



HAL
open science

DNA replication in budding yeast : link between chromatin conformation and kinetics of replication

Claire Panciatici

► **To cite this version:**

Claire Panciatici. DNA replication in budding yeast : link between chromatin conformation and kinetics of replication. Subcellular Processes [q-bio.SC]. Université Paris Saclay (COMUE), 2016. English. NNT : 2016SACLS473 . tel-01499555

HAL Id: tel-01499555

<https://theses.hal.science/tel-01499555>

Submitted on 31 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLS473

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

Service de Biologie Intégrative et Génétique Moléculaire (SBIGeM), CEA

ECOLE DOCTORALE N°577
Structure et dynamique des systèmes vivants

Spécialité : Sciences de la Vie et de la Santé (SDSV)

Par

Mlle Claire PANCIATICI

DNA replication in budding yeast: Link between chromatin
conformation and kinetics of replication

Thèse présentée et soutenue à Saclay, le 6 décembre 2016 :

Composition du Jury :

M. BLOYER, Sébastien	Professeur, Université Paris Sud	Président
M. LE TALLEC Benoit	CR2, INSERM	Rapporteur
M. VICTOR Jean-Marc	DR1, CNRS	Rapporteur
M. AUDIT Benjamin	DR2, CNRS	Examineur
M. GILBERT Nick	Professeur, Université d'Edimburgh	Examineur
M. PEREZ Javier	Responsable ligne, Soleil	Co-encadrant
M. GOLDAR Arach	Ingénieur, CEA	Directeur de thèse

Financée par LLB/Soleil et IDEX



Remerciements

Tout d'abord, je remercie Arach Goldar pour m'avoir donné l'opportunité de travailler sur ce projet et ses relectures du manuscrit. Merci aussi à Didier Lairez pour m'avoir permis de travailler sur PACE, aidé grandement dans mon apprentissage de matlab et donné l'exemple d'un esprit critique scientifique. Merci à Javier Perez pour son temps passé avec moi sur SWING et la relecture attentive de mon manuscrit. Je remercie aussi toute l'équipe de SWING pour leur bonne humeur et leur accueil chaleureux, malgré mes visites en pointillé.

Merci à tous les gens qui ont marqué mon quotidien au labo : Elodie et Gwen, mes ex-collègues de bureau et toute l'équipe de Julie Soutourina. Merci en particulier à Thomas qui m'a aidé à créer la souche MNase. Merci aussi à Christelle et Rosa qui, grâce à leur travail, rendent la vie au labo plus facile. Enfin un grand merci à Bénédicte. Tu as été présente tout au long de ma thèse et tes conseils d'experte (de la levure comme des hommes) ont été un précieux soutien.

Merci à Vincent et à mes amis qui ont su ponctuer ces années de convivialité. Enfin, comment faire le bilan de ces trois ans sans penser à toi Diane. J'essaye de m'inspirer de ta force et de me conformer à les haute idée que tu avais de moi.

Abstract

Genetic information carried in the cell nucleus must be faithfully duplicated to be transmitted to daughter cells during cell division. In order to orchestrate their division, cells go through a reproducible 4 stages cycle called «cell cycle». The preparation and execution of the DNA replication program is restricted to specific phases and implies many proteic and structural regulators. In particular, DNA replication occurs on a complex template of DNA associated with proteins. The latter is both influencing and influenced by DNA replication. This work aims at investigating the link between chromatin conformation and the kinetics of DNA replication. In order to do so, we combine several techniques.

Using flow cytometry, we follow the evolution of a cell population with regards to their DNA content. This drives us to develop a methodology that deciphers the population averaged temporal program of DNA replication from a simple FACS histogram of an exponentially growing cell population. By analysing the cell cycle of 3 *Saccharomyces cerevisiae* mutants, where the regulation of Ribonucleotide reductase is challenged, we show that the replication origin firing is slowed down if the speed of DNA synthesis (v) is increased. This observation leads us to predict that cell control mechanisms (checkpoints) are not activated if $\frac{v_0}{4} < v < v_0$ where v_0 is the speed of replicative helicases.

Combining small-angle scattering of neutrons (SANS) and X-rays (SAXS), we show that the large-scale chromatin conformation cannot be directly assessed from SANS data but that SAXS data provide information on the local organisation of protein and DNA *in vivo*. Our data can be interpreted as a liquid crystal with a nematic order and a short correlation length, which suggests that yeast chromatin *in vivo* is predominantly devoid of 30 nm fibres organisation. In this prospect, chromatin inside the nucleus *in vivo* could be organised as 10nm fibres distant of 25 to 30nm. This distance, which is constant during DNA replication, is mainly imposed by the physico-chemical properties of the yeast nucleus. During interphase however, the correlation length is rising which indicates an increasing organisation of the chromatin before division.

On the other hand, we performed DNA combing to study the replication program in single cells. We reproduce previously obtained result showing that distance

between replicated tracks is of ~60kb which corresponds to the distance between known origins of replication. However, studying the behaviour of initiation, we propose that the initiation events are more frequent than previously measured and correspond to distances between MCMs proteins loaded on the genome.

**Réplication de l'ADN chez la levure
de boulanger : lien entre la
conformation de la chromatine et la
cinétique de réplication**

Résumé

L'information génétique contenue dans le noyau de la cellule doit être dupliquée fidèlement afin d'être transmise aux cellules filles pendant la division cellulaire. Pour organiser leur division, les cellules suivent un cycle reproductible composé de quatre étapes appelé cycle cellulaire. La préparation et l'exécution du programme de réplication de l'ADN ont lieu pendant des phases spécifiques du cycle grâce à l'intervention de multiples partenaires protéiques et de régulateurs structuraux. En particulier, la réplication de l'ADN s'effectue sur une matrice complexe constituée d'ADN associé à des protéines appelée chromatine. Cette dernière influence et est influencée par la réplication de l'ADN. Le travail présenté ici a pour objectif de faire le lien entre la conformation de la chromatine et la cinétique de réplication de l'ADN. Pour ce faire, nous combinons plusieurs techniques.

La cytométrie de flux nous permet de suivre la quantité d'ADN présent dans une population de cellules. Ceci nous a conduit à développer une méthode pour extraire le programme de réplication moyen de la population de cellules à partir de l'histogramme de FACS d'une population de cellules en croissance exponentielle. L'analyse du cycle cellulaire de trois mutants de la levure *Saccharomyces Cerevisiae* dans lesquels la régulation de la ribonucléotide réductase est altérée, montre que le déclenchement des origines de réplication est ralenti si la vitesse de synthèse de l'ADN (v) augmente. Cette observation nous conduit à prédire que les mécanismes de contrôle du cycle cellulaire (checkpoints) ne sont pas activés si $\frac{v_0}{4} < v < v_0$ où v_0 est la vitesse des hélicases.

En combinant les techniques de diffusion aux petits angles des neutrons (SANS) et rayons X (SAXS), nous montrons que la conformation de la chromatine à grande échelle ne peut pas être établie directement par des mesures de SANS. En revanche les données de SAXS fournissent des informations sur l'organisation locale des protéines et de l'ADN *in vivo*. Nos données peuvent être interprétées comme un cristal liquide avec un ordre nématique et une faible longueur de corrélation, ce qui suggère que la chromatine de la levure est majoritairement dépourvue d'une organisation en fibre de 30nm *in vivo*. Dans cette perspective, la chromatine dans le noyau serait organisée en fibres de 10nm distantes de 25 à 30nm. Cette distance, constante au cours de la réplication de l'ADN, est principalement imposée par

les propriétés physico-chimiques du noyau de la levure. Au cours de l'interphase cependant, la longueur de corrélation augmente, ce qui indique une organisation de la chromatine de plus en plus structurée avant la division cellulaire.

Par ailleurs, nous avons réalisé une étude du programme de réplication en molécules uniques grâce à la méthode de peignage d'ADN. Nous reproduisons les résultats précédemment obtenus montrant que la distance entre zones répliquées est d'environ ~60kb qui correspond à la distance entre des origines de réplication identifiées. Cependant, d'après l'étude du comportement dynamique de l'initiation, nous proposons que les initiations sont plus fréquentes que ce qui a été mesuré précédemment et correspondent à la distance entre les protéines MCM disposées sur le génome.

Contents

Remerciements	i
Abstract	iii
Résumé	vii
Contents	ix
List of Figures	xv
List of Tables	xix
Nomenclature	xix
Multiscale Significance of the Replication process	1
Introduction	3
1 Molecular mechanism of DNA replication	5
1.1 Replication of double-stranded DNA	5
1.1.1 Double helix structure and DNA duplication	5
1.1.2 Stepwise assembly of the replisome	13
1.1.3 Conservation of DNA replication	17
1.2 Replication of chromatin	21
1.2.1 Nucleosome as the structural unit of chromatin	21
1.2.2 Chromatin higher order organisation in vivo is still under debate	24
1.2.3 Chromatin is reconstructed as DNA is replicated	27

2	4th dimension of DNA replication	31
2.1	Spatio-temporal program of DNA replication	31
2.1.1	Replication profile	31
2.1.2	From defined replication profiles to cell to cell variability . .	34
2.1.3	Simulations and mathematical models to interpret the replication profiles	36
2.2	Establishment and execution of the temporal program	40
2.2.1	Potential origins recruitment in a nuclear structured frame .	40
2.2.2	Stochastic execution	42
2.2.3	Influence of transcription and chromatin landscape	44
2.3	Conservation and significance of the timing	47
2.3.1	Conservation of the replication program	47
2.3.2	Protection of genomic integrity	49
2.3.3	Transcription, epigenetic state and replication : causes and consequences	50
3	Replication in the yeast nucleus	55
3.1	Organisation of the yeast interphase nucleus	55
3.1.1	Rabl structure	55
3.1.2	Nuclear domains	60
3.1.3	Nucleosome organisation	63
3.2	Replication in the interphase nucleus	65
3.2.1	Subnuclear organisation of timing	65
3.2.2	Chromatin contacts and replication factories	66
3.3	Several techniques gave insight on the 3D chromatin organisation and dynamic	68
3.3.1	Usual techniques to study genome organisation	68
3.3.2	Small-angle scattering	70
	Material and Methods	77
4	Yeast strains and cell cycle	79
4.1	Strains	79
4.1.1	Yeast strains used	79
4.1.1.1	BY4741	79
4.1.1.2	MCM869	79
4.1.1.3	BYMNase	80
4.1.2	Growth	85
4.1.2.1	Growth conditions	85
4.1.2.2	Thymidine analogs	86

4.1.3	Synchronisation	87
4.1.3.1	α -factor	87
4.1.3.2	Nocodazole	87
4.2	FACS	88
4.2.1	Sample preparation	88
4.2.2	Data acquisition	89
4.2.2.1	Device settings	89
4.2.2.2	Acquisition	89
4.2.2.3	Data reading	89
4.2.3	Data analysis	90
4.2.3.1	Normalisation	90
4.2.3.2	Data fitting	91
5	SAS	95
5.1	Sample preparation	95
5.2	Device	97
5.2.1	Neutrons	97
5.2.2	X-Rays	97
5.3	Data analysis	98
6	DNA combing	101
6.1	DNA purification in plugs	101
6.1.1	Cells sample	101
6.1.2	Plugs preparation	101
6.1.3	DNA purification	102
6.2	DNA combing and staining	102
6.2.1	Converslips cleaning and silanisation	102
6.2.2	DNA combing	103
6.2.3	BrdU and total DNA staining	103
6.3	DNA acquisition and quatification	103
6.3.1	Image acquisition and treatment	103
6.3.2	Obtaining and correcting binary values	105
6.3.3	Computation of replication parameters	106
	Results and Discussion	111
7	Evolution of cell cycle	113
7.1	Cell synchronisation	113
7.1.1	Cell cycle block	113
7.1.2	Effect on S phase	117

7.2	Progression in S phase	124
7.2.1	FACS profile contains information on three main populations of cells	124
7.2.2	Average S phase progression of MCM869 cells	126
7.2.3	Effect of growth conditions	127
7.3	Deciphering the replication program from FACS data	133
8	Using Small-Angle Scattering to study yeast nuclear organisation	145
8.1	Experimental conditions	145
8.1.1	Sample preparation	145
8.1.2	Radiation damages	149
8.2	Data analysis	152
8.2.1	Qualitative analysis	152
8.2.2	Evolution with time	155
8.2.3	Data fitting	157
8.3	Origin of the signal	160
8.3.1	SAXS and SANS comparison	160
8.3.2	Link to replication	163
8.3.3	MNase digestion	168
9	Temporal and spatial replication program	171
9.1	Doing the spadework for DNA combing	171
9.2	Kinetics of DNA replication	174
9.2.1	Replication dynamic	174
9.2.2	Origin distribution	176
9.2.3	Timing	180
	Conclusions and perspectives	183
	Annexes	191
A	Evolution of the cell cycle - supplementary informations	193
B	Yeast nuclear organisation - supplementary informations	195
C	DNA combing - supplementary informations	201
D	Replication time of an eye	203

Bibliography	205
Résumé français	247

List of Figures

1.1.1	From X-Ray Image to structure	6
1.1.2	Unidirectionality of polymerases imposes discontinuous synthesis on the lagging strand	8
1.1.3	helicase activity of the CMG complex	9
1.1.4	Top1 release supercoiling induced by DNA replication	10
1.1.5	Several secondary structures of DNA exist in vivo	11
1.1.6	Replisome architecture	12
1.1.7	Schematic view of the cell cycle	13
1.1.8	Loading and firing of a replication origin	15
1.1.9	Regulation of budding yeast cyclins during the cell cycle	17
1.1.10	Replicon model	18
1.1.11	The universal replication fork	19
1.1.12	Distribution of DNA replication proteins across eukaryotic supergroups	20
1.2.1	Nucleosome composition	22
1.2.2	PTMs on histone tails	23
1.2.3	Lexicon of nucleosome positioning	23
1.2.4	Two models for chromatin fibre	24
1.2.5	Loop extrusion model	26
1.2.6	Histone reassembly after replication fork passage	28
2.1.1	Construction of a replication profile	32
2.1.2	Replication profiles of several eukaryotes	33
2.1.3	Identification of origins of replication by transition density of Okazaki fragments	34
2.1.4	Average behaviour of single cells recapitulates the population behaviour	35
2.1.5	Identifiability	36
2.1.6	Illustration of the domino-like initiation in human TTRs	38
2.1.7	Abundance of potential origin as an explanation for inhomogenous efficiency	38

LIST OF FIGURES

2.2.1	Replication domains model	42
2.2.2	Origin tethering organised by multiple proteic factors	43
2.2.3	Collision between replication and transcription	45
2.2.4	Model of replication and transcription co-directionality	45
2.3.1	Similarity of the replication profiles of related species	48
2.3.2	Different regulation of developmentally regulated and constitutive domains	51
2.3.3	Model of multiple influences of genes expression, chromatin conformation and replication timing	52
3.1.1	Rabl configuration of yeast nucleus	56
3.1.2	Computational models of yeast nucleus organisation	58
3.1.3	Reorganisation of the chromatin structure in quiescence	59
3.1.4	Chromatin contacts in yeast genome	61
3.1.5	Genes clusters	62
3.1.6	Contact between adjacent nucleosomes	64
3.2.1	Organisation in replication factories	67
3.3.1	Schematic methodology of 3C and derivative techniques	69
3.3.2	Measure of scattering intensity	71
3.3.3	Principle of contrast	72
3.3.4	Compared SAXS signal from chicken or rabbit erythrocyte	73
3.3.5	SAXS signal from HeLa nuclei	74
4.1.1	Plasmid pGZ136(pRS406-pREB1-3FLAG-MNase-NLS) and its integration mode	80
4.1.2	Western Blot of MNase protein	82
4.1.3	1% agarose gel showing DNA digestion after MNase induction	84
6.3.1	Fluorescence Spectra of fluorescent dyes used for DNA combing	104
6.3.2	λ -DNA stretching by DNA combing	105
7.1.1	Schematic effect of mating pheromone	114
7.1.2	Effect of α -factor on two yeast strains	114
7.1.3	Nocodazole effect on BY4741 cell	115
7.1.4	Effect of nocodazole treatment on MCM869 cells morphology	116
7.1.5	DNA replication after nocodazole treatment on BY4741 cells	118
7.1.6	Influence of nocodazole on DNA replication	119
7.1.7	Blocking duration by α -factor in MCM869 cells influences the synchronisation in G1	120
7.1.8	Cell cycle after various duration of treatment by α -factor in MCM869 cells	121
7.1.9	S-phase parameters after various α -factor treatment duration	121

7.1.10	Evolution of replicated fraction during two S phases in BY4741 cells	122
7.2.1	Fluorescence level to DNA content	124
7.2.2	Fit of normalised FACS data	125
7.2.3	High quality fit on averaged data allow to extract cycling cells behaviour	126
7.2.4	Evolution of replicated fraction of cells growing in YPD or CASA	128
7.2.5	Thymidine and its analogs	129
7.2.6	Effect of the variation of concentration of thymidine analog in YPD	130
7.2.7	Effect of the concentration of thymidine analog in CASA vs YPD	130
7.2.8	Incorporation of EdU in CASA vs YPD medium	131
7.2.9	EdU incorporation during replication in YPD	132
8.1.1	Schematic view of the stirrer	146
8.1.2	Stability and evolution of yeast cells in SWING incubator	147
8.1.3	SAXS signal from MCM869 cells, spheroplasts and isolated nuclei	148
8.1.4	SAXS curves from yeast cells after successive expositions	151
8.2.1	SAXS and SANS compared profiles from yeast and chicken erythrocytes	154
8.2.2	Evolution of the SAXS peak with time after release from α -factor in pellet or liquid	155
8.2.3	Global evolution of SAXS amplitude with time after release from α -factor in liquid samples	156
8.3.1	Application of SAXS/SANS contrast difference to SANS fit curve	161
8.3.2	Peak amplitude in HU	163
8.3.3	Correlation of peak area with replicated fraction	164
8.3.4	Fit of the peak zone of the liquid sample SAXS curve	166
8.3.5	DNA digestion by MNase after calcium induction	168
8.3.6	Effect of DNA digestion by MNase on SAXS profiles	169
9.1.1	Replicated fraction during combing experiment	173
9.2.1	Replication parameters extracted from DNA combing	175
9.2.2	Creation and merge of forks	176
9.2.3	Eye-to-eye distance	177
9.2.4	Biases impose minimum and maximum measurable ETED (Fig.)	178
9.2.5	ACS and MCM distances distributions	179
9.2.6	Probability of initiation linked to square MCM density	181
A.1	Logarithmic state of MCM869 and BY4741 cells	193
A.2	Logarithmic growth of BY4741	194
A.3	Goodness of the fit for FACS histograms and rreplicated fraction	194

LIST OF FIGURES

B.1	Cell, spheroplasts and nuclei normalised SAXS data	196
B.2	Krakty plot of SANS data	197
B.3	Maximum position converted into characteristic distances for all SAXS experiments	198
B.4	SAXS and SANS curves of yeasts spheroplasts fit with the descri- bed models	199
B.5	Goodness of the fit of the peak area in SAXS data	200
B.6	F(t) of MCM869 and BY4741 cells	200
C.1	Example of combed fibres	202
D.1	Schematic composition of and eye composed of n replicons	203

List of Tables

1.1	Replication enzyme fidelity assessed by the frequency of single base substitution	8
3.1	Parameters and constraint on the models	58
4.1	Composition of buffers for in vivo DNA digestion by MNase enzyme	84
4.2	Composition of growth media	85
5.1	Composition of buffers for spheroplasting	96
5.2	Composition of buffers for nuclei isolation	96
6.1	Architecture of the structure used to store information about DNA combing	107
8.1	SANS and SAXS density and contrast	160

Liste des symboles

Å	Ångströms
3C	Chromatin conformation capture
A	Adenine
ACS	ARS consensus sequence
APC	Anaphase promoting complex
ARS	Autonomously replicating sequence
ASF1	Antisilencing function 1
bp	Base pair
CAF-1	Chromatin assembly factor
C	Cytosine
cdc6	Cell division cycle 6
CDK	Cyclin-dependent kinase
Cdt1	Chromatin licensing and DNA replication factor 1
CID	Chromatin interaction domains
CMG	Cdc45-MCM-GINS
CTCF	CCCTC-binding factor
CTRs	Constant timing regions
DDK	Dfb4-dependent kinase
DNA	Deoxyribonucleic Acid

LIST OF TABLES

dNTP	Deoxyribonucleotids
FEN1	Flap endonuclease 1
Fkh	Forkhead transcription factor
FPC	Fork protection complex
G4	G quadruplex
G	Guanine
H1	Histone protein 1
HU	Hydroxyurea
MCM	Mini chromosome maintenance protein
NAP1	Nucleosome assembly protein 1
NDR/NFR	Nucleosome depleted/free regions
NE	Nuclear envelope
ODP	Origin decision point
ORC	Origin-recognition complex
PP1	Protein phosphatase 1
pre-RC	Pre-replicative complex
rDNA	Ribosomal DNA
Rif1	Rap-interacting factor 1
RPA	Replication protein A
S.c	Saccharomyces Cerevisiae
S.p	Schizosaccharomyces pombe
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
Sir	Silent information regulatory complex

SPB	Spindle pole body
TADs	Topologically associated domains
TDP	Timing decision point
TSSs	Transcription starting sites
T	Thymine
TTRs	Transition timing region

Multiscale Significance of the Replication process

Genetic information carried in the nucleus of the cell give instructions to build, maintain and reproduce the cell. Therefore it is of great importance to duplicate these instructions without errors in order to obtain two cells with proper genetic information after cellular division. DNA is the molecular template of the genetic information and the process of its duplication is called replication.

The story of the DNA replication study started more than fifty years ago. The discovery of the DNA structure allowed to imagine the way it was duplicated, unlocked the comprehension of enzymes and rise many new questions. At many scales, DNA replication process is linked to the structure of the template and linear replication is undoubtedly organised by the double helix structure. DNA needs to undergoes several level of packing in order to fit of the nucleus. However, the hyperstructure of the chromatin is an obstacle for the propagation of replication forks and requires many proteic partners to be dealt with. Nucleosome positioning is of great importance for DNA replication and particularly for choosing the positions of initiation. Last, 3D organisation of the nucleus and subnuclear position of a DNA part is one of the main influent of its timing of replication. Replication is therefore a multi-scale process and the organisation of the template at several scale is significant to understand its proceeding. Moreover, as in eukaryotes DNA replication starts at different time in different points of the genome, time must be considered as the fourth dimension of the replication process.

We will see in a first part how experimental and conceptual discoveries gives us a clear picture on the molecular mechanism of double strand DNA and chromatin duplication. Next we will describe more in depth the spatio-temporal program of DNA replication. At last we will focus on the three-dimensional organisation of the genome of the budding yeast which will be our model organism to study the DNA replication process.

Chapter 1

Molecular mechanism of DNA replication

The replication of the genome is a delicate process, that needs to be tightly regulated in order to guarantee its correct and total duplication before cell division. In addition, duplication occurs on a condensed template that is not directly accessible. Indeed, the Deoxyribonucleic Acid (DNA) molecule is negatively charged and cannot be completely folded into itself spontaneously. Thus in order to fit inside the nucleus, it associates with proteins in a complex called chromatin.

This first chapter takes an historical point of view to describe how molecular actors that enable the DNA replication process were discovered. In the first part, all the components necessary for the polymerisation of DNA are introduced and the evolutionary conservation of these components in all eukaryotes is emphasised. Then the packing of DNA into chromatin is described in the second part as well as the preservation of this organisation during DNA replication.

1.1 Replication of double-stranded DNA

1.1.1 Double helix structure and DNA duplication

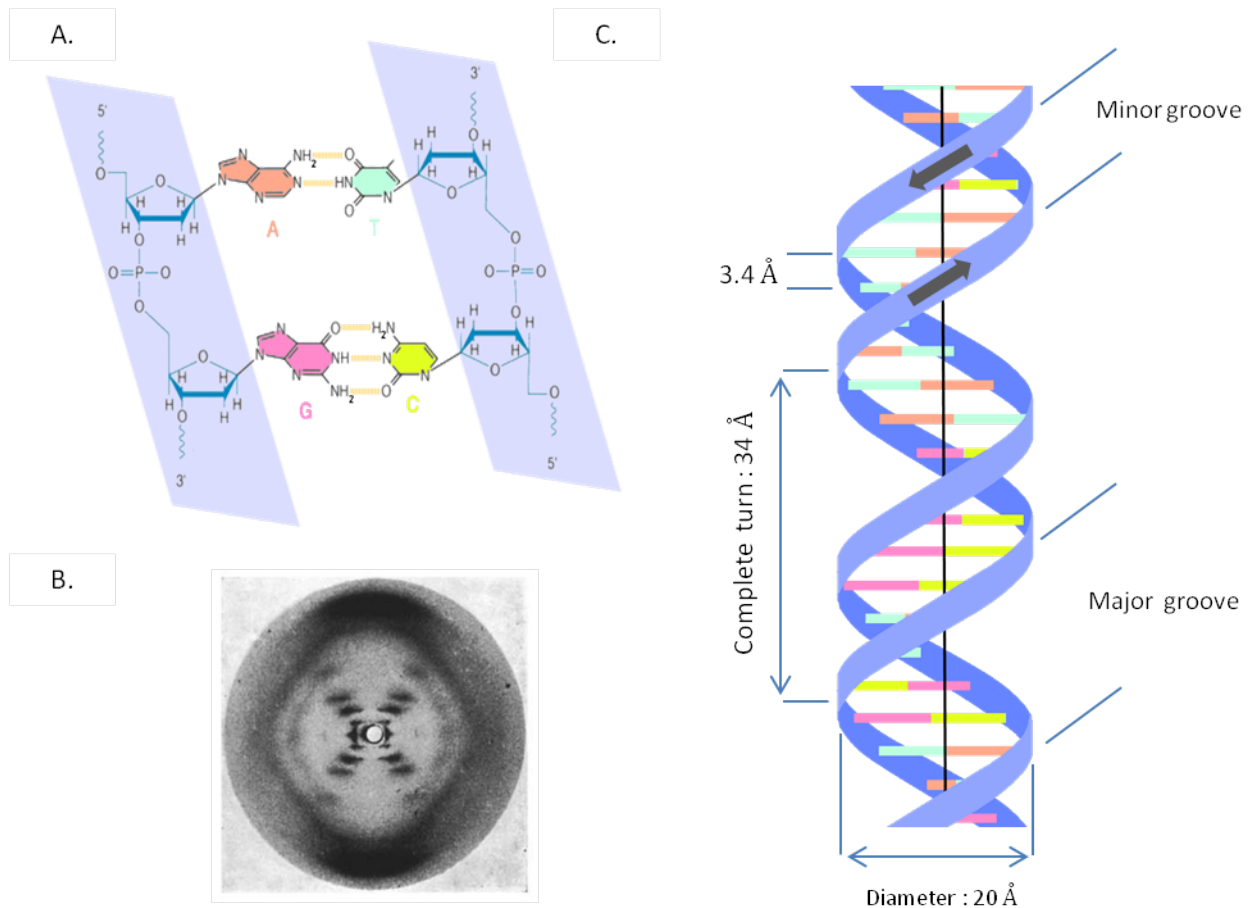
The resolution of the DNA structure triggered many discoveries. In particular, it raised questions that lead to the uncovering of major molecular actors of its duplication machinery.

Canonical form of DNA and semi-conservative replication

DNA is a polymer of nucleotides composed of a backbone of alternating sugar (deoxyribose) and phosphate groups. One of four possible nucleic bases : Adenine

(A), Thymine (T), Guanine (G) or Cytosine (C) is associated to each sugar. In 1951, Chargaff discovered the rules of base pairing based on their chemical structure : two hydrogen bonds link A and T and three link G and C (Fig. 1.1.1 A.) (Chargaff 1989). Two strands of DNA are paired between complementary bases in an antiparallel manner and coil around a common axis. The structure of this double-stranded DNA was first published by James D. Watson and Francis Crick in 1953 (Watson and Crick 1953b) based on a X-Ray diffraction image of DNA (Fig. 1.1.1 B.) from Rosalind Franklin in 1952 (Gosling and Franklin 1953). The most common structure, called B-DNA, is 20 ångstroms (Å) wide and extends 34Å per 10bp (base pair) (Menderlkern et al. 1981) (Fig. 1.1.1 C.).

Figure 1.1.1: From X-Ray Image to structure



- A. Nucleic bases are paired thanks to their chemical affinities. Adenine forms two hydrogens bonds with thymine while guanine forms three bonds with cytosine.
 B. Photograph 51 is a famous image of diffraction from DNA published by R. Franklin team. (Gosling and Franklin 1953)
 C. The most common structure adopted by DNA in vivo called B-DNA consists of two strands of DNA coiling around a common axis in an antiparallel manner.

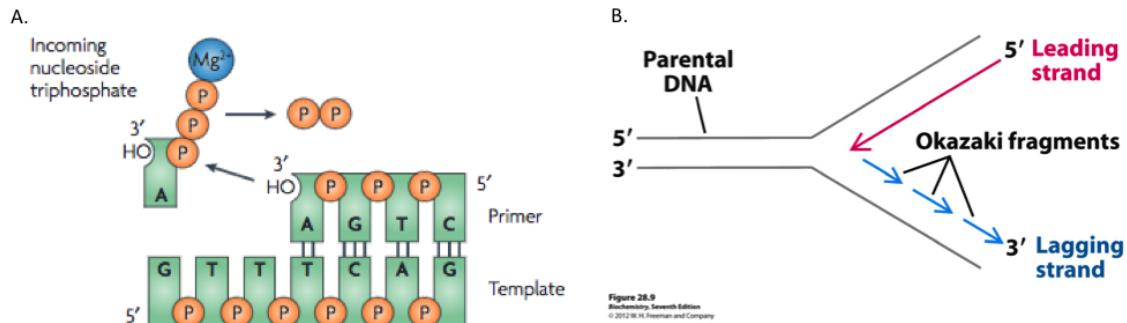
Following their discovery, Watson and Crick commented on its genetic implications (Watson and Crick 1953a). Indeed the presence of two complementary strands paired only by weak chemical bounds, suggests that they could be separated to serve as template for duplication. In this hypothesis, each strand could be individually duplicated therefore half of the parent DNA molecule would be conserved in the daughter cell. Semi-conservative replication was later proven by Meselson and Stahl in 1958 (Meselson and Stahl 1958). In their experiment, *E.Coli* bacteria were grown in medium containing heavy nitrogen and then released in medium containing normal nitrogen. DNA was isolated after one, two and three generations. Due to the difference of weight of nitrogen, double helix of heavy DNA, double helix of light DNA and composite double helix of light and heavy DNA can be separated by their density in a cesium chloride gradient. The experiment showed that after one generation most DNA is present as a double helix containing one parental strand and one newly synthesised strand confirming the semi-replicative hallmark of DNA replication.

Unidirectionnality of DNA polymerases and asymmetric DNA replication

DNA replication as suggested by Watson and Crick, could have been a spontaneous mechanism with the alignment of bases of high affinity. However, Kornberg, a biochemist who studied the synthesis of nucleotides, postulated that an enzyme was necessary to assemble the DNA. In 1956, his team found such an enzyme in bacteria but only capable of catalysing the polymerisation of deoxyribonucleotids (dNTP are the association of base with the sugar and a triphosphate group) in 5' to 3' direction (Mitra and Kornberg 1966). Indeed, the chemical reaction proceeds between dNTP activated on their 5' position and the 3'OH group of the previously incorporated nucleotide (Fig. 1.1.2 A.). Since the two strands are antiparallel they cannot be both synthesised continuously. Okazaki discovered in 1966 (Sakabe and Okazaki 1966) that small fragments of DNA were actually synthesised in 3' to 5' direction (Fig. 1.1.2 B.) to be combined afterwards (Sugino et al. 1972). The synthesis of DNA on 3' to 5' strand is slightly delayed compared to the 5' to 3' and they are therefore termed respectively as lagging and leading strands (Lee et al. 1998).

Many DNA polymerases are needed *in vivo* for replication and as well for the repair of damaged DNA (Loeb and Monnat 2008). Three of them, Pol α , Pol δ and Pol ϵ are essential for the eukaryotic DNA replication (Johnson and O'Donnell 2005). None of them however can perform polymerisation *de novo* as they need a 3'OH group to start the chemical reaction. Therefore replication is started by the polymerisation of a small RNA fragment of about 10 ribonucleotides called primer. Pol α -primase is a complex of two polymerases and two primases and is able to start the DNA synthesis on both leading and lagging strands. Pol ϵ

Figure 1.1.2: Unidirectionality of polymerases imposes discontinuous synthesis on the lagging strand



A. The activation of dNTP on the 5' position imposes the direction of DNA polymerisation. Figure 2b from Loeb and Monnat 2008 reused with permission.
 B. While the synthesis of the leading strand is continuous, the lagging strand is synthesised in multiple fragments.

catalyses the leading strand synthesis while Pol δ synthesises DNA on the lagging strand (Burgers 2009). Pol α only synthesises few ribonucleotides compared to Pol δ and ϵ (Chilkova et al. 2007). Highest processivity of polymerases δ and ϵ is explained by their physical interaction to the Proliferating Cell Nuclear Antigen protein (PCNA) which has a ring shape and encircles DNA (Moldovan et al. 2007).

