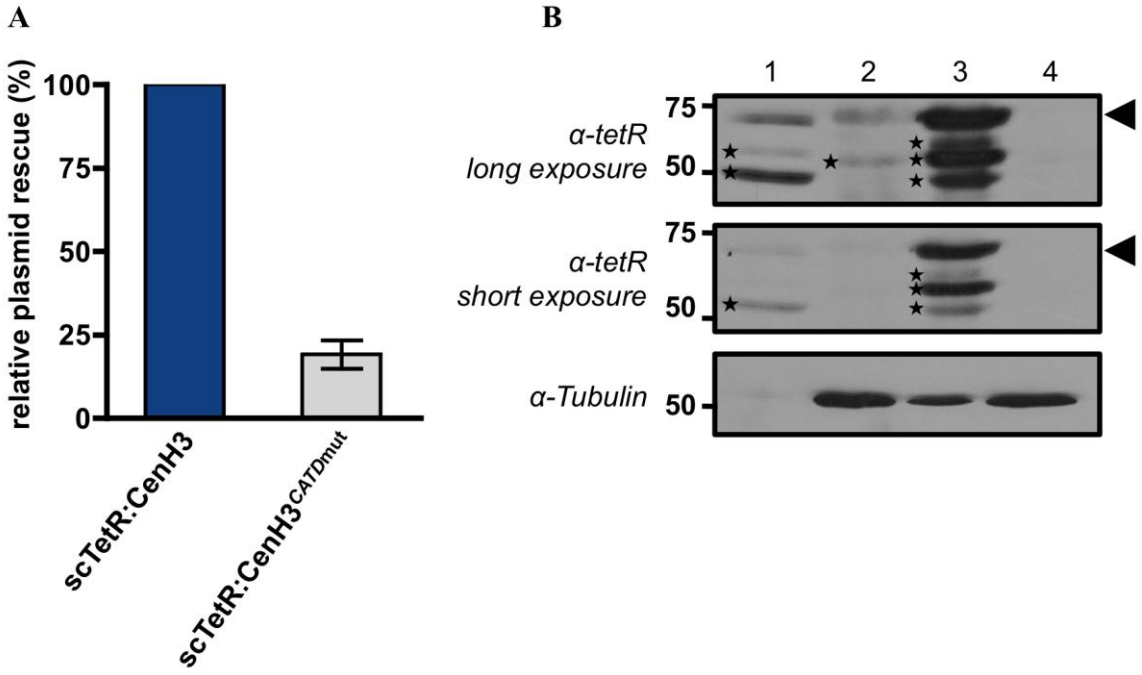


Figure 5.9 Mutation of the CATD of CENH3 leads to a massive decrease in retained pDNA vectors.

A) Plasmid rescue assays were performed as described earlier (chapter 3.1.4.1). Experiments in cells expressing a mutant version of CENH3^{CENP-A} *CATD* were only able to rescue 19 % of the amount of pDNA vectors of wildtype CENH3^{CENP-A} (n = 2, +/- SD). **B)** Representative Western Blot of the transiently expressed *scTetR*:CENH3^{CENP-A} and *scTetR*:CENH3^{CENP-A}*CATD*^{mut} from experiment II 48h post transfection. Cell lysates of different cell lines were loaded. Lane 1: CENH3^{CENP-A} high (positive control, stable cell line), lane 2: CENH3^{CENP-A}-wildtype transient transfection; lane 3: CENH3^{CENP-A}*CATD*^{mut}, transient transfection, lane 4: HEK 293 EBNA1 (negative control). Black arrowheads indicate the signal of the *scTetR*:CENH3^{CENP-A} fusion protein. Tubulin was used as loading control and asterisks indicate unspecific bands.

As expected, plasmid rescue assay experiments proved the importance of the *CATD* domain for the functionality of the *pCONactive* model system. The number of colonies obtained with transient expression of the *scTetR*:CENH3^{CENP-A} was set to 100 %. With the *scTetR*:CENH3^{CENP-A} *CATD*^{mut} only 19 % of the wildtype CENH3^{CENP-A} level was observed (Figure 5.9 A). This finding suggests that a functional *CATD* domain is required for stable pDNA reporter vector segregation, as a non-functional *CATD* domain does not support neocentromere establishment and a quicker loss of the pDNA vectors from the cell during cell division.

5.4.1.1 Genetic integrity of rescued pDNA vectors

To check if the pDNA reporter vectors remain stable in the cells during the selection period I tested the genetic integrity of rescued pDNA vectors with an analytical restriction digest. I picked single clones from plates of plasmid rescue assays after three weeks under selection (150 µg/ml) and digested the DNA with the restriction enzyme *Pst*I. The

predicted fragment sizes for the 40 *TetOperator* pDNA reporter vector after *Pst*I digest are shown in Table 5.1 and a vector map with the cut sites is depicted in Figure 5.10.

Table 5.1 Fragment prediction of the 40 TetOperator cut with PstI.

Predicted fragment size (bp)	***Annotation***
5719	Amp resistance cassette, *TetOperator* sites, DS
2146	part of hygromycin resistance cassette, eGFP
1071	part of hygromycin resistance cassette, VBB[38]

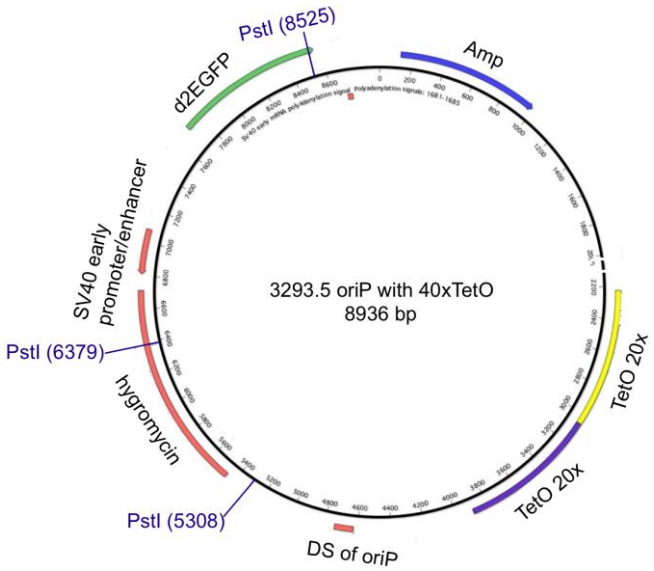


Figure 5.10 pDNA vector map of the 40 TetOperator vector with PstI restriction sites.

*Pst*I cuts the 40 *TetOperator* pDNA reporter vector three times, once in the pDNA vector backbone, once in the hygromycin resistance cassette and once in the eGFP sequence. The predicted fragment sizes and annotations of the vectors after digest are given in Table 5.1.

In the analysis three clones derived from both cell lines (mutant and wildtype) were tested (Figure 5.11). The input plasmid was loaded in undigested (Figure 5.11, lane 1) and digested status (Figure 5.11, lane 2) as reference. All clones rescued from *scTetR*:CENH3$^{CENP\text{-}A}$ cells showed three bands according to the predicted fragment size, with the largest fragment migrating slightly higher than expected (Figure 5.11, lanes 3-5). Clones rescued from the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} cells did not reflect the predicted fragment pattern (Figure 5.11, lanes 6-8). Clone 6 is not well visible on the gel due to a low DNA content in the sample and clones 7 and 8 displayed a dominant pattern, that was also obtained with clones from experiments in *scTetR*:H3.3 cells (fragment sizes: ~ 1200 bp, ~ 1400 bp, ~ 2700 bp). After sequencing analysis of this vector it turned out that all of

[38] Prokaryotic vector backbone

these vectors display the same sequence (see appendix 9.1 and 9.2) and that the rescued vector is a contamination of the sample with a foreign plasmid. However, the contaminating plasmid is only detected in plasmid rescue assays performed with *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} cells. This indicates that the colony number of the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} experiment might be indeed even lower than measured in Figure 5.11. However, the experiment has to be repeated to confirm this assumption.

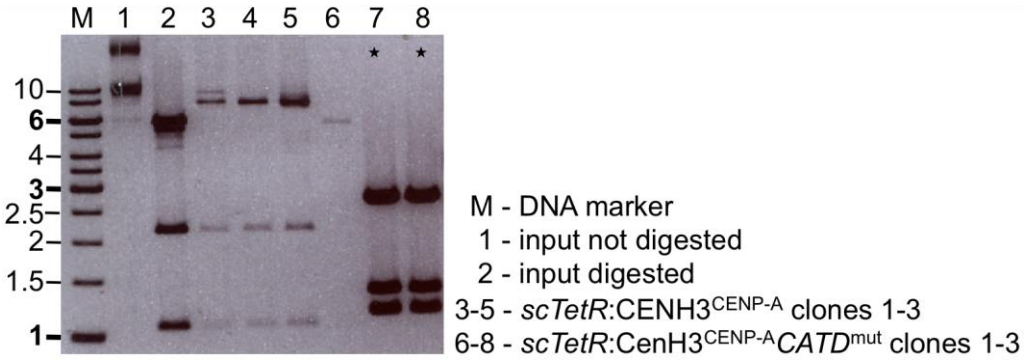


Figure 5.11 Analytical restriction digest of the 40 TetOperator pDNA reporter vector rescued from scTetR:CENH3$^{CENP\text{-}A}$ and scTetR:CENH3$^{CENP\text{-}A}$CATD^{mut} cells with PstI.

DNA was isolated from colonies taken from plasmid rescue assay plates and digested with *Pst*I to analyze the stability of the transfected pDNA reporter vector after three weeks under selection. Predicted fragment sizes for this pDNA vector after *Pst*I digest were: 5719 bp, 2146 bp and 1071 bp. The dominant fragment pattern of plasmids marked with * represented the pattern of a contaminating plasmid (lanes 7 and 8). Marker bands indicate kbp sizes.

5.4.2 The histone variant H3.3 does not support pDNA vector retention

The histone 3 variant H3.3 is a specialised variant among all histone 3s expressed cell cycle-independently. It comprises only a few differences in amino acid composition compared to Histone H3.1 (Elsaesser *et al.*, 2010). H3.3 incorporates predominantly at active chromatin regions in a replication-independent mechanism (Elsaesser *et al.*, 2010). It was demonstrated that H3.3 might even be responsible for initiation of gene transcription of IFN[39]-inducible genes (Tamura *et al.*, 2009). During S-phase, no CENH3$^{CENP\text{-}A}$ but H3.3 is incorporated at centromeres, leading to a dilution of CENH3$^{CENP\text{-}A}$ during the cell cycle. The centromeric H3.3 is only replaced afterwards in the following G_1 phase (Dunleavy *et al.*, 2011).

To compare the two histone H3 variants I performed plasmid rescue assays with cells expressing *scTetR*-H3.3 and *scTetR:*CENH3$^{CENP\text{-}A}$ during a period of ten weeks. In addition I investigated the role of selection pressure in the experiment by keeping one set

[39] Interferon

of cells in the presence and a second set of cells in the absence of selection pressure (150 µg/ml Hygromycin). Plasmid rescue assays were performed as described before.

5.4.2.1 Generation of scTetR-H3.3 positive cells

HEK 293 EBNA1 cells were co-transfected with a plasmid encoding for the *scTetR*:H3.3 protein under the control of an E-Cadherin promoter as well as a puromycin resistance cassette as described in chapter 3.1.5 and selected with puromycin (300 ng/ml) for two to three weeks. Single cell clones were isolated from the plates and analysed for the stable expression of *scTetR*:H3.3 by Western Blot using a *TetR*-specific antibody (Figure 5.12). The *scTetR*:H3.3 protein was detected at 65 kDa. As positive control I used a *scTetR*:HMGA1a stable cell line (Figure 5.12, lane 1).

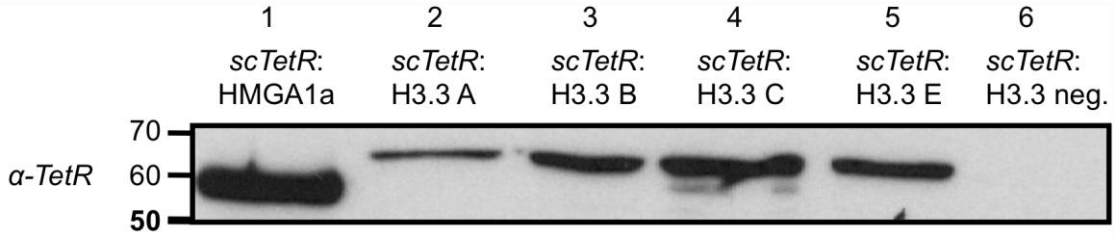


Figure 5.12 scTetR-H3.3 expression levels in different stable cell lines.

Stable expression of *scTetR*:H3.3 in HEK 293 cells was achieved by co-transfection of linearised expression plasmids encoding for *scTetR*:H3.3 and a puromycin resistance cassette. After two to three weeks of puromycin selection (300 ng/ml puromycin) single clones were analysed for the expression of *scTetR*:H3.3 using a *TetR*–specific antibody. *scTetR*:H3.3 was detected at 65 kDa. As a control I loaded a *scTetR* positive cell line in lane1. Cell lysates from four *scTetR*-H3.3 positive clones with different expression levels were loaded in lanes 2-5 and the lysate from a *scTetR*-negative clone was loaded in lane 6.

5.4.2.2 Ability of H3.3 to support pDNA vector retention

I started these experiments with the initial idea to use histone H3.3 as a negative control for the model of *scTetR*:CENH3$^{CENP\text{-}A}$ induced neocentromere formation and active pDNA vector segregation. I expected H3.3 to be non-functional in supporting pDNA vector retention due to the fact that the *CATD* domain of CENH3$^{CENP\text{-}A}$, which is missing in H3.3, is specifically required for the pDNA vector retention function of CENH3$^{CENP\text{-}A}$. As outlined above, this domain is the major structural difference between CENH3$^{CENP\text{-}A}$ and H3.3 and the loss-of-function *CATD* mutant reduced the efficacy of the system significantly to 19 % (Figure 5.9).

I tested the wt-*oriP* reporter vectors as well as the two *TetOperator* pDNA reporter vectors (Figure 5.4) in plasmid rescue assays with HEK 293 EBNA1-*scTetR*:H3.3 cells. I transfected cells with the pDNA reporter vectors and cultivated them under selection (150

µg/ml Hygromycin) for three weeks. Afterwards I divided the culture in two groups, one of which was kept in antibiotic-free medium and the other group under continuous selective pressure until the end of the experiment. Samples were taken in addition at three and seven weeks (overall duration of the experiment 10 weeks) and plasmid rescue assays were performed.

In contrast to my expectations the results I obtained were surprising, as the *scTetR*:H3.3 fusion allowed the rescue of a high number of pDNA vectors after three weeks (Figure 5.13, black bars). After the initial three weeks of selection the amount of rescued pDNA vectors was similar between the EBNA1-*oriP* system and the *TetOperator* pDNA reporter vectors targeted by *scTetR*:H3.3.