Polymerases must have a high fidelity (Table 1.1, McCulloch and Kunkel 2008). Contrary to Pol α , Pol δ and Pol ϵ possess a proofreading mechanism as their exonuclease activity in the 3'-5' direction allows to remove mismatched bases. Despite the presence of a post DNA replication complementary mechanism to prevent errors called mismatch repair (MMR) (Li et al. 2016), the low fidelity of Pol α is an issue as it seems to synthesise a substantial proportion of lagging strand (10 bases every 100-200 bases).

Polymerase	Fidelity
Pol α	9.6×10^{-5}
Pol δ	$\leq 1.3 \times 10^{-5}$
Pol ϵ	$\leq 0.2 \times 10^{-5}$

Table 1.1: Replication enzyme fidelity assessed by the frequency of single base substitution

This problem is most probably sorted during the step of Okazaki fragments maturation which consists on the removal of RNA fragments and ligation of DNA parts. Pol δ provokes a displacement of the next Okazaki fragment (Burgers 2009). This segment is therefore synthesised again with high fidelity. The displaced part

contains the RNA primer that can be also removed by Ribonuclease H (RNase H) as well as the DNA portion produced by polymerase α (Pavlov et al. 2006) and is processed by the flap endonuclease 1 (FEN1) (Goulian et al. 1990). DNA ligase I then completes the remaining nick (Olivera et al. 1968). The length of this displaced portion is proposed to be controlled by the presence of nucleosomes (Smith and Whitehouse 2012, Yadav and Whitehouse 2016) that are rapidly reassembled on the newly synthesised DNA (see subsection 1.2.3).

Helicase opening the double helix

Polymerases need to gain access to the bases therefore the double helix must be untwisted and the two strands separated. First helicase was discovered in bacteria by Hoffman-Berling in 1976 (Abdel-Monem et al. 1976). Many helicases exists *in vivo* with several applications (Tuteja and Tuteja 2004). In eukaryotes, the hexamer composed of MCM 2 to 7 proteins (named for their essential role in minichromosome maintenance) performs the helicase activity required for initiation and elongation step (Lohman, 1993). MCM complexes are installed as head-to-head double hexamer during the G1 phase of the cell cycle among with other factor and, as it will be discussed in subsection 1.1.2, their activation is the key step to the initiation of the replication (Riera et al. 2014). To be activated, MCM complex conformation changes (Bochman and Schwacha 2009) from encircling double-stranded DNA to single-stranded DNA thanks to a tight interaction with protein cdc45 and the GINS complex to form the CMG complex (Costa et al. 2011) (Fig. 1.1.3).

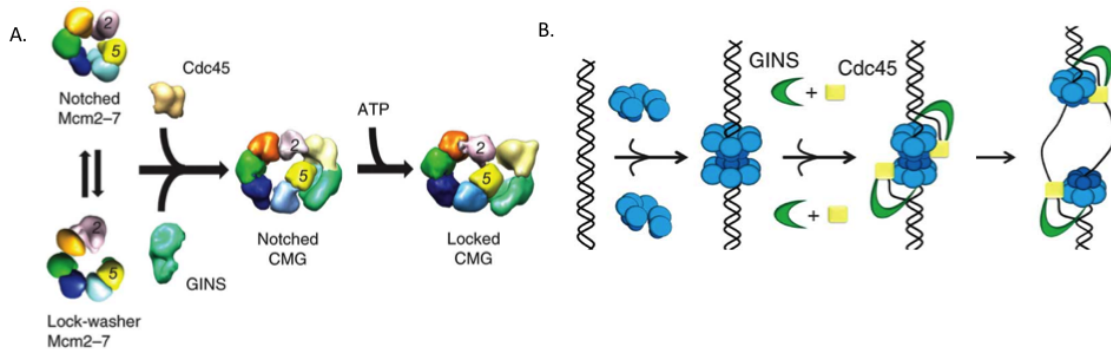


Figure 1.1.3: helicase activity of the CMG complex

A. MCM2-7 complex can adopt an open and a close configuration. The notched configuration is stabilised by the interaction with Cdc45 and GINS. ATP reinforce the interaction of subunits 2 and 5 and grant the helicase activity of the complex.

B. Two MCM complexes are installed in tandem on ds-DNA. GINS and Cdc45 changes the configuration of MCM to encircle only one strand of DNA.

Figure 8 from Costa et al. 2011 reused with permission.

Helicase progression is an ATP-dependant process and several translocation

mechanisms have been proposed (Patel and Donmez 2006). According to new structures deduced from single molecule electron microscopy, the model for MCM progression and DNA unwinding is a inchworm translocation where the two strands are separated by steric exclusion (O'Donnell and Li 2016).

The unwinding of DNA leads to the apparition of single-stranded DNA where replication protein A (RPA) are recruited in order to prevent re-pairing of separated strands, formation of hairpin via self-pairing and endogenous aggressions (Fanning 2006).

Supercoils and secondary structures

As early as 1954 Max Delbrück thought that DNA replication of a double helix would imply some topological issues (Delbrück, 1954). Indeed, the opening of the twisted molecule would lead to a even greater level of coiling named supercoiling. (Fig. 1.1.4 A.).

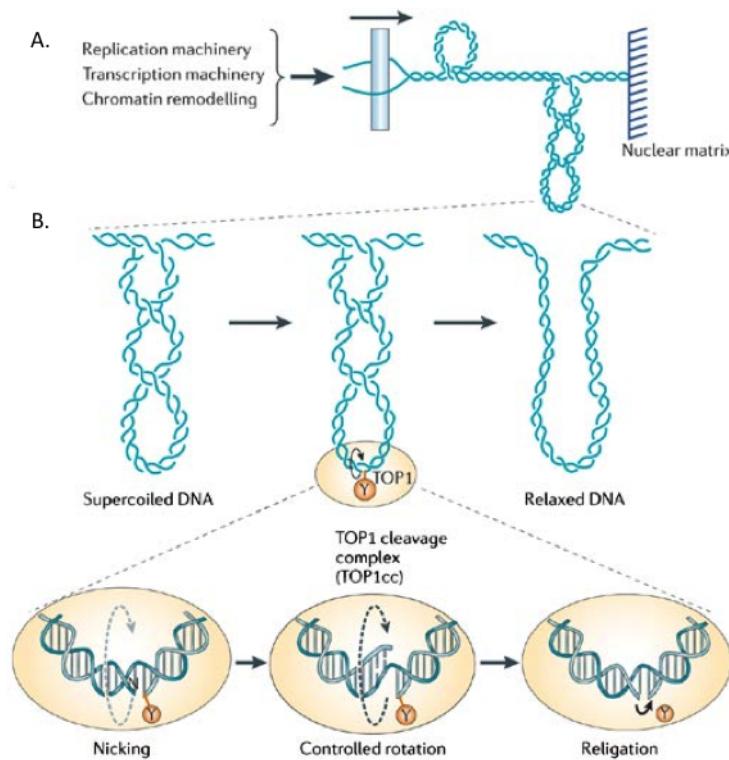


Figure 1.1.4: Top1 release supercoiling induced by DNA replication

A. Biological processes acting on ss-DNA such as DNA replication create supercoiling.

B. Topoisomerases enzymes I and II relax these topological tensions.

Figure 1 from Pommier 2006 reused with permission

James C. Wang discovered DNA topoisomerases while studying supercoiled

dismantle these structures to allow the correct progression of the replication machinery (León-Ortiz et al. 2014). However the resolution of such structures as well as proteins bound to DNA requires the pause of the replication machinery (Labib and Hodgson 2007).

Dynamics of the replisome

The gathering of all the necessary proteins for replication is called replisome or replication fork. Its assembly and progression are tightly regulated (Kurth and O'Donnell 2013). Indeed, it is important that the helicase and the polymerase stay coordinated to avoid important formation of single-stranded DNA. Most enzymes of the replisome are therefore in direct interaction to prevent such an effect. The exact organisation of the replisome in eukaryotes is not totally defined but fast progresses are made in this direction (O'Donnell and Li 2016) and the CMG is known to interact with the polymerases α and ϵ (Sun et al. 2015). However no interaction with the polymerase δ which would grant coordination of leading and lagging strand was detected yet. The smooth progression of the replication fork suggests some mechanism for coordination of the two strands synthesis. The clamp loader RFC, responsible of PCNA loading could act as a coordinator of both polymerases ϵ and δ (Indiani and O'Donnell 2006) explaining their comparable speed when bound to PCNA (Stodola and Burgers 2016) without a direct contact with the helicase.

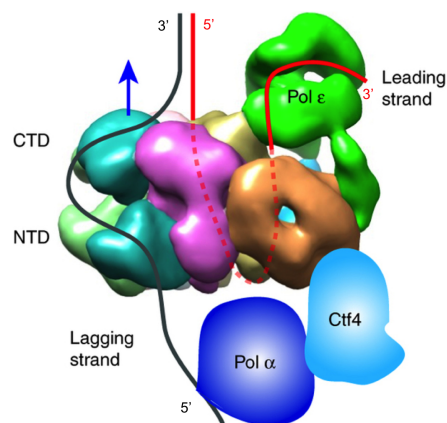


Figure 1.1.6: Replisome architecture

Red and black lines illustrate possible leading and lagging strand DNA path into the replisome. The blue arrow indicates the direction of replisome movement on DNA. The diagram indicates a long path of leading-strand DNA through the entire MCM ring and then bending back up to Pol ϵ , requiring about 40 nucleotides of ss-DNA. Leading ss-DNA is illustrated as going completely through the MCM2-7 complex and then bending up through the second 'accessory' channel of CMG, but this path is speculative. Figure 6 from Sun et al. 2015 reused with permission.

Challenges to the replication from internal or external sources are generally

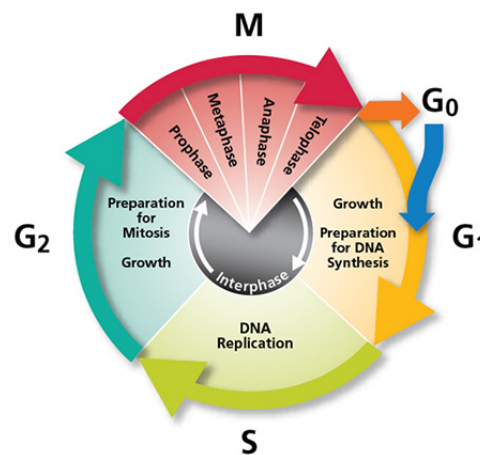
termed as replication stress (Zeman and Cimprich 2013). Despite the replisome interactions, replication stress can induce the uncoupling of double helix unwinding and DNA synthesis as well as the pause or stalling of fork. During such pauses, replication forks are protected by the interaction of several proteins forming the fork protection complex (FPC) (Leman and Noguchi 2012) to be able to resume the replication once the barrier is removed. This requires factors that are known to be physically associated with forks, even though the precise mode of interaction is not yet described (Calzada et al. 2005). Moreover, in case of the encountering of damaged DNA (alteration such as break in DNA strand, missing or chemically changed base), replisome contributes to the recruitment of DNA repair factors (Drissi et al. 2015, Haye and Gammie 2015).

All the coordination needed for normal or challenged progression of the replisome outlines a complex and dynamic process for molecular mechanism of DNA polymerisation. From the basis of the DNA structure and experiment after experiment, scientists revealed how molecular actors tackle each issues. Although some of the actors are probably still unknown, the focus is now made on deciphering the sequential interactions between this great number of proteic partners whose coordination is as crucial as their action to preserve genome integrity.

1.1.2 Stepwise assembly of the replisome

At the scale of the cell, one additional challenge of DNA replication is to complete once and only once the replication of the totality of its genome before division. The life of the cell is a cycle of four alternating phases : G1 phase during which the replication is prepared, S phase when DNA is synthesised, G2 phase when quality control are made and division prepared, followed by the M phase when the division occurs and produces two daughter cells ready to perform the same cycle. G1, S and G2 phases, called interphase, constitute the crucial step for DNA replication.

Figure 1.1.7: Schematic view of the cell cycle



G1 preparation and MCM paradox

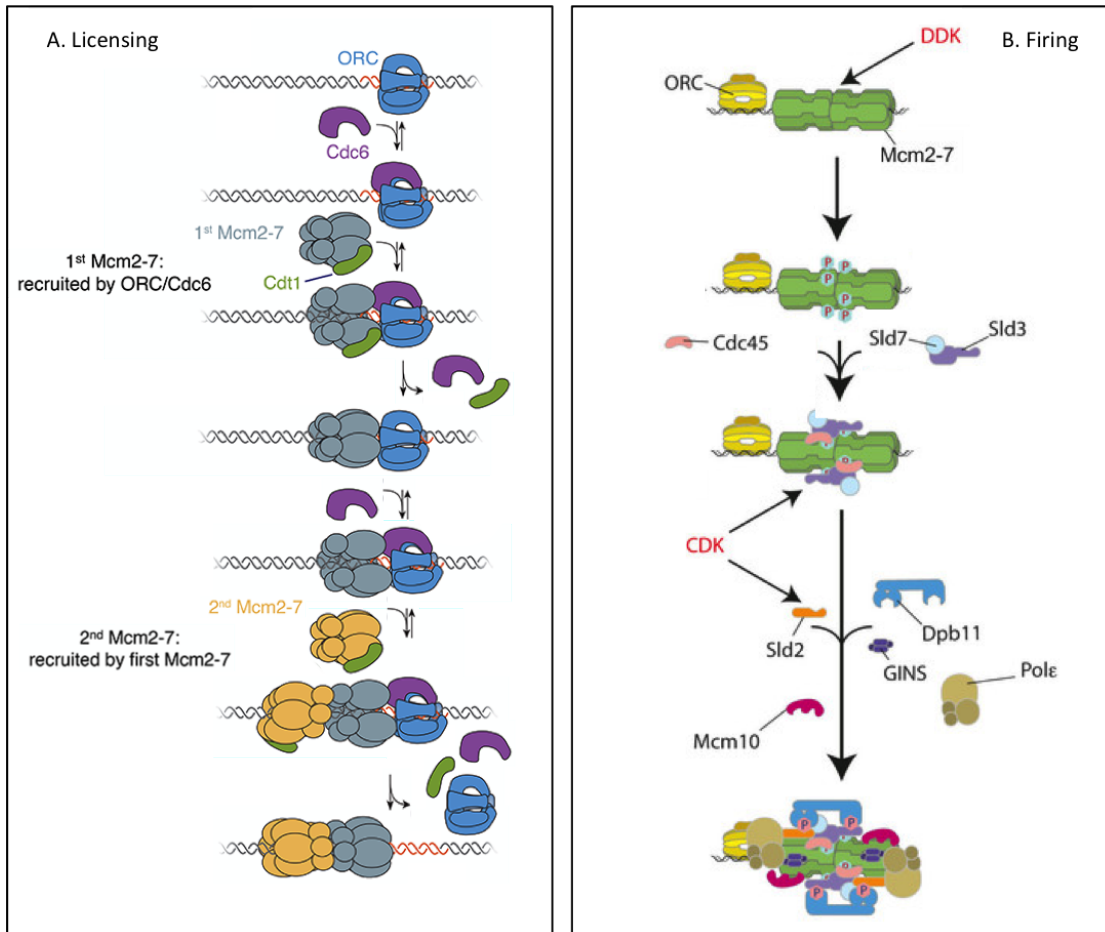
The preparation of DNA replication in G1 phase consists of the recruitment of the pre-replication complex (pre-RC) in several positions of the genome (defined in next subsection and next chapter). This process, called licensing, is precisely described in recent reviews (Riera and Speck (2016), Deegan and Diffley 2016) and illustrated in Fig. 1.1.8 A.

Briefly, the origin-recognition complex (ORC) composed of six subunits (1-6) binds the DNA. ORC then recruits cell division cycle 6 (*cdc6*) and chromatin licensing and DNA replication factor 1 attached to MCM (Cdt1-MCM). Interaction of MCM with Cdt1 is crucial to allow its import to the nucleus (Tanaka and Diffley 2002) and its interaction with ORC (Cook et al. 2004). Both ORC and Cdc6 possess ATPase activity that is necessary for the loading and subsequent activation of two MCM 2-7 complexes (Randell et al. 2006, Coster et al. 2014). The pre-RC complex was often considered to include ORC, *cdc6*, Cdt1 and MCM. However, such a complex is not the form that is subsequently activated in G1 phase and the preRC complex in recent publications mostly refer to the couple of MCM helicases (Siddiqui et al. 2013, Rivera-Mulia and Gilbert 2016b). Indeed, it has been shown that Cdc6 and Cdt1 are disengaged from the preRC before its activation (Chang et al. 2015). Also ORC has been shown to be non necessary during S phase after the loading of MCM (Gibson et al. 2006) and many MCM not bound to ORC are present and have been shown to be functional (Woodward et al. 2006).

The great abundance of MCM compared to ORC was referred as the MCM paradox as their distribution seemed inconsistent with their presumed helicase activity (Hyrien et al. 2003, Forsburg 2004). The excess of MCM is now known to be required to complete DNA replication in case of replication stress (Ge et al. 2007, Ibarra et al. 2008) and MCM has been shown to be involved in other processes as well (Das et al. 2014). Moreover, recent models suggest that the local excess of MCM could increase the probability of firing in a given region (Das et al. 2015) (see next chapter).

While efforts were made on understanding the role of excess MCM, very few information is known about how these MCM are loaded (Das and Rhind 2016). The possibility that MCM binds chromatin without being loaded by ORC is extremely unlikely as it is an active process. Moreover ORC is essential for cell viability and DNA replication (Randell et al. 2006, Yeeles et al. 2015). One bound-ORC could be able to load several MCM double hexamers (Wu and Nurse 2009) or it could load MCM double hexamer, unbind and move to a different location before loading another one (Bowers et al. 2004, Das and Rhind 2016). The double hexamer could then slide along double-stranded DNA as it has been shown *in vitro* (Edwards 2002, Remus et al. 2009, Evrin et al. 2009, Gros et al. 2015) particularly by being displaced by transcription machinery (Gros et al. 2015).

Figure 1.1.8: Loading and firing of a replication origin



A. Licensing is the sequence of events that lead to the loading of MCM double hexamer during G1. After binding to DNA, ORC recruits Cdc6. This complex then recruits the Mcm2-7 helicase bound to CDT by binding the C-terminal regions. Sequential release of Cdc6 and Cdt1 results in loading of the first Mcm2-7 helicase. ORC then binds a second Cdc6, the interactions between the N-terminal domains of the two Mcm2-7s recruit the second helicase. Arrival of the second Mcm2-7 drives release of Cdc6, followed by Cdt1 and ORC. Figure 3 from S. Bell lab website (<http://web.mit.edu/bell-lab/>) modified from Ticau et al. 2015.

B. The second step occurs in S phase and consists on the activation of the Mcm2-7 complex to trigger its helicase activity. It requires several firing factors (Sld2, Sld7, Sld3, Dpb11, Cdc45, GINS and the DNA polymerase ϵ), and it is dependent on the activity of two kinases, DDK and CDK. Figure 6 from Yeeles et al. 2015 reused with permission.

A recent papers in *Drosophila* strongly supports the model of several MCM loaded by one ORC and displaced by transcription (Powell et al. 2015) and is consistent with observations in human cells (Petryk et al. 2016).

Process in S phase is mainly regulated by initiation

The helicase loaded in G1 are inactive and several steps are needed to form the pre-initiation complex that will unwind the DNA double helix and allow the recruitment of the replisome (Fig. 1.1.8 B.) (Wu et al. 2014).

The Dbf4-dependant kinase Cdc7 (DDK) phosphorylates multiple MCM subunits (2, 4 and 6) leading to the recruitment of Cdc45 and of Sld3 and Sld7 (Yabuuchi et al. 2006). Once phosphorylated by cyclin dependent kinases Clb5-Cdc28/Cdk2 (CDK) (Takeda and Dutta, 2005), Sld3 and 7 proteins are stabilised as a complex with Sld2 and the tetrameric GINS complex (Takayama et al. 2003) by the adaptor protein Dpb11 (Yeeles et al. 2015). The tight association between MCM, Cdc45 and GINS forms the CMG complex and allows the helicase activity (described in 1.1.1). The opening of the double helix induces the recruitment of the replisome and the processive replication by polymerases (described in 1.1.1). Other factor such as MCM10 are known to be necessary for initiation without an identified attributed role (Perez-Arnaiz et al. 2016).

Once formed, the replisome, also called replication fork, progresses at constant speed (Sekedat et al. 2010, Duzdevich et al. 2015) and stops when encountering another replication fork. Helicases are principally attached to the leading strand and therefore they are approaching each other on opposite strands in case of fork convergence. CMG is dismantled only after fork converged and the nascent double strands is fully ligated (Dewar et al. 2015). The CMG is therefore unloaded from double-stranded DNA by a poorly understood process that involves polyubiquitylation (Moreno and Gambus 2015).

Keeping the two steps apart

The strict separation of licensing and firing is crucial to avoid re-replication as no new pre-RC should be loaded on already replicated DNA. This insulation is allowed by a temporal regulation of anaphase-promoting complex (APC) and cyclins. Cyclins levels do not only regulate the DNA replication initiation but all cell cycle processes (Bloom and Cross 2007) by activating and often giving substrate specificity to partners cyclin-dependant kinases. In short, G1-phase cyclin-CDKs (G1-CDKs) phosphorylate proteins to promote S-phase entry, S-phase cyclin-CDKs (S-CDKs) are required to activate DNA replication, and mitotic cyclin-CDKs (M-CDKs) accurately regulate chromosome segregation through mitosis (Fig. 1.1.9). Regulation of DNA replication is reviewed in DePamphilis et al. 2012

and Siddiqui et al. 2013.

Briefly, G1 phase is characterised by a high level of APC activity which is promoting the degradation of Dbf4 and a low level of CDKs activity thanks to a cyclin inhibitor (CKI). The phosphatase Cdc14 has multiple roles such as protecting CKI from degradation, inducing S/M cyclins degradation and promoting the transcription of CKI and *cdc6*.

While the preRC formation is promoted, initiation is inhibited by the phosphorylation of Sld3 and Dbf4 (Zegerman and Diffley 2010). The G1/S transition, is marked by the removal of the inhibition on important transcription factors and the subsequent increase of G1/S-cyclins as well as replication proteins. The firstly produced G1-CDKs will counteract the effect of Cdc14 and promote the degradation of CKI. Therefore Sld2 and Sld3 can be phosphorylated to promote initiation (Zegerman and Diffley 2007). The diminution of APC/C activity then induces a burst of Dbf4. Concomitantly, Cdc6 and Cdt1 are degraded to prevent formation of new pre-RC.

Despite the prevention mechanisms, re-replication can occur and triggers activation of a checkpoint (Arias and Walter 2007). Re-replication is probably sensed by colliding forks that induce the formation of double strand breaks (Davidson et al. 2006). Depending on the cellular background, this leads to the activation of checkpoint pathways inducing G2/ M arrest, senescence or apoptosis (Truong and Wu 2011).

1.1.3 Conservation of DNA replication

Origin of replication from procaryotes to eukaryotes

The first replication model, called replicon, was set in 1963 by Jacob and Brenner. In the replicon model, replication was regulated by a positive interaction between an initiator protein, which is encoded by a structural gene, and a specific

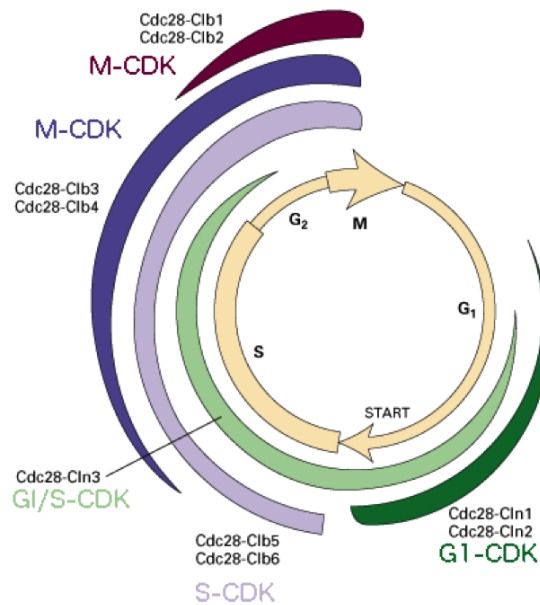


Figure 1.1.9: Regulation of budding yeast cyclins during the cell cycle

genetic element of recognition, which is known as the replicator. The model was proposed for bacteria that duplicate their circular chromosomes from a single position. Although in eukaryotes, DNA replication starts from several positions, the replicon model was adapted by considering the interaction of an initiator produced in more important quantity to several possible replicators.

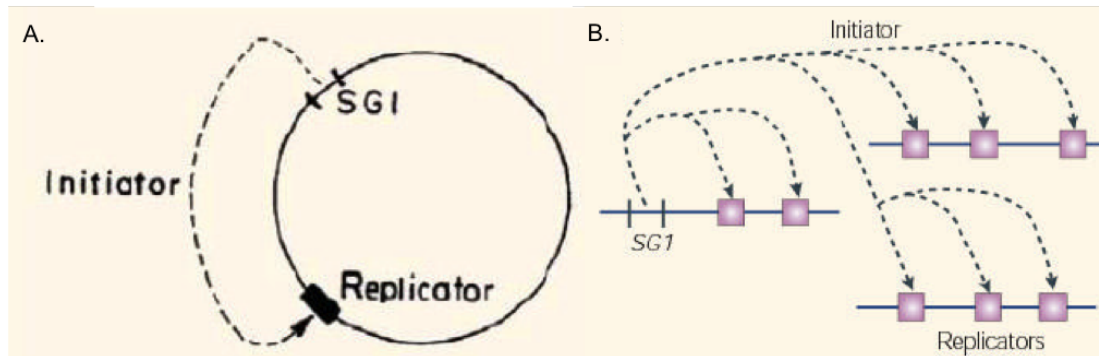


Figure 1.1.10: Replicon model

A. In circular genome of prokaryotes, an initiator is produced from a gene to activate the single replicator. B. This model can be adapted to eukaryotes that would possess multiple replicators all along their genomes. Figure 1 from Gilbert 2004 reused with permission.

Indeed, sequences that can give the ability to initiate the replication to a plasmid called autonomously replicating sequence (ARS) could be considered as replicator. Such sequences have been isolated from the *Saccharomyces cerevisiae* (*S.c*) (Struhl et al. 1979) and *Schizosaccharomyces pombe* (*S.p*) (Maundrell et al. 1988) genomes. In *S.c*, ARS elements are ~100 bp in length and are characterised by a required consensus sequence (5'-[A/T]TTTAT[A/G]TTT[A/T]-3') (Theis and Newlon 1997) as well as two redundant, non-conserved, auxiliary elements (Weinreich et al. 2004) called ARS-consensus sequence (ACS). The initiator has then been identified as ORC, that binds to this consensus sequence. However in other eukaryotes and despite years of investigation, no consensus sequences have been found for ORC binding. Latest studies also reveal that even in *S.c* ORC is able to bind unspecific sequences (Hoggard et al. 2013) but has a strong affinity for identified consensus sequence (Duzdevich et al. 2015) and the residues required for binding are conserved in all eukaryotes (Kawakami et al. 2015). Moreover, cdc6 binding could also be important in the recognition of the sequence (Speck and Stillman 2007) and could therefore help the loading of helicase at particular locations only, consistent with a poor efficiency of ORC non-specifically bound compared to ARS-bound ORC (Duzdevich et al. 2015).

The positions of initiation are called origins of replication. However, as the term was chosen based on the first replicon model it is now vitiated by some ambiguity. In order to remove this confusion, I distinguish three concepts often

equally referred as replication origins in the literature :

- ORC-binding sites : genomic sequences/loci where ORC binds the DNA and starts the sequential recruitment of the replication machinery (other site of ORC binding that do not lead to the recruitment of replication factors are not considered in this report)
- Potential origins : positions that are competent for the initiation of DNA replication (which, according to previous discussion, corresponds to double hexamers of MCM)
- Origins : positions where replication actually started in a particular cell during a particular cell cycle

Indeed, we saw in previous section that MCM are much more abundant than ORC therefore all potential origins as I termed them are not strictly ORC-binding sites. Moreover, as it will be detailed in the next chapter, not all potential origins fire in each cell at each cell cycle.

Conservation of the molecular machinery and models

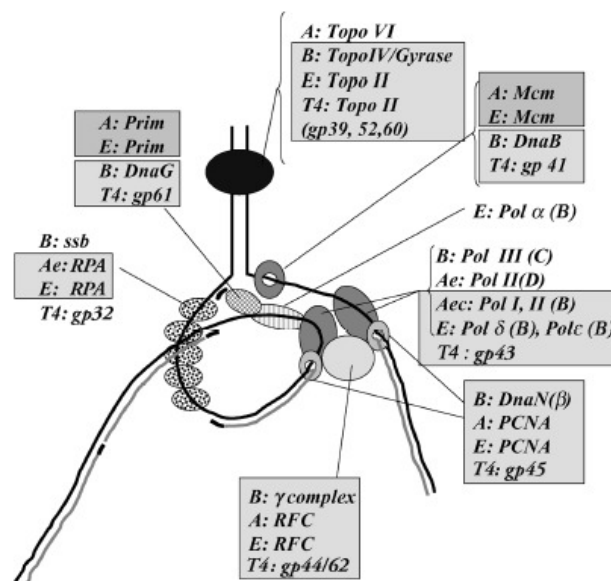


Figure 1.1.11: The universal replication fork

A= Archaea (Ae=euryarchaea, Ac=crenarchaea), B=Bacteria, E=Eukarya, and bacteriophage T4. Homologous proteins performing the same function are framed together. Letters in brackets indicate DNA polymerase families. Figure 5 from Forterre et al. 2013.

The replisome is particularly well conserved with strong a homology between its constituents in bacteria, archea and all eukaryotes (Fig 1.1.11) (Leipe et al. 1999).

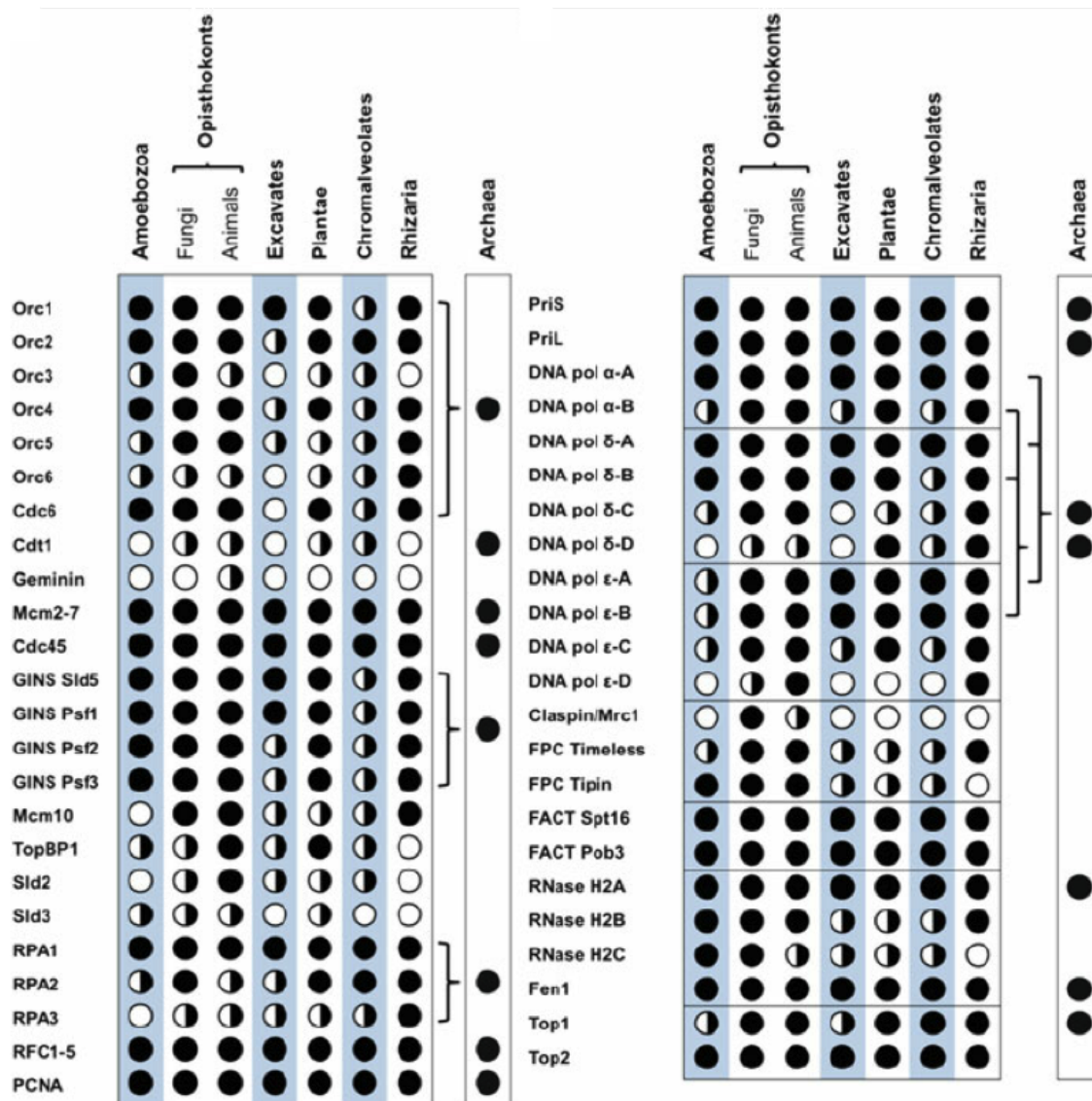


Figure 1.1.12: Distribution of DNA replication proteins across eukaryotic supergroups

Black dot indicates proteins present in all species; black/white dot indicates proteins present in some species; white dot indicates undetected proteins. Replication proteins with established archaeal homologues are indicated (final column: black dots). Lines connecting the eukaryote rows to the Archaea rows indicate paralogue relationships. First column is dedicated for proteins involved in licensing while second describes proteins of initiation of DNA replication. Figure 2.2 reused from Aves et al. 2012 with permission.

The main particularity of the bacteria replisome is the clear coupling of leading and lagging strand synthesis that is granted by the physical interaction of the clamp loader with the helicase. Such an interaction has not been found yet in eukaryotes where MCM proteins are already surrounded by direct partners for activation, regulation and chromatin maintenance. However, as described in subsection 1.1.1, other coupling interactions are very likely to occur in eukaryotes as well.

Due to the major difference on the number of origins, some regulators of the licensing and initiation are not present in bacteria but exist both in archea and eukaryotes (Aves et al. 2012) (Fig. 1.1.12). The cell cycle regulation by CDKs is also conserved (Zegerman 2015).

In addition to unicellular organisms (e.g. *S.c*, *S.p*), several *in vitro* models have been developed from purified proteins providing precise and accurate information (Mehanna and Diffley 2012, Riera et al. 2014, Ticaud et al. 2015). Another long used *in vitro* system to study DNA replication is the *Xenopus leavis* eggs extract. Isolated nuclei from frog sperm in presence of cytosol extract from eggs perform a naturally synchronous, fast and total replication of their DNA (Gillespie et al. 2012). Moreover, other DNA can be replicated in the same cytosolic extract (Lebofsky et al. 2009).

The relevance of unicellular organisms such as the budding yeast to get general conclusions for all eukaryotes is debatable and has to be reserved to delimited areas. DNA replication is a field where simple model organisms have provided and will still provide precious insight. Indeed, in spite of some variations (Errico and Costanzo 2010) DNA replication is highly conserved in eukaryotes and beyond. Moreover unicellular organism and *in vitro* systems cost less in time and money and allow to set the hypothesis to be confirmed in higher organisms.

1.2 Replication of chromatin

1.2.1 Nucleosome as the structural unit of chromatin

DNA wrapping around histones proteins core

Several experimental observations such as the x-ray diffraction pattern of DNA (Wilkins et al. 1959), the ratio of histones proteins, the visualisation of a characteristic size on digested DNA (Hewish and Burgoyne 1973), as well as beads on a string aspect of first electron microscopy images of DNA, suggested the presence of a structural unit (Kornberg 1974). This “coiled DNA molecules jointed by histones bridge” (Wilkins et al. 1959) was named nucleosome (Oudet et al. 1975) and

its structure was later resolved with high resolution (Luger and Richmond 1998).

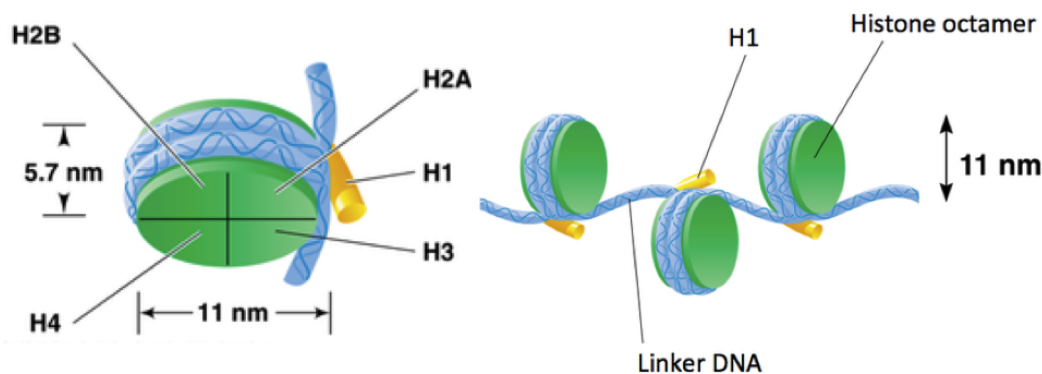


Figure 1.2.1: Nucleosome composition

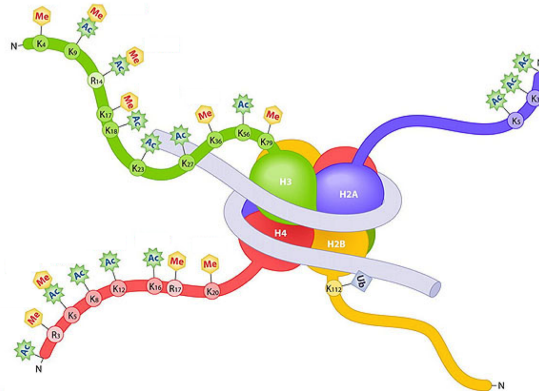
Nucleosome is composed of an octamer of histones (two repeats of histones H2a, H2b, H3, H4) encircled by 1.7 fold of DNA which corresponds to 146bp forming a left-handed helix and a free DNA part called linker whose size varies between 20 and 100 bp (McGinty and Tan 2015). The eight-histones core forms a flat cylinder of 10nm diameter and 6nm thickness. Histone proteins have a particular structure with a domain called histone fold which is highly conserved and less structured tails. These tails are not involved in the DNA stabilisation in the nucleosome but mostly intervene in interaction with neighbouring nucleosomes and other proteic partners. Histones tails can carry post-translational modifications (PTM) which thus have an influence on the level of compaction of chromatin.

Post-translational modification of histones

At least 8 classes of chemical modifications have been identified, each of which could exist on several possible positions (Kouzarides 2007). The best described and characterised are the acetylation and methylation that can be deposited and removed by respective enzymes : histone acetyltransferase (HAT), histone deacetyltransferase (HDAC), histone methyltransferase (HMT) and histone demethyltransferase (HDM) (Cortini et al. 2015).

Acetylation of a lysine which is possible on all core histones but more frequent on lysine 16 of H4 (H4K16ac), neutralises partially the positive charges of histones and contribute to a diminution of their affinity for DNA leading to more open region associated with genes transcription (Rice and Allis 2001). On the contrary, mono-, di- or tri- methylation of lysines and arginines give more complex result with some modification being linked to open chromatin (H3K4me(1/2/3), H3K36me(1/2/3), H4K20me1, H3R17me(1/2)) and other to repressive or closed

Figure 1.2.2: PTMs on histone tails



state (H3K9me(1/2/3), H3K27me(2/3), H4K20me3, H3R2me2). Notably, H3K9me3 leads to the recruitment of HP1 and H3K27me3 of polycomb architectural complex which both have further role in the chromatin silencing.

Although a tremendous diversity of marks is possible, only few combinations actually exist in vivo with some histone modifications being always associated together giving only few chromatin types (Filion et al. 2010, Roudier et al. 2011, Ernst et al. 2011).

Nucleosomes positioning on DNA chain

Besides being a packaging unit, nucleosomes play a crucial role in cellular process by regulating DNA accessibility. The position of nucleosome is therefore essential and must be regulated, but the global factors regulating nucleosome positioning are still unclear. Nucleosome mapping revealed that DNA sequences is important to determine nucleosome position due to differential histone-DNA affinity and physical properties of the sequence such as flexibility (Vailant et al. 2007).

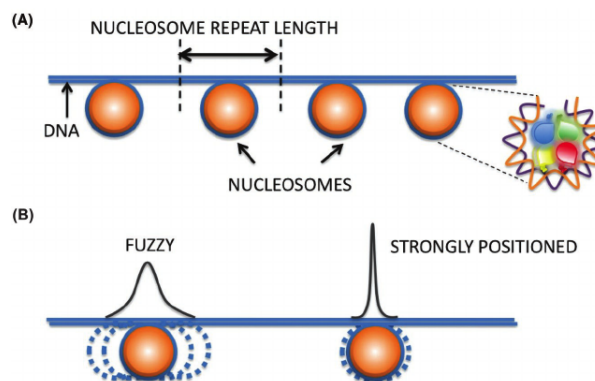


Figure 1.2.3: Lexicon of nucleosome positioning
Schematic of nucleosome locations on DNA depicting the nucleosome repeat length (A), nucleosome fuzziness (B). Figure 1 from Arya et al. 2010.

Despite a fuzziness of the nucleosome positioning (Cole et al. 2012), some areas

preferentially lack nucleosome and are referred as nucleosome free or depleted regions (NFR, NDR) while other are preferentially covered. *In vitro* and *in vivo* nucleosome positioning is significantly more similar than random expectation showing the contribution of DNA sequence (Struhl and Segal 2013). However at particular positions, such as gene promoters, *in vitro* and *in vivo* data diverge showing a strong implication of other factors. The action of proteins called nucleosome remodellers allows to reproduce these patterns but no mechanism is known so far. Indeed, these factors are able to displace nucleosomes and favour a thermodynamical equilibrium but are not able to specifically position nucleosome or address them to particular region (Arya et al. 2010).

In addition to positioning, the two remodellers Isw1 and Chd1 in yeast contribute to establish the nucleosome spacing. It has been observed that the linker has preferential length possibly due to the folding of nucleosomes into higher order structures that could impose nucleosomes spacing (Jansen and Verstrepen 2011, Grigoryev 2012).

1.2.2 Chromatin higher order organisation *in vivo* is still under debate

Nucleosome packing into 30nm fibre

The interaction of histone protein H1 at the entrance and exit of DNA in histones core allows the nucleosomes to get more compact configurations (Thoma and Koller 1977). Notably, a 30nm fibre is observed by electron microscopy and its internal structure has been debated since its first observation.

Two principal family of chromatin compaction models exist : the solenoid models (Finch and Klug 1976) where consecutive nucleosomes form a single helix while the linker DNAs are bent in the interior of the

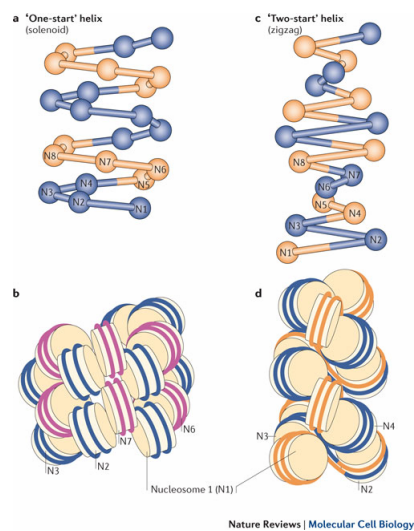


Figure 1.2.4: Two models for chromatin fibre

The solenoid model is characterised by interactions between consecutive nucleosomes ($n, n + 1$; a,b), whereas the zigzag model implies interactions between alternate nucleosomes ($n, n + 2$; c,d). Figure 4 reused from Luger et al. 2012 reproduced from Li and Reinberg 2011 with permission.

fibre and the zigzag (Williams et al. 1986) or cross-linker models which features two separate nucleosomal stacks with linker DNAs crossing the fibre core thus consecutive nucleosome in the helix are not directly stacked. Several experimental data support one model or the other (Szerlong and Hansen 2011) suggesting that DNA could adopt both configuration alternatively. Recent model on the zigzag family allows to reproduce the experimental features observed at various salt conditions and in particular the link between compaction and linker length (Song et al. 2014, Wu et al. 2016). The model shows interdigitating nucleosome in such fibre consistent with overlapping of nucleosomal territories (Chereji and Morozov 2014)

Existence of the structure *in vivo*

As it was extensively studied *in vitro*, the debate has been strong on the existence of such a structure *in vivo* (Van Holde and Zlatanova 1995, Fussner et al. 2011, Hansen 2012, Maeshima et al. 2010). The structure was believed to be an artefact due to non physiological salt conditions *in vitro* and to cell fixation and heavy staining protocol *in vivo*. Direct measurement by small-angle scattering (described in chapter 3) in isolated nuclei showed difference between cell types and were interpreted as the presence of a 30nm structure in some interphase nuclei such as chicken erythrocytes (Langmore and others 1983) and not in other such as HeLa cells (Maeshima et al. 2014). However, data on HeLa cells are consistent with the absence of 30nm fibre detected by cryo-electron microscopy (Eltsov et al. 2008) but not with recently observed fibre in several sample preparation protocols with electron tomography (Li et al. 2015).

Although the debate on the existence of 30nm fibre *in vivo* is not yet closed, the dogmatic view of all genome being folded into a rigid very well defined fibre is now reserved to textbooks. Conformation capture experiment (described in chapter 3) indeed described a more extended fibre with local increase in compaction in yeast (Dekker 2008). A recent experiment called EM-assisted nucleosome interaction capture (EMANIC) studies chromatin architecture in HeLa interphase nuclei and metaphase chromosome, showing the persistence of a zigzag motif but organised as open and irregular arrays. A less distinct diameter compared to canonical zigzag model would explain the lack of 30nm signal with other techniques. Moreover, the organisation of chromatin changes in metaphase with an increase of long range interactions, consistent with intra-fibre looping (Grigoryev et al. 2016).

Higher order compaction

Many chromatin functions require the gathering of distant genomic position by so called chromatin loops (Kadauke and Blobel 2009). Loops have variable stability

and it is not clear if they constitute a distinct hierarchical level (Woodcock and Ghosh 2010). Indeed, a model of regularly folded loops does not explain the high variability in distances between genomic loci (Trask et al. 1993, Berger et al. 2008) and the multiple possible contact between one locus to several other loci revealed by conformation capture techniques (Simonis et al. 2006, Lieberman-Aiden et al. 2009). Polymer forming random loops (Yokota et al. 1995) however provide a good model for global chromatin behaviour, showing the importance of polymer physics to describe chromatin (Fudenberg and Mirny 2012, Imakaev et al. 2015) as it will be discussed in last chapter of the introduction.

The fractal or crumpled globule model (Grosberg et al. 1993, Lieberman-Aiden et al. 2009) is an unknotted and out-of-equilibrium fractal structure. In budding yeast where the ratio between the nucleus size and the genome length is greater than in metazoans, this structure can reach an equilibrium to form an equilibrium globule (Mirny 2011, Wang et al. 2015). However, increasing resolution of the chromatin conformation capture technique (Rao et al. 2014) reveals that the fractal globule does not reproduce every feature of the human genome folding on every scale. Indeed, human chromosome fold into chromosome territories (Cremer et al. 2006) which are not accurately represented by the fractal globule.

This led to the construction of a new model called extrusion loop (Sanborn et al. 2015). In addition to a good consistency with chromosome contacts and domains formation (Fudenberg et al. 2016), it also provides a biological context by explaining the mechanism of loop formation with the involvement of CCCTC-binding factor (CTCF) and cohesin (Sanborn et al. 2015). This unknotted model also describes the compaction into

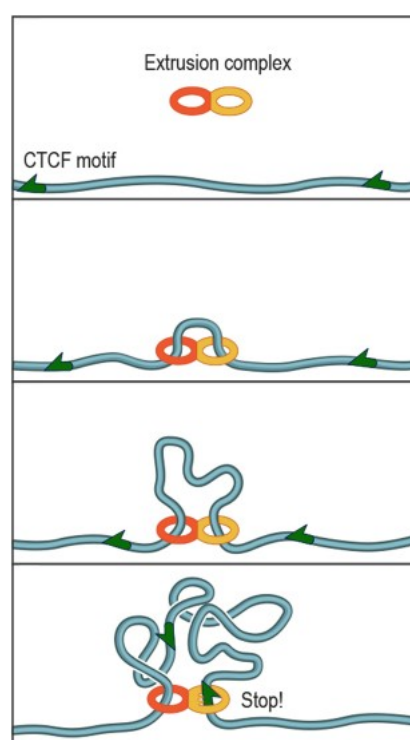


Figure 1.2.5: Loop extrusion model

Extrusion complex loads onto the fibre at a random locus, forming an extremely short-range loop. As the two subunits move in opposite directions along the fibre, the loop grows and the extruded fibre forms a domain. When a subunit detects a motif on the appropriate strand, it can stop sliding. The territory thereby formed include only one chromosome. Figure 5. A. from Sanborn et al. 2015.

chromosomes (Goloborodko et al. 2016b) and the chromatide segregation during cell division (Goloborodko et al. 2016a). Overall this new model is an exciting view of active genome organisation (Yardımcı and Noble 2015) already referred as loopy globule (Grosberg 2016).

1.2.3 Chromatin is reconstructed as DNA is replicated

Nucleosomes have to be removed to allow passage of the replication fork and have to be replaced on the two newly formed DNA strands. The cell seems to have found an efficient way to organise this reassembly as chromatin is almost immediately reconstructed after the passage of the fork (Sogo et al. 1986) which means that two new nucleosomes are assembled every 5 to 10 s. Moreover, contrary to DNA replication, no template is available to reproduce the nucleosome distribution and post-translational modifications, which raise a lot of questions tackled in several reviews (Margueron and Reinberg 2010, Alabert and Groth 2012, MacAlpine and Almouzni 2013).

New nucleosome assembly after replication

DNA replication disrupts nucleosomes in a close vicinity of the fork possibly as a result of a collision with the replicative helicase or due to supercoiling ahead of the fork (Alabert and Groth 2012). H2A and H2B have a dynamic behaviour during the entire cell cycle with an important turn-over in nucleosome composition (Kimura and Cook 2001). H3-H4, however, are much more stable and a segregation occurs with parental H3-H4 tetramers being repositioned behind replication fork and other synthesised *de novo* (Riley and Weintraub 1979). Chromatin assembly factor 1 (CAF-1) acts as histone chaperone and promotes deposition of histones H3-H4 onto replicating DNA (Smith and Stillman 1989). Antisilencing function 1 (ASF1) also acts as a chaperone to facilitate chromatin assembly via interaction with CAF-1 (Tyler et al. 1999, Mello et al. 2002). The subsequent addition of histones H2A – H2B involves the nucleosome assembly protein 1 (NAP1) chaperone (Zlatanova et al. 2007) and the FACT (facilitates chromatin transcription) complex that acts as an H2A – H2B chaperone in many cellular processes.

To couple histone deposition with DNA synthesis, CAF-1 interacts with the replication fork through PCNA (Moggs et al. 2000, Zhang et al. 2016) while MCM-2 binds parental or new H3-H4 and seems to act as a platform to keep H3-H4 in close proximity of the fork where they are taken care of by ASF1. H3-H4 are both recycled from parental nucleosome and synthesised *de novo* (Alabert et al. 2015). Experimental data do not allow to solve whether parental tetramers are split to be completed by *de novo* H3-H4 dimers or not (reviewed in Annunziato 2015). The resolution of the structure of the complex MCM-H3-H4-ASF1 reveals that H3-H4

can bind MCM as a tetramer or a dimer in presence of ASF1 (Richet et al. 2015, Huang et al. 2015) suggesting a mechanism for both mixed or direct re-association (Clément and Almouzni 2015).

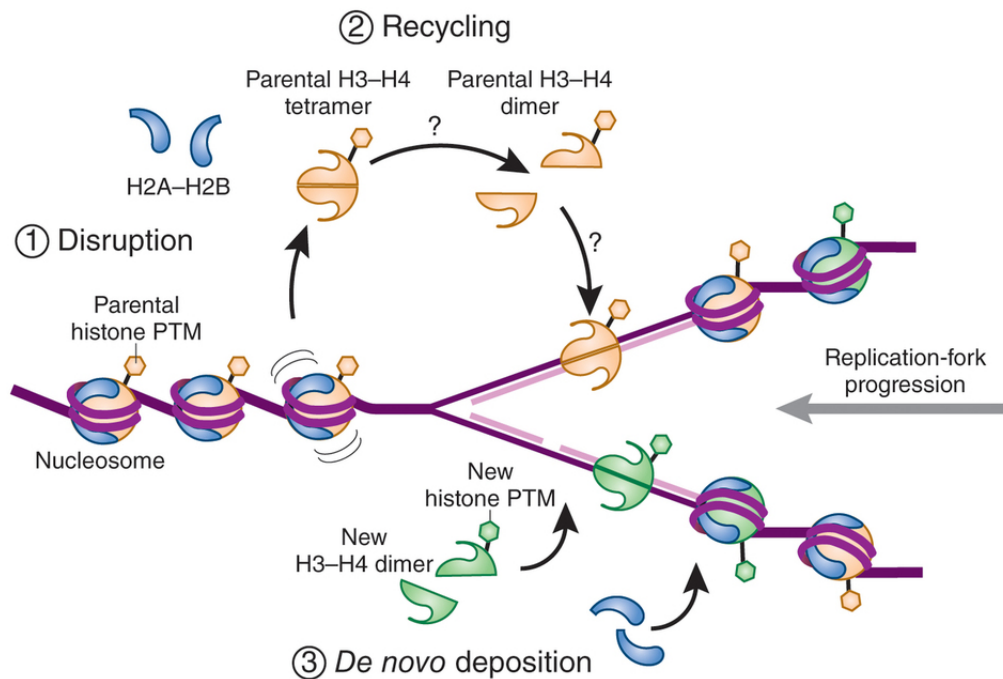


Figure 1.2.6: Histone reassembly after replication fork passage

For each parental nucleosome disrupted by replication-fork passage (indicated by the grey arrow), an H3–H4 tetramer or two H3–H4 dimers are made available. The histones are in turn recycled on newly synthesised DNA either directly as a tetramer or as two dimers. New histones are deposited on nascent DNA to ensure a full complement of nucleosomes on the nascent DNA. Recycling of parental histones and de novo deposition are thought to occur randomly on both the leading and the lagging strands. Here, for clarity, de novo deposition is depicted on the bottom strand. Figure 1. from Clément and Almouzni 2015 reused with permission.

Redeposition of post-translational marks

Replication also disturbs the post-translational modifications (PTM) pattern. Indeed, the deposition of preRC promotes the acetylation of neighbouring nucleosomes in order to grant access to the replication machinery (Liu et al. 2012). Moreover, DNA replication progression also creates a wave of acetylation upstream of the fork (Bar-Ziv et al. 2016).

PTM on recycled histones are conserved when they are incorporated to new nucleosomes (Alabert et al. 2015). New histones however, present PTM specific of pre-chromatin assembly such as diacetylation for H4 (Allis et al. 1985) and acetylation and monomethylation (H3K14ac, H3K18/K23ac, H3K9me1) for H3 (Loyola et al. 2006) and need to be modified. In both human and budding yeast,

new histones acquire most their PTM without delay to mirror the parental histones suggesting a replication coupled mechanism (Bar-Ziv et al. 2016, Alabert et al. 2015). However in both organism, trimethylation marks (H3K9me3 and H3K27me3) are not immediately established and show a significant delay spreading over several cell generations. The level of methylation is however maintained by the combined methylation of both new and old histones (Alabert et al. 2015).

Nucleosome positioning and initiation

Moreover, positioning of nucleosome already described in a previous subsection is disrupted by the formation of pre-RC complex and must also be reestablished after the fork passage. ORC, as well as other DNA binding factors, have the ability of changing the nucleosome positioning (Lipford and Bell 2001, Wasson and Hartemink 2009, Belsky et al. 2015). It has been observed that ORC binding causes the NFR to widen and the adjacent nucleosome to be much strongly positioned (Eaton et al. 2010, Berbenetz et al. 2010). In addition, recent findings about MCM sliding *in vivo* (Powell et al. 2015) suggest that they require a displacement or rearrangement of neighbouring nucleosomes (Das and Rhind 2016). The main hypothesis is that nucleosomes rebind after letting MCM slide, which is supported by the tight association of well positioned MCM and origin-flanking nucleosomes (Belsky et al. 2015).

It has been observed that not only the nucleosomes are very rapidly redeposited on new DNA but their position is also restored extremely fast (Lucchini et al. 2001, Yadav and Whitehouse 2016). However, *in vitro* replication with purified histones chaperones fails to reestablish nucleosome spacing and periodicity (Struhl and Segal 2013). Study of nucleosome deposition on Okazaki fragment shows that the repositioning of the nucleosome is an ATP-dependant process directly coupled with the loading of the histones by a cooperation with histones chaperones (Yadav and Whitehouse 2016).

Although much is still unknown about the dynamic reassembly of chromatin after fork passage, the process seems to be both fast and efficient to replace nucleosomes in their correct positions with appropriate PTMs. Overall this part describes a very well regulated DNA replication process with a huge cohort of proteins involved in ensuring its correct proceeding as well as intervening on possible hitches.

Chapter 2

4th dimension of DNA replication

As described in chapter one, DNA replication is enabled in the three dimensional space by multiple proteic partners that guarantee the faithful reproduction of the linear DNA composition, as well as the correct reorganisation of the nucleosomes after the passage of the replication fork. In eukaryotes, DNA replication starts from multiple sites and the orchestration of the initiation of DNA replication at these different sites also needs to be regulated.

This chapter explains how the DNA replication is organised through time. After introducing the characteristics of the timing of DNA replication, few mechanisms proposed as regulators of this timing are described and the potential benefit for the cell are discussed.

2.1 Spatio-temporal program of DNA replication

2.1.1 Replication profile

Temporal program

All origins do not fire at the same time. Instead, different part of the genome are replicated at different time during S phase. This observation lead to hypothesise that this sequential duplication was regulated, forming a temporal program (Taylor 1960). First observations were made around active genes and showed that they replicate early during S phase (Goldman et al. 1984, Hatton et al. 1988) suggesting some functional significance for this temporal program.

The first map of the timing of replication of genomic positions was realised in yeast. Identified ARS could then be sorted as early or late according to the time at which they are duplicated during S phase (Fangman and Brewer 1992). Such an approach was not so successfully applied to other species than yeast (reviewed

in Raghuraman and Brewer 2010). However, new techniques developed during the last fifteen years allowed genome wide mapping of the replication timing for many organisms (Yabuki et al. 2002, Schübeler et al. 2002, Woodfine 2003, review in Rhind and Gilbert 2013).

Contrary to what the terminology implies, initiation does not happen in two waves of early and late but more as a continuum through S-phase (Bechhoefer and Rhind 2012). In yeast, the notions of early and late origins now refers respectively to potential origins that fire or not in presence of hydroxyurea (HU), that blocks the dNTP production and the S phase progression (Alvino et al. 2007).

Constructing replication profiles

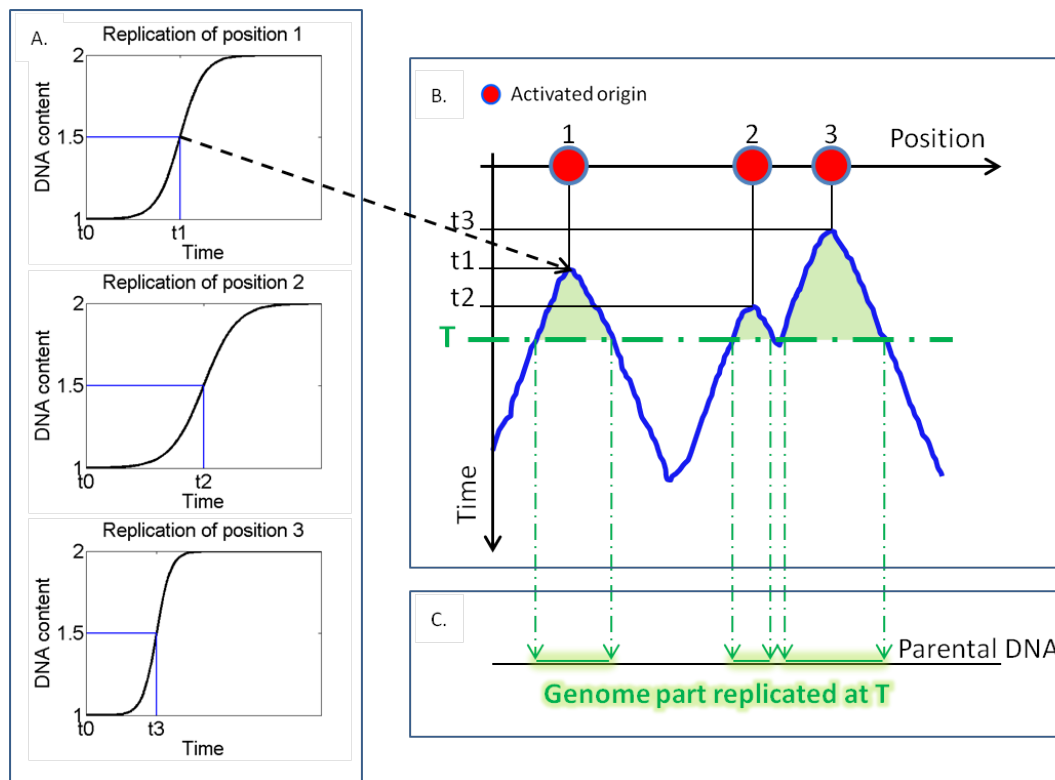


Figure 2.1.1: Construction of a replication profile

A. DNA content of genomic position as a function of time. The sigmoidal shape of the curve is due to small asynchrony and variability in the firing time. This variability increases for late origins for example 2 compare to 3. B. Replication profile. The time at which a position is half replicated is used to construct the replication profile. Measures on more positions result in higher resolution profiles that are usually smoothed by a mobile mean. The time axis is inverted to represent early replicating regions as peaks. C. The genome parts replicated at a time T on average in a cell population can be deduced from the replication profile. As a comparison these portions can be also measured in single cells by DNA combing.

In order to obtain the population average replication timing of a position, cells must be synchronised in the cell cycle. A useful method in budding yeast is to block the cells somewhere in the G1-phase of the cell cycle with the pheromone α -factor. This, as well as other synchronisation methods, leads the population of cell to enter the S-phase in a synchronous way. However, the synchrony is not perfect and the percentage of replication of a position with time is not a pure switch from 0 to 1 but a sigmoid curve (Fig. 2.1.1 A.). Many plausible causes can induce synchrony defects such as a variability in the blocking efficiency and the dynamics of cell cycle re-entering. Furthermore, the synchrony is influenced by experimental settings.

In addition to the lack of synchrony, it has been measured that late replicating regions have a higher variability in their firing time (i.e. a flatter sigmoid curve) (Yang et al. 2010). Replication profiles represent the time after the beginning of S phase at which a position is duplicated in 50% of a cell population as a function of the genomic position (Jeon et al. 2005). The time axis is inverted with the start of S phase being the maximum and the end of S phase the origin of the axis (Fig. 2.1.1 B.).

Informations on replication profiles

Such a representation provides direct information on the average behaviour of replication program.

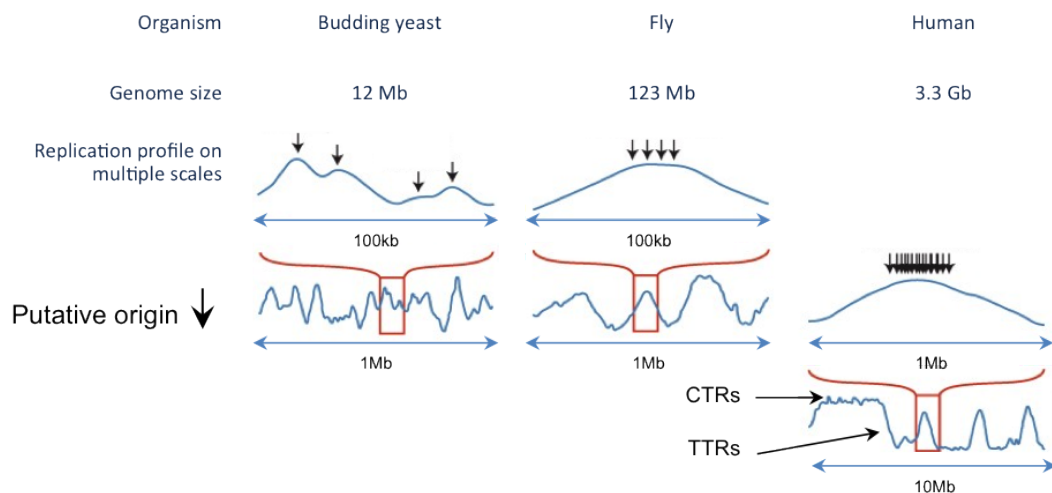


Figure 2.1.2: Replication profiles of several eukaryotes

Eukaryotes replication profiles present similar pattern but on different scales to match the genome size. In yeast, peaks correspond to one defined ORC-binding site. In higher eukaryotes, broader regions of initiation probably encompass many origins and form constant timing regions (CTRs) delimited by transition timing regions (TTRs). Figure adapted from Rhind and Gilbert 2013.