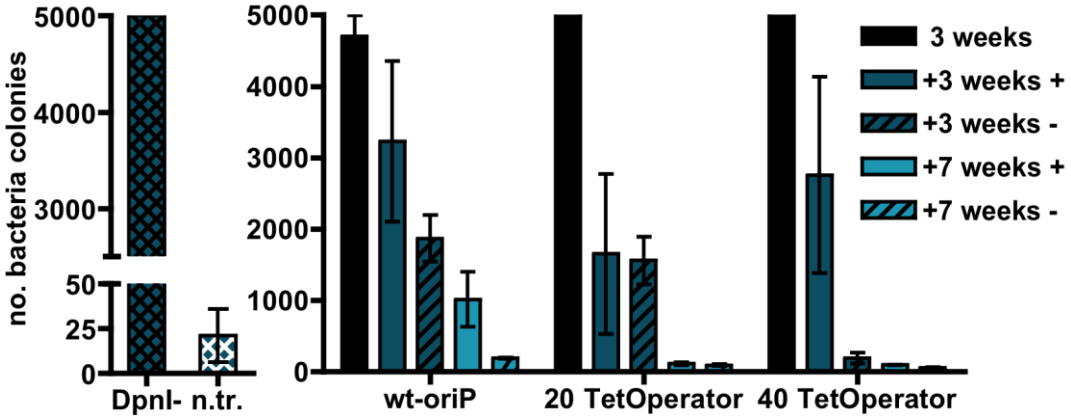


Figure 5.13 Results of plasmid rescue assays performed with scTetR-H3.3 expressing cells.

Plasmid rescue experiments were performed as indicated before (chapter 3.4.1). Cells were selected for an initial period of three weeks, afterwards the culture was divided in two parts, from which one part was further kept under constitutive selection (plain bars) and the other part was cultured without selection antibiotics (dashed bars). Cells were harvested after three and seven additional weeks and analysed in plasmid rescue assays ($n = 3$, +/- SEM). Control experiments were performed without *Dpn*I digest and not transfected cells ($n = 3$, +/- SEM). +: with selection, -: without selection

The wt-*oriP* vector functions *via* a *scTetR*:H3.3 independent mechanism, only reliant on EBNA1. The result that I obtained for the wt-*oriP* vector was according to previous experiments in our lab, even though the pDNA vector loss happened more slowly than expected. In the presence of hygromycin (150 µg/ml), the amount of rescued pDNA vectors dropped to 20 % between the first and last time point. In the absence of selection the wt-*oriP* vector loss was faster and reduced to about 95 % after ten weeks (Figure 5.13, *oriP*-panel). I expected a rapid loss of the 20- and 40 *TetOperator* reporters in *scTetR*:H3.3 cells, as the mechanism of active pDNA vector retention is not functional in these cells. Very high numbers of pDNA reporter vectors were rescued after three weeks similar to the

oriP vector (compare black bars). After three weeks of cultivation in the presence of hygromycin (150 µg/ml) the number of rescued 20 *TetOperator* pDNA reporter vectors dropped rapidly to approx. 50 % of the initial number. This observation was even more pronounced after ten weeks, when more than 95 % of the initial pDNA vectors were lost from cells. Interestingly, the withdrawal of selection had no major impact on the number of rescued pDNA vectors (Figure 5.13, 20 *TetOperator* panel, compare filled and dashed bars). 40 *TetOperator* pDNA reporter vectors showed different loss behaviour after six weeks with and without selection. Cells under selection maintained the pDNA vectors in comparable amounts to the wt-*oriP* vector, but cells without selection completely rejected the pDNA vector. Ten weeks after transfection all pDNA vectors were lost from the cells for both cultivation conditions. Control experiments with non-transfected cells revealed that the background of the experiments was on average 25 bacteria colonies (Figure 5.13, left panel). Summing up the experiment I conclude that H3.3 is not able to substitute the function of $CENH3^{CENP\text{-}A}$. In fact, the characteristics of pDNA vectors in *scTetR*:H3.3 cells are very similar to the results obtained with wt-*oriP* vectors. This indicates for a different retention mechanism compared to the *TetOperator* pDNA vectors with very likely have a functional neocentromere. Most likely this is also a passive *piggyback* mechanism, which leads to the loss of pDNA reporter vectors after approx. two months. In comparison pDNA vectors containing a functional neocentromere are stably maintained over a period of five months.

5.4.2.3 Genetic integrity of rescued pDNA vectors

To test the genetic integrity of rescued pDNA vectors, vector DNA was isolated and digested with *Pst*I. Fragment sizes of the pDNA reporter vectors are shown in Table 5.2.

Table 5.2 Fragment prediction of plasmid digest with PstI.

Reporter type	***Predicted fragment size***	***Annotation***
wt-*oriP* vector	5602	Amp resistance, *oriP*
	2146	Hygromycin resistance
	1071	prokaryotic vector backbone
20 *TetOperator* pDNA vector	5466	Amp resistance, *TetOperator*
	2146	Hygromycin resistance
	1071	prokaryotic vector backbone
40 *TetOperator* pDNA vector	5719	Amp resistance, *TetOperator*
	2146	Hygromycin resistance
	1071	prokaryotic vector backbone

The vector maps with the respective restriction sites of *Pst*I are shown below in Figure 5.14.

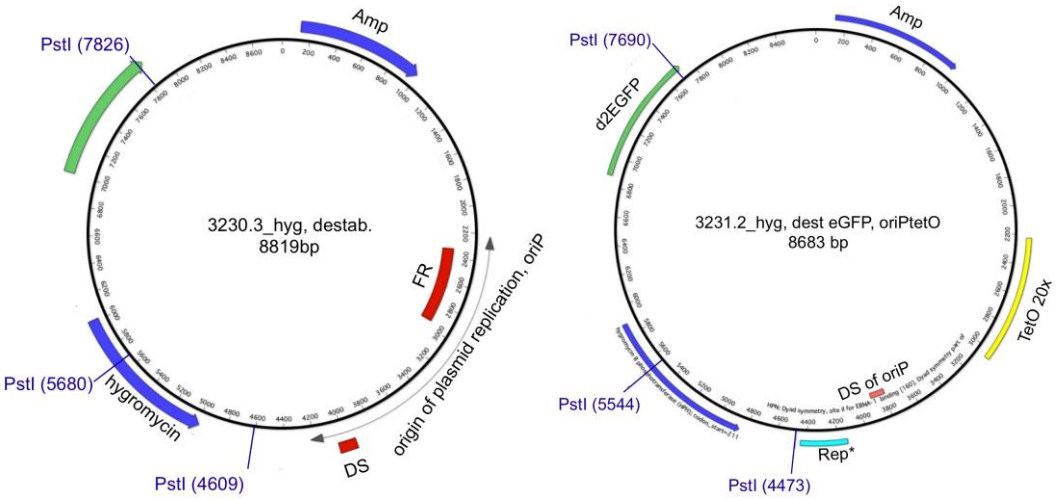


Figure 5.14 pDNA vector map of the wt-oriP vector and the 20 TetOperator vector with PstI restriction sites.

pDNA vectors (20 *TetOperator* reporter vector, right panel and wt-*oriP* vector, left panel) are both cut by *Pst*I three times, once in the vector backbone, once in the hygromycin resistance cassette and once in the eGFP sequence. The predicted fragment sizes and annotations of the pDNA reporter vectors after the digest are given in Table 5.2 and Figure 5.15 illustrates the analytical restriction digest.

After the restriction digest samples were analysed on an agarose gel. One reporter was analysed six days post transfection in the presence and absence of hygromycin (+ and - selection) and three reporters were analysed ten weeks post transfection (+ and - selection) (Figure 5.15). Restriction enzyme analysis confirmed all analysed wt-*oriP* clones as genetically stable for the entire period of the experiment (Figure 5.15 A). The input plasmids were loaded undigested (Figure 5.15, lane 1) and digested (Figure 5.15, lane 2) as reference. Among the analysed clones only two out of seven of the 20 *TetOperator* pDNA reporter vectors depicted the expected DNA fragment pattern (Figure 5.15 B, lanes 4 and 5). Furthermore, only one out of nine of the analysed 40 *TetOperator* pDNA reporter vectors displayed the predicted pattern (Figure 5.15 C, lane 7). Again sequencing analysis revealed, that these colonies were not recombined pDNA vectors but that they stem from a contamination with a foreign plasmid, possibly gained during the preparation of the sample. The respective vectors are labelled with an asterisk in Figure 5.15.

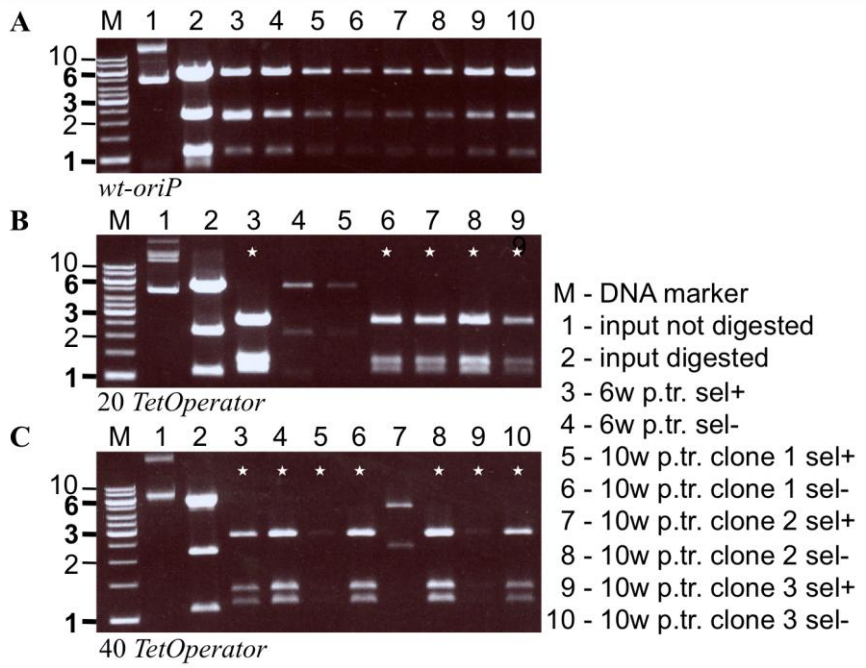


Figure 5.15 Analytical restriction digest with PstI of clones rescued from scTetR-H3.3 cells.

Single colonies of plates from plasmid rescue assay were picked, DNA was isolated and digested with *Pst*I. Analyses were performed with the wt-*oriP* vector in **A)** the 20 *TetOperator* pDNA reporter vector in **B)** and the 40 *TetOperator* pDNA reporter vector in **C)**. Predicted fragment sizes after *Pst*I digest are shown in Table 5.3. Marker bands indicate kbp sizes. A dominant fragment pattern found in several samples does not correspond to the fragment prediction, but represent the pattern of a contaminating vector. The respective lanes are marked with asterisks; w: weeks, p. tr.: post transfection, sel+: selection pressure applied, sel-: no selection pressure applied.

5.5 All-in-one pDNA vectors

To further improve our assay in terms of safety issues and to simplify its application I aimed to circumvent the need for cells stably expressing the artificial *scTetR*:CENH3^{CENP-A} protein. Instead I used an *all-in-one* pDNA vector, containing both the reporter *TetOperator* sites (20x) as well as the *scTetR*:CENH3^{CENP-A} fusion protein expression cassette. As shown before (Figure 4.10) an initial pulse of *scTetR*:CENH3^{CID} should suffice to establish the epigenetic mark on the pDNA vector for a long time in *D.melanogaster* (Mendiburo *et al.*, 2011) and supposedly this holds also true in *H.sapiens*.

For my experiments I transfected the *all-in-one* pDNA vector and the wt-*oriP* vector in HEK 293 EBNA1 cells. For comparison of the efficacy of this approach I again tested 20 *TetOperator* and 40 *TetOperator* pDNA reporter vectors in *scTetR*:CENH3^{CENP-A} high expressing cells.

The results obtained with the pDNA reporter vectors described before (20 *TetOperator*- and 40 *TetOperator* reporter and wt-*oriP* vector) and the pDNA vector retention of a

reporter containing both features, *i.e.* the *TetOperator* sites and the *scTetR*:CENH3^{CENP-A} expression cassette (*all-in-one* pDNA vector) are shown in Figure 5.16.

No difference was detectable between the two systems. In comparison to the wt-*oriP* positive control all other pDNA reporter vectors were rescued from cells at a level of approx. 75 %.

In the future the *all-in-one* pDNA vector system has to be investigated in more detail. Most importantly all rescued pDNA vectors have to be checked in restriction digest analysis as this was not yet done and the level of protein expression obtained with the *all-in-one* pDNA vector has to be checked with Western Blot analysis.

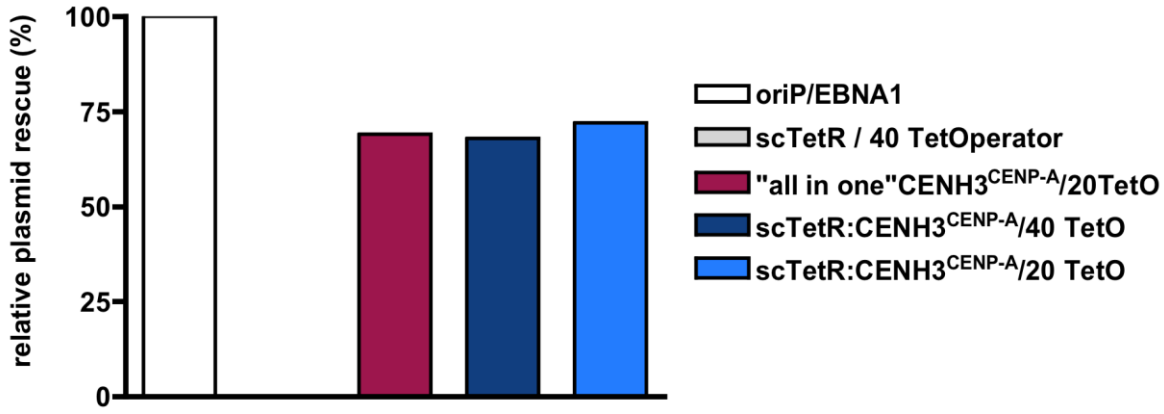


Figure 5.16 Application of the all-in-one pDNA vector leads to similar vector retention efficiencies as transfection of CENH3^{CENP-A} positive cell lines with TetOperator pDNA reporter vectors.

Plasmid rescue assays were performed as previously described with *scTetR*:CENH3^{CENP-A} EBNA1 cells transfected with 20 *TetOperator* and 40 *TetOperator* pDNA reporter vectors (blue bars) and HEK 293 EBNA1 cells transfected with the *all-in-one* reporter vector (purple bar). The wt-*oriP* vector transfected in HEK 293 EBNA1 cells served as positive control of the system (white bar), the 40 *TetOperator* vector transfected into *scTetR* cells served as negative control (grey bar). No difference was detected between the efficiencies of the *all-in-one* reporter vector and the *TetOperator* vectors transfected into stable cell lines (n = 1).