A peak corresponds to a point replicated before neighbouring regions while a valley indicates termination. The descent can be due to simple fork progression which mean that the slope would be dependent on the fork speed or by a biased polarity of the forks that can be due to a domino-like initiation of origins (Fig. 2.1.6) (Guilbaud et al. 2011). In budding yeast, rather sharp peaks are correlated with defined ORC-binding positions (ARS) whereas in higher eukaryotes broad constant timing regions (CTRs) also called replication domains (Hiratani et al. 2008) are separated by progressive slopes called timing transition regions (TTRs) (Fig. 2.1.2) (Rhind and Gilbert 2013).

Despite being informative on a global behaviour, replication profiles do not allow to decipher the behaviour of single cells.

2.1.2 From defined replication profiles to cell to cell variability

First picture of the temporal program of DNA replication was highly deterministic. The notion of temporal program indeed suggests that defined positions are always duplicated in the same order, even in presence of perturbations (Alvino et al. 2007, McCune et al. 2008, Koren et al. 2010) and this temporal program was supposedly transmitted to daughter cells (Raghuraman et al. 2001, Knott et al. 2009). However multiple evidences suggest a more complex picture.

Origin efficiency

In 2D gels assays, restriction fragments containing replication intermediates are resolved according to their mass and shape and allow to discriminate between replication bubbles, due to an initiation event and fork signals synonym of a passive replication of the region by a fork incoming from a neighbouring origin (van Brabant et al. 1998). These assays allowed to discover that regions that act as origins in some cells are actually passively replicated in other (Friedman et al. 1997, Yamashita et al. 1997), suggesting that different cells use different cohorts of origins. This has been confirmed later by sequencing Okazaki frag-

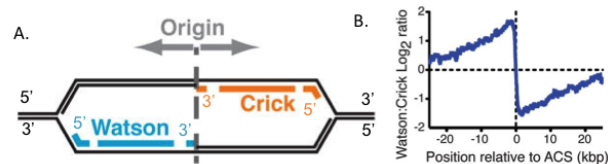


Figure 2.1.3: Identification of origins of replication by transition density of Okazaki fragments

A. From an origin of replication Okazaki fragments are synthesised to form the Watson (5' to 3') and Crick (3' to 5') strands. Repression of DNA ligase allows to retrieve short DNA fragments that can be sequenced to quantify the proportion of Watson or Crick fragments. B. Ratio of reads (usually in log base 2) from Watson and Crick strands shows that the strand composition is particularly biased around replication origin. Figure 2 b and c from Smith and Whitehouse 2012 reused with permission.

ments. Indeed, the density of Okazaki fragments on each strand presents a transition at the initiation site (Fig. 2.1.3) (Smith and Whitehouse 2012). Therefore it allows to detect initiation and termination sites by identifying the fork directionality (McGuffee et al. 2013).

Origin efficiency is the proportion of cells in a population in which of each replication started at a given position. Due to the ambiguous usage of the term origin (discussed in the first chapter), the definition of origin efficiency is also equivocal, as it seems to refer to the ability of a potential origin to fire before being passively replicated. However, this requires that all cells possess a potential origin in this position which has not been verified. The notion of competence, defined as the ability of a given region to load and maintain a viable preRC, need to be introduced (Fig. 2.1.5 A.). Thus, the efficiency of a position is both influenced by its competence and the ability of this possibly potential origin to fire before being passively replicated (Raghuraman and Brewer 2010).

Single cell analysis of DNA replication variability in origin usage

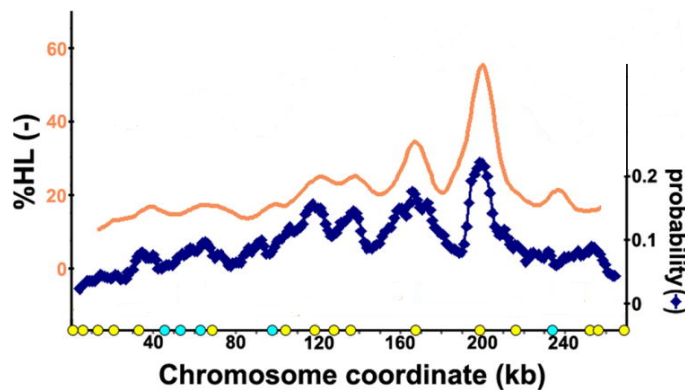


Figure 2.1.4: Average behaviour of single cells recapitulates the population behaviour

Proportion of replicated DNA as a function of chromosomal coordinate obtained from a cell population 10 min after release in S-phase by Alvino et al. 2007 (orange line, scale on the left) (%HL means percentage of heavy-light which correspond to replicated DNA in density transfer method comparable with the meselson experiment described in first chapter). On the same plot, the probability to be replicated measured on individually combed chromosomes by Czajkowsky et al. 2008 at the same time in S phase (~10 min) (blue dots, scale on the right) shows a similar pattern. Figure 2 from Czajkowsky et al. 2008 adapted with permission.

Fibre analysis have provided great insight on the origin usage of single cells (Tuduri et al. 2010), in particular DNA combing, that consists of stretching DNA fibres on silanized coverslips to be visualised (Michalet et al. 1997). The incorporation of nucleotides analogs during replication in vivo allows to specifically label with fluorescent probes the positions replicated at a given time (Herrick et al. 2000,

Lengronne et al. 2001). These experiments showed that despite the defined firing times on average population, different subset of origins are activated in each cell and with a different timing pattern (Patel et al. 2006, Czajkowsky et al. 2008). Moreover, the average of single-cell signal recapitulates the population replication profile (Fig. 2.1.4) (Czajkowsky et al. 2008, Rhind et al. 2010) which proves that the result both on population and single cell are consistent.

In addition to not being used by all cell in a population, the same cohort of origins is not used in two consecutive cell cycles (unpublished data commented in Tuduri et al. 2010).

2.1.3 Simulations and mathematical models to interpret the replication profiles

This variability, as well as the difficulty to decipher the parameters responsible for efficiency (termed identifiability in de Moura et al. 2010) (Fig. 2.1.5 A.), makes the use of mathematical models and simulation extremely valuable (Hyrien and Goldar 2010). Moreover, in addition to investigating the parameter of a given position, models allow to propose and test several hypothesis on the dynamics of the regulation of initiation. With this tools, parameters can be set in order to confront the simulated data to experimental ones.

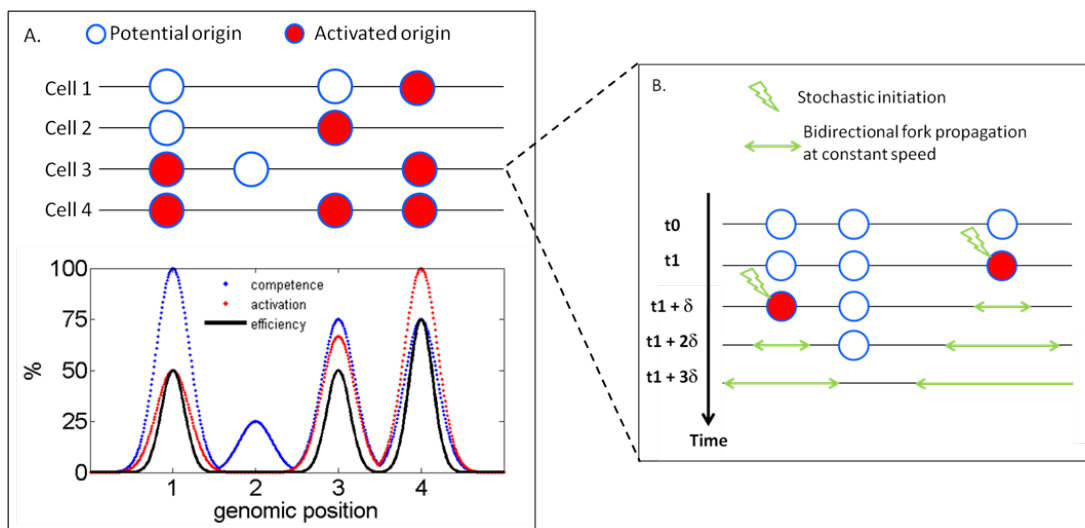


Figure 2.1.5: Identifiability

A. Schematic view of efficiency in a population of four cells. Not all four positions are competent to initiate replication in all cells. In addition some potential origins are not activated therefore two positions with different competence and frequency of activation can have the same efficiency.

B. Given the position and the efficiency of each position, mathematical models recapitulates the replication timing by assuming stochastic initiation, constant fork velocity and an increase of an increase of firing probability with time.

Theoretically, the replication process resembles a nucleation and growth process in one dimension of a crystal (Jun and Bechhoefer 2005, Jun et al. 2005, Retkute et al. 2011, Retkute et al. 2012). Indeed, nucleation can be assimilated to initiation, growth speed to fork velocity and coalescence to termination. The rate of initiation or frequency of initiation, computed as the number of initiation per time and per unreplicated length, is a crucial parameter of the temporal program of DNA replication (Herrick et al. 2002) because the only knowledge of the initiation rate and the fork velocity allows to reproduce temporal DNA profiles of DNA replication.

Stochastic execution of the program

First, numerical simulations confirmed that the fork velocity is independent of time in yeast (Yang et al. 2010) and for the main part of metazoans (Gauthier et al. 2012) in agreement with experimental data (Sekedat et al. 2010, Guilbaud et al. 2011, Chagin et al. 2016). Assuming that the termination does not slow the fork progression, and that the fork velocity is constant and loci-independent, the kinetics of initiation characterise the whole temporal program.

The simulation of the stochastic initiation of positioned potential origins of defined efficiency allowed to reproduce the replication profile of the budding yeast chromosome VI (de Moura et al. 2010) and of the *S.p* genome (Lygeros et al. 2008). Another stochastic model recapitulates the initiation rate in *S.c* as well as the firing time distribution (Yang et al. 2010) inferred from microarrays data (McCune et al. 2008). To be able to reproduce the data, these models need that the probability of initiation increases with time to insure that inefficient late positions will be replicated and prevent large gaps from not being replicated on time. This has been biological linked with the presence of limiting factors of initiation (Goldar et al. 2009, discussed in next section) that could be recycled and/or become more abundant.

The spatiotemporal program of DNA replication is entirely described by the position and efficiency of excess potential origins. However, these models did not investigate how is the efficiency of a position defined.

Influence of neighbouring initiation

In addition, these models were assuming that initiation of two distinct positions was not correlated (Bechhoefer and Rhind 2012). Other models suggest that neighbouring initiation or fork progression can increase locally the probability of initiation (Maya-Mendoza et al. 2010, Ma et al. 2012, Guilbaud et al. 2011, Löb et al. 2016) in a domino-like model. Domino-like models allow to create clusters of timing observed in many eukaryotes and it reproduces well the temporal pro-

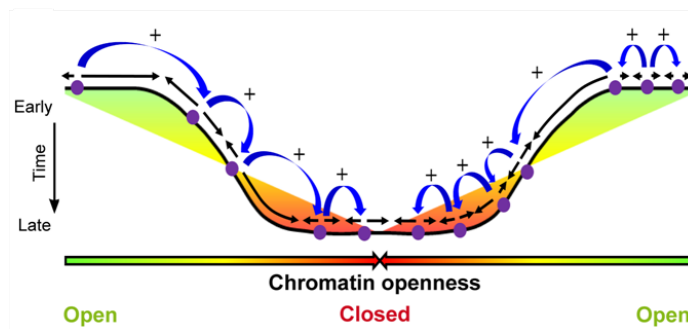


Figure 2.1.6: Illustration of the domino-like initiation in human TTRs
 The slope between two constant timing regions could be due to a cascade activation. Figure 12 from Guilbaud et al. 2011 reused with permission.

gram without any assumption on the position or efficiency of individual origins. In order to also reproduce the spatial organisation of the replication timing, some assumption have to be made on the causes of inhomogenous efficiency.

Modelling inhomogenous efficiency

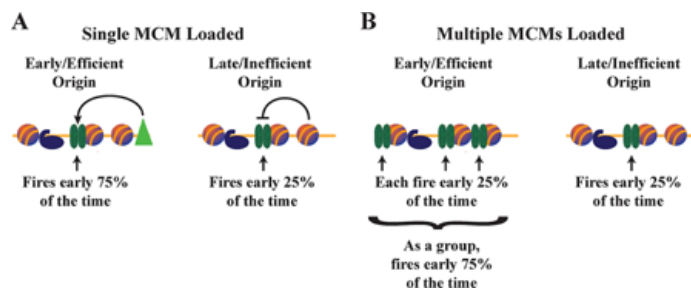


Figure 2.1.7: Abundance of potential origin as an explanation for inhomogenous efficiency

A. In the model of single MCM loading, a transacting factor is necessary to explaining inhomogenous efficiency of potential origins by promoting or repressing the initiation.
 B. In the multiple MCM loading model, no molecular mechanism is necessary to induce the inhomogenous efficiency as an accumulation of MCMs around a defined position leads to a higher local efficiency.
 Figure from Das and Rhind 2016 reused with permission.

Models that account for the variable efficiency of genomic positions show that the main determinant in human cells is the spatial distribution of initiation sites (Gindin et al. 2014) that was not imposed from available experimental data of potential origin positions but inferred from genomic landmark and more specifically from DNase-hypersensitive sites. A differential probability in open or close chromatin compartment also allows to reproduce the global replication profile pattern of HeLa cells (Löb et al. 2016).

It has also been shown that more MCM are loaded at early compare to late origins in budding yeast (Das et al. 2015). This suggests a common mechanism where potential origins in late or early parts of the genome have the same probability to fire but a much greater probability of a encounter between a limiting factor and potential origin in a early part of the genome due to a much higher density of potential origins (Rhind 2014). This elegant hypothesis has not been extensively investigated yet and other work did not find any correlation between MCM density and timing (McGuffee et al. 2013, Belsky et al. 2015). Moreover, some late origin are very efficient (Ferguson et al. 1991) suggesting a combination of several mechanisms of regulation.

Global chromatin organisation

Instead of focusing on local properties of chromatin a recent mathematical model proposed to consider only the global conformation of the chromatin in the nucleus as a frame in which initiation factors can diffuse (Goldar et al. 2016). Initiation occurs when an initiation factor encounters one of the potential origins that are randomly distributed on the genome. In this minimal model the probability of initiation depends on the global chromatin conformation, defined as a polymer globule, and on the random diffusion of the initiation factor. With this model, fork density and initiation rate can be reproduced for both yeast and HeLa cells (Goldar et al. 2016) suggesting that the spatial conformation of chromatin strongly constrains the kinetics of replication in all eukaryotes.

Eukaryotes genome is duplicated according to a spatio-temporal program which creates replication profiles characteristic of a cell type population. Models and simulations are compatible with a stochastic execution of the temporal program defined by random initiation, propagation of fork at a constant speed and termination mainly induced by fork encounter. In order to duplicate all DNA in a limited time, probability of initiation must increase during the S phase to allow potential origin to fire in late replicating parts of the genome. Frequency of the usage of one position as origin in a population (efficiency) depends on the competence of this position and its propensity to be passively replicated by replication fork emanating from neighbouring origins. The efficiency of an area is likely regulated by the number of potential origin. Several other biological phenomenon are correlated with the timing of replication and could support this view of a spatiotemporal program:

- established by chromatin conformation and origin distribution,
- executed by the diffusion of limiting replication factors.

2.2 Establishment and execution of the temporal program

2.2.1 Potential origins recruitment in a nuclear structured frame

Sequences determinant of origin positioning

The presence or absence of a potential origin in a specific position determines its competence and is therefore essential to define its efficiency. Origin positioning have been deeply investigated, therefore specification of origin positioning (Aladjem et al. 2006, Hamlin et al. 2008, Méchali et al. 2013, Smith and Aladjem 2014, Hyrien 2015) and methods to detect origins (Urban et al. 2015) have been extensively reviewed.

As already discussed, budding yeast possesses sites of high ORC-binding affinity that contain a consensus sequence. In other yeast cells, AT rich regions are also favoured for ORC binding but without the identification of a consensus sequence for most of them. In metazoans, all sequences are able to support initiation with an efficiency linked to their size rather than their base composition (Heinzel et al. 1991). However, some sites have been identified in metazoans to possess a high efficiency (Aladjem et al. 2006) which imply a high competence and therefore a very well positioned potential origin. Moreover, instead of a direct sequence affinity, ORC could be targeted to specific sites by cofactors or stabilised by interaction with other proteins of the preRC (Hyrien 2015).

Globally, metazoan origins tend to be enriched in GC (Woodfine 2003) and contain CpG sites (Cayrou et al. 2012). The high GC content elements has the potential to form G-quadruplex (described in the first chapter, Fig. 1.1.5). These structures have been correlated with highly efficient origins (Cayrou et al. 2011, Bessnard et al. 2012, Hizume et al. 2013, Valton et al. 2014). However, G4 structures are not necessarily correlated with origins and reciprocally (Mesner et al. 2013) and some technical biases have been evoked to explain this enrichment (Foult et al. 2015 reviewed in Urban et al. 2015). Overall, G4 may affect origin efficiency in certain contexts (Rivera-Mulia and Gilbert 2016a) and particularly by the association with Rif1 protein that shape the frame of replication (Kano et al. 2015). Consistent with the fact that gene rich regions have a high GC-content, ORC binding has also be reported at transcription start sites (TSSs) (Dellino et al. 2013).

However, apparent specificity could also be due to origins opportunistically occurring in reproducibly available sites determined by other features and processes (Urban et al. 2015).

Nucleosome and origin positioning

Chromatin state could be determinant in this regard. As already described, ORC binding is linked to nucleosome free regions and the ability to exclude nucleosomes could be the most important feature of the genomic sequence to promote ORC binding (Méchali et al. 2013). In addition, ORC has been shown to interact directly with nucleosomes in human cells (Hizume et al. 2013) as well as in yeast defining “chromatin-dependant” origins in addition to the well known “sequence-dependant” ones (Hoggard et al. 2013). Moreover, identification of origins of replication in yeast *Candida Albicans*, thanks to machine learning algorithm, also identified sequence dependent and independent origins of replication (Tsai et al. 2014).

ORC binding is crucial to define the origin position but it could also load several MCM that have the ability to slide on DNA even in the presence of nucleosomes (discussed in the first chapter, review in Das and Rhind 2016). These potential origins could also be displaced during S phase due to transcription (Powell et al. 2015) and potentially be accumulated on barrier regions such as G4 (Rivera-Mulia and Gilbert 2016a) creating efficient areas.

Timing decision point

Cells need a recovery time after the division to establish the temporal program during the G1 phase (Wu and Gilbert 1996). Without a necessary delay in G1, the replication program is not established. This leads to the terminology of a “Timing Decision Point” (TDP) to define the unknown biological event occurring in G1 which determines the execution of the replication program. The TDP is set before the Origin Decision Point (ODP) which is the recruitment of pre-RC, suggesting that timing is set independently of individual origins (Rhind and Gilbert 2013). The timing decision point is likely to be regulated by global nuclear structuring (Dimitrova and Gilbert 1999) that would form a template in which ORC could preferentially bind certain areas to set the potential origins (Kara et al. 2015). Chromatin conformation capture allows to define topologically associated domains (TADs) in mammalian genomes (Dixon et al. 2012). These domains are correlated with replication domains, and in particular, their boundaries coincide with timing transition regions (Fig. 2.2.1) (Pope et al. 2014). The establishment of the topologically-associated domains also occurs during early G1, concomitantly to the establishment of the TDP (Dileep et al. 2015).

Overall, it seems that the binding of ORC is not only dictated by its local affinity but also by the global organisation of chromatin, therefore defining potential origin positioning.

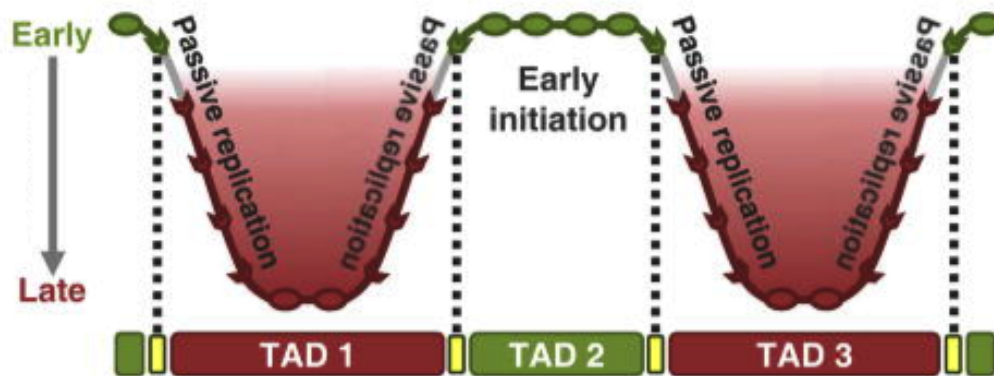


Figure 2.2.1: Replication domains model

Model of the spatial organisation of replication timing. Topologically associated domains (TADs) are spatial compartment that are correlated with cell function. Domains boundaries coincides with the boundaries of replication domains. Reused from Rivera-Mulia et al. 2015

2.2.2 Stochastic execution

Origin tethering

Early observations by fluorescence microscopy (Ma et al. 1998) lead to conclude that origins are organised into spatial clusters or factories (Berezney et al. 2000, Kitamura et al. 2006, Gillespie and Blow 2010). These clusters appear to be stochastically formed from neighbouring replicons (Saner et al. 2013) and they are highly dynamic (Meister et al. 2006). Chromatin conformation capture also confirmed that early origins are in close contact in budding yeast (Duan et al. 2010) as well as in higher eukaryotes (Yaffe et al. 2010).

Several biological factors have been proposed to impose a structuring of the genome able to tether origin together (Fig. 2.2.2). Forkhead transcription factor (Fkh1 and Fkh2) are correlated with early replication (Knott et al. 2012) and could be able to bring origins together via loop formation by dimerisation (Aparicio 2013, Ostrow et al. 2014).

Recently, Fkh1/2 have also been implied in the establishment of the timing decision point in yeast (Peace et al. 2016). Indeed, the timing in double mutant Fkh1/2 can be reprogrammed by a fkh1 overexpression in late G1. This observation is counterintuitive with the discussed picture that efficiency is mainly due to an abundance of MCM (Das et al. 2015), although the distribution of MCM has not been checked in the case of the Fkh1/2 mutant. On one hand, the two views can be reconciled if, as in human (Symeonidou et al. 2013), most MCM are loaded in late G1 in yeast. This means that Fkh1/2 could be acting on the distribution of the potential origins. On the other hand, as the timing of ARS affected by Fkh1/2 is not directly correlated to the MCM number (Das et al. 2015), it is also

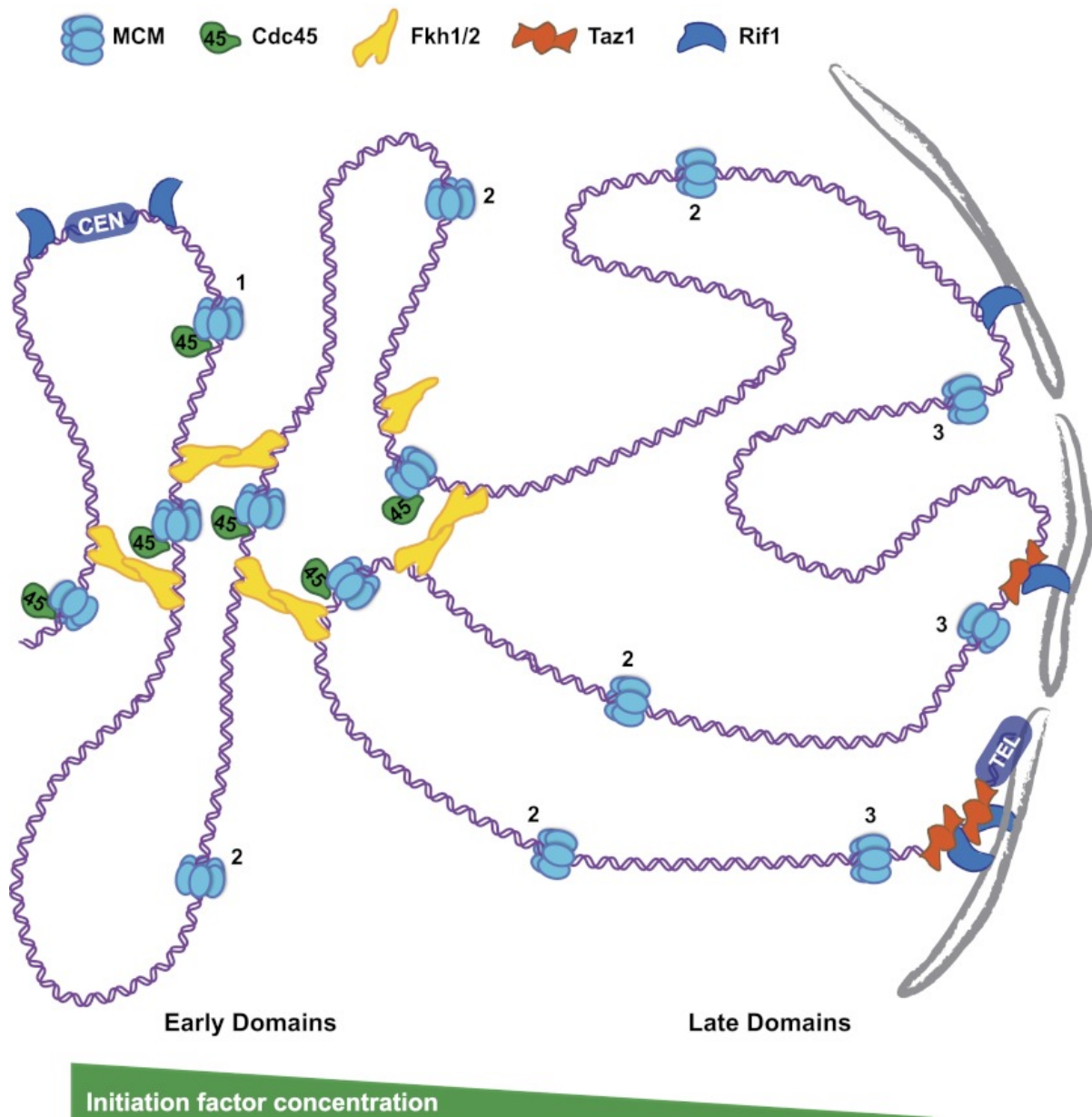


Figure 2.2.2: Origin tethering organised by multiple proteic factors
 Model of the spatial organisation of timing domains. Early domains in yeast are created by the tethering of potential origins inside a zone of high concentration of initiation factor by Fkh1/2. Late domains are granted by the clustering of late origin at the nuclear periphery by Rif1 and its association with other proteins. Figure from Aparicio 2013 reused with permission.

possible that Fkh1/2 provides an additional layer of regulation. Indeed, the effect of Fkh1/2 on clustering could be determinant in late G1 for temporal initiation because it seems to be able to promote the recruitment to *cdc45* to potential origins (Knott et al. 2012) which is characteristic of very early origins (Aparicio et al. 1997).

Another factors, rap-interacting factor 1 (Rif1) seems to promote late initiation by tethering chromatin at the nuclear periphery (Yamazaki et al. 2012, Yamazaki et al. 2013, Kanoh et al. 2015, Foti et al. 2016). Moreover, it helps to prevent initiation by counteracting DDK activity through the recruitment of the protein phosphatase 1 (PP1) (Davé et al. 2014, Mattarocci et al. 2014). Moreover, in human cells, other interactions exist between chromatin and nuclear matrix that induce tethering of late replicating regions at the nuclear periphery (Hutchison et al. 1994, Kalverda et al. 2008).

Identity of the limiting factors

The encounter of initiation factors with a potential origin is the limiting step of the initiation process. Indeed, many initiation factors are present in limiting quantities (Tanaka et al. 2011, Köhler et al. 2016) therefore their overexpression lead to an advance of the timing of late origins in yeast (Mantiero et al. 2011). Moreover, early replicating origins have been reported to be linked to the recruitment of *sld3* and *cdc45* in G1 phase (Kamimura et al. 2001, Kanemaki and Labib 2006), leaving only the CDK activation and Dpb11 recruitment for the S phase and therefore increasing greatly the probability of initiation of these origins.

Limiting factors could be recycled to initiate new potential origins after one initiation or after termination in the case of *cd45* that travels with the replication fork (Aparicio et al. 1997). It has been proposed that the genome organisation in domains could allow a concentration of limiting factors to be efficiently redistributed at the sites of replication stress (Rivera-Mulia and Gilbert 2016a).

In addition, dNTP pool is also limiting (Poli et al. 2012). Its tight regulation seems crucial: both imbalance between the dNTP levels (Kumar et al. 2010) and the increase of the pool (Chabes and Stillman 2007) is deleterious for the cell as both seem to reduce polymerase fidelity (Buckland 2014).

2.2.3 Influence of transcription and chromatin landscape

Replication transcription conflict

The competition of replication and transcription over the same DNA strand can produce head-on as well as co-directional collisions (Brambati et al. 2015).

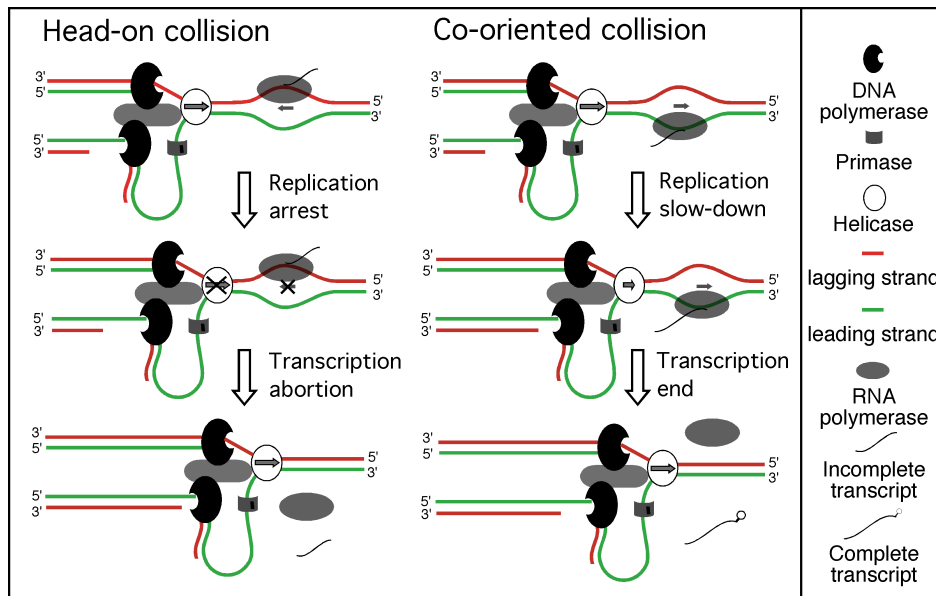


Figure 2.2.3: Collision between replication and transcription

Head-to-head collision of the replication fork with the RNA polymerase lead to a fork arrest (left column). Co-directionality of transcription machinery with replication fork can also lead to collision (right column) however less deleterious for the cell. Image from Eduardo Rocha website (<http://www.wabi.snv.jussieu.fr/erocha/research/order.html>).

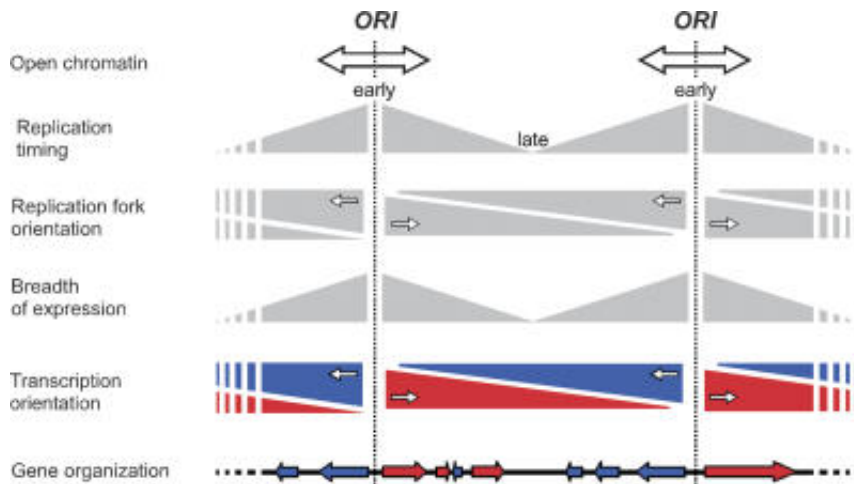


Figure 2.2.4: Model of replication and transcription co-directionality

The presence of origins in open chromatin regions lead to a strong fork directionality. Genes organisation lead to a co-directionality of transcription machinery with replication forks. Figure 6 from Huvet et al. 2007 reused with permission.