5.6 Summary – Results II

With this second set of experiments performed in human cells I confirmed the previous results obtained in the *Drosophila* system. I was able to show that targeting of CENH3^{CENP-A} to pDNA reporter vectors enhances their stability and retention in the cells most likely also by neocentromere formation. Over a period of five months I was able to rescue steady amounts of pDNA reporter vectors independently of the presence of selection pressure. This certifies the epigenetic character of the centromeric mark on the pDNA vector. It was furthermore demonstrated in this chapter, that wt-*oriP* vectors and

TetOperator reporters establish with different efficiencies in cells and I tested the functionality of CENH3$^{CENP\text{-}A}$s *CATD* domain as well as the role of H3.3. However, due to the fact, that samples prepared from the *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} and *scTetR*:H3.3 cells contained a contaminating plasmid the validity of these results is diminished. According to my results obtained with the *all-in-one* pDNA vectors this strategy seems to be the most promising for future developments. Artificial transgene expression is reduced to a minimum level and only one transfection procedure is necessary in the complete protocol.

6 DISCUSSION

The field of gene therapy is still seen as a young field of science, although people have been treated with gene therapeutic approaches for more than twenty years. Although a lot of progress has been made concerning therapies with virus-derived gene vectors, the over-all slow adoption towards a daily-routine application is mostly due to the lack of appropriate vehicles for transgenes and guaranteed stable transgene expression (Glover *et al.*, 2005). At the moment almost exclusively viral vectors are applied in clinics. Non-viral episomal pDNA vectors play a minor role although they offer several advantages compared to their virus-based counterparts. Among these advantages are *e.g.* the occurrence of less pathogenic effects, easier production, a decreased risk of insertional mutagenesis and larger transgene size capacities (Glover *et al.*, 2005). However, the main problem with non-viral vectors is, that they do not sufficiently satisfy the need for stable vector maintenance in the cell and reliable long-term transgene expression. The requirement for novel, non-viral but persistent gene vectors is getting more and more important since the currently available virus-derived vectors do not satisfactorily fulfil the needs for routine medical tools; mainly because they harbour the risk of insertional mutagenesis and strong immune responses of the patients. During my thesis I aimed to elucidate a novel retention mechanism for non-viral gene vectors with the goal to create a pDNA vector with prolonged maintenance in transfected cells.

Due to several reasons, like *e.g.* extrachromosomal maintenance and faithful segregation to daughter cells, Epstein-Barr virus is a much-valued blueprint for the development of extra-chromosomal gene vectors. However, whole EBV as well as its smaller version mini-EBV, in which most lytic genes are depleted, cannot be used as a gene vector as both of them lead to transformation of cells (Kaye *et al.*, 1999; Kempkes *et al.*, 1995). For its segregation to daughter cells EBV depends on the viral transactivator EBNA1 and the FR element of *oriP*. The viral genome is attached to the chromosome symmetrically *via* EBNA1 during mitosis and partitioned equally to daughter cells (Kanda *et al.*, 2007). Although no progeny virus is produced the number of EBV genome remains fairly stable in an infected cell clone during latency (Miller, 1982), although the number of EBV particles in a cell can vary between different clones. 10 - 400 EBV molecules were detected among different cell lines and no correlation between the number of EBV and EBNA1 molecules per cell was observed (Sternas *et al.*, 1990). Also EBV-derived *oriP* vectors are replicated together with and attached to the host cell chromosome using a passive *piggyback* mechanism (Sears *et al.*, 2004). Replication of *oriP* vectors is, in contrast to EBV, solely dependent on *oriP*'s DS element. Plasmids are replicated once each cell cycle using the endogenous replication machinery (Schepers *et al.*, 2001) but EBV-derived plasmids get lost from cells over time. This is also due to the fact that a rather high number of EBV-derived plasmids (16 %) is not duplicated properly (Nanbo *et al.*, 2007).

In this work I aimed to establish a new pDNA vector segregation mechanism for non-viral, extra-chromosomal *oriP*-derived vectors based on the principle of the *pCON* system (Figure 1.4). To this end I changed the passive distribution mechanism into an active segregation system, in which the pDNA vector gains a neocentromere by *scTetR*:CENH3 targeting, which turns the pDNA vectors into "quasi-chromosomes". These autonomous entities are incorporated into the mitotic distribution system of chromosomes. Active segregation should ideally enhance pDNA vector stability in the cell in a selection-independent manner. The main questions I addressed in my thesis are:

(1) Is it possible to prolong pDNA vector retention in a cell by turning the pDNA vector into a "quasi-chromosome" by CENH3 targeting?

(2) What are the differences concerning pDNA vector loss between active and passive pDNA vector segregation?

(3) How strong is the impact of the actively segregating pDNA vector on the host cell?

(4) What possibilities are offered by pCONactive in future pDNA gene vector development?

6.1 pCONactive – a new generation plasmid-based gene vector

The *pCON* system is a well-established gene vector system based on the origin of latent replication of Epstein-Barr virus, *oriP*. DNA replication and/or vector retention can be either provided by the two functional domains of *oriP*, namely the dyad symmetry element (DNA replication) and the family of repeats (vector retention) in an EBNA1-dependent mechanism, or they are substitutes by an endogenous, cellular protein (Pich *et al.*, 2008). Since years our lab is interested in testing endogenous cellular proteins for their ability to substitute the functions of DNA replication and/or pDNA vector retention. By replacing the wildtype elements DS and FR with *TetOperator* sites and targeting of *scTetR*-fused endogenous proteins one is able to circumvent the need for EBNA1, the viral transactivator of *oriP*. This is important, because it was described that EBNA1 has carcinogenic potential in transgenic mice (Tsimbouri *et al.*, 2002). Although Humme *et al.* have demonstrated that this does not hold true in human cells (Humme *et al.*, 2003), EBNA1 is considered to be a factor driving oncogenesis possibly by production of reactive oxygen species and therefore induced genomic instability (Gruhne *et al.*, 2009). In this respect the circumvention of EBNA1 in the *pCON* system is clearly an advantage and an ultimate goal is to get rid off of any viral component from the vector system.

Furthermore, by the replacement of EBNA1 with endogenous proteins, it is possible to investigate the functions of DNA replication and pDNA vector retention separately. For testing the replicative and/or retention function of proteins one has to create fusion proteins containing *scTetR* elements fused to the protein of choice. *Via* a strong *scTetR-TetOperator* binding the protein is targeted to *TetOperator* binding sites on the pDNA reporter vector replacing DS and/or FR. Plasmid rescue assays or Southern Blot analysis enable to determine how efficient the protein of choice can rescue the function of the EBNA1-DS and EBNA1-FR interaction. In earlier projects it was found out, that the proteins Orc6 and HMGA1a substitute the replicative potential of the DS element (Thomae *et al.*, 2008). Also the function of retention has been investigated and HMGA1a (Thomae *et al.*, 2008) as well

as HP1α were proven to be capable to confer the function of pDNA vector retention (Deutsch M., unpublished).

During this work, I did not just test another protein for the function of retention, but also changed the mechanism of retention from passive to active. The resulting *pCONactive* pDNA vector system offers for the first time a vector segregation mode that is active and host-chromosome independent. The results demonstrate that the obtainment of the centromeric structure - and therefore the segregation to the daughter cells in a host chromosomes-independent mechanism - leads to higher pDNA vector retention efficiencies, especially in long-term applications. This in turn might enable a prolonged transgene expression in gene therapeutic approaches with non-viral vectors; so far one of the most important obstacles in the field. The new system exhibits new qualities compared to other plasmid-based gene vectors, illustrated in the model depicted in Figure 6.1.

6.1.1 scTetR:CENH3^{CENP-A} and CENH3^{CID}:GFP:LacI support pDNA vector retention

In this work, I report the establishment of a new pDNA gene vector system, which segregates *via* an active mechanism. It was shown with immunofluorescence experiments, that the formation of neocentromeres is achieved by targeting of CENH3^{CID}:GFP:*LacI* to *LacOperator* sites on a pDNA vector (Figure 4.9). Based on these experiments performed in *D.melanogaster* I assume, that the same mechanism of neocentromere induction holds true for *TetOperator* pDNA vectors after a *scTetR*:CENH3^{CENP-A} protein has been targeted to the binding sites in human HEK 293 cells.

With the induction of a neocentromere on *LacOperator* and *TetOperator* (*pCON*) vectors I was able to increase the pDNA vector retention in transfected *Drosophila* Schneider S2 and human HEK293 cells, respectively. The expression of *scTetR*:CENH3^{CENP-A} in HEK 293 cells and the expression of CENH3^{CID}:GFP:*LacI* in *Drosophila* Schneider S2 cells and the subsequent formation of neocentromeres on the pDNA vectors led to a prolonged pDNA vector retention in cells. pDNA vectors with no neocentromeres were lost from the cells rapidly after transfection and subsequently the cells died due to selection pressure (Figure 4.3 and Figure 5.2).

For experiments performed in *D.melanogaster* I used a pDNA vector with *LacOperator* sites and cells expressing CENH3 fused to *LacI* and GFP. It has to be mentioned that this system is not based on the *pCON* model. Only pDNA vectors used in the human system

are *pCON* vectors and the new pDNA vectors system following the active segregation mode, are called *pCONactive* pDNA vectors throughout this work.

6.1.1.1 The pCONactive retention model

A model describing the retention profiles of *pCONactive* vectors in comparison to *wt-oriP* vectors is given in Figure 6.1. My model of *pCONactive* is based on the results I obtained in the long-term experiments with *pCONactive* pDNA vectors in human cells (Figure 5.7). In the model I am comparing the behaviour of *pCONactive* pDNA vectors (Figure 6.1, blue line) to the behaviour of wt-*oriP* vectors (Figure 6.1, grey line) as it is described in the literature, *i.e.* a rather fast lost of ≥25 % of the vector during the first two weeks after transfection (Figure 6.1, dark blue area) and a subsequently slowed down plasmid loss rate after successful establishment of the vector in the cell (Figure 6.1, bright blue area) (Leight and Sugden, 2001). Below a certain threshold level a complete vector loss cannot be prohibited (Nanbo et al., 2007). Compared to *pCONactive* pDNA vectors wt-*oriP* vectors are found with much higher copy numbers in cells and get lost over time, whereas *pCONactive* pDNA vectors establish with lower copy numbers but remain stable over a period of several months.

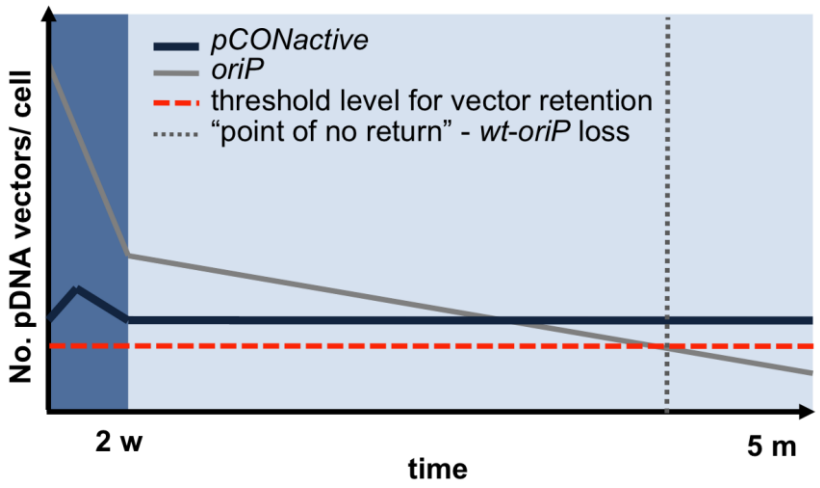


Figure 6.1 Model of retention profiles of pCONactive and oriP vectors in human cells.

pCONactive establishes similarly efficient but with lower copy numbers than wt-*oriP* vectors shortly after transfection in the target cells. From two weeks after transfection *pCONactive* vectors remain stable regarding copy numbers for a period of more than five months. In contrast, wt-*oriP* vectors are lost quickly after transfection and are lost beyond recall from cells after the number of pDNA vectors has fallen below the threshold level. The copy number of the vector has to be above that threshold level to ensure long-term pDNA vector maintenance in the cells (Nanbo *et al.*, 2007). Dark blue line: *pCONactive*, grey line: *oriP*, red line: threshold level for pDNA vector retention, dotted grey line: "point-of-no-return", dark blue: time from transfection until establishment, light blue: time after successful vector establishment, w: weeks, m: months

Nanbo *et al.* demonstrated that a certain threshold of pDNA vectors has to be maintained in a cell under selection pressure to ensure stable retention of the vector in the cell. According to this publication the majority of pDNA vectors cannot maintain the selective advantage and is subsequently lost from cells, if the number of pDNA vectors is falling below the threshold level (Nanbo *et al.*, 2007). The threshold level (Figure 6.1, red line) is one possible explanation of the differences observed between the retention of wt-*oriP* vectors and *pCONactive* pDNA vectors over a prolonged period of time. *pCONactive* might represent a pDNA vector type, which is able to keep the copy number in cells above the threshold level. Subsequently, due to its stable maintenance *pCONactive* represents a tool to overcome the vector loss problem and guarantees long-term maintenance. wt-*oriP* vectors, however, are constantly lost from cells at rates of 3 - 5 %; until a certain "point-of-no-return" is reached (dotted blue line, Figure 6.1). By reaching this "point-of-no-return" complete vector loss cannot be prohibited.

Table 6.1 illustrates the role of selection pressure on *pCONactive*. The number of rescued pDNA vectors was comparable for cells kept under selection and cells cultivated in medium without selection pressure. As pointed out before the fact that *pCONactive* is not dependent on selection criteria demonstrates the ability of *scTetR*:CENH3$^{CENP\text{-}A}$ to establish self-propagation, arguing for an artificial and inherited epigenetic centromeric mark. After an initial targeting of *scTetR*:CENH3$^{CENP\text{-}A}$ to the pDNA reporter vector the neocentromere is inherited by the daughter pDNA vector.

Table 6.1 Results of long-term plasmid rescue assays with and without selection.

Numbers indicate the amount of colonies.

pDNA vector type	***3 weeks***	***5 months + selection***	***5 months - selection***
20 *TetOperator*	128	165	193
40 *TetOperator*	106	88	140

6.1.1.2 wt-oriP vector loss in scTetR:CENH3$^{CENP\text{-}A}$ expressing cells

In my experiments the majority of wt-*oriP* vectors was lost from cells five months post transfection, no matter if they are kept under selection or not (Figure 5.7). The fact that wt-*oriP* vectors are lost from cells over time makes these results not unexpected as five months represent a rather long experimental set-up. However, as pointed out before, the pace of wt-*oriP* vector loss especially in the beginning of my experiment is unusual. Commonly wt-*oriP* vectors are lost faster from cells cultivated without selection pressure,

but in my experiment the vector loss happened during a similar time under both conditions. After a high number of wt-*oriP* vectors was rescued after three weeks under selection, this number dropped to approx. 150 and 300 rescued vectors after one month for both conditions. As stated above, the vector loss rate after establishment is usually at 3 -5 % per generation (Leight and Sugden, 2001). The experiment with wt-*oriP* transfected into *scTetR*:H3.3 cells illustrated in Figure 5.13 shows a more likely plasmid loss behaviour of this vector type. According to these results one can speculate that the overexpression of *scTetR*:CENH3^{CENP-A} might influence wt-*oriP* vector loss or that technical issues, like *e.g.* the DNA preparation, influenced the results.

6.1.1.3 Dose-dependencies in cis and trans

Experiments conducted by Thomae *et al.* showed a dose-dependency of *pCON* vectors in *cis* when testing the replicative potential of HMGA1a. This means more *TetOperator* sites on the pDNA vector led to an increased replicative potential of HMGA1a (Thomae *et al.*, 2008). However, proteins that have been tested for the support of DNA retention, like HMGA1a (Thomae *et al.*, 2008) and HP1α (data not published) did not show a dose-dependent behaviour.

Also in the case of *scTetR*:CENH3^{CENP-A} I did not detect a dose-dependency in *cis, i.e.* increased numbers of colonies in the plasmid rescue assay when using pDNA reporter vectors with 40 *TetOperator* sites instead of 20 *TetOperator* sites (Figure 5.5). More colonies would mean that more pDNA vectors were rescued from transfected cells, which in turn stands for the fact, that more pDNA vector are present in the cells.

Double the amount of *TetOperator* sites on the pDNA reporter vector can possibly lead to double the amount of *scTetR*:CENH3^{CENP-A} being recruited to the respective site on the pDNA reporter vector. One conclusion of such a scenario is, that more than one neocentromere establishes at the *TetOperator* sites on the pDNA vector. However, also on monocentric chromosomes areas of 10 - 50 kbp of centromeric DNA with incorporated CENH3 are present, building the base for only one centromere/kinetochore. It was shown that the actual size of CENH3 occupied areas at inner kinetochore plates of endogenous centromeres and neocentromeres are similar (Marshall *et al.*, 2008b) although the amount of CENH3 at neocentromeres is greatly reduced (Irvine *et al.*, 2004).

The second possibility of a dose-dependency, namely a higher expression of the s*cTetR*:CENH3^{CENP-A} fusion protein (dose-dependency in *trans*), did also not lead to differences in results of plasmid rescue assays (Figure 5.5). My results lead me to the

assumption that only a certain and constant amount of *scTetR*:CENH3^{CENP-A} is recruited to *TetOperator* binding sites, independently of how much protein is available in the cell. This means that with a certain amount of protein the system is saturated and an increase in CENH3^{CENP-A} expression does not lead to an increased amount of rescued pDNA vectors any more.

In the, presumably very rare, case of the establishment of more than one neocentromere/kinetochore per pDNA vector, I am not able to detect this phenomenon with my experimental approach. If more than one centromere/kinetochore exists on the pDNA vector this will lead to severe segregation complications during mitosis. More than one microtubule attachment site per pDNA vector leads to its complete disruption and ultimately pDNA vector loss from the cells. Eventually no increased amount of rescued pDNA vectors is detectable in the plasmid rescue assay.

6.1.1.4 DNA replication of pDNA vectors

pDNA vector replication is a bidirectional process following the theta replication principle and using mainly the replication machinery of the host (del Solar *et al.*, 1998; Gahn and Schildkraut, 1989). It is known that for achieving proper vector maintenance DNA replication is less important than vector retention (Pich *et al.*, 2008). However it is interesting to know, if the expression of CENH3^{CID}:GFP:*LacI* influences the pDNA vector replication in cells. The fact that cells with and without CENH3^{CID}:GFP:*LacI* both led to increased and comparable amounts of bacteria colonies during the first week after transfection already indicated, that pDNA vectors are able to replicate even in the absence of CENH3^{CID}:GFP:*LacI* (Figure 4.3 grey bars until day 7).

The hypothesis, that replication of the pDNA vector is CENH3 independent, was checked by performing Meselson-Stahl experiments. In this assay active DNA replication is detected by the incorporation of BrdU and the subsequent weight differences of the DNA molecule, which is detectable in a density gradient (Figure 4.8).

The data I gained indicate, that DNA replication of the pDNA vector is integrated into the cell cycle machinery of the host cell. The experiments were performed in the *Drosophila* cell system, which does not reflect the *pCONactive* experimental set-up (*scTetR*:*TetO* targeting) and a DS element for DNA replication. Instead replication of the *LacOperator* pDNA vector in these cells happens sequence independently, possibly starting in the proximity of CpG islands on the pDNA vector as it was described to happen on chromosomes (Delgado *et al.*, 1998). Once-per-cell-cycle mutual and semi-conservative

replication of genomic DNA and pDNA vectors occurs and I assume that this hold also true in the human system. The only difference is that in this case, DNA replication of *pCONactive* pDNA vectors depends on the DS element of *oriP*, which recruits the components of the cellular replication machinery (Schepers *et al.*, 2001). I conclude that for both vector types (*LacOperator*, *TetOperator*) replication is autonomous of CENH3^{CID}:GFP:*LacI* and *scTetR:*CENH3^{CENP-A} expression, respectively, in the cell.

6.1.2 Implications of pCONactive on the cellular level

pCONactive is an improved version of *pCON*, designed to serve as a new generation pDNA vector. Eventually the retention mechanism of *pCONactive* shall be utilised as a helpful tool in the design of non-viral pDNA vectors and help improve current plasmid-derived vectors by the enabling of active vector segregation. A careful consideration of cellular consequences following pDNA vector application is indispensable.

6.1.2.1 Nuclear localization of the pDNA vector

An interesting issue is the localization of *pCONactive* pDNA vectors in the cells, especially in regards to specific higher-order nuclear domains in the nucleus. Also the epigenetic status of the pDNA vector itself needs to be analysed in the future. It is known, that virus genomes prefer certain areas of the nucleus when residing in the cell during interphase. *E.g.* Epstein-Barr virus genomes in interphase are preferentially found at the open-structure perichromatin. Perichromatin constitutes the area between higher order chromatin and interchromatic regions. It was demonstrated by Immuno-FISH experiments in our lab that EBV was detected at the same sites as EBNA1, indicating that the presence of the protein is crucial for the stability of the virus.

Also for EBV-derived pDNA vectors the epigenetic localization is important, since successful establishment of the vector only happens when it is located in a specific region of the nucleus. It has been demonstrated further in our lab with immunofluorescence experiments, that EBV-derived wt-*oriP* vectors are preferentially tethered to H3K4me3[40] and H3K9ac[41] enriched sites, *i.e.* active chromatin, during interphase. The results were confirmed by ChIP experiments at wt-*oriP*, which revealed the same modification pattern at wt-*oriP* on the vector (Deutsch *et al.*, 2010). The episomal gene vector *pEPI* also locates preferentially at certain areas of the nucleus. The fact, that the episome, once it is

[40] Histone 3 lysin 4 trimethylation
[41] Histone 3 lysin 9 acetylation

established in a cell, is stably maintained also without selection pressure indicates for epigenetic influence on it retention (Piechaczek *et al.*, 1999). Indeed, it was proven that *pEPI* is found at highly transcribed and early replicating chromatin, especially enriched in H3K9ac and H3K14ac[42] (Schaarschmidt *et al.*, 2004). *pEPI* is located at the interchromatin compartment or at perichromatin close to condensed domains (Stehle *et al.*, 2007).
In the case of *oriP* vectors DS plays only a minor role regarding vector localization in the cell; FR, however, is essential for tethering to specific loci (Deutsch *et al.*, 2010). On the one hand, since FR is missing on *pCONactive* pDNA vectors it is questionable if the same pattern is found on *pCONactive* pDNA vectors. On the other hand also these pDNA vectors might use the same nuclear compartment for successful establishment in the cell, maybe through another underlying mechanism. Yet, the exact epigenetic pattern on the *pCONactive* vectors remains to be elucidated.

6.1.2.2 Transfection of foreign DNA

Clearly it also has an impact, when foreign DNA is transfected into a cell. First of all, it can happen that the pDNA particle integrate into the host's genome. The integration rate of episomal pDNA vectors is generally expected to be very low, especially when the transfected DNA is circular and not linearized. The latter conformation is disadvantageous not only because it is more susceptible to integration but also because following this event genes on these vectors are mostly silenced. Interestingly, *e.g.* integrated *pEPI* vectors were not found at active chromatin compartments like their circular and episomal counterparts (Stehle *et al.*, 2007) (chapter 6.1.2.1). It was furthermore found out, that cells that presented an integrated copy of *pEPI* usually contained only one to two copies of the vector, whereas cells containing episomal *pEPI* vectors displayed a higher copy number (Stehle *et al.*, 2007). Quite likely circular, supercoiled *pCONactive* pDNA vectors are not prone to integration into the cell's genome, but the possibility cannot be entirely excluded.
Additionally to the problem of integration, transfection of foreign, bacterial DNA leads to immunological reactions in the cell. On the one hand, the presence of unmethylated CpG pairs on the bacterial vector backbone activated the TLR9-mediated immune response (Hemmi *et al.*, 2000) (chapter 1.1.1). On the other hand, CpG pairs can cause silencing of the transgene when they are methylated. One possibility to overcome this problem is the depletion of CpG sequences in the pDNA vector backbone, as *e.g.* seen in the *pEPito* vector (Haase *et al.*, 2010) and minicircles (Kobelt *et al.*, 2013) (chapter 6.4.2).

[42] Histone 3 lysin 14 acetylation

6.1.2.3 Artificial protein expression

pCONactive and *LacI*:*LacO* mediated targeting for active pDNA vector segregation are based on two main functional components: a reporter vector with *TetOperator* and *LacOperator* sites as *cis*-acting element and the cellular expression of *scTetR*:CENH3^{CENP-A} and CENH3^{CID}:GFP:*LacI* fusion proteins, respectively, as *trans*-acting factor. However, CENH3 overexpression in cells results in severe effects on chromosome segregation as CENH3^{CID} incorporates at ectopic sites resulting in mitotic defects and chromosome aberrations (Heun *et al.*, 2006). The fact, that an initial CENH3^{CID}:GFP:*LacI* expression pulse is sufficient for neocentromere formation and it's subsequent epigenetic inheritance (Mendiburo *et al.*, 2011) offers the possibility of lower or even exclusively transient protein expression concomitant with pDNA vector application. After the optimal duration of protein expression has been determined expression of the *scTetR*:CENH3^{CENP-A} and CENH3^{CID}:GFP:*LacI* protein can be stopped, *e.g.* with the help of induced promoter silencing. Another possibility is the transient co-transfection of an expression plasmid encoding for the *scTetR* and *LacI* fusion protein, respectively, as it was shown before in Figure 4.10 (Mendiburo *et al.*, 2011).

A second protein that needs to be expressed in cells in the case of the *pCONactive* system is EBNA1. EBNA1 is essential for DNA replication provided by DS on the *pCONactive* pDNA vector, but the presence of the protein is connected to tumour development (Gruhne *et al.*, 2009). *pCON* as well as *pCONactive* are modular pDNA vector systems in which the EBNA1-dependent functions of DNA replication and retention can both be substituted by endogenous protein. The DNA replication function can be substituted *e.g.* with HMGA1a (Thomae *et al.*, 2008) whilst retention is substituted with CENH3. The only requirement for such a double-substitution is the usage of two different targeting systems for both functions. In this way the expression of EBNA1 is obsolete and DNA replication and retention of the pDNA vector are both supported by endogenous proteins. After all, the expression of artificial protein tags for tethering (*scTetR*, *LacI*) cannot be circumvented in the system.

6.1.3 Prolonged maintenance of the pDNA vector is only provided by scTetR:CENH3^{CENP-A} but not by scTetR:H3.3 and scTetR:CENH3^{CENP-A} CATD^{mut}

CENH3^{CENP-A} is a histone variant that differs from canonical histone H3.3 only in one domain – the CENP-A targeting domain (*CATD*). This domain is responsible for CENH3^{CENP-A}'s crucial function in centromere identity and needed for the proteins targeting (Black *et al.*, 2007). In my experiments I used H3.3 but also a plasmid encoding for CENH3^{CENP-A} with three point mutations at amino acids 89, 91 and 92 inside of the *CATD*. These mutations represent the sequences of H3.1/2, which do not possess the targeting function and do not ensure proper CENH3^{CENP-A} localization (Black *et al.*, 2007). This of course has an immense negative impact on the functionality of *pCONactive*. Without this functional domain, 81 % of pDNA vectors are lost after three weeks compared to the situation when the CENH3^{CENP-A} wildtype protein is expressed with the *scTetR* tag in cells (Figure 5.9). This illustrates, that the initial targeting of *scTetR*:CENH3^{CENP-A} to the *TetOperator* sites and the resulting neocentromere are really the driving forces in this process. Due to the fact that a part of the rescued pDNA vectors turned out to originate from a foreign plasmid stock, which has most likely happened to contaminate the sample during the preparation step the experiment has to be evaluated carefully. Interestingly, the contamination was only detectable in pDNA vectors rescued from mutant *scTetR*-fused CENH3^{CENP-A} cells, but not from cells expressing the *scTetR*:CENH3^{CENP-A} protein.

The results that I obtained with *scTetR*:H3.3 cells in this context were surprising. The experiments with the *CATD* mutant revealed a drastic loss of pDNA vectors from the cells. For the reason that the *CATD* is the only difference between CENH3^{CENP-A} and H3.3 I expected a similar result for the experiments I performed with the H3.3 variant, but as pointed out before in chapter 5.4.2.2 I obtained ambiguous results. After three weeks I detected no difference in plasmid rescue assays performed with the wildtype *scTetR*:CENH3^{CENP-A} and the *scTetR*:H3.3 fusion proteins, but over a prolonged period of time scTetR:H3.3 cells lost the reporter pDNA vectors completely. Also here, the contaminating plasmid was detected in the rescued clones from *scTetR*:H3.3 cells. The contamination was only detected in *TetOperator* samples but not in the samples of the wt-*oriP* vector. For the latter all rescued pDNA vectors showed the expected fragment sizes after digestion with the restriction enzyme.

6.1.4 Genetic stability of the pDNA vector

It is important to ensure that the genetic stability of the pDNA vector is maintained throughout the experimental procedure, *i.e.* during propagation in cells. To check if the rescued pDNA vectors had lost certain elements or gained *e.g.* endogenous sequences I performed restriction digest analyses on the DNA isolated form the rescued clones. I compared the obtained pattern with the input plasmid, which was transfected in the beginning of the experiment.