Highly transcribed genes have been shown to impede fork progression in yeast (Azvolinsky et al. 2009). In particular, the frontal encounter can cause genomic damage (Lin and Pasero 2012) and it has been observed that the directionality

of fork progression often favours co-directionality of the two machineries (Fig. 2.2.4) (Huvet et al. 2007). However, genes that are not actively transcribed during S phase also show this particular orientation regarding to the initiation zones (Necsulea et al. 2009). This raised question on whether this was the result of a selection to reduce the number of collisions during replication or not. One possibility is that genes are clustered according to their expression pattern (Zaghloul et al. 2012) inducing a co-directionality of genes and replication fork progression even when they are not highly transcribed during S-phase.

Moreover, timing of origin firing and cell-cycle regulation of the transcription has been observed to be separated for genes necessary for S-phase specific functions such as nucleosome assembly, DNA repair (Meryet-Figuere et al. 2014) and initiation of replication (Kylie et al. 2016) and has been proposed to be imposed by changing chromatin states during the cell cycle (Kylie et al. 2016).

Multiple correlation

The influence of transcription and chromatin landscape on timing have been extensively discussed (see the last review of the series Rivera-Mulia and Gilbert 2016b). Correlation between highly transcribed regions, open chromatin and early replication have been established for all multicellular organisms (Woodfine 2003, Schübeler et al. 2002,). The case of yeast has been less investigated and the timing of replication has been correlated to mRNA levels (Fraser 2013) but show no correlation on transcriptional activity of most studied gene class except for transcription of histone coding genes that is coupled with early replication (Raghuraman et al. 2001). Even in metazoans, replication timing and transcription level seem to be correlated overall (Hiratani et al. 2008, Desprat et al. 2009, Dellino et al. 2013) but many exceptions exist, suggesting an indirect link (Maric and Prioleau 2010). Moreover, replication timing can be predicted by DNase I hypersensitivity better than by TSSs (Gindin et al. 2014). This suggests that the correlation between replication timing and transcription is mediated by chromatin state (Lubelsky et al. 2014).

Influence of the chromatin landscape

The correlation of timing with chromatin state was observed with open accessible chromatin (euchromatin) being replicated early while closed chromatin (heterochromatin) was replicated late (Weinreich et al. 2004). The subject has been extensively reviewed (Smith and Aladjem 2014). The influence of chromatin state on the timing of replication could appear strong due to a double action on the potential origin distribution (discussed in the first subsection) as well as on the accessibility of these potential origins to limiting replication factors (Ding and

MacAlpine 2011, Dorn and Cook 2011).

Replication timing is correlated with multiple elements involving proteic partners as well as other cellular processes but no general rules have emerged yet. The key of a comprehensive view of DNA replication will be to decipher the intricacy of all correlated processes without overlooking the complexity. Recent advances have allow to propose two key steps in the regulation of the temporal program:

- its establishment during G1 by a non homogenous distribution of the potential origins due to higher accessibility of certain areas, the higher affinity of ORC-binding with some sequences or structures and the further mobility of these potential origins,
- its execution by the concentration of limiting factor in particular nuclear localisations thanks to a global chromatin organisation into spatial domains, the tethering of early replicating DNA by Fkh1/2 and the sequestration of late replicating DNA at the periphery by Rif1.

2.3 Conservation and significance of the timing

2.3.1 Conservation of the replication program

Different cell lines or strains of human, mouse, drosophila, chicken, arabidopsis (Costas et al. 2011), maize (Bass et al. 2015), *yeast* and other parasites (Stanojeic et al. 2016) have been studied and they all present a replication profile reproducible in a cell population (reviewed in Hyrien et al. 2013). Possessing a replication timing program seems to be a conserved trait of all organisms that replicate their DNA from multiple origins. The process that lead to this program also seems universal as the initiation rate present the same dynamics in many eukaryotes (Goldar et al. 2009).

The spatio-temporal program is highly conserved

Initiation sites present asymmetric nucleotides composition (the number of G is different of the number of C and the number of A is different of the number of T) due to the fact that the transition rate of A→G is greater than T→C on the leading strand. This creates a skew in GC and AT content in leading and lagging strand centred in the position of an initiation (Touchon et al. 2005, Chen et al. 2011). Only an accumulation over long evolutionary time can allow to detect such change meaning that the position of origins are conserved.

The spatiotemporal program itself is highly conserved. Indeed, replication profiles of several yeast strains show pattern as similar as two data set from the *S.c* specie (Fig. 2.3.1 A.) (Müller and Nieduszynski 2012).

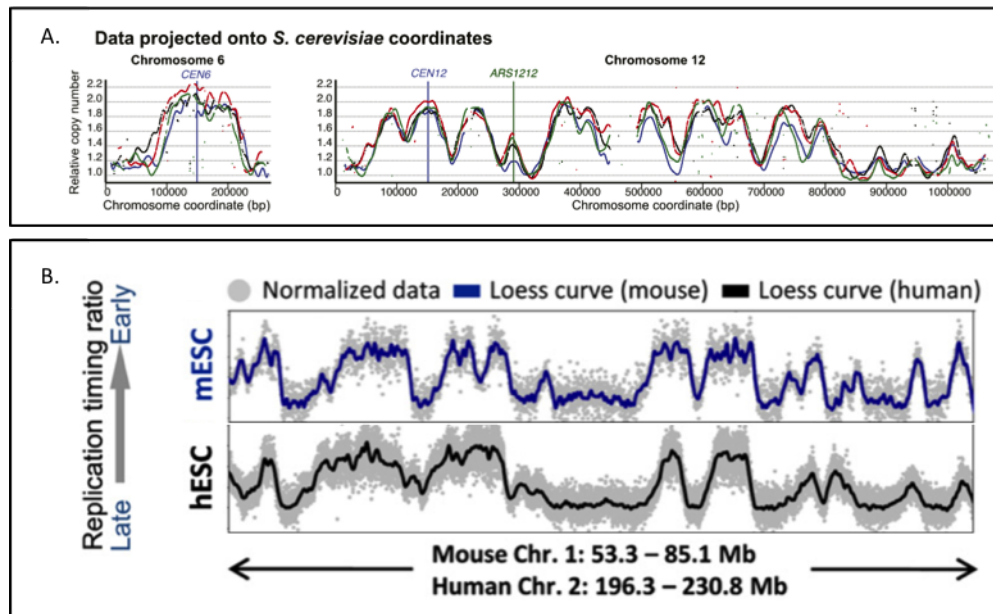


Figure 2.3.1: Similarity of the replication profiles of related species

A. Superimposed replication profiles of several yeast strains. Figure 2 for Müller and Nieduszynski 2012 reused with permission.

B. Replication profiles from mouse (top) and human (bottom) epithelial stem cells show striking similarities. Figure 3 A. from Ryba et al. 2010 reused with permission.

The observation has also been made that the spatio-temporal program is identical in primary, immortalised and transformed mammalian cells (Dimitrova 2002). Moreover, between mammals, replication profiles have been shown to be also strikingly conserved, notably between mouse and human embryonic stem cell (Ryba et al. 2010). The distinction of genomic area that change their replication timing with differentiation lead to the definition of constitutive versus switching replication timing. About 50-70% of the human genome is composed of constitutive timing that is therefore conserved in all studied cell types (Rivera-Mulia et al. 2015).

Saccharomyces cerevisiae : a specific case ?

The identification of consensus sequences for ORC binding in *S.c* is rather unique feature among eukaryotes. A consensus sequence has been identified only in few other yeast species such as *Candida Glabrata* (Descorps-Declère et al. 2015) that presents a similar ACS despite being evolutionary rather far from *S.c* . These

well defined sequences where ORC binds with a really high affinity, lead to a stronger positioning and therefore a greater efficiency of potential origins in *S.c* compared to any other eukaryotes. Therefore, the replication program in *S.c* was described as deterministic and intrinsically different of the timing program of *S.p* and higher eukaryotes which was referred as stochastic. However, ARS are not necessary for DNA replication (Gros et al. 2014) and ORC can bind to other site, dependent on the chromatin context. Furthermore, the identification of many inefficient origins in *S.c* have proven that there is no difference in the nature of the process between *S.c* and other eukaryotes but rather some different adjustment of parameters with ORC affinity being more preponderant in the distribution of potential origins (On et al. 2014) than in some other species.

2.3.2 Protection of genomic integrity

The replication timing shows a striking level of conservation in all eukaryotes which lead to question its significance. Why establishing such a program ? This remains largely an open question. DNA replication is a delicate step that need to be perfectly executed to allow the cell to proceed trough the cell cycle. The evolutionary pressure on having a robust system is therefore extremely high, so the existence of a timing could be derived from the necessity of preserving the genome integrity.

Completion of genome replication

In order to insure completion of genome replication, an unfired potential origin should always be available between two propagating forks as well as between a propagating fork and the telomere. Mathematical modelling in several yeast species showed that the spacing of referenced origins was suitable to maximise the probability of replication completion in case of fork stalling (Newman et al. 2013). These potential origins are not activated all at the same time because the quantity of initiation and replication factors is not sufficient. Overexpressing these factors lead to the early activation of many normally late origins however its is deleterious for the cell (Buckland et al. 2014, Köhler et al. 2016). Limiting the number of active forks could prevent an overtaxing of the repair machinery (Rhind and Gilbert 2013) as it has been observed in bacteria where a too high fork number can saturate the mismatch repair capacity (Schaaper and Radman 1989).

In conclusion, a large number of potential origins is needed in order to complete genome replication. However these potential origins cannot be all activated at the same time because these would require a higher number of initiation factors and dNTP pools, that are not compatible with high fidelity replication and can lead in a saturation of the mismatch repair capacity of the cell. These are strong

arguments to explain why staggered initiation is an advantage to the cell but does not explain why this staggered initiation has a conserved temporal order.

Spatial concentration

Spatial organisation of the genome is one of the major determinant of replication timing. The division of large genome in spatial domains also corresponding to replication domains allows to concentrate fork activity in defined areas. This has been proposed to be a way to efficiently deal with replication stress (Rivera-Mulia and Gilbert 2016a). Indeed, activation of checkpoint in case of the presence of a high level of ss-DNA, that can be caused by numerous stalled forks, induces a global inhibition of initiation which is only promoted in actively replicating regions (Toledo et al. 2013, Yekezare et al. 2013). This allows the cell to focus on repairing damages, restoring stalled forks and ultimately completing replication in a restricted area before proceeding to other areas. In this scenario, the spatial organisation defines the replication timing thanks to spatial domains in which the initiation occurs stochastically and is influenced by all the other factors described in the previous sections.

Mutations

Another possible evolutionary pressure on the replication timing can be hypothesised from the observation that mutation rate is lower in early replicating regions (Stamatoyannopoulos et al. 2009). This could induce a pressure on replicating early gene-rich regions. However the opposite effect could also be true as early replication give more time to homologous recombination to proceed reparation.

Overall, observed correlations between replication timing and important aspects of genome maintenance may simply be mediated by the structure of chromatin (Rhind and Gilbert 2013).

2.3.3 Transcription, epigenetic state and replication : causes and consequences

Early replication timing, transcriptional activity and open chromatin are often shown to correlate but it is difficult to decipher the causes and the consequences. As previously discussed, chromatin state is the element that best correlate with replication timing suggesting that a given chromatin state promote both replication and transcriptional activity. However, DNA replication disrupts chromatin and provides therefore a window of opportunity to reorganise the cellular program.

Change of chromatin state during replication

A strong evidence of the influence of the replication timing on chromatin structure is that the time at which a plasmid is injected in a cell in S phase influence its chromatin state (Zhang et al. 2002, Lande-Diner et al. 2009). Indeed, DNA packaged with deacetylated histones becomes packaged in acetylated histones when replicated early (Lande-Diner et al. 2009), increasing at the same time its expression level (Zhang et al. 2002). This is consistent with a H4 diacetylation marks being less abundant on histones newly incorporated during late S phase (Alabert et al. 2015). Reciprocally, it has been shown that the level of acetylation influences the timing of replication (Vogelauer et al. 2002, Mantiero et al. 2011). This circular problem was examined during cell reprogramming.

Changes of RT during reprogramming

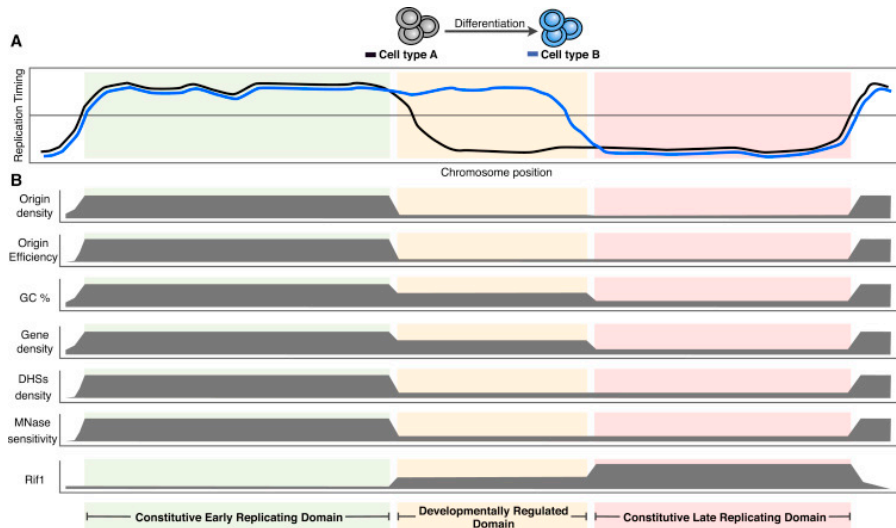


Figure 2.3.2: Different regulation of developmentally regulated and constitutive domains

Some domains are constitutively early or late and other are developmentally regulated. The developmentally regulated domains are not correlated with the same cues as constitutive domains. Figure 4. from Rivera-Mulia and Gilbert 2016a reused with permission.

Changes in the replication profiles upon differentiation have been shown to be concomitant with changes in chromatin structure and transcriptional activity (Norio et al. 2005, Hiratani et al. 2008). Interestingly, change from late to early is also correlated with a re-localisation from the periphery to the centre of the nucleus and conversely (Hiratani et al. 2008, Rivera-Mulia and Gilbert 2016b). A simple model to explain this triple correlation has not been established yet. The advance

of replication timing of a domain could lead to a chromatin state change into open or close chromatin, and allow a differential gene expression. This hypothesis would imply that the temporal program of replication is essential to define cell types. However, the correlation between replication timing and transcription is overall decreased with differentiation (Rivera-Mulia et al. 2015, Fig. 2.3.2).

Model of transcriptional control

The study of replication profiles changes during differentiation allowed to distinguish constitutive replication domains that exhibit stable replication timing in all cell lines and developmental regulated domains that switch their replication timing (Rivera-Mulia et al. 2015, Smith et al. 2016). The major outcome from this observation is that only constitutive domains really show a correlation with all genome-wide activities previously described.

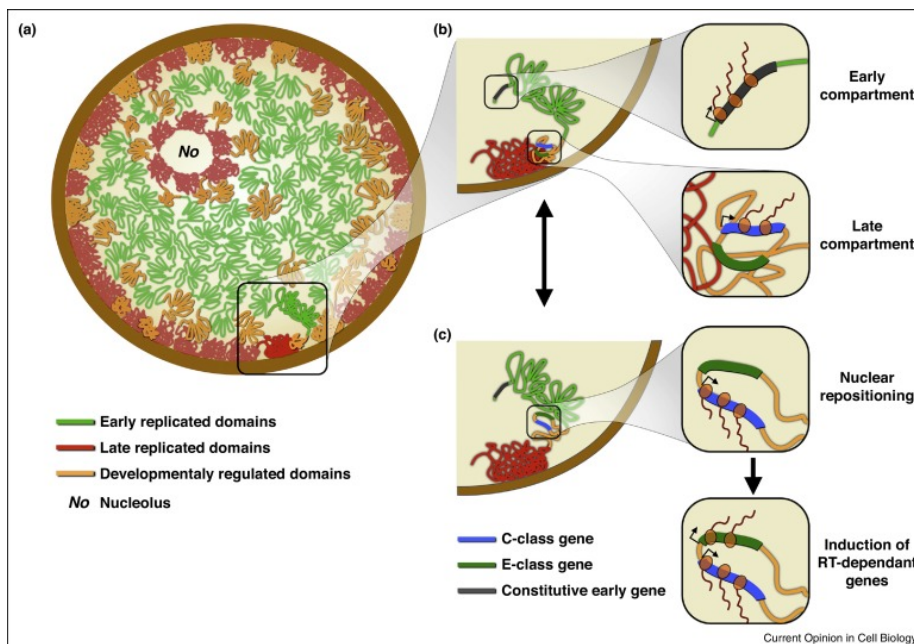


Figure 2.3.3: Model of multiple influences of genes expression, chromatin conformation and replication timing

Metazoans genome is split between early and late domains. During differentiation some domains can switch their status which is also associated with a nuclear repositioning and a change in the expression level of E-class genes. However, C-class genes switch their expression level before the timing switch which suggest a sequential reprogramming. Figure from Rivera-Mulia and Gilbert 2016b reused with permission.

In particular, transcription levels and timing are strongly correlated for the constitutive domains but not for the differential ones. A close analysis revealed several classes of genes, a majority are transcribed when in a late replication domain at

least in one cell type (C-class) and a substantial proportion (~20%) have a transcription level correlated to timing (E-class) (Rivera-Mulia et al. 2015). The lack of correlation have been attributed to the fact that the switch of expression level of the C-class genes can precede the replication timing switch by one or two intermediate stages of differentiation (Rivera-Mulia and Gilbert 2016b). This described a gradual model of the genome reorganisation with a gene changing its expression level leading to a local switch of replication that will impose open chromatin marks and define the replication timing of the domain allowing the expression of other less essential genes in the same area.

Replication timing is present in all organisms that have multiple initiation sites for DNA replication. Time imposes itself as a full dimension of the replication process. Biological significance of such a program has been difficult to address and it is still not exclude that it is only an inevitable consequence of some other factors such as chromatin landscape. However, it has been proposed that the spatial organisation of replicating domains is an evolutionary conserved trait that allow efficient preservation of genome integrity during DNA replication. Replication timing could also act at the necessary reorganisation of these domains during cell differentiation.

Chapter 3

Replication in the yeast nucleus

S.c., also referred as budding yeast, is a model organism widely used to study DNA replication. As already discussed, this specie possesses genomic locations that are very efficient to initiate replication, because they have a very strong affinity for ORC binding. Therefore, this rather small genome comprise well identified potential origins with known efficiency and timing. Moreover, a lot of genome wide data are available to correlate DNA replication timing and other informations such as transcription levels, chromatin landscape or binding of several proteins.

In addition, the maintenance of yeast population is also very well documented and numerous tools are available for its manipulation. Yeast is an ideal tool to extract the general rules that impose DNA replication timing in all eukaryotes.

As discussed in the previous chapter, global and local chromatin conformation are one of the determinant of DNA replication timing. In this work we propose to investigate to which extent the latter impacts DNA replication in budding yeast. Therefore this section will summarise what is already known about genome organisation in the interphase nucleus and recapitulate the link that has already been made between this organisation and the timing of DNA replication. This chapter will be focused on *S.c.* In addition, techniques allowing to study the nuclear organisation will be explained and discussed.

3.1 Organisation of the yeast interphase nucleus

3.1.1 Rabl structure

Budding yeast genome is divided into 16 chromosomes. These chromosomes are organised by stable interaction between chromatin and structural elements of the nucleus such as the spindle pole body (SPB), nuclear envelope (NE) and the nucleolus as reviewed in Zimmer and Fabre 2011 and Taddei and Gasser 2012.

Chromosomes are linked by their centromeres to the SPB and folded in two arms extending to the nuclear interior forming a structure called Rabl or Rabl like (Bystricky et al. 2005) hypothesised by Carl Rabl in 1885.

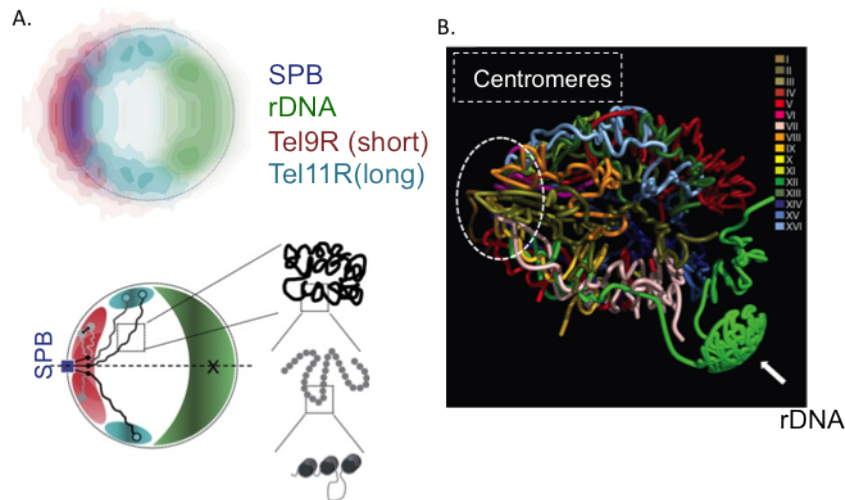


Figure 3.1.1: Rabl configuration of yeast nucleus

A. The localisation of several specific points in the nucleus as seen by fluorescent microscopy. Centromeres are linked to SPB in one pole of the cell while rDNA form a crescent shape as the opposite pole. Telomeres are tethered at the nuclear periphery and their position depend on the length of the chromosome arm. Figure 1 from Zimmer reused with permission.

B. Reconstitution of genome structure from 3C data confirms the clustering of centromeres as well as the isolation of rDNA (on chromosome 12). Figure 5 from Duan reused with permission.

Tethering of centromeres

Centromeres remain attached to the spindle pole body during the whole cell cycle. SPB is a multilayered organelle anchored in the nuclear envelope composed of multiple proteins essential for the viability of the cell (Jaspersen and Winey 2004). SPB is duplicated and the two SPBs are connected by a bridge that is subsequently cut to allow migration of one of the SPB at the opposite pole of the cell (Jaspersen and Winey 2004). Microtubules connect the centromeres with SPB. Their length vary during the cell cycle from ~150nm in interphase to ~300nm when SPB are present in opposite pole and reduce to ~30nm at the end of chromatin segregation (O'Toole et al. 1999). Several microtubules can link one centromere to the SPB (Nannas et al. 2014). This physical link induces a clustering of centromeres at one pole of the cell during interphase (Jin et al. 2000). Therefore, inter-chromosomal interactions are directly dependent on the distance to the centromere in particular in short chromosome that are interacting with other chromosomes along their full length (Duan et al. 2010).

Nucleolus

The opposite pole of the SPB is occupied by the nucleolus, a crescent-shaped compartment that occupy roughly one third of the nucleus volume (Bystricky et al. 2005). This compartment is not delimited by a membrane and is defined by its contents which is dedicated to ribosome biogenesis (Olson and Dundr 2015). Indeed, it contains rDNA which is a particular locus on chromosome XII containing a variable number tandem repeats (~200 on average) of 9.1kb sequences coding for ribosomal RNA (Taddei and Gasser 2012). The assembly of this genomic region as a compartment involves several proteins (Olson and Dundr 2015) and leads to a spatial segregation with no interaction of this region with other chromosomes (Duan et al. 2010).

Anchorage of telomeres at the nuclear periphery

Telomeres are the extremities of chromosomes and are associated with special secondary structure in order to prevent chromosome ends from being recognised as double strand breaks (Harari and Kupiec 2014). Telomeres are anchored to nuclear periphery by redundant pathways involving the protein yKu as well as the silencer protein Sir4(silent information regulatory complex) (Hediger et al. 2002, Bystricky et al. 2005). Contrary to centromeres they are not all tethered together but form dynamic groups of five to ten telomeres visible as distinct foci when marked with fluorescence (Gotta et al. 1996). The two ends of the same chromosome are in close interaction when the two arms are of similar size (Bystricky et al. 2005, Schober et al. 2008, Therizols et al. 2010). Moreover, even when not in direct interaction, the motion of the two ends of a chromosome are more correlated than between two different chromosomes (Bystricky et al. 2005).

Constraint on the global genome organisation

Multiple numerical models have been developed to mimic the nuclear organisation of chromatin inside the yeast nucleus considering polymer models (Table 3.1) (Gehlen et al. 2012, Tjong et al. 2012, Tokuda et al. 2012, Wong et al. 2012, Verdaasdonk et al. 2013, Wong et al. 2013, Avşaroğlu et al. 2014, Gürsoy et al. 2014).

They reveal that imposing the clustering of centromeres, the telomere anchorage at the periphery and an exclusion from the nucleolus zone that contain rDNA is enough to recapitulate global structure of chromatin packing inside the nucleus. The main force driving the nuclear organisation is the entropy (Vasquez et al. 2016) constrained by the nucleus boundaries itself and these anchorage at several points of the nuclear envelope.

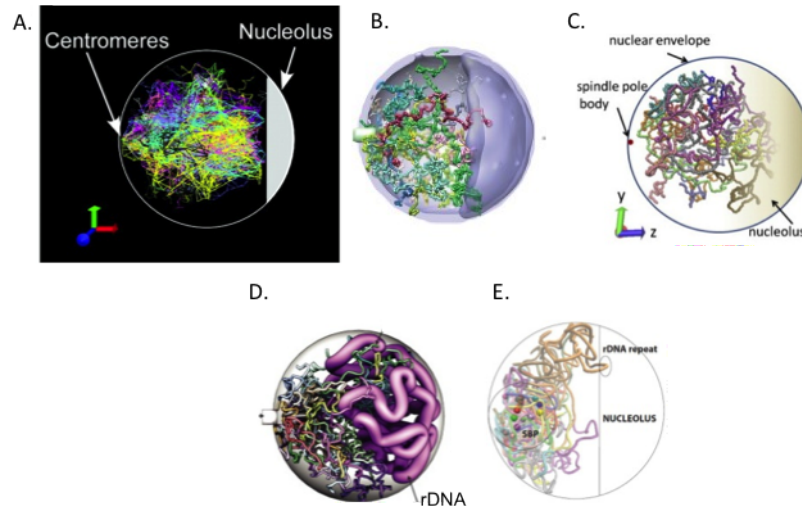


Figure 3.1.2: Computational models of yeast nucleus organisation

Several models of genome organisation in yeast. A. Gehlen et al. 2012 B. Tjong et al. 2012 C. Tokuda et al. 2012 D. Wong et al. 2012 E. Verdaasdonk et al. 2013

	compaction	persistence length
Gehlen et al. 2012	22bp/nm, 130bp/nm	30 nm, 200 nm
Tjong et al. 2012	90 bp/nm	47-72 nm
Tokuda et al. 2012	130bp/nm	170-220 nm
Wong et al. 2012	80bp/nm	60 nm

	centromeres clustering	telomeres clustering	rDNA
Gehlen et al. 2012	elastic	NE	spatial
Tjong et al. 2012	fixed	-	spatial + exclusion
Tokuda et al. 2012	elastic	-	exclusion
Wong et al. 2012	fixed	-	heteropolymer

Table 3.1: Parameters and constraint on the models

Moreover, the comparison between growing and quiescent cells is also very informative to deduce the rules that drive nuclear organisation. Quiescence is a non proliferative state, usually triggered by the lack of nutrient. The cells stop their cycle after division and do not enter G1 phase but rather a phase called G0. Cells can exit quiescent state when nutrient become available again (Gray et al. 2004). In quiescent cells, the centromeres are less clustered compared to proliferating cells (Jin et al. 1998, Laporte et al. 2013, Rutledge et al. 2015) probably because of a drastic change in the microtubules organisation (Laporte et al. 2013). The nucleolus is also displaced (Laporte et al. 2013). Moreover, the telomeres have stronger interaction (Rutledge et al. 2015) and form an hyper-cluster (Guidi et al. 2015, Laporte 2016). Significantly, this change is necessary to insure longevity during the resting phase as deleting Sir3, allows the telomeres clustering and reduces this longevity (Guidi et al. 2015). Therefore, the global layout of the genome is a regulator of chromatin conformation and is linked to functional activity of the cell.

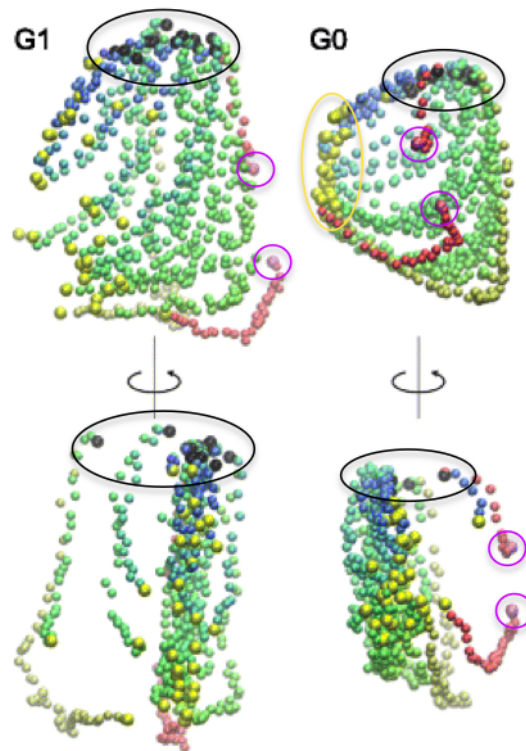


Figure 3.1.3: Reorganisation of the chromatin structure in quiescence

Average conformation of the yeast genome in active (left) and quiescent (right) cells. Centromeres are in black, telomeres in yellow. Shortest chromosome is blue and longer red while rDNA boundaries in chromosome XII is in purple. In G0, centromeres are least clustered while telomeres form an hyper-cluster. Figure 3. from Guidi et al. 2015 reused with permission.

Polymer models for global nuclear organisation

Computational models assumed subtly different polymer models : the genome is divided in multiple rigid segments that are jointed (Wong et al. 2012) or in beads or spheres that are jointed by more or less elastic segments (Tjong et al. 2012). Both reproduce very well chromatin contact data without any input on the global behaviour but assuming the properties of the segments or beads and linkers. The equilibrium globule on the other hand is a global polymer model that do not require to explicit the local property of the polymer (Fudenberg et al. 2013). This equilibrium globule model is compatible with the thermodynamical equilibrium of previous physical models (Tjong et al. 2012, Wong et al. 2012). This model has been proposed in mirror to the fractal globule proposed for the organisation of the human chromatin (Mirny 2011) and in accordance with the contact probability obtained from chromatin conformation capture experiment (Duan et al. 2010). In human, data from chromatin conformation capture at higher resolution disproved the fractal globule model (Rao et al. 2014, Sanborn et al. 2015) however no such analysis is available in yeast at the moment (Wang et al. 2015). In addition, the equilibrium globule does not describe the nuclear domains (Hsieh et al. 2015) observed in yeast.

3.1.2 Nuclear domains

Chromosomal domains

The small size of the nucleus ($<1\mu\text{m}$ radius) and the high mobility of genomic positions lead to question the existence of chromosomes territories comparable to those of higher eukaryotes (Haber and Leung 1996, Bystricky et al. 2005). Chromosome conformation capture data reveals that chromosomes have more interaction with themselves than with other chromosomes, suggesting that they form chromosome territories (Rodley et al. 2009, Duan et al. 2010). Recently, Micro-C, a conformation capture technique with increased resolution, allowed to identify chromatin interaction domains (CID) comparable to human TADs (Hsieh et al. 2015).

Interestingly, CIDs encompass one to five genes which is equivalent to the number of genes present in human TADs (Hsieh et al. 2015). However, genes being much longer in human the average size of a domain is more of $\sim 0.5\text{Mb}$ compared to $\sim 5\text{kb}$ in yeast. The boundaries of these domains are co-localised with active genes promoters and with cohesin loading complex. Authors suggest that these functional domains are created by the constraint of the boundaries rather than a preferential three-dimensional organisation inside the domain (Hsieh et al. 2015).