The genetic stability of the *LacOperator* pDNA reporter vector was provided in *D.melanogaster* cells throughout the experiment for a period of three weeks in cells with and without CENH3^{CID}:GFP:*LacI* expression (Figure 4.6). Only one minor reduction in the size of *LacOperator* cassette was detected in one out of eight analysed rescued pDNA vectors in CENH3^{CID}:GFP:*LacI* expressing cells. This difference is due to the fact that highly repetitive sequences are prone to recombination. As there are 256 *LacOperator* sites on the respective reporter vector it is not surprising that one clone has lost some of the repeats during propagation in the *Drosophila* cells. More importantly the pDNA vector backbone remained stable in all investigated clones; as this is the potential site of the pDNA vector where a transgene would be located this result is satisfying.

pDNA vectors transfected into human cells expressing *scTetR*:CENH3$^{CENP\text{-}A}$ were also not altered but displayed the same pattern as the input vector, except for a slight increase in the size of the biggest fragment (Figure 5.11, lanes 3 - 5). The results I obtained with rescued vectors from *scTetR*:CENH3$^{CENP\text{-}A}$*CATD*^{mut} and *scTetR*:H3.3 cells cannot be evaluated due to the fact that a contamination with a foreign plasmid occurred most likely during the preparation of the sample.

6.1.5 pDNA vectors segregate via microtubule attachment to the kinetochore/neocentromere and similar to chromosomes

wt-*oriP*-derived vectors are segregated *via* a passive *piggyback* mechanism attached to the host's chromosome (Sears *et al.*, 2004). As this mechanism is inefficient over a prolonged amount of time leading to a high vector loss after several weeks, we tried to establish a new and active mechanism of pDNA vector retention in the cell. To achieve this, the EBNA1-dependent mechanism of wt-*oriP* vector retention has been changed into an EBNA1-independent and active mode. I was able to achieve this goal by inducing a neocentromere on the pDNA vector, which turns the pDNA vector into a "quasi-

chromosome". An active kinetochore assembles on the neocentromere, which enables the pDNA vector integration into the cell cycle machinery, so that pDNA vectors and chromosomes use the same mechanism of segregation to daughter cells but segregate as autonomous entities during mitosis (Figure 1.1 C). Integration of the pDNA vector into the cell cycle machinery is the ideal segregation situation for pDNA reporter vectors, since this mechanism is very efficient, stable and provides long-term maintenance.

To investigate, whether the pDNA vectors are really recognised and bound by the microtubules during mitosis, and to check, if our proposed model of chromosome independent segregation is working, I transfected *Drosophila* Schneider S2 cells with *LacOperator* pDNA reporter vectors and immunofluorescence experiments were performed[43]. These experiments impressively demonstrated, that the microtubules recognize neocentromeres/kinetochores on pDNA vectors similar to the kinetochores on chromosomes during mitosis (Figure 4.9 B).

An interesting phenomenon we observed was, that pDNA vectors are able segregate prior to host cell chromosomes (Figure 4.9 C). Why chromosomes and pDNA vectors can segregate at different times during the cell cycle is not yet understood. It is possible, that pDNA vectors do not have to pass through the spindle assembly checkpoint (SAC) at the transition from metaphase to anaphase. During mitosis this checkpoint inhibits anaphase onset by prohibiting the activation of the anaphase-promoting complex (APC) through Cdc20 blocking until all chromosomes are properly attached to the spindle apparatus (Cleveland *et al.*, 2003). Possibly, pDNA vectors, which are excluded from this control mechanism, segregate independently and earlier than chromosomes. In the case, that the APC activation has not to be delayed because all chromosomes are immediately ready to segregate, differences in segregation timing between chromosomes and pDNA vectors might not be as pronounced as depicted in Figure 4.9 C. However, the lack of a proper segregation control mechanism can also mean that, even though pDNA vectors segregate like chromosomes, the long-term maintenance is less secure for pDNA vectors than for chromosomes. Additionally it can be, that microtubules reach extrachromosomal pDNA vectors earlier than chromosomes, due to the fact, that thy might not only assemble at the metaphase plate but localize also elsewhere in the cell, *e.g.* closer to the centrosome. As the APC can start just after assembly of the chromosomes at the metaphase plate and proper kinetochore binding to all chromosomes, but independently of pDNA vector

[43] All immunofluorescence experiments during this project were conducted at the MPI Freiburg in the group of Patrick Heun

assembly at the metaphase plate, this strengthens the idea that the pDNA vector is excluded from the checkpoint control. As stated above this might lower the safeness of the system, since an uncontrolled segregation mechanism can possibly lead to the loss of the pDNA vector over time due to a random segregation to daughter cells.

It was demonstrated by FISH experiments that the majority of *oriP* vectors following the *piggyback* mechanism is attached to the host's chromosomes symmetrically and vectors are found as co-localizing pairs in cells, which was detectable by a duplication of the FISH signals at these spots. After S-phase, 84 % of vectors are duplicated, from which 88 % are found as co-localised pairs in cells. These pairs of vectors are subsequently segregated symmetrically to two daughter cells. Only 16 % of EBV-derived vectors did not replicate properly and segregate randomly to only one daughter cell (Nanbo *et al.*, 2007).

It is not yet clear, if also active pDNA vector segregation (*i.e.* the pDNA vector is not attached to the host chromosome) occurs symmetrically like Figure 4.9 C suggests. Here, two signals for pDNA vectors are detected near to each of the two spindle poles. However, one has to keep in mind that this experiment represents a single cell. I hypothesise, that a symmetrical segregation of pDNA vectors in all cells is rather unlikely due to the fact, that there is no connection to the host cell chromosomes and because pDNA vectors themselves do not provide a chromosome-like structure of two equal elements (*i.e.* the chromatids), which are separated and subsequently segregated to the daughter cells symmetrically during mitosis.

pDNA vector segregation was so far only illustrated in *D.melanogaster* cells with immunofluorescence. It would be interesting so follow the pDNA vectors throughout one complete cell cycle and additionally it would be helpful to stain the centrosome of the cell to see, if the sites of pDNA vector localization really are the spindle poles. A near-term goal in this project is to determine whether or not this mechanism hold also true in human cells and to check if we can see the same effects and time course there. Experiments in this direction are currently underway.

6.2 Differences between the human and Drosophila system

For reaching my aim to investigate the ability of CENH3 to support pDNA vector retention I performed plasmid rescue assays in two different model organisms, namely *D.melanogaster* and *H.sapiens* and the bottom line of my experiments in both models is the same: targeting of CENH3 to pDNA vectors prolongs vector retention in cells.

However, I also observed some differences between the two systems, which are discussed in the following paragraphs.

When comparing results of short-term experiments over a period of one month almost nine times more bacteria colonies were formed on agar plates after electroporation with *D.melanogaster* derived samples than with samples from human HEK 293 cells seven days post transfection (3440 colonies in *D.m.* and 385 colonies in *H.s.*). During the first week post transfection pDNA vectors are still in the course of establishmen in the cell, so the more important question is, if pDNA vectors establish with different copy numbers in *D.melanogaster* compared to the human system. This question can only be solved by comparing samples from later time points.

Generally it is known, that only 1 - 5 % of episomal vectors establish efficiently in cells after transient transfection and only from these clones, cells with stable episomal vectors derive (Leight and Sugden, 2001; Stehle *et al.*, 2007). When looking at the following data, one has to remember that here the results from experiments in two different model organisms are compared and that the 256 x *LacOperator* pDNA vector does not represent a derivate of the *pCON* system like the *TetOperator pCONactive* pDNA vectors. These factors might contribute to the differences between the two systems.

Four weeks post transfection there is still a 6.6-fold difference in colony numbers between the two systems (817 colonies in *D.m.* and 123 colonies in *H.s*). The amounts of lost pDNA vectors from *Drosophila* cells were different to the amounts of lost *TetOperator* pDNA reporter vectors from HEK 293 cells. During day 7 and day 14 14.5 % of *LacOperator* pDNA vectors and 49.1 % of *TetOperator* pDNA vectors were lost from cells. Between day 14 and day 28 however, the loss of *LacOperator* pDNA vectors increased to 72.2 % (or 26.1 % per week) whereas only 37 % of *TetOperator* pDNA vectors were lost during the same time (18.6 % per week). It is important to note, that the numbers of cell generations during this time is not taken into account in the calculation of these values. This means, that these values cannot be compared to plasmid loss rates mentioned earlier in this work (*e.g.* chapter 6.1.1.1). Although these numbers suggest a better establishment of *TetOperator* pDNA vectors in human HEK 293 cells compared to *LacOperator* pDNA vectors in *Drosophila* Schneider S2 cells, one has to be careful with drawing conclusions, since this preliminary suggestion is waiting to be verified with the help of proliferation curves and generation times.

In the human cell system cells lacking the *scTetR*:CENH3$^{CENP\text{-}A}$ protein died due to the loss of the pDNA vector and therefore in response to antibiotic selection (Figure 5.2 A).

This happened much faster than in the *Drosophila* system, already after one week under selection, whereas in the latter the effect was only detected after more than two weeks under selection. The reason for this is, that the generation time of HEK 293 cells is faster and therefore also the response to the antibiotic treatment is faster than in *Drosophila* cells. It is known that the selection of *D.melanogaster* cells with G418 is a rather continuous process with a slow response of the cells (Figure 4.4).

The higher level of bacteria colonies in *Drosophila* made the background of the assay (*i.e.* colonies derived from plasmid rescue assays with not transfected cells) negligible. In HEK 293 cells the background colony number was comparable to the one obtained in *Drosophila* cells (in both cell lines around 25 colonies), however the number of rescued pDNA vectors was much lower in these cells and therefore has a stronger impact on the results. Also the fact, that pDNA vectors were rescued from *scTetR*:CENH3^{CENP-A}-lacking cells after three and four weeks was restricted to the human HEK 293 cell system (Figure 5.2 A, days 21 and 28). I think that these colonies are also background colonies and that the effect was just smaller in samples from days 7 and 14.

6.3 pCONactive all-in-one reporter vectors

The experimental set-up that I used for the confirmation that plasmids with neocentromeres remain in transfected cells for a prolonged time comprises two main functional components: a reporter plasmid with *TetOperator* and *LacOperator* sites as *cis*-acting elements, respectively, and cell lines stably expressing the *scTetR*:CENH3^{CENP-A} and CENH3^{CID}:GFP:*LacI* fusion protein, as *trans*-acting factors. Contrary to this set-up, *pCONactive all-in-one* reporter vectors encode for both segregation features on the same vector, *i.e. TetOperator* sites and the expression cassette of the *scTetR*:CENH3^{CENP-A} fusion protein.

The question was, if it is possible to establish the centromeric mark on the pDNA vector with this system. I compared *TetOperator* reporter vectors in *scTetR*:CENH3^{CENP-A} expressing cells with *pCONactive all-in-one* reporter vectors in HEK 293 EBNA1 cells in plasmid rescue assays and detected no differences between the efficiency of both systems (Figure 5.16). In fact, I was able to rescue very similar amounts of pDNA vectors from the different sets of experiments three weeks post transfection. The pDNA vector rescue level was at approx. 75 % compared to wt-*oriP* vectors rescued from the same cells. For the DNA replication function of the DS element on the *pCONactive all-in-one* vector, EBNA1

still has to be expressed in the cells, but currently a version of the pDNA vector is cloned in our lab, that additionally encodes for EBNA1. With this further development the vector provides all necessary elements for its own functionality and the tested cells obtain only the *pCONactive all-in-one* reporter and no other expression plasmids before. For example one possible development in this direction would be the construction of a vector containing the *scTetR*:CENH3 expression cassette and *TetOperator* sites replacing FR and a *LacI*:Orc6 (or HMGA1a) expression cassette and *LacOperator* sites replacing DS for DNA replication. The functions of this pDNA vector would only rely on endogenous proteins and the expression of any viral component can be avoided.

6.4 Outlook

6.4.1 Does CENH3^{CENP-A}´s chaperon HJURP confer pDNA vector retention?

In the course of this project I started to investigate the role of CENH3^{CENPA}´s chaperon HJURP[44] in regards to pDNA vector retention. HJURP is an essential chaperon for proper CENH3^{CENP-A} deposition at the centromere during late telophase/early G1 phase (Dunleavy *et al.*, 2009). It has been demonstrated, that HJURP is sufficient to establish functional *de novo* kinetochores and that a *LacI*:HJURP fusion protein targets CENH3^{CENP-A} to integrated *LacOperator* sites in the genome of U2OS cells (Barnhart *et al.*, 2011). Interestingly, the mechanism of recruitment is different for centromeric chromatin and *LacOperator* sites. The Mis18 complex is needed for the recruitment of HJURP to centromeric sites on the genome, but for HJURP-targeting to *LacOperator* site this interaction is dispensable. It was furthermore demonstrated, that CENH3^{CENP-A} is stably maintained at these sites even after the removal of *LacI*:HJURP from the system (Barnhart *et al.*, 2011). It is very likely, that HJURP fulfils the requirements for targeting endogenous CENH3^{CENP-A} to the *TetOperator* sites on the pDNA vectors and enhances pDNA vector retention in cells. If this model can be verified in the future, this will demonstrate for the first time that HJURP is sufficient for neocentromere formation on pDNA vectors.

So far I was not successful in generating a HEK 293 cell line stably expressing *scTetR*-HJURP. The reason for this might be the choice of promoter (E-Cadherin) on the *scTetR*-HJURP expression plasmid used for the transfection. I am in the process of generating an expression vector using the endogenous HJRUP promoter (1500 bp upstream of transcription start site of HJURP). This construct will then be used to generate stable HEK

[44] Holliday junction recognition protein

293 *scTetR*:HJURP cell lines as well as *all-in-one* pDNA vectors to perform plasmid rescue assays.

One has to keep in mind, that HJURP overexpression in human HeLa cells leads to mitotic defects (Mishra *et al.*, 2011). An option to reduce the overexpression of *scTetR*:HJURP protein in cells would be the transient transfection of two pDNA vectors – the *TetOperator* pDNA reporter vector and a plasmid for the transient expression of *scTetR*:HJURP. After the pDNA reporter vector has gained the centromeric mark it is stably maintained in the cell even after the loss of the *scTetR*:HJURP expression plasmid like demonstrated by Mendiburo *et al.* in *Drosophila* Schneider S2 cells (Mendiburo *et al.*, 2011). To this end I transfected the *scTetR*-HJURP expression plasmid transiently together with the pDNA reporter vector (similar to the experiment in Figure 5.9). The experiment was performed twice with quite different outcome. In the first experiment the pDNA rescue rate from *scTetR*:HJURP expressing cells was 12 % and in the second 72 % compared to *scTetR*:CENH3$^{\text{CENP-A}}$ expressing cells (data not shown). Experiments need to be repeated in *scTetR*:HJURP expressing cells, since possibly the short transient expression pulse of the fusion protein is too low for a successful establishment of the neocentromere on the pDNA vector.

6.4.2 Combining pCONactive and the minicircle technology

Many different transgene vehicles have been established over the years and each and every one of them has its advantages and disadvantages. This fact suggests itself to think about sophisticated combinations of two ore more different vector systems to increase the positive factors of gene delivery vehicles through combinatory effects. In the case of *pCONactive* a combination of the pDNA vector with the minicircle technology could improve performance and transgene expression of the system.

Minicircles are small pDNA vectors depleted of unnecessary elements for expression in human cells (Gracey Maniar *et al.*, 2013). In the introduction of this work the negative influence of CpG islands in the pDNA vector backbone on *e.g.* transgene expression and tolerance in the patient after application was discussed. The modifications of pDNA vector backbones in minicircles lead to higher transgene expression by avoiding silencing effects after CpG depletion and also immunogenic reactions of the patient are reduced. As other pDNA vectors also minicircles are usually segregated to daughter cells passively. It would

be very interesting to see, if our new technology of active segregation *via* a neocentromere is feasible on minicircles.

The production of minicircles is a labour-intensive technology, because the pDNA vectors cannot be generated in bacteria for the above-mentioned reasons. To avoid this effort and to get a faster hint if the system works a rational project to start with is the creation of a "minicircle-like" *pCONactive* pDNA vector. First this could be achieved by size-reduction, as minicircles are usually much smaller than conventional plasmid vectors. In the case of *pCONactive* a size reduction from 8683 bp for the 20 *TetOperator* reporter and 8936 bp for the 40 *TetOperator* pDNA reporter vector, respectively, can easily be done by further depletion of *TetOperator* sites. As shown in the Results II section of this work (chapter 5.2.1) I did not observe any dose-dependency in *cis* comparing 40 *TetOperator* sites and 20 *TetOperator* sites on the reporter. A further reduction of *TetOperator* binding sites is self-evident and would reduce the pDNA vector size for 42 bp per binding site. Also I observed no diminished efficiency of the system with recombined pDNA vectors that have lost *TetOperator* sites.

Second, a more "minicircle-like" status of the *pCONactive* pDNA vector system is achieved by reduction of the number of CpG pairs in the backbone of the pDNA reporter vector. This service is offered by several companies and would further decrease to size and in addition lead to better transgene expression profiles and to a smaller immunogenic reaction potential.

A comparison of this new "minicircle-like" pDNA vector to our current pDNA reporter vectors and further experiments especially in regards to transgene expression levels in different cell systems should provide an indication of the efficiency of the new system and show us, whether or not this strategy is worth to be followed up. Eventually minicircles with neocentromeres should represent a new group of pDNA gene vectors with a host chromosome independent retention mechanism, stable, long-term transgene expression, extrachromosomal maintenance and a good acceptance of the pDNA vector in the target cell.

7 CONCLUSION

EBV-derived wt-*oriP* vectors are segregated to daughter cells *via* a passive *piggyback* mechanism. wt-*oriP* vectors are targeted to the host cell chromosomes with the help of the viral transactivator EBNA1. At the chromosomes wt-*oriP* vectors remain attached and they replicate and segregate in synchrony with the host cell chromosomes during S-phase and mitosis. However, this system is leaky and during each cell division wt-*oriP* vectors are lost from cells until, after a certain number of cell generations, the vectors are not detectable in the host cells any more.

The aim of this work was to develop an entirely new pDNA vector segregation mechanism with the help of the centromere specific histone variant CENH3 (*pCONactive*). The main difference between this new model and the wildtype segregation mechanism of *oriP* vectors is, that segregation of the *pCONactive* pDNA vector is an active mechanism and independent of host chromosome segregation. Segregation of *pCONactive* happens similar to the mechanism of chromosome segregation mediated *via* a centromere-microtubule interaction during mitosis. The approach also comprises a long-term goal, namely the improvement of non-viral, plasmid-based gene vectors for gene therapeutic applications.

By targeting CENH3^{CID}:GFP:*LacI* to *LacOperator* sites and *scTetR*:CENH3^{CENP-A} to *TetOperator* sites, respectively, I aimed to induce the formation of a neocentromere on plasmid DNA. The feasibility of this mechanism was proven with immunofluorescence experiments in *D.melanogaster* and is supposed to work in a similar manner also in human

cells. The induced neocentromere on the reporter DNA turns the pDNA vector into a "quasi-chromosome" and makes it host chromosome-independent during segregation to the daughter cells in mitosis. With this mechanism I am able to increase pDNA vector stability in the cell and provide a prolonged retention of episomal pDNA vectors in target cells in comparison to passively retained pDNA vectors.

I demonstrated in *D.melanogaster* and *H.sapiens* that induced neocentromerization leads to enhanced pDNA vector retention and maintenance profiles in cells. In the human system the mechanism was followed over a period of five months and I was able to demonstrate that the number of rescued pDNA vectors is stable during this time course with and without selection pressure. In contrast to wt-*oriP* vectors, *pCONactive* pDNA vectors display much lower copy numbers in transfected cells during the first weeks after transfection. However, over an elongated time span it turns out that wt-*oriP* vectors are lost steadily, whereas *pCONactive* reporters are evenly maintained and therefore also guarantee prolonged transgene expression. Further characterization of the system revealed that CENH3 does not interfere with plasmid DNA replication, that the pDNA vectors are maintained as concatemeric entities in the cells and that the genetic integrity of the pDNA vector is stable in *scTetR*:CENH3^{CENP-A} and CENH3^{CID}:GFP:*LacI* positive cells. Furthermore I could prove the essential role of the *CATD* domain of CENH3^{CENP-A} for maintaining the functionality of the system. Relying on the results I obtained with *all-in-one* pDNA vectors, I suggest the application of this vector type in future applications, as there was no reduction in plasmid rescue efficiencies observed.

I am able to state that the *pCONactive (all-in-one)* pDNA vector system has the potential to highly impact in the field of non-viral, plasmid-based gene therapy. In this respect I expect the most promising future development in a combination of the CENH3-dependent segregation mechanism with the minicircle technology.

8 REFERENCES

Aiyar, A., Tyree, C., and Sugden, B. (1998). The plasmid replicon of EBV consists of multiple cis-acting elements that facilitate DNA synthesis by the cell and a viral maintenance element. EMBO J *17*, 6394-6403.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet *9*, 923-937.

Alonso, A., Hasson, D., Cheung, F., and Warburton, P.E. (2010). A paucity of heterochromatin at functional human neocentromeres. Epigenetics and Chromatin *3*, doi:10.1186/1756-8935-1183-1186.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. PNAS *92*, 11170-11174.

Argyros, O., Wong, S.P., Fedonidis, C., Tolmachov, O., Waddington, S.N., Howe, S.J., Niceta, M., Coutelle, C., and P., H.R. (2011). Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver. J Mol Med *89*, 515-529.

Baiker, A., Maercker, C., Piechaczek, C., Schmidt, S., Bode, J., Benham, C., and Lipps, H. (2000). Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. Nat Cell Biol *3*, 182-184.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229-243.

Baum, C., Schambach, A., Bohne, J., and Galla, M. (2006). Retrovirus vectors: toward the plentivirus? Mol Ther *13*, 1050-1063.

Bergmann, J.H., Martins, N.M., Larionov, V., Masumoto, H., and Earnshaw, W.C. (2012). HACking the centromere chromatin code: insights from human artificial chromosomes. Chromosome Res *20*, 505-519.

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods Jr., V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. Nature *430*, 578-582.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell *25*, 309-322.

Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved organisation of centromeric chromatin in flies and humans. Dev Cell *2*, 319-330.

Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. (1992). Biological significance of unwinding capability of nuclear matrix-associating DNAs. Science *255*, 195-197.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.

Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., and Henikoff, S. (1999). A histone-H3-like protein in C. elegans. Nature *401*, 547-548.

Burrack, L.S., and Berman, J. (2012). Neocentromeres and epigenetically inherited features of centromeres. Chromosome Res *20*, 607-619.

Calos, M.P. (1998). Stability without a centromere. PNAS *95*, 4084-4085.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. PNAS *109*, 1979-1984.

Chan, F.L., and Wong, L.H. (2012). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res *40*, 11178-11188.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. Nat Rev Mol Cell Bio *9*, 33-46.

Chen, Z.Y., He, C.Y., Ehrhardt, A., and Kay, M.A. (2003). Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. Mol Ther *8*, 495-500.

Choo, K.H., Vissel, B., Nagy, A., Earle, E., and Kalitsis, P. (1991). A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res *19*, 1179-1182.

Church, G., and Gilbert, W. (1984). Genomic sequencing. PNAS *81*, 1991-1995.

Cim, A., Sawyer, G.J., Zhang, X., Su, H., Collins, L., Jones, P., Antoniou, M., Reynes, J.P., Lipps, H.J., and Fabre, J.W. (2012). In vivo studies on non-viral transdifferentiation of liver cells towards pancreatic β cells. J Endocrin *214*, 277-288.

Clarke, L., Baum, M., Marschall, L., Ngan, V., and Steiner, N. (1993). Structure and function of Schizosaccharomyces pombe centromeres. Cold Spring Harb Symp Quant Biol *58*, 687-695.

Cleveland, D.W., Mao, Y., and F., S.K. (2003). Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling. Cell *112*, 407-421.

Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M.I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., *et al.* (1999). Genetic definition and sequence analysis of Arabidopsis centromeres. Science *286*, 2468-2474.

Cottarel, G., Shero, J.H., Hieter, P., and Hegemann, J.H. (1989). A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in Saccharomyces cerevisiae. Mol Cell Biol *9*, 3342-3349.

Daya, S., and Berns, K.I. (2008). Gene therapy using adeno-associated virus vectors. Clin Microbiol Rev *21*, 583-593.

De Rop, V., Padeganeh, A., and Maddox, P.S. (2012). CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. Chromosoma *121*, 527-538.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998). Replication and Control of Circular Bacterial Plasmids. Microbiol Mol Biol Rev *62*, 434-464.

Delecluse, H.J., and Hammerschmidt, W. (2000). The genetic approach to the Epstein-Barr virus: from basic virology to gene therapy. Mol Path *53*, 270-279.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J *17*, 2426-2435.

Deutsch, M.J., Ott, E., Papior, P., and Schepers, A. (2010). The latent origin of replication of Epstein-Barr virus directs viral genomes to active regions of the nucleus. J Virol *84*, 2533-2546.

Diehl, V., Henle, G., Henle, W., and Kohn, G. (1968). Demonstration of a Herpes Group Virus in Cultures of Perioheral Leukocyctes from Patients with Infectious Mononucleosis. J Virol *2*, 663-669.

Dunleavy, E.M., Almouzni, G., and Karpen, G.H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase. Nucleus *2*, 146-157.

Dunleavy, E.M., Beier, N.L., Gorgescu, W., Tang, J., Costes, S.V., and Karpen, G.H. (2012). The cell cycle timing of centromeric chromatin assembly in Drosophila meiosis is distinct from mitosis yet requires CAL1 and CENP-C. PLoS Biol *10*, e1001460.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485-497.

Earnshaw, W.C., Allshire, R.C., Black, B.E., Bloom, K., Brinkley, B.R., Brown, W., Cheeseman, I.M., Choo, K.H., Copenhaver, G.P., Deluca, J.G., *et al.* (2013). Esperanto for histones: CENP-A, not CenH3, is the centromeric histone H3 variant. Chromosome Res *21*, 101-106.

Earnshaw, W.C., and Migeon, B.R. (1985). Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. Chromosoma *92*, 290-296.

Edwards, N.S., and Murray, A.W. (2005). Identification of xenopus CENP-A and an associated centromeric DNA repeat. Mol Biol Cell *16*, 1800-1810.

Ehrhardt, A., R., H., Schepers, A., Deutsch, M.J., Lipps, H.J., and A., B. (2008). Episomal Vectors for Gene Therapy. Curr Gene Ther *8*, 147-161.

Elsaesser, S.J., Goldberg, A.D., and Allis, C.D. (2010). New functions for an old variant: no substitute for histone H3.3. Curr Opin Genet Dev *20*, 110-117.

Escors, D., and Breckpot, K. (2010). Lentiviral vectors in gene therapy: their current status and future potential. Arch Immunol Ther Exp (Warsz) *58*, 107-119.

Gahn, T.A., and Schildkraut, C.L. (1989). The Epstein-Barr virus origin of plasmid replication, oriP, contains both the initiation and termination sites of DNA replication. Cell *58*, 527-535.

Gerhardt, J., Jafar, S., Spindler, M.P., Ott, E., and Schepers, A. (2006). Identification of new human origins of DNA replication by an origin-trapping assay. Mol Cell Biol *26*, 7731-7746.

Gill, D.R., Pringle, I.A., and Hyde, S.C. (2009). Progress and prospects: the design and production of plasmid vectors. Gene Ther *16*, 165-171.

Glover, D.J., Lipps, H.J., and Jans, D.A. (2005). Towards safe, non-viral therapeutic gene expression in humans. Nat Rev Genet *6*, 299-310.

Gotlieb-Stematsky, T., Vonsover, A., Ramot, B., Zaizov, R., Nordan, U., Aghal, E., Kende, G., and Modan, M. (1975). Antibodies to Epstein-Barr virus in patients with Hodgkin´s disease and leukemia. Cancer *36*, 1640-1645.

Gracey Maniar, L.E., Maniar, J.M., Chen, Z.Y., Lu, J., Fire, A.Z., and Kay, M.A. (2013). Minicircle DNA vectors achieve sustained expression reflected by active chromatin and transcriptional level. Mol Ther *21*, 131-138.

Gruhne, B., Sompallae, R., Marescotti, D., Kamranvar, S.A., Gastaldello, S., and Masucci, M.G. (2009). The Epstein-Barr virus nuclear antigen-1 promotes genomic instability via induction of reactive oxygen species. PNAS *106*, 2313-2318.

Haase, R., Argyros, O., Wong, S.P., Harbottle, R.P., Lipps, H.J., Ogris, M., Magnusson, T., Vizoso Pinto, M.G., Haas, J., and Baiker, A. (2010). pEPito: a significantly improved non-viral episomal expression vector for mammalian cells. BMC Biotechnol *10*, doi: 10.1186/1472-6750-1110-1120.

Haase, R., Magnusson, T., Su, B., Kopp, F., Wagner, E., Lipps, H., Baiker, A., and Ogris, M. (2013). Generation of a tumor- and tissue-specific episomal non-viral vector system. BMC Biotechnol *13*, 49.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol *166*, 557-580.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., *et al.* (2000). A Toll-like receptor recognizes bacterial DNA. Nature *408*, 740-745.

Henikoff, S., Ahmad, K., Platero, J.S., and B., v.S. (2000). Heterochromatic deposition of centromeric histone H3-like proteins. PNAS *97*, 716-721.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell *10*, 303-315.