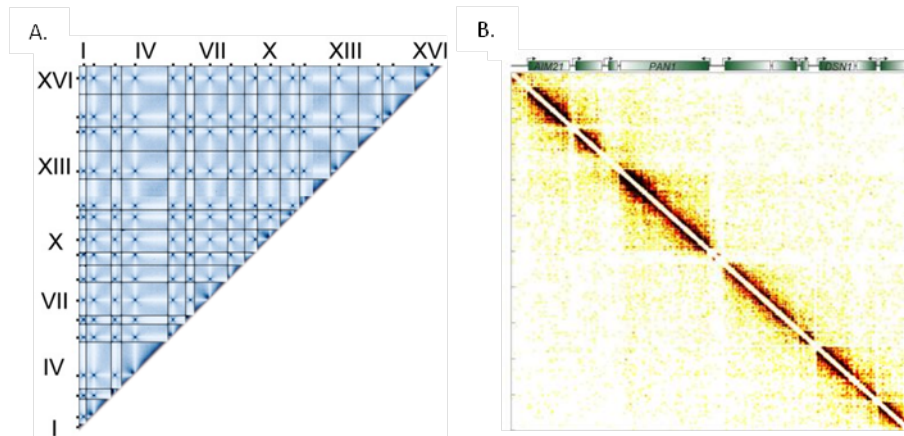


Figure 3.1.4: Chromatin contacts in yeast genome

Contact frequency in yeast genome. Darker points mark stronger interaction.

A. All genome contact frequency obtained from Hi-C in proliferating cells show the strong interaction of centromeres (marked on each chromosome by a small black rectangle) between them but not with the rest of DNA. Figure 1. B. from Rutledge et al. 2015 reused with permission.

B. Higher resolution of contact frequency map of a part of chromosome IX obtained with Micro-C presenting the genes on the top of the map. Genome high resolution maps reveal the presence of chromosomal domains encompassing one to five genes. Figure 1. from Hsieh et al. 2015 reused with permission.

Genes clusters

Genes also occupy smaller defined territories (Berger et al. 2008). An interesting study of the differences between *S.p* and *S.c* using a similar computational model, in Tjong et al. 2012, shows that gene territories are much more defined in *S.c* (Gong et al. 2015). *S.p* has an overall similar nuclear architecture and an equivalent genome size but which is divided in only three long chromosomes. This suggest that the volume exclusion effect of the 16 chromosomes tethered at the centromere strongly contribute to the spatial organisation of the nucleus in *S.c* (Gong et al. 2015).

In both species, highly transcribed genes are on average closer than the lowly transcribed genes and they occupy distinct nuclear localisation respectively on the nuclear interior and on the nuclear periphery (Gong et al. 2015). In addition, genes related to similar function are also closer than functionally unrelated genes. Random conformation subject to geometric constraint explains in part the functional organisation of the genome (Gong et al. 2015).

Dynamic

Specific genomic positions can be marked by fluorescence to track their motion *in-vivo* inside the nucleus. The spatial fluctuation of the observed positions is characteristic of anomalous diffusion (Cabal et al. 2006, Albert et al. 2013, Hajjoul

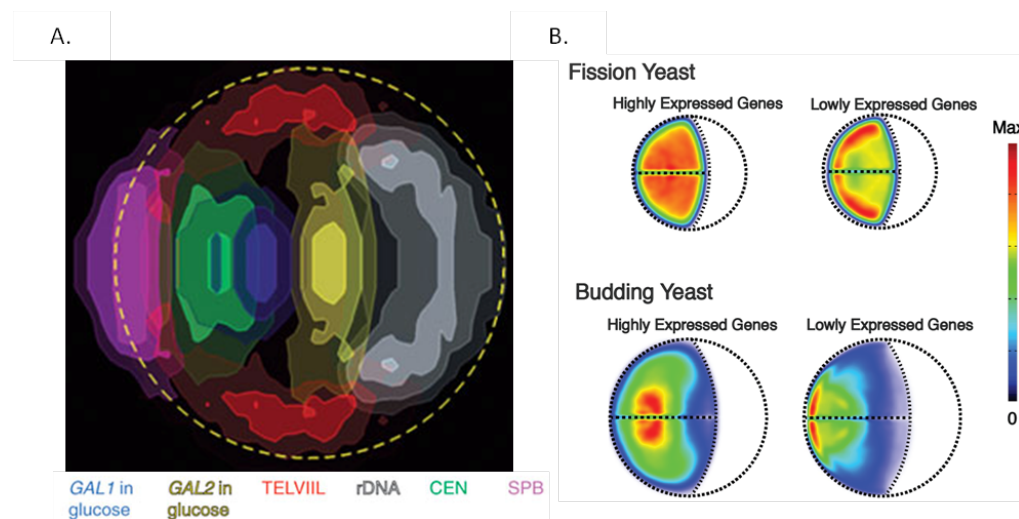


Figure 3.1.5: Genes clusters

A. Average localisation of genes observed by fluorescent microscopy. Position of SPB, centromeres and rDNA allows to average the position between different cells. Figure 4. a. from Berger reused with permission.
 B. Position of genes according to their expression in model of fission and budding yeast. In both case expressed genes are present in the interior of the cells while lowly expressed genes are at the periphery of the cell however the localisation is much stronger in budding yeast compared to fission yeast. Figure 4. C. from Gong et al. 2015 reused with permission.

et al. 2013, Backlund et al. 2014). This means that the mean square displacement does not evolve linearly with time as it would for a random walk (Wang et al. 2015). Instead, mean square displacement and time have a power law relation. The exponent is inferior to 1 which mean that the displacement is hindered, the smaller the exponent the stronger the hindrance. Rouse model which considers the diffusion of chains by brownian motion of beads connected by harmonic springs, describes anomalous diffusion with an exponent close to 0.5 (Hajjoul et al. 2013). However, multiple studies fail to obtain a good agreement on the exponent value between 0.4 and 0.75 that could be due to differences between yeast strains (Backlund et al. 2014) but nevertheless cast doubt on the precise mode of diffusion in yeast nucleus (Récamier 2013). However, inside each study, the exponent is similar in all the nuclear volume (Hajjoul et al. 2013), except in the nucleolus (Albert et al. 2013). Modelling diffusion in a confined domain allows to find domain size of few hundred nanometers (Amitai and Holcman 2013) consistent with experimental measurements (Berger et al. 2008, Albert et al. 2013, Hajjoul et al. 2013).

In particular, the two telomeres of a chromosome move in a partially coordinated manner. Therefore, it was proposed that the restriction of a position in a confined area was influenced by the chromatin continuity (Bystricky et al. 2005). Modelling the motion of a tethered polymer confirmed that the dynamic fluctuations are dictated by tethering at centromeres (Verdaasdonk et al. 2013). Conversely, the

statistical analysis of a single locus trajectory recapitulates the local interaction on chromatin (Amitai et al. 2015).

3.1.3 Nucleosome organisation

Nucleosome positioning

Several genome-wide studies of nucleosome positioning in yeast revealed that about 80% of the genome is covered and an even higher proportion of genes is covered (Lee et al. 2007, Eaton et al. 2010, Deniz et al. 2016). These studies show that most nucleosomes are well positioned, in particular at TSS or at ORC-binding sites. Interestingly, the proportion of well positioned (Fig. 1.2.3 (B)) nucleosomes decreases from 65-70% in G1, G2 and M to 50% in S phase (Deniz 2014). Moreover, the size of the linker DNA between two nucleosomes core is short (~20bp) and present less variability in size than is higher eukaryotes.

Specificity of yeast heterochromatin

Most chromatin is present as open euchromatin (Bi 2014). Heterochromatin is only present at telomeres, rDNA and in silenced mating type loci. Indeed, budding yeast are able to switch their mating type from MAT_a to MAT_α and conversely. Therefore, in addition of the active MAT locus, two loci containing a silent copy of each type is present and their expression is repressed thanks to heterochromatin (Braunstein et al. 1996).

Contrary to other eukaryotes, *S.c* does not possess active post-translational marks of repression on histones but instead, deacetylation of H4K16 (Cortini et al. 2015) and demethylation of H3K29 (van Leeuwen et al. 2002, Ryu and Ahn 2014) promote the silencing mainly via the interaction of silent information regulatory complexes (Sir) with chromatin (Oppikofer et al. 2013). In addition, in other eukaryotes, centromeres are also associated with heterochromatin which is not the case of budding yeast. Indeed, budding yeast centromeres are called point centromeres as they consist in a small sequence (~125bp) rather than a repetition of sequence that is present in other species and favours the formation of heterochromatin (Taddei and Gasser 2012).

A loose zigzag chromatin fibre

Micro-C is a chromatin conformation capture technique that analyse interactions between fragment of DNA cut between nucleosomes, and therefore allows to obtain information on nucleosome interaction (Hsieh et al. 2015). Interaction between the nucleosome N and N+1 is as strong as between N and N+2, suggesting the presence of a motif characteristic of two-start zigzag fibre (Hsieh et al. 2015).

However no interaction was observed after N+4 suggesting that no chromatin fibre is present in vivo or is not seized by the short range cross-link (Hsieh et al. 2015) or lost but averaging different fibre organisations in the asynchronous population (Mozziconacci and Koszul 2015).

Overall this view is extremely consistent with the observation of fibres formed by reconstituted nucleosomes arrays by cryogenic electro microscopy (Song et al. 2014). Indeed, the unit of the fibre is composed of four nucleosomes that are tightly stacked therefore preventing interaction between nucleosomes N1 and nucleosome N5 and N6. Zigzag organisation of four nucleosomes unit without higher regular folding is also possible (Grigoryev et al. 2016) and more consistent with the view of an extended fibre with regional variation in compaction already described based in chromatin conformation capture data (Dekker 2008).

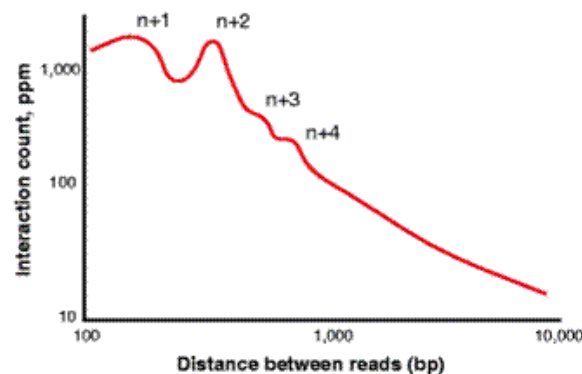


Figure 3.1.6: Contact between adjacent nucleosomes

Contact between adjacent nucleosomes reveals that interaction with the N+2 nucleosome is as strong as the N+1 which suggest a higher order organisation of the nucleosome fiber. However no interaction was visible after the N+4 nucleosome which could mean that the fibre is composed of disorganised quadrinucleosomes units or that the methods cannot detect higher order organisation. Figure 1. c. from Mozziconacci reproduced from figure S3 from Hsieh reused with permission.

The linker histone H1, which is necessary to form chromatin fibre in higher eukaryotes, is not present in yeast as the homologous protein Hhop1 is not essential and do not seem to impact the accessibility to MNase (Patterton et al. 1998). Recently, the protein HMO1 has been proposed to function as a linker histone in yeast (Panday and Grove 2016). Isolated chromatin from yeast have been shown to form 30-nm filament comparable to the organisation of chromatin in chicken erythrocyte (Lowary and Widom 1989). Moreover, the different estimations of the persistence length from experimental data are not in agreement and give a stiffness that is consistent (~ 200 nm Bystricky et al. 2004) or not (< 30 nm Hajjoul et al. 2013) with the existence of a regularly folded chromatin fibre. The most convincing computational models picture chromatin with a persistence length around 60nm (Table 3.1) which could combine several state of chromatin folding.

Yeast nucleus is structured by the physical interaction of the genome at anchoring points of the nucleus. The genome is well described by a polymer whose organisation and dynamic are dictated by the tethering of centromeres at SPB, of telomeres at nuclear periphery and the exclusion of most DNA from the nucleolus compartment. Recently, Micro-C technique shed some light on the the organisation of chromatin at the intermediary scale between molecular and micron scale. It revealed a strong interaction between four adjacent nucleosomes, suggesting a stacking compatible with an higher order organisation. Moreover, it also confirmed the existence of chromosomal domains. These domains seem to be imposed by physical tethering due to global genome layout which causes anomalous diffusion of genomic positions. Overall, the global chromatin conformation seems to be determinant for functional organisation that rise the question of the impact on DNA replication timing.

3.2 Replication in the interphase nucleus

3.2.1 Subnuclear organisation of timing

Early replication at centromeres

Centromeres are replicating early (McCarroll and Fangman 1988) and they promote early replication in their surroundings even when artificially displaced (Pohl et al. 2012). However, in order to do so, they need to be functional and linked to the SPB. Therefore, the action of the centromere to promote early replication could be due to a high density of initiation factor near the SPB, where centromeres are clustered (Pohl et al. 2012). Proper replication timing at centromere is crucial to allow a correct segregation of chromosomes (Feng et al. 2009).

Late replication at telomeres

On the contrary, telomeres are replicating late (McCarroll and Fangman 1988). Telomeres are shortening at each round of replication because the lagging strand synthesis cannot be completed, therefore telomeres can exist in a cell as short or long. A specific reverse transcriptase enzymes, called telomerase, counters this effect by adding repeated sequence in the leading strand but do not always access telomeres and are not always active. The shortening of telomeres is a cause of diseases and one of the mechanism of ageing.

Interestingly, normal long telomeres replicate late but short telomeres replicate early (Bianchi and Shore 2007). The early timing is granted by association with the protein Tel1 (Cooley et al. 2014), potentially by counteracting Rif1 tethering

effect, which advance the timing of replication while promoting the recruitment of telomerase (Sridhar et al. 2014).

Global positioning

Overall early origins are present in the interior of the nucleus while late origin are more present close to the nuclear periphery (Heun et al. 2001). However, displacing an early origin at the nuclear periphery is not enough to change its timing (Ebrahimi et al. 2010). Moreover, excising a late origin from its position at nuclear periphery leads to its random positioning in the nucleus but without losing its timing (Heun et al. 2001), consistent to observation in other organisms (Therizols et al. 2015).

3.2.2 Chromatin contacts and replication factories

Organisation in factories

Fluorescent dNTP can be incorporated in yeast nuclei during DNA replication. Cells are observed at several stages of DNA replication, and the replicated positions can be directly visualised by a local accumulation of fluorescence. Up to 15-20 discrete foci can be observed in a nucleus (Pasero et al. 1997). Fluorescent foci are also observed when components of the replisome such as PCNA are fluorescently labelled in fixed cells (Ohya et al. 2002) and in vivo (Kitamura et al. 2006, Saner et al. 2013). The number of foci is significantly lower than the estimated number of forks, meaning that each foci could contain several forks and even several replicons (portion of the genome replicated from the same origin).

Strong evidences from microscopy suggest that the two forks emanating from the same origin stay in interaction during replication (Kitamura et al. 2006, Saner et al. 2013) although the interaction is not necessary to perform DNA replication in vitro (Yardimci et al. 2010). The association between two different origins in a factory is not necessary and most measured factories contain only two forks suggesting they could contain a single replicon (Saner et al. 2013).

However, the mobility of chromatin can lead to the stochastic interaction of potential origins that stay associated for several minutes (Saner et al. 2013). The probability to be in the same factory for two distinct origins is proportional to their distance proving that this association is not deterministic but is more probably established by random contact that are stabilised (Saner et al. 2013). Stable interaction could be provided by the same interaction between replisome components that stabilise sister forks together, or by additional factors forming loops such as Fkh1/2 transcription factor described in the second chapter.

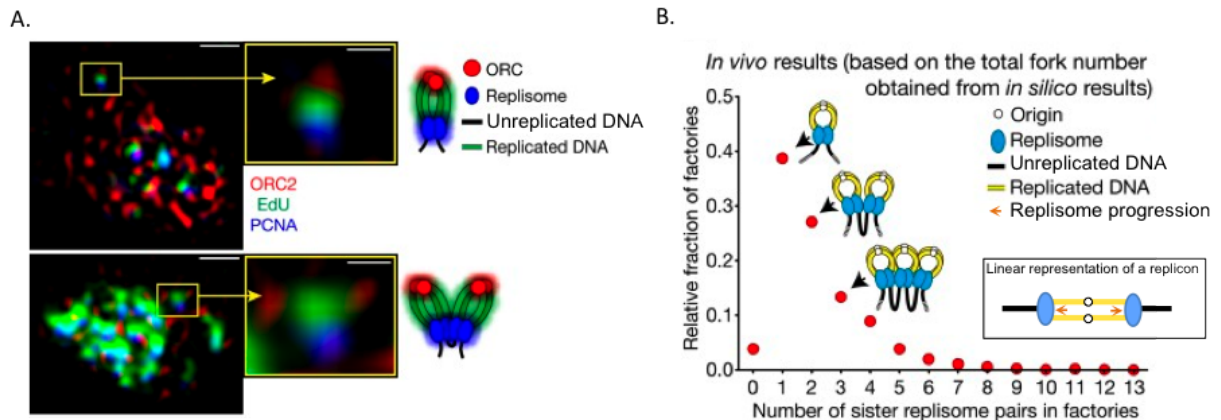


Figure 3.2.1: Organisation in replication factories

A. Replication factories are visible by fluorescent microscopy. The schematic on the right show the fluorophore association. Some factories contain one replicon (top) while other contain several (bottom).

B. Most factories contain only the two forks from the same replicon as deduced from the fluorescence intensity and the number of forks estimated from simulation.

Figure 3. F. and E. from Saner reused with permission.

Spatial clustering of early origins

More generally, positions of early replication are in closer contact than random positions, whereas positions of late replication are more dispersed (Duan et al. 2010, Cournac et al. 2012, Rutledge et al. 2015). However, this preferential interaction is probably not driven by formation of factories, as similar observation can be done in quiescent cells that do not undergo DNA replication (Rutledge et al. 2015) expect for early origins close to centromeres that happen to be significantly less clustered in quiescent cells. Moreover, computational models show that the structure of the nucleus is sufficient to impose a preferential interaction of early origins without introducing any looping or clustering mechanism (Tjong et al. 2012). Therefore, the spatial proximity of early origin is probably due to global chromatin organisation with a clustering of the centromeres and preferential contact along chromosomes arms.

Even if factories are stable during ~10min, they are stochastically formed and mainly composed of a unique ORC (Saner et al. 2013). Overall origin position is highly dynamic during G1 and S-phase (Heun et al. 2001) therefore the formation of factories containing multiple replicons in some cells is probably too diluted to be visible on the interactions maps from asynchronous population.

The layout of yeast nucleus is reflected in the timing of replication. Particularly, replication timing at centromeres and telomeres is crucial for the proper cell division and maintenance. The formation of factories that is visible in single living cells by microscopy is not detectable by chromatin conformation capture techniques. The functional importance of these clusters is still to be determined. In particular, it remains unclear if an active looping mechanism is necessary or if natural dynamic motion only can lead to the formation of these stochastic factories.

3.3 Several techniques gave insight on the 3D chromatin organisation and dynamic

3.3.1 Usual techniques to study genome organisation

Microscopy

Studies of genome organisation have greatly took advantages of microscopy and in particular fluorescent microscopy (Bystricky 2015). Development of the technique and of its spatial and temporal resolution allows to study the organisation of genome inside the nucleus. Moreover, microscopy can be performed in living cells, therefore allowing to gain insight on the dynamic of its organisation. In addition, microscopy allows to study single cells therefore rare events otherwise, lost in the average behaviour, can be detected.

The drawback of the method lies in its low throughput which is also often decreasing with the resolution. In addition, only a limited number of positions can be studied at the same time. Development of high throughput tools for image acquisition and image analysis and their democratisation are the current challenge of microscopy to study genome conformation (Shachar et al. 2015, Pegoraro and Misteli 2016).

Chromatin conformation capture

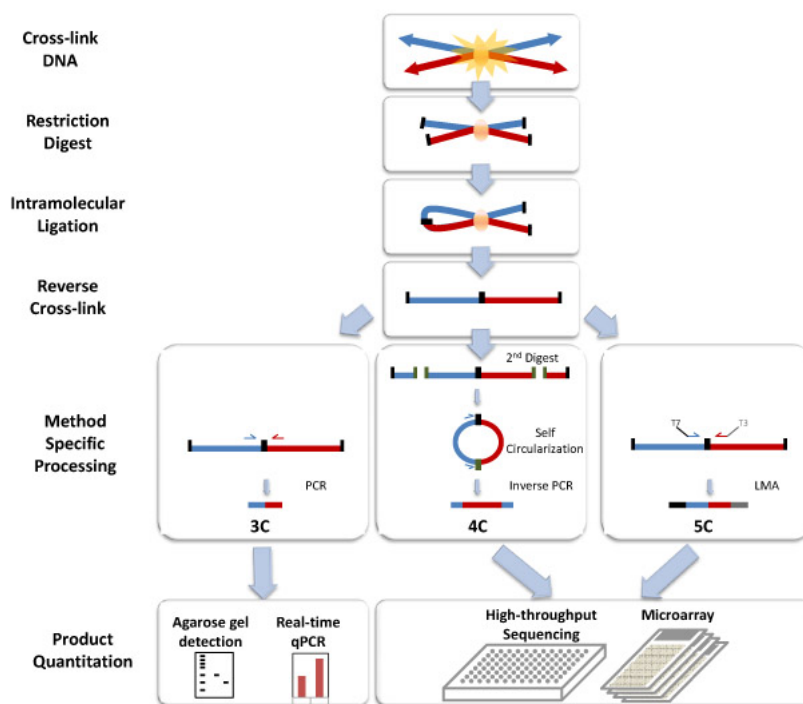
Chromatin conformation capture (3C) gave a precious insight on global chromatin conformation inside the nucleus at unprecedented resolution (de Wit and de Laat 2012, Dekker et al. 2013, Denker and de Laat 2016). The technique was first developed in 2002 in Kleckner lab (Dekker et al. 2002) and followed by a development of several variants.

For 3C, cells are cross-linked with formaldehyde to link covalently chromatin segments that are in close spatial proximity. Next, chromatin is fragmented by restriction digestion or sonication. Cross-linked fragments are then ligated by a

3.3 Several techniques gave insight on the 3D chromatin organisation and dynamic

ligase enzyme to form unique hybrid DNA molecules. Finally, each DNA hybrid molecule corresponding to an interaction is detected by several methods. In the original 3C method, single ligation products were detected by PCR one at the time using locus-specific primers but they can be now directly sequenced using modern deep-sequencing platform therefore leading to a higher coverage of the genome. Other variant have been developed (Dekker et al. 2013, Barutcu et al. 2016) including Hi-C that differs because DNA segments ends are biotinylated before ligation to facilitate purification of jointed ligation products that can be directly sequenced.

Figure 3.3.1: Schematic methodology of 3C and derivative techniques



Several normalisation steps are necessary to remove random interactions (Dekker 2006, Cournac et al. 2012, Ay and Noble 2015). One of the greatest limit to high resolution was the heterogenous distribution of the restriction site on the genome (Hahn and Kim 2015) and is skewed by GC-content (Mozziconacci and Koszul 2015). As small fragment are less likely to have been cross-linked, the resolution was $>1\text{kb}$ and contained biases linked to the genomic sequences. The new technique called Micro-C use micrococcal nuclease (MNase) instead of restriction enzyme to cut the genome (Hsieh et al. 2015). MNase cuts the DNA between nucleosomes, that are particularly evenly spaced in yeast, leading to fragment of

constant size (Mozziconacci and Koszul 2015). Micro-C technique is complementary to Hi-C because it has an increased resolution but poorly captures long-distance interactions.

3C-based techniques provide population averaged data although single-cell Hi-C have been performed on human cells (Nagano et al. 2013) showing the same organisation of genome in domains compared to population data but stochastic organisation of these domains at larger scale. Moreover, if the techniques cannot tackle highly dynamic questions, large reorganisation such as the quiescent state (Guidi et al. 2015, Rutledge et al. 2015) or a switching of mating type (Belton et al. 2015) can be assessed.

However the relation between probability of interaction and genomic distance is not trivial and requires a physical model (Hahn and Kim 2013, Lesne et al. 2014). Moreover, cell variability has to be considered to propose a relevant model of genome organisation (Sekelja et al. 2016). This and the still non-negligible cost of sequencing are the remaining impediment to a wider democratisation of the technique.

Computational models

Both microscopy and 3C techniques call on polymer models for the interpretation of the data (Huet et al. 2014). Simulations of several constraints can be carried out on these polymers to decipher the potential contribution of each constraint. From the simulation, several thousand of structures can be obtained and treated as real cells to try to recapitulate microscopy and 3C data. The model can encompass several variables such as polymer parameters that can be optimised to fit the experimental data. These approaches have been extremely successful in particular in modelling yeast genome which is small and possesses a well documented layout (Iyer et al. 2011). The next challenge for the field is to build multi-scale models with more precise descriptions of the small scales in order to investigate precise interactions (Wang et al. 2015).

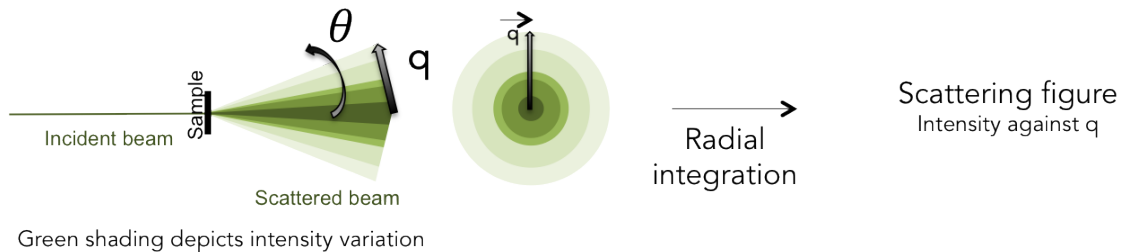
3.3.2 Small-angle scattering

Small angle scattering (SAS) is the collective name given to the techniques of small angle neutron (SANS), X-ray (SAXS) and light (LS) scattering. In each of these techniques radiation is scattered by a sample and the resulting scattering pattern is analysed to provide information about the size, shape and orientation of some component of the sample. This scattering is assumed to be devoid of energy loss and referred to as elastic. In biological laboratories, SAXS is mainly known to provide information on the structure of macromolecules in solution and is used to determine the structure of proteins or of protein-DNA complexes such as

isolated nucleosomes (Luger and Richmond 1998). The principle of the technique is the same than in crystallography : a beam interacts with the sample and part of the beam is deviated by irregularities in the sample. The size of the deviation is inversely proportional to the scale of observation therefore, SAS techniques, that study deviation at small angle, are interested in the organisation at large scales. More precisely, SANS and SAXS can give information of the sample organisation over a range between 10-1000nm which is an intermediary scale between the range observed by molecular and microscopy techniques.

Scattering intensity (Higgins and Benoît 1994)

Figure 3.3.2: Measure of scattering intensity



Each particle is scattered by the sample by an angle θ which corresponds to a momentum transfer q that depends on the wavelength on the beam λ .

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right)$$

The number of particles and their deviation from the direct beam are detected which gives the scattering intensity as a function of q . The formula of $I(q)$ derives from the fact that the scattered intensity $I(q)$ is the Fourier Transform of $g(r)$, the correlation function of the scattering density $\rho(r)$, which corresponds to the probability to find a scatterer at position $r-r_0$ in the sample if another scatterer is located at position r_0 : elastic scattering experiments reveals the spatial correlations in the sample.

The scattering intensity per unit volume depends linearly on the number of scatterers by unit of volume N , the contrast $\Delta\rho^2$, their shape and size described by the form factor $P(q)$ and their interaction relative to each other described by the structure factor $S(q)$.

$$I(q) = N\Delta\rho^2 \times P(q) \times S(q)$$

The contrast defines how well a scatterer scatters the particle compared to the solvent by comparing their scattering length density.

$$\Delta\rho^2 = (\rho - \rho_s)^2$$

The scattering length density depends on the nature of the beam.

Specificity of beam and solvent matching

X-rays used for SAXS interact with the electrons and the scattering length density is proportional to the electronic density of the sample. Neutrons interact with nucleons inside the atom nuclei and the scattering length density has no intuitive rule but can be experimentally measured.

Because of this difference contribution of atoms to SAXS and SANS signal is different. Notably, hydrogen and deuterium are identical to SAXS while very different to SANS. Therefore, in SANS, changing the deuteration level of the solvent allows to change the contrast.

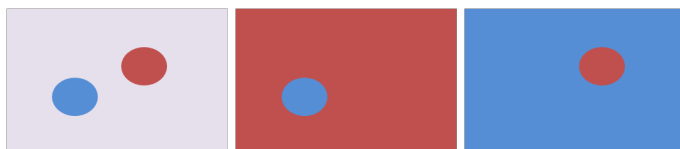


Figure 3.3.3: Principle of contrast

Two particles (blue and red) are visible in a grey solvent. However, changing the solvent to match one of the colors allow to focus on one of the element only.

Advantages and limits of the method to study chromatin organisation

As already describe, SAS allows to investigate scales that are precisely the one where the organisation of chromatin remains poorly understood. The sample do not need to be particularly fixed or stained which is also a great advantage. The risk exist with X-rays but not with neutrons that the irradiation will damage the sample therefore the two methods are complementary.

However, SAS main challenge lie in the data analysis. Indeed, intrinsically some information about the structure is lost as only the intensity but not the phase of the Fourier transform is known. This is already true for a solution of particles homogenous in size and shape of known concentration. The problem is obviously greater for cellular compartment or entire cells which possess many elements that contribute to the signal and are prone to interfere making the analysis very difficult. However, relevant qualitative information have already be inferred from cellular compartments such as isolated nuclei or entire cells. Moreover, SAS signal can now be produced from numerical models reflecting the organisation of DNA to assess the validity of the conclusions.

Ambiguity of the signal from cell compartments

In the 80s, many studies have been conducted to investigate the organisation of the chromatin fibre for isolated chromatin or isolated nuclei by SANS and SAXS (Ibel et al. 1983, Langmore and others 1983, Notbohm 1986, Bordas et al. 1986). In SAXS signal, a peak at $q \sim 0.15\text{-}0.2 \text{ nm}^{-1}$ corresponding to a compact structure with a repetition of $\sim 30 \text{ nm}$ was identified consistently with the existence of a chromatin fibre in isolated chromatin or nuclei.

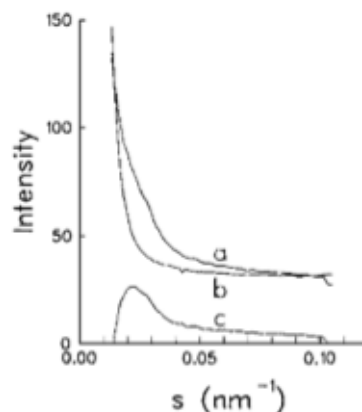
Only one study however was truly dedicated to investigate whether this signal could be attributed to chromatin or not (Langmore and others 1983). Several experiments were made in various cell types and lead to the conclusion that this peak as well as few other probably corresponding to nucleosome interaction can be attributed to chromatin.

However, the $\sim 30\text{nm}$ peak was not present in some cell type such as ursine sperm where chromatin is believed to be very compact due to an absence of transcription. Therefore the authors emphasised the need of controls to attribute the presence of a peak at 30nm to chromatin (Langmore and others 1983).

First, the peak is present in chicken erythrocyte but not in rabbit erythrocyte that do not contain nucleus (Fig. 3.3.4) (Langmore and others 1983). In addition, signal from entire cells and cell nuclei are qualitatively similar at least for species with a small cytoplasm volume such as mouse lymphocytes or ursine sperm (Langmore and others 1983). Therefore, the signal is emanating from the nucleus, at least for these species. In addition chicken erythrocyte isolated nuclei treated with DNase I loose the peaks revealing that they are caused by the structure of DNA (Langmore and others 1983). The signal of HeLa interphase nuclei or HeLa metaphase chromosomes is also similar which suggests a very robust configuration (Langmore and others 1983).

As already discussed in the first chapter, the existence of a chromatin fibre in vivo is still under debate. Indeed, for some cell types such as chicken erythrocytes

Figure 3.3.4: Compared SAXS signal from chicken or rabbit erythrocyte

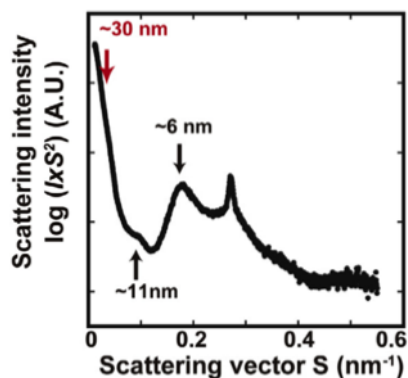


a. SAXS signal from chicken erythrocyte, b. signal from rabbit erythrocyte (that do not contain nucleus) c. subtraction of the curves gives a peak with a maximum at $s = \frac{q}{2\pi} \sim 0.025 \text{ nm}^{-1}$ corresponding to a characteristic distance of $\sim 40\text{nm}$

Figure 2. from Langmore and others 1983

experimental data are in accordance but for HeLa cells, images from electron microscopy do not support the existence of a chromatin fibre in vivo (Eltsov et al. 2008). Instead, ribosomes form aggregation of high electronic density on HeLa isolated nuclei (Nishino et al. 2012).

Figure 3.3.5: SAXS signal from HeLa nuclei



SAXS signal from HeLa isolated nuclei after ribosomes washing

Based on these observation, Maeshima team recently reproduced the SAXS data for chicken erythrocyte nuclei, HeLa interphase nuclei and HeLa metaphase chromosomes (Joti et al. 2012, Nishino et al. 2012). However the peak at ~30nm disappears after washing ribosomes that are artefactually stuck on the HeLa nuclear membrane and on isolated chromosomes after isolation procedure. Indeed, aggregation of ribosomes only also produces a peak corresponding to a characteristic distance of ~30nm (Nishino et al. 2012). These conclusions enlighten the difficulty to interpret SAXS data and the need to establish control for each specie.

Information on large scale organisation

SAXS on HeLa mitotic chromosomes present no peak over a long range (50-1000nm) suggesting a disordered structure (Nishino et al. 2012). A power law relation in $I(q)$ vs q can have multiple interpretations depending on the exponent that varies between 1 and 4 (Hammouda 2010). Entire exponents are often linked to particular structures such as -2 which could imply the presence of a gaussian chain or a disc, -1 is characteristic of stick shape, while -4 is characteristic of a surface. The exponent -3 however does not corresponds to a particular structure as well as non entire values which suggest a scale invariance also called fractal structure (Hammouda 2010).

SANS study in interphase chicken erythrocyte nuclei also lead to propose a fractal organisation of chromatin (Lebedev et al. 2005). However, SANS signal does not present a peak corresponding to the characteristic distance of ~30nm which is strongly supported by multiple evidences in this cell type but this discordance was not discussed.

3.3 Several techniques gave insight on the 3D chromatin organisation and dynamic

Several techniques provided insight that help build the knowledge we have about chromatin organisation in the nucleus. However, the intermediary scale between molecular and microscopic scale was not possible to investigate until the recent Micro-C experiment and the organisation of the genome at this scale is still elusive. Therefore, SAS technique could be an interesting alternative to explore this window, given the fact that they already produced interesting, yet controversial results.

Material and Methods

"Success is the ability to go
from failure to failure without
losing your enthusiasm."

(Anonymous)

Chapter 4

Yeast strains and cell cycle

4.1 Strains

4.1.1 Yeast strains used

4.1.1.1 BY4741

BY4741 is yeast strain of genotype (MATa his3 Δ 1leu2 Δ 0met15 Δ 0ura3 Δ 0) derived from S288C with deletion of selectable marker genes (Winston et al. 1995, Brachmann et al. 1998). For this reason, it will be used as a background for strain construction (see BYMNase strain). BY4741 yeasts from the laboratory stored at -80°C were scraped on YPD solid plates, grown for 2 days at 30°C and stored at 4°C. Fresh plates were prepared from the previous plate every 3 to 4 weeks. For liquid growing, one or few colonies were diluted in YPD medium and grown over night or over the week-end. At this stage, cells are in stationary phase and can be diluted to obtain logarithmic growing cells (see next subsection).

4.1.1.2 MCM869

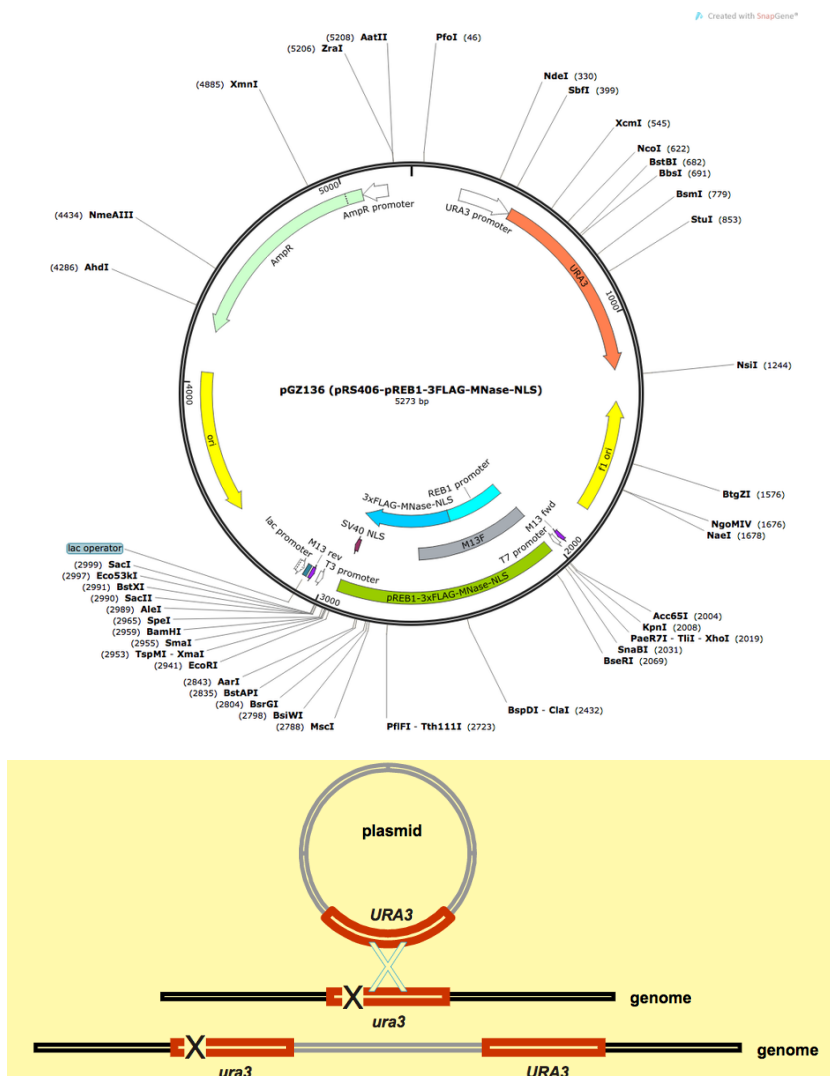
MCM869 (MATa ade2-1 trp1-1 can1-100 leu2-3, his3-11,15 URA3 GPD-TK7x AuR1c ADH-hENT1 bar1 Δ LEU2 cdc21 Δ kanMX) is a derivative of the E1000 strain (Lengronne et al. 2001) in which the CDC21 gene (encoding thymidylate synthase) was disrupted to increase BrdU incorporation (Vernis 2003). MCM869 yeasts from the laboratory stored at -80°C were scraped on YPD solid plates containing 100 μ M of thymidine, grown for 2 days at 30°C and stored at 4°C. Fresh plates were prepared from the previous plate every 2 to 3 weeks. For liquid growing, one or few colonies were diluted in YPD medium containing 100 μ M thymidine and grown over the week-end. At this stage, cells are in stationary phase and can be diluted to obtain logarithmic growing cells (see next subsection).

4.1.1.3 BYMNase

Plasmid

The plasmid pGZ136(pRS406-pREB1-3FLAG-MNase-NLS) was purchased from addgene and delivered inside bacteria. Bacteria were grown at 37°C in sterile LB medium and 4 mL were used to purify the plasmid using two Sigma GenEluteTM Plasmid Miniprep Kits. The DNA quantity in the final volume of 100µL was measured with Nanodrop ND-1000 giving a value of 57.2ng/µL.

Figure 4.1.1: Plasmid pGZ136(pRS406-pREB1-3FLAG-MNase-NLS) and its integration mode



As a quick verification, 30 μ L of the obtained DNA were digested at 37°C overnight with 2 μ L of EcoRI and XhoI enzymes and 3 μ L of smart cut buffer all purchased from New England Biolabs. To remove the salts and enzymes, 1 mL of ethanol 95% was added and mixed by gently turning the tube and left at -20°C during 30min. The DNA was then precipitated by 15 min centrifugation at max speed at 4°C (Eppendorf 5415R rotor FA-45-24-11). Supernatant is removed and pellet is rinsed with 500 μ L of ethanol 70%. After 15min centrifugation at max speed at 4°C, supernatant is removed by pipetting and ethanol is completely removed with another quick centrifugation followed by the pipetting of the remaining microliters and tubes are let to dry during 15min to allow remaining ethanol to evaporate. The DNA pellet is then resuspended in 50 μ L milliQ water and 5 μ L are deposited in 1% agarose gel TBE 0.5% containing ethidium bromide (3 μ L in 100mL gel) and migrated for approximately 30min at 90V. The gel contains 3 wells : 1kb+ DNA ladder (Thermo Fisher Scientific), full plasmid linearised by opening by NcoI enzyme with the same protocol, plasmid digested with EcoRI and XhoI.

To prepare plasmid for transformation, 8mL of saturated bacteria were used to purify plasmid with 4 kits and the obtained plasmid in 200 μ L were linearised by overnight digestion at 37°C with 2 μ L NcoI and 20 μ L of cut smart buffer.

Transformation

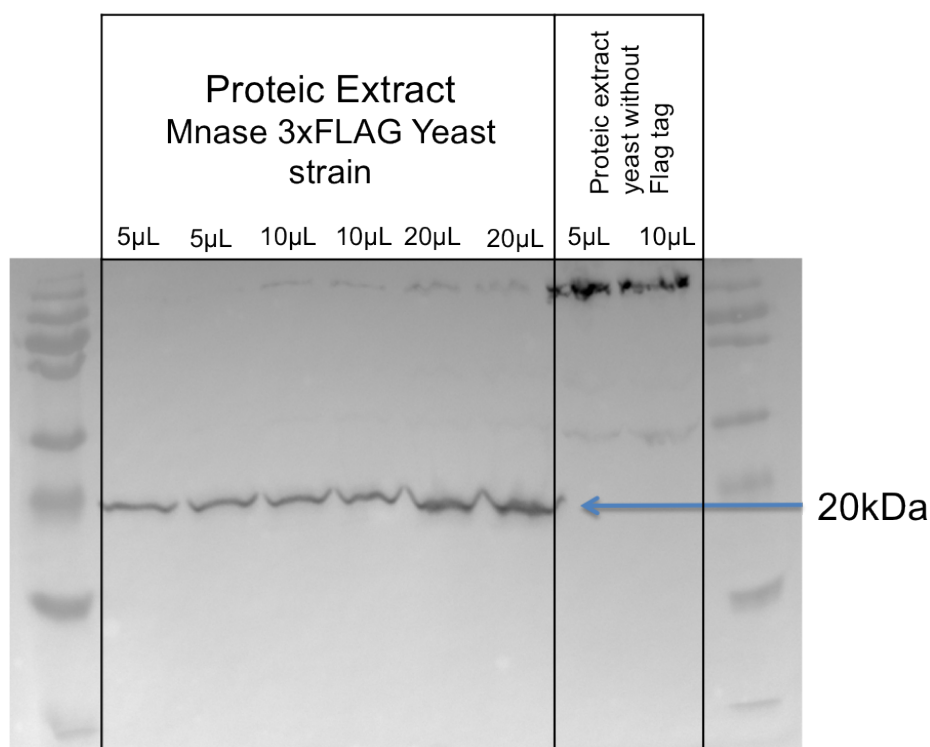
10mL of BY4741 cells in logarithmic growing phase were collected by 2min 4000rpm centrifugation (Eppendorf 5810 rotor A-4-62) and washed with 1mL of sterile water by the same centrifugation. Pellets are resuspended in 1mL of TEAcLi (10 ml sterile water, 100 μ L TE 100x, 1mL filtered (0.22) AcLi 10x) and left 30min at room temperature for permeabilisation. Cells were then centrifuged 1min at 6000rpm (Eppendorf 5415R rotor FA-45-24-11) to avoid breaking the cells. Cells were then resuspended in 50 μ L TEAcLi and 25 μ L of DNA carrier Sigma-Aldrich, 20 μ L of plasmid solution and 300 μ L of PEG-TEAcLi (90 μ L sterile water, 800 μ L PEG 50%, 100 μ L AcLi 10x and 10 μ L TE 100x) were subsequently added. The mix was vortexed quickly and the tube was put in a water bath at 30°C for 3 hours. The tube was then transferred in a water bath at 42°C during 20min to create a thermic shock. Cells are collected by 1min centrifugation at 6000rpm and the supernatant is carefully removed to remove the viscous PEG solution as much as possible. Cells are then resuspended in 200 μ L sterile water and spread on a plate containing growth medium lacking uracyle (locus producing uracyle is impaired in BY4741 cells and will be restored by the introduction of the plasmid by homologous recombination therefore only the cells that have incorporated the plasmid in their chromosomal DNA will be able to grow). Cells were allowed to grow for three days at 30°C and very small colonies developed. Despite being very small, these colonies were not false positives as they were not present on the control box

(same protocol but without the addition of the plasmid). The biggest colony was spread again on a plate lacking tryptophan and grew normally despite pretty small colonies.

Western blot

The production of the MNase enzyme was checked by western blot. Proteins were extracted with Trichloroacetic acid (TCA) by the following protocol. 5mL of yeast culture at saturation were collected and left at -20°C overnight. Cells were unfrozen and washed with PBS 1x once. Cells were then resuspended in 200µL TCA 20% with one spoon of glass beads. Samples were vortexed 15min. The supernatant was washed with 200µL TCA 5%. Supernatant was centrifuged 15min at maximum speed (Eppendorf 5415R rotor FA-45-24-11). Sample was resuspended in 200µL Laemli mix (500µL Laemli 2x, 500µL Tris HCl pH 8.8, 20µL β-mercaptoethanol) and warmed at 10min 95°C. Finally sample is centrifuged 2 min at max speed, supernatant are collected and warmed at 95°C.

Figure 4.1.2: Western Blot of MNase protein



Samples were analysed by western blot. Migration was performed in a 14% polyacrylamide gel and proteins were transferred on a nitrocellulose membrane.

MNase was specifically targeted thanks to its tagging by FLAG (see plasmid) with an overnight incubation at 4°C with a primary antibody against FLAG M2 (Sigma) 1:5000 in 2.5% milk and during 1h with a secondary antibody against mouse coupled with HRP (Promega). Signal was revealed with a short incubation with Clean-Blot™ IP Detection Kit in a Fusion Fx7 device. Non specific signal was detected with another pool of proteins from yeast that do not contain an FLAG tag.

Digestion

Mnase enzymatic activity is induced by a calcium input with requires a mild permeabilisation of yeast cellular envelope. The protocol followed was the one described in Zentner et al. 2015 except separately aliquoted anti-proteases were used in buffer A instead of the anti-proteases cocktail and calcium is used at a concentration of 10mM instead of 2mM. In brief, 50mL of cells in logarithmic state are collected and washed twice with 1 mL buffer A 1x (see table 4.1 for composition) by centrifugation 2 min at 4000rpm (eppendorf 5810 rotor A-4-62). Cells are re-suspended in buffer A 1x, transferred in 1.5mL eppendorf and washed once 1 min at 9000 rpm (eppendorf 5415R rotor FA-45-24-11). Cells are resuspended in 600 µL of buffer A 1x containing 0.1% digitonin (obtained from equal mix of buffer A 2x and digitonin 0.2%) and warm at 30°C for 5min. 3µL of CaCl₂ 2M was then added to the mix. Several time points can be sampled by pipetting up and down to mix the sedimented cells before sampling 100µL. Reaction is stopped by mixing the sample with 100µL of stop buffer (see table 4.1 for composition).

Agarose gel

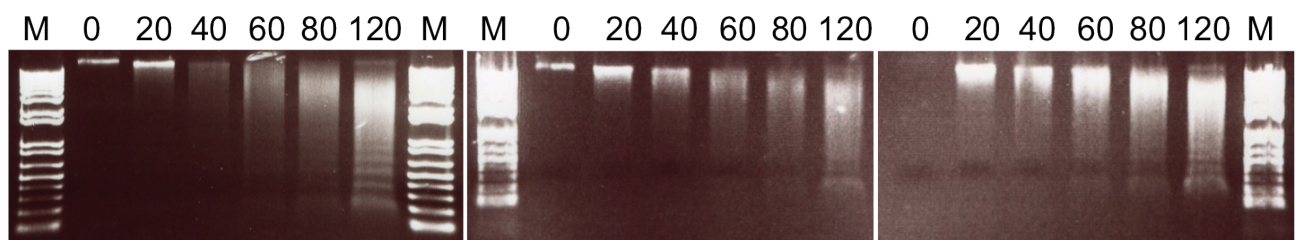
To check the DNA degradation, DNA was extracted by phenol-chloroform and precipitated in ethanol. The composition of the sample mixed with stop buffer corresponds to a composition close to the usual lysis buffer used to extract DNA. 200µL of Phenol:Chloroform:Isoamyl 25:24:1 (PCI) and glass beads are added to the mix and vortexed strongly during 4min. 200µL of TE 1X pH 8 is added and cells are centrifuged under the hood for 5 min (Hettick Zentrifuger max 15000 rpm). Supernatant is carefully sampled in 1.5mL eppendorf avoiding the proteins pellets (~300µL can be sampled). 1mL of ice-cold ethanol is added, tube is turned up and down to mix gently and is placed at -20°C for at least 20min and preferably overnight. DNA is precipitated by centrifugation and ethanol is removed as explained in the plasmid paragraph. DNA pellets are resuspended in 50µL of TE 1X pH 8. 1µL of RNase A 12.5mg/mL boiled to deactivate potential DNase enzymatic activity is added to the sample which is incubated 30min to 1h at 37°C. DNA quality and concentration is checked with Nanodrop ND-1000 and 4µL to

Buffer A 2X	STOP buffer
<ul style="list-style-type: none"> • 160mM KCl • 30mM Tris-HCl pH 7.5 • 0.4mM spermine • 1mM spermidine • 2mM PMSF and other anti-proteases • 0.2mM EGTA • qsp sterile H_2O 	<ul style="list-style-type: none"> • 0.4M NaCl • SDS 1% • 20mM EDTA • 4mM EGTA • qsp sterile H_2O

Table 4.1: Composition of buffers for in vivo DNA digestion by MNase enzyme

16 μ L (depending on the DNA concentration) of sample mixed with loading dye 6X (the dye must be initially diluted because concentrated dye causes a shadow when revealing the gel that masks the nucleosome and multiple nucleosomes bands) is deposited on 1% agarose gel containing Ethidium Bromide. Migration at \sim 90V lasts \sim 30min and the bands are revealed by Herolab transilluminator.

Figure 4.1.3: 1% agarose gel showing DNA digestion after MNase induction



The digestion protocol was applied three times from exponentially growing cells of the same batch, the DNA isolated from time series sample is migrated on 1% agarose gel and revealed with ethidium bromide

To verify that cells are digested with the same efficiency in different experiments, the experiment is done three times on exponentially growing cells from the same initial batch and the DNA is purified directly after the sampling.

4.1.2 Growth

4.1.2.1 Growth conditions

External conditions

Cells are grown in an HT infors multitron at 30°C and 180 rpm in sterile plastic erlenmeyers. The correct oxygenation is granted by using erlenmeyers of a capacity 5 times bigger than the culture volume and not tightening the cap. Media (YPD and CASA) are prepared in house (see compisition table 4.2) as well as the amino-acids used to complete CASA medium.

Table 4.2: Composition of growth media

YPD	CASA
<ul style="list-style-type: none"> • Yeast extract (10g/L) • Bactopeptone (20g/L) • D-glucose (20g/L) • in milliQ water 	<ul style="list-style-type: none"> • Yeast nitrogen base (6.7g/L) • D-Glucose (20g/L) • Casaminoacid (caseine hydrolysate) (1g/L) • in milliQ water

Concentration

From stationary phase, cells are diluted to low optical density (OD) (~0.05-0.2) and their growth is monitored by the evolution of the OD with time. After a lag phase, cells enter the logarithmic state where the relation between OD and time is exponential. The curve of $\log(\text{OD})$ as a function of time was fitted by a linear curve in Excel. The slope of this curve (growth coefficient) was used to determine the OD of the cells at t_0 (OD_i) necessary to obtain cells in the logarithmic state ($\text{OD}_f \approx 0.5$) the next morning ($\Delta t = t_f - t_0$) by the formula :

$$\text{OD}_i = \exp(\log(\text{OD}_f) - \text{growth}_{coef} \times \Delta t)$$

Cells in logarithmic state can be diluted to OD_i in the volume needed and grown overnight. This method is the best to obtain a population of cells in logarithmic phase and leaving a whole day to perform experiments. However, most of the time very small volumes must be sampled and despite agitation, yeast are prone to

form aggregates leading to important errors. Therefore it is preferable to perform a strong dilution of the stationary cells in order to be able to sample volumes $\geq 20\mu\text{L}$.

When necessary, stationary cells can also be diluted to an OD of $\sim 0.2-0.4$ and grown until the exit of the lag phase ($\sim 30-60$ min for BY strains and $\sim 90-120$ min for MCM869 strain).

Contaminants such as bacteria can artificially increase the OD and absence of contamination was regularly checked by visible light microscopy (Zeiss Standard 20 with a 40 objective).

4.1.2.2 Thymidine analogs

Products

Thymidine and 5-Bromo-2'-deoxyuridine (BrdU) are purchased from Sigma Aldrich as powders and stored at -20°C in 1 or 0.5mL aliquots at 100mM concentration. EdU is purchased inside Click-it® EdU Alexa Fluor® 488 imaging kit.

EdU click it reaction

Click-it reaction is performed on fixed cells : 1mL of cells sampled from the growing medium are mixed with 1mL of ice-cold 95% ethanol and stored at 4°C overnight. Cells are concentrated and washed twice with PBS 1X SVF 3% by centrifuging 2 min at 9000rpm (eppendorf 5415R rotor FA-45-24-11). Cells are permeabilised 20min maximum at room temperature in PBS containing Triton 100x 0.5%. Cells are washed again twice with PBS 1X SVF 3% by centrifuging 2 min at 9000rpm. Cells are resuspended in 100 μL of the click-it reaction mix containing the reconstituted products of the kit (for 500 μL : 438 μL PBS, 10mL copper, 2.5 μL Dye Azide Alexa fluor 488, 50 μL reaction buffer additive) and the reaction is performed at room temperature during 30min protected from light. Cells are washed again twice with PBS 1X SVF 3% by centrifuging 2 min at 9000rpm and incubated 10min with 200 μL PBS containing 0.2 μL Hoetsch or DAPI. Cells are eventually washed again twice with PBS 1X SVF 3% by centrifuging 2 min at 9000rpm, resuspended in PBS and stored at 4°C .

Fluorescence measure and quantification

10 μL of cells were deposited on a microscope slide and covered by a glass coverslip. After a short equilibration period, the thin layer of solvent between the slide and the coverslip is just enough to contain one layer of cells therefore the hydrated cells are all present in the same focal plan. Images were taken using a Leica

epifluorescence microscope equipped with a 100 oil immersion objective and a charge-coupled device camera.

Images were analysed with imageJ as followed. Nucleus were localised in the DAPI field and the selected area was reported in the EdU field. The intensity of fluorescence was measured and reported to an Excel template were average values were computed for all conditions.

In on of the experiment described in the result, click it reaction was performed on cells already used for FACS experiment whose DNA was stained with PI (see next section). However, PI does not allow a proper nucleus detection therefore only the number of cells with an incorporation of EdU (localised signal versus equal signal in all the cell) were counted and divided by the number of observed cells in bright field.

4.1.3 Synchronisation

4.1.3.1 α -factor

Product

The α -factor used was purchased as a peptide in powder (WHWLQLKPGQPMY) at IGBMC Lab in Strasbourg. Stock solution at a concentration of 10^{-3} M in H₂O was stored in 1mL aliquots at -20°C .

Usage

Determined concentration of α -factor (10^{-7} M for MCM869 and 10^{-6} M for BY4741, see first chapter of results and discussion) is added to the growth medium of cells in logarithmic state (OD \sim 0.5 see previous section). Cells are treated during 2 to 4 hours (see first chapter of results and discussion) in normal growing condition. Cells are then collected by 3min centrifugation at 4000rpm (eppendorf 5810 A-4-62) and washed three times with YPD by 2min centrifugation at 4000rpm to remove α -factor. Cells are then resuspended in the appropriate medium depending on the experiment.

4.1.3.2 Nocodazole

Product

Nocodazole was purchased from Sigma as a powder and stored at 4°C . Stock solution at a concentration of 3mg/mL in DMSO was stored in 0.5mL aliquots at -20°C .

Usage

In its typical usage, 15mg/ml of frozen nocodazole is directly added to the warming medium prior to the experiment and left under normal growing conditions for 20 to 45 min to ensure a correct defrosting and solubilisation.

In the specific case of the measure of its blocking abilities on BY4741 cells presented in the part 7.1.1, Nocodazole in unfrozen and 15 $\mu\text{g}/\text{ml}$ are added to the medium containing cells in logarithmic state. In addition, every hour, half the volume quantity initially used is added to the medium. Note that the stock of nocodazole was all unfrozen and left at 30°C at the beginning of experiment (RT was too low to allow DMSO to unfrozen) which probably leads to an inactivation or a degradation of the product.

4.2 FACS

4.2.1 Sample preparation

Sampling

During experiment, 1mL from the experiment volume is sampled in 2mL eppendorf and mixed with 1mL 90% Ethanol by turning the eppendorf twice. The sample are stored on ice until the e of the experiment and stored at 4°C at least overnight. Samples in this state were stored for a maximum of 2-3 weeks but can be kept virtually indefinitely.

Staining with propidium iodide

The mixture of growth medium and ethanol is removed by 2 min centrifugation at 9000rpm (FA-45-24-11). Supernatants are poured off and pellet is washed once with 0.5mL PBS 1X with 2min centrifugation at 9000rpm. Pellets are then resuspended in 100 μL PBS with 0.25mg/mL RNase A. Cells are incubated at least one hour at 50°C. 100 μL of 50 $\mu\text{g}/\text{mL}$ filtered PI in PBS are added to each sample and incubated at room temperature hidden from light for 15min. 800 μL of PBS were added to dilute the PI and samples were kept at room temperature for an extra 45min. Samples were kept at 4°C at least overnight. At this stage samples were kept 1 to few days before analysis but are viable to be analysed for at least 3-4 weeks.

Sonication

Cells are sonicated exactly five seconds to separate cells (Sonics Materials Vibra Cell). Despite a precisely measured time of sonication, cells at the late time points

are more prone to form aggregates.

4.2.2 Data acquisition

4.2.2.1 Device settings

Device and software

The device used is a FACScalibur (Becton and Dickinson) and the data were acquired in CellQuestPro software (version 5.2).

Template

Already existing template named cycleJM is used for the data acquisition. The parametered windows are a representation of $\log(\text{SSC-H})$ as a function of $\log(\text{FSC-H})$ and a rectangular area defines the events counted as cells. The representation of FL2-A as a function of FL2-H is approximately linear for unique rounds cells which allows to check for abnormal aggregations. Then counts of FL2 presents the actual data that will be analysed later.

4.2.2.2 Acquisition

Cells are diluted twice in filtered PI 5mg/mL usually using 250 or 500 μL of sample and are well mixed by pipetting back and forth. Cells are passed at high speed leading to a count of approximately 7000 events per second until 100 000 events are counted. Voltage on FL2 is arbitrary set in order to obtain the first peak at around 200. Settings are usually done on the time 0 after α -factor release sample which is supposed to contain only one peak.

4.2.2.3 Data reading

Files are .fcs that can be read in Matlab using the function `fca_readfcs.m` (code in annexe). The home-made function `extract_data.m` reconstitutes the histogram based on the data exported via `fca_readfcs.m`. Adjustable boundaries can be set according to the setting of the voltage during the acquisition and are essential to remove free propidium iodure as well as aggregated cells. The column containing the relevant data is also changing depending on the voltage adjustment thus it is an adjustable parameter with a default value set at 4 as it is the column containing the relevant data with the most common voltage adjustment.

This function returns the binned data as well as the fraction of cells among all the event counts, the duration necessary for the analysis and the number of cells counted by time unit which is proportional to the concentration assuming a

constant flow rate. Moreover, in order to compare easily the data, the count value is normalised by the number of cells obtained by under curve integration.

4.2.3 Data analysis

The analysis of the data is made by a home-made package of three functions :

- `get_dataNpeaks.m` : uses `extract_data.m` to read all the data files in a folder, normalises the signal by the number of cells, performs the detection of G1 and G2 peaks for normalisation and saves the truncated data set as well as the peaks position
- `rep_frac_comp.m` : reads the data saved from `get_dataNpeaks.m`, normalises them by the peak position, computes the replicated fraction and stores normalised data and replicated fraction
- `FACS_fit.m` : uses the normalised data saved from `rep_frac_comp.m`, fits them with three gaussians and computes the associated replicated fraction and population proportion

4.2.3.1 Normalisation

Peak detection

The function `find_peaks.m` of Matlab is used to propose peaks position visible on a plot of the data which can be confirmed manually by clicking on the proposed position or elsewhere. Indeed, data contains a certain level of noise that makes the detection of a unique value for the G1 and G2 peaks difficult and variable between two data points and experiments. In consequence, the manual confirmation of automatically proposed peaks position is a robust semi-automatic way to identify the peaks position.

Two positions must be proposed manually then these two values are treated as such : minimum value is the G1 peak except if this value is higher than 6 times the lower boundary used to select the data then the G1 peak is half the G2 peak and maximum value is the G2 peak except if this value is less than $\frac{4}{3}$ of the minimum value the the G2 peak value is twice the minimum value. This conditions are made for the cases where only one peak position is explicit (no automatic value proposed).

Normalisation

The G1 peak correspond of a DNA quantity of 1 and the G2 peak of 2. Therefore, knowing G1 and G2 corresponding fluorescence levels allows to normalised the

data by the formula :

$$\text{DNA content} = \frac{\text{Level}_{\text{Fluorescence}} - G1_{\text{position}}}{G2_{\text{position}} - G1_{\text{position}}} + 1$$

Extraction and fit of replicated fraction

The replicated fraction (F) can be then easily obtained if x is DNA content and y the normalised cell count by :

$$F_{\text{norm}} = \sum x_i \times y_i - 1$$

F is computed for each time point and plot as a function of time after release from α -factor block. $F_{\text{norm}}(t)$ can be fitted with a Boltzmann sigmoid to obtain the S phase duration $4 \times \text{slope}$.

$$F_{\text{norm}}(t) = \text{min} + \frac{\text{max} - \text{min}}{1 + \exp\left(\frac{t_{\text{mid}} - t}{\text{slope}}\right)}$$

The fit is performed with the Fit.m function using the Levenberg-Marquardt algorithm also based on the lsqnonlin.m Matlab solver.

Consecutive cycles

The sigmoid function describe the lagging phase during which the cell reorganises after the release from treatment, the increase of DNA content during S phase and the plateau due to nocodazole or simply G2 phase. However when cells are analysed during longer times in absence of nocodazole, replicated fraction decreases and increases again. Therefore three Boltzmann sigmoids (Boltz, see first paragraph) with their associated parameters are used to fit the data. To avoid interference between the two first function, two unit step function (H, matlab heaviside function) with a set cut value T where introduced.

$$F_{\text{norm}}(t) = \text{Boltz1}(t, P_1) \times H(t - T) + \text{Boltz2}(t, P_2) \times H(t - T) + \text{Boltz3}(t, P_3)$$

4.2.3.2 Data fitting

Fitting of normalised data

FACS data contain three populations : G1 cells which is the first peak, G2 cells which is the second peak and S phase that are the cells between the two. The level of fluorescence (and therefore apparent DNA quantity in normalised data) in one of these population can be represented as gaussians (gauss) because it is not exactly the same in all cells. The Matlab function normpdf.m returns the value of a gaussian function depending on the mean μ and the standard deviation σ :

$$y = \exp\left(-\frac{0.5 \times \left(\frac{x-\mu}{\sigma}\right)^2}{\sqrt{(2\pi) \times \sigma}}\right)$$

Three of these functions modulated by different amplitudes and a constant are used to define the signal obtained in FACS to represent respectively, G1, S and G2 populations. G1 and G2 population are linked as the quantity of DNA double between G1 and G2 therefore the parameters of the third gaussian are linked to those of the first.

$$y = A_1 \times \left(\text{gauss}(\mu_1, \sigma_1) + A_2 \times \text{gauss}(\mu_2, \sigma_2) + A_3 \times \text{gauss}(2\mu_1, \sqrt{2}\sigma_1) \right) + cst$$

Data are fitted using the Fit.m function and in particular the constrained trust-region-reflective algorithm that is based on the lsqnonlin.m Matlab solver inside constrained boundaries. The boundaries are necessary to attribute population to a particular peak. The boundaries are defined to force the second gaussian representing S population to have its mean always included between the first and the last.