Hirt, B. (1966). Evidence for semiconservative replication of circular polyoma DNA. PNAS *55*, 997-1004.

Houben, A., and Schubert, I. (2003). DNA and proteins of plant centromeres. Curr Opin Plant Biol *6*, 554-560.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H.A. (2000). Early disruption of centromeric chromatin organisation in centromere protein A (Cenpa) null mice. PNAS *97*, 1148-1153.

Humme, S., Reisbach, G., Feederle, R., Delecluse, H.J., Bousset, K., Hammerschmidt, W., and Schepers, A. (2003). The EBV nuclear antigen 1 (EBNA1) enhances B cell immortalization several thousandfold. PNAS *100*, 10989-10994.

Hung, S.C., Kang, M.S., and Kieff, E. (2001). Maintenance of Epstein-Barr virus (EBV) oriP-based episomes requires EBV-encoded nuclear antigen-1 chromosome-binding domains, which can be replaced by high-mobility group-I or histone H1. PNAS *98*, 1865-1870.

Irvine, D.V., Amor, D.J., Perry, J., Sirvent, N., Pedeutour, F., Choo, K.H., and Saffery, R. (2004). Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. Chromosome Res *12*, 508-815.

Jansen, L.E., Black, B.E., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol *176*, 795-805.

Jenke, B.H.C., Fetzer, C.P., Stehle, I.M., Jönsson, F., Fackelmayer, F.O., Conradt, H., Bode, J., and Lipps, H.J. (2002). An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. EMBO Reports *3*, 349-354.

Kanda, T., Kamiya, M., Maruo, S., Iwakiri, D., and Takada, K. (2007). Symmetrical localization of extrachromosomally replicating viral genomes on sister chromatids. J Cell Sci *120*, 1529-1539.

Kaye, K.M., Izumi, K.M., Li, H., Johannsen, E., Davidson, D., Longnecker, R., and Kieff, E. (1999). An Epstein-Barr Virus That Expresses Only the First 231 LMP1 Amino Acids Efficiently Initiates Primary B-Lymphocye Growth Transformation. J Virol *73*, 10525-10530.

Kazuki, Y., Hoshiya, H., Takiguchi, M., Abe, S., Iida, Y., Osaki, M., Katoh, M., Hiratsuka, M., Shirayoshi, Y., Hiramatsu, K., *et al.* (2011). Refined human artificial chromosome vectors for gene therapy and animal transgenesis. Gene Ther *18*, 384-393.

Kempkes, B., Pich, D., Zeidler, R., and Hammerschmidt, W. (1995). Immortalization of human primary B lymphocytes in vitro with DNA. PNAS *92*, 5875-5879.

Kim, J.H., Kononenkoa, A., Erliandria, I., Kimb, T.A., Nakanoc, M., Iidad, Y., Barrette, J.C., Oshimurad, M., Masumotoc, H., Earnshawf, W.C., *et al.* (2011). Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. PNAS *108*, 20048-20053.

Kobelt, D., Schleef, M., Schmeer, M., Aumann, J., Schlag, P.M., and Walther, W. (2013). Performance of High Quality Minicircle DNA for In Vitro and In Vivo Gene Transfer. Mol Biotechnol *53*, 80-89.

Krysan, P., Haase, S.B., and Calos, M.P. (1989). Isolation of Human Sequences That Replicate Autonomously in Human Cells. Mol Cell Biol *9*, 1026-1033.

Kurumizaka, H., Horikoshi, N., Tachiwana, H., & Kagawa, W. (2012). Current progress on structural studies of nucleosomes containing histone H3 variants. Curr Opin Struct Biol *23*, 109-115.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Larin Monaco, Z., and Moralli, D. (2006). Progress in artificial chromosome technology. Biochem Society Transactions *34*, 324-327.

Leight, E.R., and Sugden, B. (2001). Establishment of an oriP replicon is dependent upon an infrequent, epigenetic event. Mol Cell Biol *21*, 4149-4161.

Lipps, H.J., Jenke, A.C.W., Nehlsen, K., Scintele, M.F., Stehle, I.M., and Bode, J. (2003). Chromosome-based vectors for gene therapy. Gene *304*, 23-33.

Lufino, M.M., Edser, P.A., and Wade-Martins, R. (2008). Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. Mol Ther *16*, 1525-1538.

Lufino, M.M., Manservigi, R., and Wade-Martins, R. (2007). An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency. Nucleic Acids Res *35*, e98.

Maggert, K.A., and Karpen, G.H. (2001). The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere. Genetics *158*, 1615-1628.

Malik, H.S., and Henikoff, S. (2000). Adaptive Evolution of Cid, a Centromere-Specific Histone in Drosophila. Genetics *157*, 1293-1298.

Manzini, S., Vargiolu, A., Stehle, I.M., Bacci, M.L., Cerrito, M.G., Giovannoni, R., Zannoni, A., Bianco, M.R., Forni, M., Donini, P., *et al.* (2006). Genetically modified pigs produced with a nonviral episomal vector. PNAS *103*, 17672-17677.

Maresca, T.J. (2011). Chromosome Segregation: A Kinetochore Missing Link Is Found. Curr Biol *21*, 261-263.

Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H. (2008a). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet *82*, 261-282.

Marshall, O.J., Marshall, A.T., and Choo, K.H. (2008b). Three-dimensional localization of CENP-A suggests a complex higher order structure of centromeric chromatin. J Cell Biol *183*, 1193-1202.

Martin Dani, G., and Zakian, V.A. (1983). Mitotic and meiotic stability of linear plasmids in yeast. PNAS *80*, 3406-3410.

Mayerhofer, P., Schleef, M., and Jechlinger, W. (2009). Use of Minicircle Plasmids for Gene Therapy. Gene Ther Cancer *542*, 87-104.

Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I., and Karpen, G.H. (2011). Assembly of Drosophila centromeric chromatin proteins during mitosis. PLoS Genet *7*, e1002068.

Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 is sufficient for centromere formation. Science *334*, 686-690.

Middleton, T., and Sugden, B. (1992). EBNA1 can link the enhancer element to the initiator element of the Epstein-Barr virus plasmid origin of DNA replication. J Virol *66*, 489-495.

Miller, G. (1982). Immortalization of Human Lymphocytes by Epstein-Barr Virus. Yale J Biol Med *55*, 305-310.

Mishra, P.K., Au, W.C., Choy, J.S., Kuich, P.H., Baker, R.E., Foltz, D.R., and Basrai, M.A. (2011). Misregulation of Scm3p/HJURP causes chromosome instability in Saccharomyces cerevisiae and human cells. PLoS Genet *7*, e1002303.

Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., DeKelver, R.C., Gregory, P.D., Urnov, F.D., and Holmes, M.C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. PNAS *104*, 3055-3060.

Morissette, G., and Flamand, L. (2010). Herpesviruses and chromosomal integration. J Virol *84*, 12100-12109.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51*, 263-273.

Murata, M. (2013). Arabidopsis Centromeres. (Wiley-Blackwell).

Nakano, M., Cardinale, S., Noskov, V.N., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., Larionov, V., Earnshwas, W.C., and Masumoto, H. (2008). Inactivation of a Human Kinetochore by Specific Targeting of Chromatin Modifiers. Dev Cell *14*, 407-522.

Nanbo, A., Sugden, A., and Sugden, B. (2007). The coupling of synthesis and partitioning of EBV´s plasmid replicon is revealed in live cells. EMBO J *26*, 4252-4262.

Nehlsen, K., Broll, S., and J., B. (2006). Replicating minicircles: Generation of nonvirol episomes for the efficient midification of dividing cells. Gene Ther Mol Biol *10*, 233-244.

Nishikawa, M., and Huang, L. (2001a). Nonviral vectors in the New Millenium: Delivery Barriers in Gene Transfer. Hum Gene Ther *12*, 861-870.

Nishikawa, M., and Huang, L. (2001b). Nonviral Vectors in the New Millenium: Delivery Barriers in Gene Transfer. Hum Gene Ther *12*, 861-870.

Nonoyama, M., Huang, C.H., Pagano, J.S., Klein, G., and Singh, S. (1973). DNA of Epstein-Barr Virus Detected in Tissue of Burkitt´s Lymphoma and Nasopharyngeal Carcinoma. PNAS *70*, 3265-3268.

Okada, M., Okawa, K., Isobe, T., and Fukagawa, T. (2009). CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. Mol Biol Cell *20*, 3986-3995.

Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., Saccani, S., and Heun, P. (2011). Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nat Cell Biol *13*, 799-808.

Palmer, D.K., O´Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. PNAS *88*, 3734-3738.

Pich, D., Humme, S., Spindler, M.P., Schepers, A., and Hammerschmidt, W. (2008). Conditional gene vectors regulated in cis. Nucleic Acids Res *36*, e83.

Piechaczek, C., Fetzer, C., Baiker, A., Bode, J., and Lipps, H.J. (1999). A vector based on the SV40 origin of replication and chromosomal S/MARs replicate episomally in CHO cells. Nucleic Acids Res *27*, 426-428.

Prud'homme, G.J., Draghia-Akli, R., and Wang, Q. (2007). Plasmid-based gene therapy of diabetes mellitus. Gene Ther *14*, 553-564.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M., and Glover, D.M. (2011). CENP-C is a structural platform for kinetochore assembly. Curr Biol *21*, 399-405.

Rattner, J.B. (1991). The structure of the mammalian centromere. Bioessays *13*, 51-56.

Ribeiro, S.C., Monteiro, G.A., and Prazeres, D.M.F. (2008). Evaluation of the effect of non-B DNA structures on plasmid integrity via accelerated stability studies. J Pharm Sci *98*, 1400-1408.

Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O., and Stanyon, R. (2012). Centromere repositioning in mammals. Heredity *108*, 59-67.

Saitoh, Y., and Laemmli, U.K. (1994). Metaphase chromosome structure: Bands arise from a differential folding path of the highly AT-rich scaffold. Cell *76*, 609-622.

Sambrook, J., and Russell, D.W. (2001). Molecular cloning : a laboratory manual., 3rd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Santaguida, S., and Musacchio, A. (2009). The life and miracles of kinetochores. EMBO J *28*, 2511-2531.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nat Rev Mol Cell Biol *6*, 139-149.

Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. EMBO J *23*, 191-201.

Schakowski, F., Gorschlüter, M., Junghans, C., Schroff, M., Buttgereit, P., Ziske, C., Schöttker, B., König-Merediz, S.A., Sauerbruch, T., Wittig, B., *et al.* (2001). A Novel Minimal-Size Vector (MIDGE) Improves Transgene Expression in Colon Carcinoma Cells and Avoids Transfection of Undesired DNA. Mol Ther *3*, 793-800.

Schepers, A., Ritzi, M., Bousset, K., Kremmer, E., Yates, J.L., Harwood, J., Diffley, J.F.X., and Hammerschmidt, W. (2001). Human origin recognition comples binds to the region of the latent origin of DNA replication of Epstein-Barr virus. EMBO J *20*, 4588-4602.

Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol *17*, 237-243.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E., and Musacchio, A. (2011). Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. Curr Biol *21*, 391-398.

Sears, J., Kolman, J., Wahl, G.M., and Aiyar, A. (2003). Metaphase Chromosome Tethering Is Necessary for the DNA Synthesis and Maintenance of oriP Plasmids but Is Insufficient for Transcription Activation by Epstein-Barr Nuclear Antigen 1. J Virol *77*, 11767-11780.

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J., and Aiyar, A. (2004). The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. J Virol *78*, 11487-11505.

Shen, M.H., Ross, A., Yang, J., de las Heras, J.I., and Cooke, H. (2001). Neo-centromere formation on a 2.6 Mb mini-chromosome in DT40 cells. Chromosoma *110*, 421-429.

Shibata, D., and Weiss, L.M. (1992). Epstein-Barr Virus-associated Gastric Adenocarcinoma. Am J Pathol *140*, 769-774.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophpresis. J Mol Biol *98*, 503-517.

Stehle, I.M., Postberg, J., Rupprecht, S., Cremer, T., Jackson, D.A., and Lipps, H.J. (2007). Establishment and mitotic stability of an extra-chromosomal mammalian replicon. BMC Cell Biol *8*, 33.

Stehle, I.M., Scinteie, M.F., Baiker, A., Jenke, A.C., and Lipps, H.J. (2003). Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. Chromosome Res *11*, 413-421.

Stellfox, M.E., Bailey, A.O., and Foltz, D.R. (2012). Putting CENP-A in its place. Cell Mol Life Sci *70*, 387-406.

Sternas, L., Middleton, T., and Sugden, B. (1990). The average number of molecules of Epstein-Barr nuclear antigen 1 per cell does not correlate with the average number of Epstein-Barr virus (EBV) DNA molecules per cell among different clones of EBV-immortalized cells. J Virol *64*, 2407-2410.

Stimpson, K.M., and Sullivan, B.A. (2010). Epigenomics of centromere assembly and function. Curr Opin Cell Biol *22*, 772-780.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev *9*, 573-586.

Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol *11*, 1076-1083.

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581-592.

Sun, X., Wahlstrom, J., and Karpen, G. (1997). Molecular Structure of a Functional Drosophila Centromere. Cell *91*, 1007-1019.

Talbert, P.B., Ahmad, K., Almouzni, G., Ausió, J., Berger, F., Bhalla, P.L., Bonner, W.M., Cande, W.Z., Chadwick, B.P., Chan, S.W., *et al.* (2012). A unified phylogeny-based nomenclature for histone variants. Epigenet & Chromatin *5*, doi: 10.1186/1756-8935-1185-1187.

Tamura, T., Smith, M., Kanno, T., Dasenbrock, H., Nishiyama, A., and Ozato, K. (2009). Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. J Biol Chem *284*, 12217-12225.

Telenius, H., Szeles, A., Keresö, J., Csonka, E., Praznovszky, T., Imreh, S., Maxwell, A., Perez, C.F., Drayer, J.I., and Hadlaczky, G. (1999). Stability of a functional murine satellite DNA-based artificial chromosome across mammalian species. Chromosome Res *7*, 3-7.

Teo, C.H., Lermontova, I., Houben, A., Mette, M.F., and Schubert, I. (2013). De novo generation of plant centromeres at tandem repeats. Chromosoma *122*, 233-241.

Thomae, A.W., Baltin, J., Pich, D., Deutsch, M.J., Ravasz, M., Zeller, K., Gossen, M., Hammerschmidt, W., and Schepers, A. (2011). Different roles of the human Orc6 protein in the replication initiation process. Cell Mol Life Sci *68*, 3741-3756.

Thomae, A.W., Pich, D., Brocher, J., Spindler, M.P., Berens, C., Hock, R., Hammerschmidt, W., and Schepers, A. (2008). Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. PNAS *105*, 1692-1697.

Thomas, C.E., Ehrhardt, A., and Kay, M.A. (2003). Progress and problems with the use of viral vectors for gene therapy. Nat Rev Genet *4*, 346-358.

Torras-Llort, M., Moreno-Moreno, O., and Azorin, F. (2009). Focus on the centre: the role of chromatin on the regulation of centromere identity and function. EMBO J *28*, 2337-2348.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. PNAS *76*, 4350-4354.

Tsimbouri, P., Drotar, M.E., Coy, J.L., and Wilson, J.B. (2002). bcl-xL and RAG genes are induced and the response to IL-2 enhanced in EmuEBNA-1 transgenic mouse lymphocytes. Oncogene *21*, 5182-5187.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H.A. (1993). A Functional Marker Centromere with No Detectable Alpha-Satellite, Satellite III, or CENP-B Protein: Activation of a Latent Centromere? Am J Hum Genet *52*, 1153-1163.

Waehler, R., Russell, S.J., and Curial, D.T. (2007). Engineering targeted viral vectors for gene therapy. Nat Rev Genet *8*, 573-587.

Warburton, P.E. (2004). Chromosomal dynamics of human neocentromere formation. Chromosome Res *12*, 617-626.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., *et al.* (1997). Immunolocalization of CENP-A suggest a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol *7*, 901-904.

Waye, J.S., and Willard, H.F. (1987). Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucleic Acids Res *15*, 7549-7569.

Welburn, J.P., and Cheeseman, I.M. (2008). Toward a molecular structure of the eukaryotic kinetochore. Dev Cell *15*, 645-655.

White, R.E., Wade-Martins, R., and James, M.R. (2002). Infectious Delivery of 120-Kilobase Genomic DNA by an Epstein-Barr Virus Amplicon Vector. Mol Ther *5*, 427-435.

Wiens, G.R., and Sorger, P.K. (1998). Centromeric chromatin and epigenetic effects in kinetochore assembly. Cell *93*, 313-316.

Wong, S.P., Argyros, O., Coutelle, C., and Harbottle, R.P. (2011). Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage. Gene Ther *18*, 82-87.

Wutzler, P., Färber, I., Sauerbrei, A., Helbig, B., Wutke, K., Rüdiger, K.D., Scheibner, K., Brichàcek, B., and Vonka, V. (1986). Demonstration of Epstein-Barr virus in malignant non-Hodgkin´s lymphomas. Oncology *43*, 224-229.

Wysokensko, D.A., and Yates, J.L. (1989). Multiple EBNA1-Binding Sites Are Required To Form an EBNA1-Dependent Enhancer and To Activate a Minimal Replicative Origin with oriP of Epstein-Barr Virus. J Virol *63*, 2657-2666.

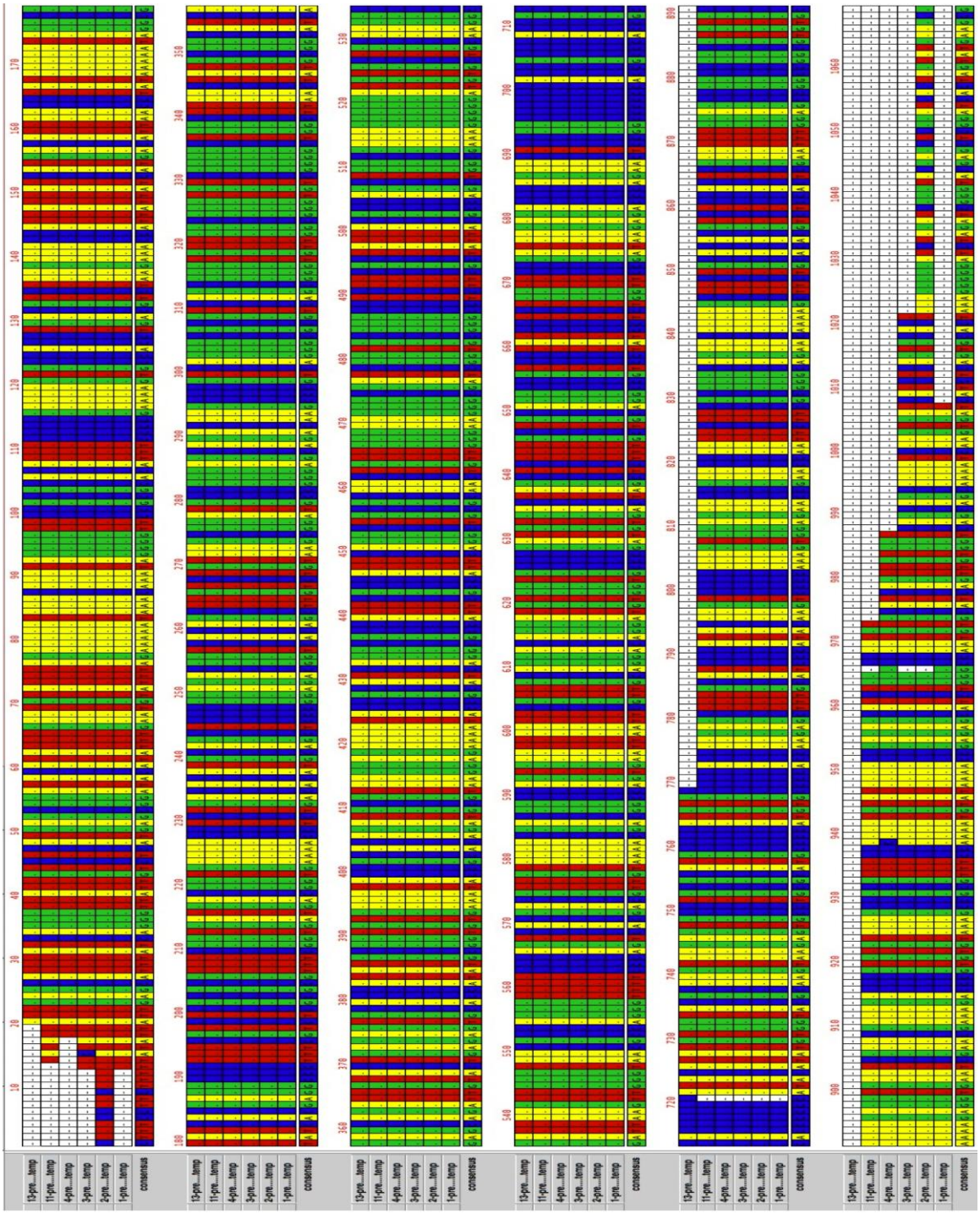
Yates, J.L., Camiolo, S.M., and Bashaw, J.M. (2000). The Minimal Replicator of Epstein-Barr Virus oriP. J Virol *74*, 4512-4522.

Zeitlin, S.G., Baker, N.M., Chapados, B.R., Soutoglou, E., Wang, J.Y., Berns, M.W., and Cleveland, D.W. (2009). Double-strand DNA breaks recruit the centromeric histone CENP-A. PNAS *106*, 15762-15767.

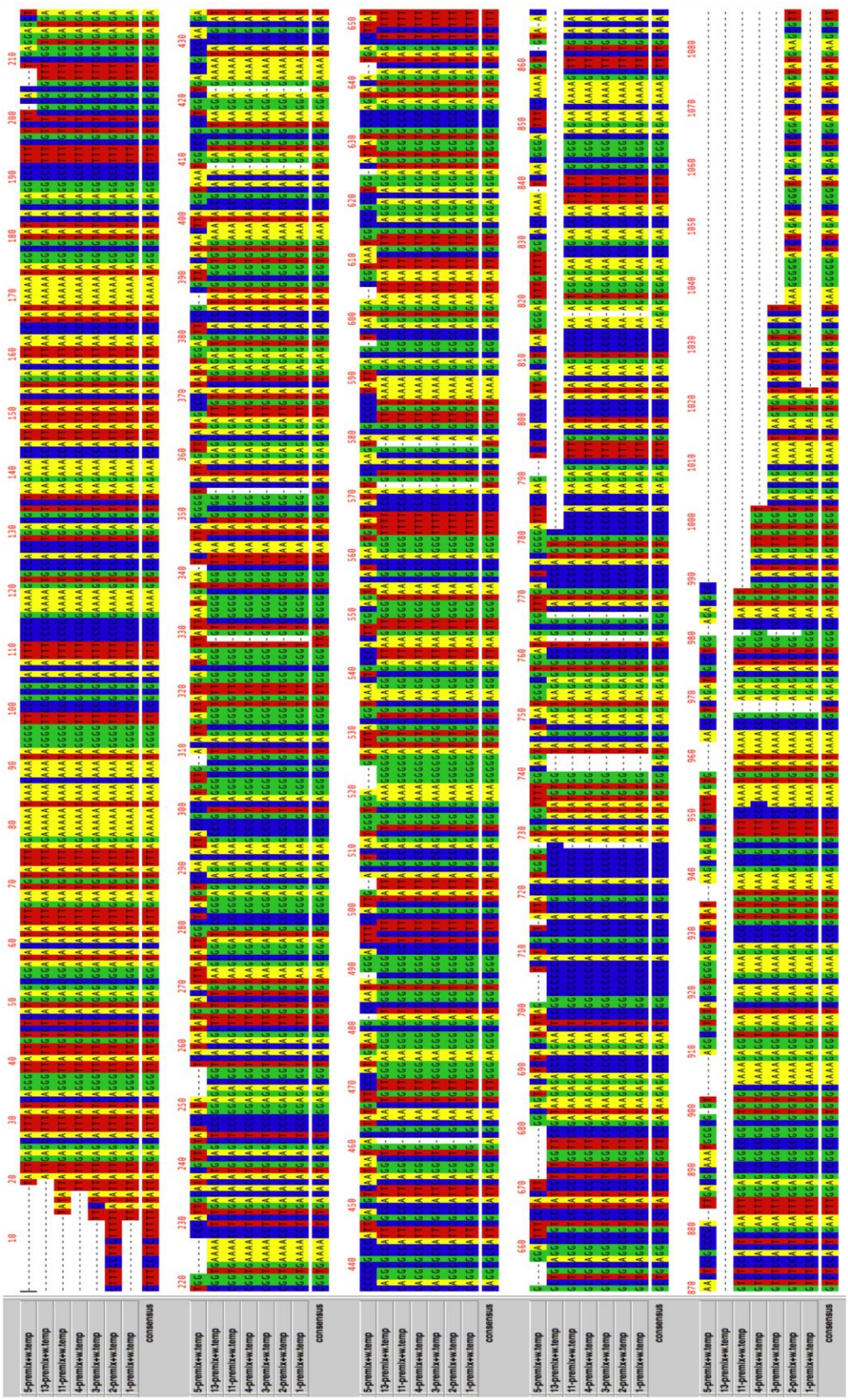
Zuo, Y., Wu, J., Xu, Z., Yang, S., Yan, H., Tan, L., Meng, X., Ying, X., Liu, R., Kang, T., *et al.* (2011). Minicircle-oriP-IFNgamma: a novel targeted gene therapeutic system for EBV positive human nasopharyngeal carcinoma. PLoS One *6*, e19407.

9 APPENDIX

9.1 ClustalW alignment of sequenced clones with Primer 255



9.2 ClustalW alignment of sequenced clones with Primer 256



9.3 Kinetochore proteins in H.sapiens and D.melanogaster

(Murata, 2013)

	Species	
Localization network	***H.sapiens***	***D.melanogaster***
inner centromere	INCENP	INCENP
	Aurora B	IPLI
	CENP-A	CID
	CENP-B	-
inner kinetochore – CCAN[45]	CENP-C	CENP-C
	CENP-H	
	CENP-I	
	CENP-K	
	CENP-L	
	CENP-M	
	CENP-N	
	CENP-O	
	CENP-P	
	CENP-Q	
	CENP-R	
	CENP-U	
	CENP-S	
	CENP-X	
	CENP-T	
	CENP-W	
outer kinetochore – KMN network[46]	hMis12	CG18156
	DSN1	
	NNF1	CGI13434
	NSL1	CGI1558
	Hec1/NDC80	Ndc80
	NUF2	Nuf2
	SPC24	(GI12063)
	SPC25	CG7242
	KNL1	
facultative	CENP-E	CENP-meta
	CENP-F	Spn
	CENP-V	
CENP-A chromatin establishment	HJURP	(CAL1?)

[45] constitutive centromere-associated network
[46] KNL1/Mis12 complex/Ndc80 complex network

	Mis18α	
	Mis18β	
	M18BP1	
	RbAp48	RbAp48
	RbAp46	

PUBLICATIONS, PRESENTATIONS

Parts of this work were published in a scientific article and presented at national conferences.

Publications

- Mendiburo MJ, Padeken J, Fülöp S, Schepers A, Heun P. (2011). Drosophila CENH3 is sufficient for centromere formation. *Science* **334:**686-690.

Oral presentations

- 22nd Annual Meeting of the Society for Virology (selected abstract)
  Essen, March 2012
  Titel: *CENH3 enhances gene vector maintenance by neocentromeres formation.*

- 19th Annual Meeting German Society of Gene Therapy (invited speaker)
  Hamburg, February 2013
  Titel: *The centromere-specific histone CENH3 enhances gene vector maintenance by neocentromere formation.*

DANKSAGUNG

Mein besonderer Dank gilt Herrn PD Dr. Aloys Schepers für das interessante Projekt, die ausgezeichnete wissenschaftliche Betreuung und die große Unterstützung beim Verfassen dieser Arbeit.

Herrn PD Dr. Josef Mautner danke ich sehr für die Übernahme der offiziellen Betreuung der Arbeit an der LMU München.

Herrn Prof. Dr. Peter Becker und Frau Dr. Sandra Hake danke ich dafür, mich als Mitglieder meines Thesis Advisory Committee, während der gesamten Zeit meiner Doktorarbeit unterstützt zu haben. Herrn Prof. Dr. Peter Becker danke ich außerdem für die Übernahme des Zweitgutachtens.

Herzlich bedanken möchte ich mich bei Dr. Christoph-Erik Mayer, Dr. Manuel Deutsch, Dr. Peer Papior, Máté Ravasz, Krisztina Zeller, Veronika Ulrich, Mala Amman, Julia Damaschke, Lara Schneider und Nina Kirstein für die tatkräftige Unterstützung und das angenehme Arbeitsklima im Labor.

Ich danke meinen Eltern und Großeltern, die mir soviel ermöglicht haben und ohne deren einzigartig emotionale sowie finanzielle Unterstützung mein Studium und diese Dissertation nicht möglich gewesen wären. Außerdem danke ich meiner Schwester Susi und Markus.

Ein herzlicher Dank gilt auch meinen Freunden Anne, Bettina, Gitta, Vera, Fabian, Kai und Johannes, die mir während der letzten vier Jahre motivierend und ratgebend zu Seite standen.

ERKLÄRUNG

Hiermit erkläre ich, dass die vorliegende Arbeit mit dem Titel

„Development of a CENH3-based, non-viral vector system“

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

München, 30. Juli 2013

Stefanie Fülöp