Extraction of replicated fraction and populations

From the fit, the position of each population peak μ and the number of cells in these peaks N_1, N_2, N_3 obtained by under curve integration are used to compute replicated fraction.

$$f_{fit} = \frac{N_1 \times \mu_1 + N_2 \times \mu_2 + N_3 \times 2\mu_1}{N_1 + N_2 + N_3}$$

f_{fit} is normalised by the position of the first peak that correspond to a replicated fraction of zeros to obtain the replicated fraction F_{fit} .

$$F_{fit} = \frac{f_{fit} - \mu_1}{\mu_1}$$

In addition, the proportion of cells in each population is computed simply by dividing the number of cell in one population N by the sum of the cells in all three populations.

Cycling cells

The number of cycling cell in each population is computed by subtracting at each time point :

- cells that are still in G1 at time 90 min (corresponding at the end of S phase) from G1 population
- cells that are in G2 after α -factor treatment (non synchronised cells) from the G2 population

The replicated fraction is then re-computed as explained in previous paragraph but with these corrected values for N_1 and N_3 .

Chapter 5

SAS

5.1 Sample preparation

Spheroplasts

Cells synchronous or asynchronous in growing phase are collected by 5 min centrifugation at 4000rpm. The supernatant is discarded and the wet pellet is weighed. The pellet is then resuspended in 3.5 ml TE 1x pH8 per g of cells and 17.5 μ l of β -mercaptoethanol per g of cells is added. The cells are incubated in a hot water bath at 30°C during 15min with gentle shaking. Cells are collected by a 5min centrifugation at 4000rpm and resuspended in 4mL/g of cells of buffer S (see table) and 1x zymolyase is added from fresh 50x zymolyase (see table). Cells are incubated in a hot water bath at 30°C for ~45min with regular checks under microscope for the spheroplasts formation characterised by their round shape. To monitor the spheroplasts formation, the OD of cells mixed with water or 1% SDS can be monitored and will drastically drop when the majority of cells is spheroplasted.

Spheroplasts are then washed three times by 3min centrifugation at 4000 rpm at 4°C and stored at 4°C or -20°C for longer times.

Nuclei isolation

Nuclear isolation mainly follows Hanh Lab protocol for nuclear extract adapted by my lab colleague Bénédicte Barrault.

Spheroplasts are made with the previously described protocol from 200mL of exponentially growing cells at OD ~0.5. Spheroplasts are resuspended in 5mL of buffer A (see table) complemented by 50 μ L PMSF 100x, 100 μ L complete anti-proteases mix 50x and 15 μ L DTT. Spheroplasts are physically crushed by 3 dounces (using a B dounce) on ice. The mix is then centrifuged three consecutive times at 11000 rpm (Beckman TL-100) for 8min at 4°C to remove cells fragments by discarding

S buffer (100mL)	zymolyase 50x
<ul style="list-style-type: none"> • 1M sorbitol • 20mM EDTA • 10mM Tris ph 7.5 • qsp sterile H_2O 	<ul style="list-style-type: none"> • 11mg zymolyase 100T • 10μL sodium phosphate 1M • 690μL glycerol 100% • 690μL sterile H_2O

Table 5.1: Composition of buffers for spheroplasting

the pellet. The supernatant is centrifuged again at 22000 rpm (Beckman TL-100) for 30 min and the pellet collected and resuspended in 500 μ L of B buffer also complemented with anti-proteases, DTT and PMSF (see table).

buffer A	buffer B
<ul style="list-style-type: none"> • 18% Ficoll 400 in H_2O • 10mM Tris pH 7.5 • 20mM K Acetate • 5mM Mg Acetate • 1mM EDTA • 0.5mM spermidine • 0.15mM spermine 	<ul style="list-style-type: none"> • 100mM Tris pH 8 • 50mM K acetate • 10 mM $MgSO_4$ • 20% glycerol in H_2O • 2mM EDTA

Table 5.2: Composition of buffers for nuclei isolation

5.2 Device

5.2.1 Neutrons

Sample holder

Spheroplasts from 150mL of culture are resuspended in 200 μ L D_2O containing 1M sorbitol. the 200 μ L are injected in 1mm quartz cells are left at 4°C to sediment overnight.

For the comparison with chicken erythrocyte SANS data, spheroplasts are resuspended in 0.1% agarose prepared in D_2O containing 1M sorbitol.

Data acquisition and reduction

Data were acquired on PACE spectrometer at LB (CEA-Saclay). Two spectrometer configurations (sample to detector distance/wavelength) were used: 4.6 m/12 Å and 1m/4.5 Å with a respective acquisition time of 2 hours and 1 hour. Data reduction (including normalisation by water, cadmium and empty cell) was made with Pasinet software which uses matlab by the local contact.

5.2.2 X-Rays

Sample in quartz capillaries

10ml of cells from growing culture are washed with TE 1x pH 8 and resuspended in a small volume of TE pH 8. For spheroplasts and nuclei, they are resuspended and their respective buffers.

20 to 50 μ L of cells, spheroplasts or nuclei are deposited in quartz capillary. Box containing capillaries is gently agitated so the volume is at the bottom of the capillary. Tubes are left at 4°C over night to obtain homogenous sedimentation.

Sample in stirrer

Data acquisition and reduction

SAXS data were acquired in SWING beam line at Synchrotron Soleil (Saint-Aubin). All settings were made by SWING local contact. The distance used was ~4m with an energy of 12keV and a wavelength of 1.03 Å. Several frames (5 and 10 respectively for pellet and liquid conditions) were acquired with 100ms, 40ms and 250ms exposure time respectively for yeast spheroplasts pellets, yeast cells pellets and yeasts in liquid conditions. Data acquisition and reduction was made via the software foxtrot (3.3.2) and the data collected as text files. Data reduction was also performed with home-made matlab scripts on raw data files

generated by foxtrot. The data reduction included, removing bad frames (measure in a bubble for example), averaging the frame and subtracting the buffer which was measure in the supernatant for capillaries and before the acquisition for liquid samples.

5.3 Data analysis

Peak parameters

For easier peak detection, the peak parameters were measured on Kratky plots ($I \times q^2(q)$) and in a truncated area around the peak position in a home-made matlab function `peak_param.m`. The area under the peak is computed with matlab function `polyarea.m` between the two closest minimum upstream and downstream the peak position and is normalised by the distance between these two minima.

The evolution of this normalised area ($NA-NA_0$) is computed and plot as a function of the replicated fraction extracted from FACS normalised data (see annexe B).

Data fitting

Data fitting with matlab is sensitive to initial values, in particular with the fitting of SAXS data that have many points and small error bars. Therefore, to be sure that the unsuccessful fit (no converging result after a certain number of iterations) was not due to the initial value selection, the fit was performed in a matlab home-made script using matlab global search algorithm that rely on the matlab solver `fmincon` and perform the fit multiple times from random starting points. In order to save computing time, boundaries were imposed to avoid values negative, too high or equal to zero in the selection of the starting points.

The peak area however was successfully fitted with the Fit function previously used for data fitting using quasi-newton algorithm based on the matlab solver `fminunc`.

Comparison of correlation length

The evolution of correlation length $\Delta\xi = \xi - \xi_0$ is computed fro pellet and liquid conditions. In order to compare the two data sets, the data points of the liquid conditions are averaged in bin centred in the time points value of the pellet condition. The difference between the two data sets taking their error bars into account is computed :

$$\chi^2 = \frac{\sum_i \frac{\Delta\xi_{p_i} - \Delta\xi_{l_i}}{\sigma_{p_i} + \sigma_{l_i}}}{N}$$

Where N is the number of data points.

Chapter 6

DNA combing

6.1 DNA purification in plugs

6.1.1 Cells sample

MCM869 cells (described in first chapter of material and methods) in logarithmic phase were treated 2 hours with 10^{-7} M α -factor in YPD containing 100 μ M dT. Cells were then washed three times with YPD by 3min centrifugation at 4000rpm (eppendorf 5810 rotor A-4-62) and resuspended in pre-warmed YPD medium containing 100 μ M BrdU and 15 μ g/mL nocodazole. At each time point, 10mL of cells are sampled, washed once with YPD once and resuspended in warm YPD medium containing 100 μ M dT and 15 μ g/mL nocodazole. After 2 hours, samples were harvested by 5min centrifugation at 4000rpm. Samples were resuspended in 1mL TE 1X pH 7.5 and transferred to 2mL eppendorfs. Samples were then washed three times with TE 1X pH 7.5 and once with EDTA 50mM by 2min centrifugation 9000rpm (eppendorf 5415R rotor FA-45-24-11).

6.1.2 Plugs preparation

Pellets were kept at 45°C in a hot water bath. Fresh zymolyase 50x solution was prepared by mixing 11mg of zymolyase 100T (amsbio), 10 μ L of sodium phosphate 1M pH 7.6 with 1.38 mL of glycerol 50% . The mix for plug containing low melting agarose 2%, 1m sorbitol and 50mM EDTA was warmed in the microwave until all agarose was dissolved and kept at 45°C in a hot water bath. 1/1000 vol. of β -mercaptoethanol and 1/50 vol. of zymolyase 50x were added to the mix at the very last moment. 200 μ L of the mix were used to resuspend each pellet with cut tips to avoid pipetting hindrance due to agarose. Plugs were formed with 100 μ L of the suspension in the appropriate plug mould. Plugs were kept in the plug mould

to set completely for few minutes at room temperature.

6.1.3 DNA purification

Plugs were put in 12-well plates in 2.5mL per well of spheroplasting buffer containing 1M sorbitol, 20mM EDTA, 10mM Tris pH 7.5. Plugs were incubated in this buffer 2 hours at 37°C. Plugs were then incubated 1 hour at 37°C with 2.5mL per well of lysis buffer containing 1% SDS, 100mM EDTA and 10mM Tris-HCl pH 8. Plugs are then incubated overnight (13h20) with fresh lysis buffer in the same conditions. Plugs were washed three times 30min at 50°C and three times 30min at room temperature with 2.5mL of TE 1X pH8 per well. Plugs were incubated three times (6h, 14h30, 7h30) at 50°C in 2.5 mL per well of Proteinase K solution containing 1% SDS, 100mM EDTA, 10mM Tris-HCl pH 8 and 200mM NaCl with the addition at the very last moment of 0.2mg/mL of proteinase K. Plugs were finally incubated overnight (16h15) at 37°C in 2.5mL per well of Tris-HCl pH 8 containing 2µg/mL of RNase A. Plugs were then washed three times 30min at room temperature with 2.5mL per well of TE 1X pH 6.5. At all steps, plate was covered by three layers of parafilm to avoid evaporation. Plugs were kept at 4°C in 2.5mL per well of TE 1X pH 6.5 with the plate covered by one layer of parafilm.

6.2 DNA combing and staining

6.2.1 Coverslips cleaning and silanisation

18x18mm glass coverslips were rinsed in deionized water three times, soaked in ultra pure chloroform (Sigma) and let to dry under the hood. Coverslips were deeply cleansed by piranha treatment (1/3 vol. H₂O₂ (Sigma-aldrich) and 2/3 vol. sulphuric acid (Sigma-aldrich) for 20min at constant temperature (about 50°C). Coverslips were rinsed three times with deionized water and sonicated in a water bath for 15min in 50/50 vol. water/methanol (Sigma-aldrich). Coverslips were then rinsed three times with deionized water and sonicated in a water bath for 5min in ultra pure chloroform. Dry cover slips were soaked in heptane and treated with trimethoxysilane (Sigma-aldrich) 1µL/ml in heptane (Aldrich) solvent. The reaction was performed overnight under the hood inside a desiccator to prevent humidity. For final cleaning, coverslips were sonicated in a water bath for 5min in heptane. Coverslips were rinsed three times with deionized water and sonicated in a water bath for 15min in 50/50 vol. water/methanol. Coverslips were then rinsed three times with deionized water and sonicated in a water bath for 5min in ultra pure chloroform. Dried coverslips were cut in half, pairs were wrapped in aluminium foil and stored at 4°C.

6.2.2 DNA combing

Plugs were put in 200 μ L of TE 1X pH 6.5 and were melted at 68°C during 20min. Molten agarose was equilibrated at 42°C during 10min and 2 μ L of β -agarase was added and incubated at 42°C overnight (\approx 16h). DNA was stained with 1 μ L of YOYO-1 overnight and kept away from light. 50 μ L of the DNA mix was added to 3.5mL MES 50mM pH 5.8 in a clean Teflon support (cleaning by boiling in milliQ water and ultrapure chloroform left to evaporate). Clean silanized coverslips were fixed to the motorized support, dipped into the liquid for 300s and removed at 25mm/min. Coverslips were taped on both side to a clean microscope slide and YOYO-1 incorporated in combed fiber was observed using a Leica epifluorescence microscope equipped with a 100 oil immersion objective to check the correct combing (straightness, density). Oil was gently removed from the coverslip and the coverslip was glued to a clean slide holder by putting super glue on the face that received oil. Slides were put at 60°C for one hour in a humid chamber.

6.2.3 BrdU and total DNA staining

Slides were dehydrated with three successive 3min baths in 70%, 90% and 100% ethanol. DNA was denature by a bath of exactly 25min in freshly made and filtered NaOH 0.5M. Neutralisation was conducted by three successive 3min baths in filtered PBS 1X. Three successive 3min baths in 70%, 90% and 100% ethanol allowed to fix DNA. Slides were bathed three times 3min in 4 filtered PBS with 0.1% Triton-100X. To prevent non specific fixation of antibodies 20 μ L of filtered PBS 1X containing 0.1% Triton-100X and 11mg/mL of milk powder was deposited on each coverslip for 15min. Slides were then washed three times 3min by baths in filtered PBS 1X. Antibody staining was done by incubation during 45min at 37°C in a humid chamber of 20 μ L of filtered PBS 1X containing 0.1% Triton-100X and 11mg/mL of milk powder containing 1:300 Mouse α -ssDNA MA18 Chemicon BU-75 and 1:20 Rat α -BrdU M299 Seralab BU-75. The same step was repeated with secondary antibodies : 1:500 Goat anti-mouse Alexa 546 M298 and 1:50 Goat anti-rat Alexa 488 M297. Finally, slides are washed by three 3min baths in filtered PBS 1X. A protective 22x22mm coverslip is mounted on top of the coverslips with DPX. After 1 or 2 hours drying the slides are kept at -20°C.

6.3 DNA acquisition and quatification

6.3.1 Image acquisition and treatment

Fiber selection

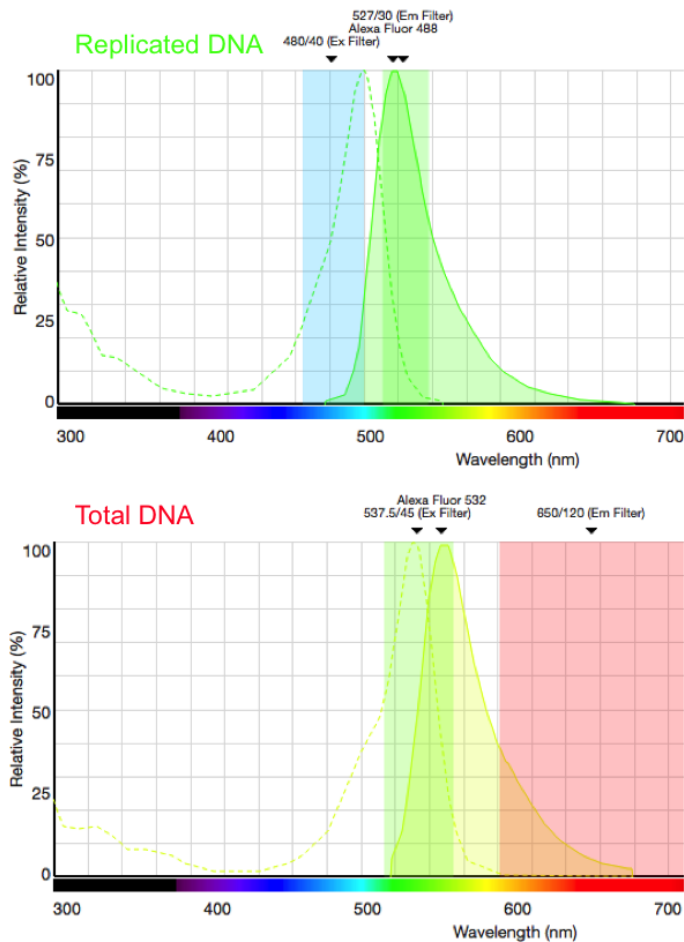
Labelled fibres were imaged using a Leica epifluorescence microscope equipped with a 100 oil immersion objective and a charge-coupled device camera.

The total magnification is measured by imaging a standard giving $922\text{px} = 40\mu\text{m}$ and therefore $1\text{px} = 0.0434\mu\text{m}$.

Two channels are recorded using the filters described in the image on the right. The two channels are merged using image J using green for replicated DNA and red for total DNA to visualize entire fiber.

When necessary, overlapping fields are measured and superimposed with the Gimp software. Images are rotated to obtain horizontal fibers and are cropped to remove all background. An equivalent surface is cropped as close as possible to the fiber in order obtain linear background.

Figure 6.3.1: Fluorescence Spectra of fluorescent dyes used for DNA combing



Spectra of fluorescent dyes used to mark replicated DNA (top) and total DNA (bottom) realised in ThermoFisher SpectraViewer website.

Fluorescence intensity

The linear intensity of the fiber and the background are obtained by averaging the fluorescent intensity of the green channel vertically using Image J «plot profile» tool. The value of intensity for each pixel is saved in a .txt file.

6.3.2 Obtaining and correcting binary values

Binarisation

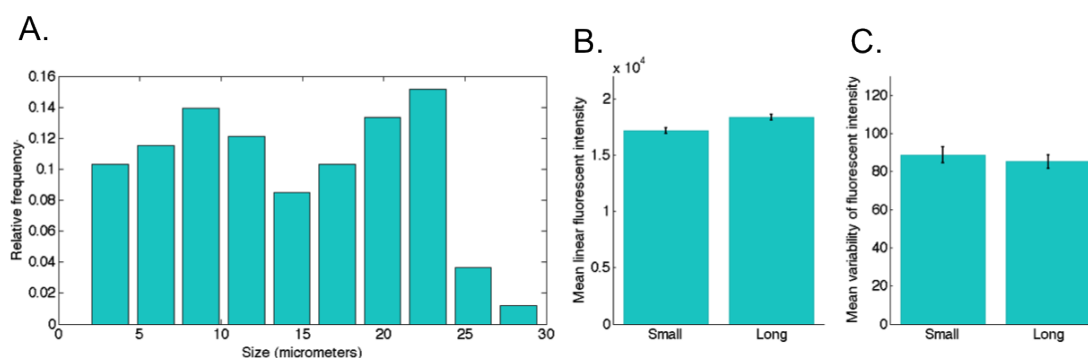
The array of fluorescent values is transformed into an array of 0 (non replicated) and 1 (replicated) using a home-made function in matlab called `binar_maker.m`. First the values of the fluorescent background are subtracted. The threshold value to determine not replicated and replicated parts is set to obtain the greater difference of average replicated fraction between the fibers of the first time point ($t=10\text{min}$) and the last ($t=120\text{min}$).

False negative and false positives

False negative and false positive values are removed in a home-made matlab function called `binar_cleaner.m`. Because the BrdU antibody is not giving a continuous signal, gaps smaller than 50 pixels are removed. On the other hand, because of the optical diffraction limit, tracks smaller than $1\mu\text{m}$ are also removed.

Correspondance between combing and genomic length

Figure 6.3.2: λ -DNA stretching by DNA combing



A. Distribution of the fiber size B. Mean linear fluorescent intensity of small ($<16\mu\text{m}$) and big ($\geq 16\mu\text{m}$) fibers with standard deviation of the mean as errorbars C. Mean variability of fluorescence intensity of small ($<16\mu\text{m}$) and big ($\geq 16\mu\text{m}$) fibers with standard deviation of the mean as errorbars

λ -DNA (Sigma) ($50\mu\text{L}$) is incubated with YOYO-1 ($8\mu\text{L}$) at 4°C overnight protected from light. $5\mu\text{L}$ of stained DNA is mixed with 3.5mL of MES adjusted to pH 5.82 with NaOH 1M and combed at $25\text{mm}/\text{min}$.

Fiber size, total fluorescent intensity and fluorescence variability along the fiber are measured in Image J. The distribution of the size of the fibers is presented in Fig. 6.3.2 A. To check if the small fibers were due to incorrect combing or cuts, average linear fluorescence (total fluorescence divided by the fiber length) (Fig.

6.3.2 B.) and average variability of fluorescence along the fiber (Fig. 6.3.2 C.) are computed for small fibers ($<16\mu\text{m}$) and long fibers ($\geq 16\mu\text{m}$).

The conversion value is obtained by dividing the genomic size of λ -DNA (48.5kb) by the average size of long fibers ($21.2\mu\text{m}$) which results in a conversion of $2.3\text{kb}/\mu\text{m}$.

Time mapping

Replication parameters are obtained as a function of the replicated fraction of individual fibers. To obtain the link between replicated fraction and time, the replicated fraction from raw FACS data $F_r(t)$ is normalised between zero and one :

$$F(t) = \frac{F_r(t) - \min(F_r(t))}{\max(F_r(t)) - \min(F_r(t))}$$

Time corresponding to the experimental replicated fractions t_i is obtained by a linear interpolation of $F(t)$ (matlab function `interp1`). Therefore, replication parameters can be plotted as a function of t_i .

6.3.3 Computation of replication parameters

Identification of track and gaps

All matrix of 1 and 0 are converted to a matlab structre containing of the parameters of the fibers by a home-made matlab function `data_storage.m`. The storage structure is architected as described in Table 6.1.

In particular, the length and positions and tracks and gaps as well as the fork number and density is extracted from binary files with a home-made matlab function named `trackNgap.m` that identify the track borders by looking for the transition of 0 to 1 or conversely. The replicated fraction of a fiber is the replicated length divided by the unreplicated length. The initiation rate is computed as the number of initiation divided by the unreplicated length and dt . Note that the values obtained here are for individual fibers and are not the one that are used for the plots of fork number and initiation rate.

Fork number

Fibers are sorted by their replicated fraction with a chosen bin of 10% (for example each fiber with a replicated fraction between 0 and 0.1 are in the first bin). The density of forks as a function of the replicated fraction $N(f)$ is the number of forks in all fibers of a bin divided by the total length of the same fibers. $N(f)$ is then mapped on time as explained in the previous section to obtain $N(t)$. Because the replication is considered to be spatially homogenous over the size of the fibers, the

Table 6.1: Architecture of the structure used to store information about DNA combing

exp_param	<ul style="list-style-type: none"> • time • time_unit • pixel_size • pixel_size_unit 	<ul style="list-style-type: none"> • experimental time of each fiber • min • 0.0434 • micrometer
fiber	<ul style="list-style-type: none"> • template • length_kb • length_pixel 	<ul style="list-style-type: none"> • initial binary template • length of all fibers in kb • length of all fibers in pixel
analysis	<ul style="list-style-type: none"> • replicated_frac • replicated_length_pixel • unreplicated_length_pixel • unreplicated_length_kb • track_nb • track_length • track_position • ete_dist • gap_nb • gap_length • gap_position • fork_dens • fork_nb • init_nb • init_pos • init_rate 	<ul style="list-style-type: none"> • replicated fraction of each fiber • replicated length of each fiber in pixel • unreplicated length of each fiber in pixel • unreplicated length of each fiber in kb • number of track per fiber • length of tracks of each fiber • tracks positions of each fiber • eye-tot-eye distance of each fiber • number of gaps in each fiber • length of gaps of each fiber • gaps positions of each fiber • fork density of each fiber • fork number for each fiber • number of initiation per fiber • positions of initiation in each fiber • initiation rate of each fiber
analysis_param	<ul style="list-style-type: none"> • combing stretching • combing_stretching_unit • gap_limit • gap_limit_unit • init_limit_size • init_limit_size_unit • dt • dt_unit • units 	<ul style="list-style-type: none"> • conversion value • kb/μm • value of false positive • pixel • size of tracks considered as initiation • unit • interval of time correspondign to this size • unit • list of non explicit units of data from analysis

fork density multiplied by the genome size gives the number of active forks in a cell.

Initiation rate

All tracks <5kb are considered to be created by a single origin and not by the merging of forks from several origin. Therefore the number of tracks smaller than 5kb $T_{<5}$ is the number of initiations that occurred during the time necessary to replicate 5kb : dt (1.78 min considering a constant fork velocity of 1.68kb/min (Sekedat et al. 2010)). Thus the initiation rate as a function of the replicated fraction I(f) is :

$$I(f) = \frac{T_{<5}(f)}{dt * U(f)}$$

where U(f) is the total unreplicated length of the fibers of a given replicated fraction. I(f) is obtained for fibers sorted by replicated fraction bins of 10% then mapped on time as explain on the previous section to obtain I(t).

The global number of initiation per length unit is obtained by numerical integration (matlab function trapz) :

$$I_r = \int_0^t I(t) \times (1 - F(t))$$

Considering that the replication process is spatially homogenous over the size of the fibers, I_r multiplied by the genome size gives the number of origin activated in a cell for a cell cycle.

Created and merged forks

The initiation rate can gives the number of created forks per time. Indeed, each initiation creates two forks and considering that the replication is spatially homogenous, the unreplicated length U can be linked to the genome size G thanks to the replicated fraction F(t) :

$$U(t) = G \times (1 - F(t))$$

The number of created forks per time unit is then obtained by :

$$N_{c,t}(t) = I(t) \times U(t) \times 2$$

The number of created forks between two time points $N_c(t)$ is obtained by integrating $N_{c,t}(t)$ between two time points :

$$N_c(t_{i+1}) = \frac{N_{c,t}(t_{i+1}) + N_{c,t}(t_i)}{2} \times (t_{i+1} - t_i)$$

In order to retain the same number of data points, the number of created forks is initialised to $N_{c,t}(t = 0min) = 0$.

The number of active forks at a given time t_0 is the number of forks created N_c until this time minus the number of forks merged N_m during the same time. Therefore the cumulative number of merged forks can be computed :

$$\sum_0^{t_0} N_c = \sum_0^{t_0} N_m - N(t_0)$$

The sum does not have to be continuous because as it was computed N_c already represents the sum of the forks created between two time points. The number of forks merged between two time points N_m is the subtraction of a time point with the previous time point.

To compute the percentage of forks merged between two time points, the number of merged forks N_m is divided by the sum of created forks N_c during the same interval added to the remaining forks from the previous time interval N .

The maximal distance D travelled by two forks coming up to each other during a time interval Δt is :

$$D = 2 \times \Delta t \times v_{fork}$$

v_{fork} is assumed to be constant and equal to 1.68kb/min (Sekedat et al. 2010).

Origin distribution

The mean of eye-to-eye distance (ETED) of all fibers in replicated fraction bins of 10% is mapped on time as previously explain. The the total origin distribution plot, all ETED are pooled and an histogram is realised with matlab hist.m function with a bin of 5kb. The same ETED with a bin of 10kb is compared to the distance between ACS and between MCM. The positions of ACS-ORC are extracted from Eaton et al. 2010 by taking the middle of the given coordinates and the positions of all MCMs directly from Xu et al. 2006. The distance between two adjacent ACS or MCM is computed chromosome per chromosome by making the difference between the position of adjacent elements. Histograms are also obtained with 10kb bins with hist.m function.

MCM distribution and timing

DNA content over time data measured with high resoltuion over the genome by Hawkins et al. 2013. For each measured position, the DNA as a function of time was fitted with the same boltzman sigmoïd function used to fit replicaated fraction as a function of time extracted from FACS data. Normalised χ^2 and Δt was obtain for each fit and Δt were only considered for $\chi^2 < 1$. The mean Δt and MCM density was computed in 10kb bins. The median value and standard deviation of to Δt was computed for each MCM density value.

Results and Discussion

"Taking a step backward after taking a step forward is not a disaster : it's cha-cha."

(R.Brault)

Chapter 7

Evolution of cell cycle

Freely growing yeast cells are distributed on all the possible phases of the cell cycle. To study DNA replication, it is useful to synchronise this population in the cell cycle in order to have an homogenous behaviour. To analyse how the population progress in the cell cycle, fluorescent dye intercalating inside the ds-DNA with a regular stoichiometry (one dye every 4-5 base pair) such as propidium iodide (PI) can be used to quantify the total amount of DNA present in a cell. Indeed, Fluorescence-Activated Cell Sorter (FACS) allows to obtain the histogram of fluorescence level per cell in a cell population (Van Dilla et al. 1968) which is proportional to the DNA content (Rabinovitch 1994). Tests of two methods of whole-culture synchronisation are detailed in the first section : α -factor pheromone which blocks cells in G1 and nocodazole which blocks cells in G2. Then, a precise analysis of the DNA quantity during cell cycle measured by FACS allows to quantify the progression of S phase in a cell population. This chapter build the basis to characterise cell populations that will be studied with other methods in the subsequent chapters.

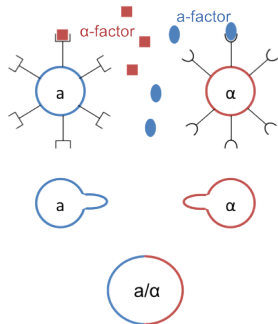
7.1 Cell synchronisation

7.1.1 Cell cycle block

α -factor treatment

Yeast have two mating types Mat a and Mat α . Cells can reproduce as haploid by budding or mate with the other type to form diploid (Fig. 7.1.1). The other mating type pheromone therefore triggers many effect on haploid cells such as change in gene expression and an arrest in G1-phase of the cell cycle.

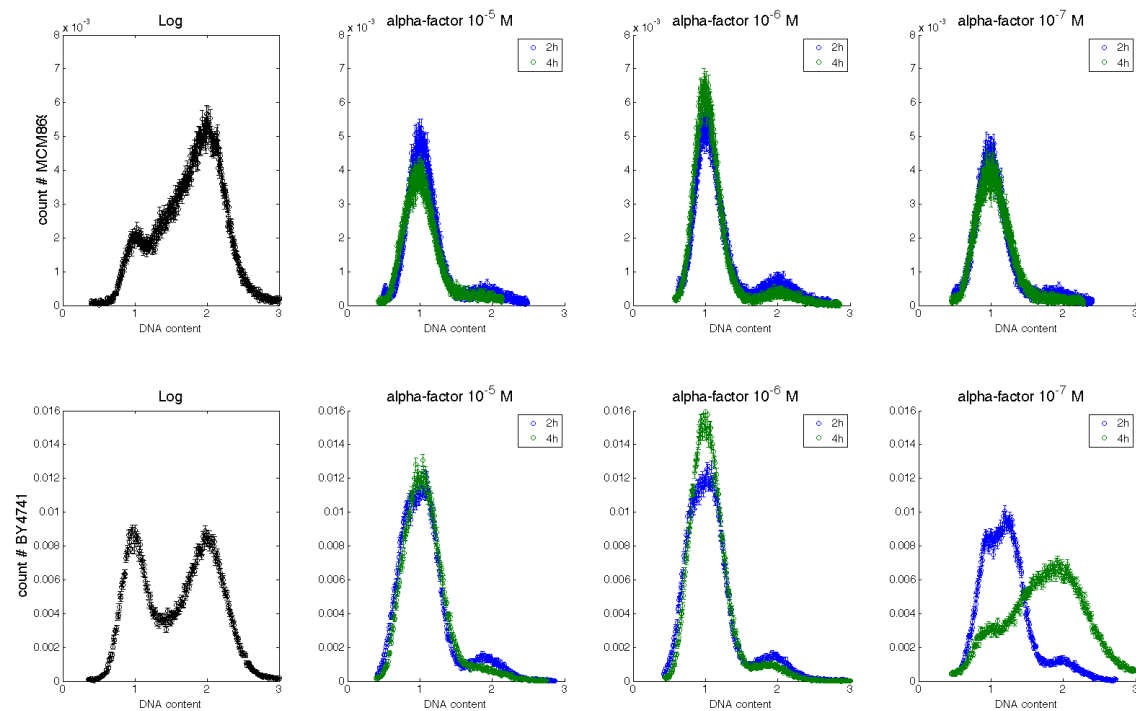
Figure 7.1.1: Schematic effect of mating pheromone



In response to the opposite type pheromone, yeasts develop a shmoo necessary for mating.

In the subsequent experiments, all yeast strains used have a Mat a genotype and thus stop in G1 in presence of α -factor pheromone (Bücking-Throm et al. 1973). When most cells are synchronised, the pheromone can be washed away and the cells resume the cell cycle. The concentration of α -factor as well as the time needed to block the cells was determined in both cell strains used using FACS. Cells with a DNA content of 1 are in G1 while cells in G2-M have a doubled amount of DNA.

Figure 7.1.2: Effect of α -factor on two yeast strains



Normalised count of cells as the function of the DNA content that is derived from the normalised fluorescence intensity (described material and methods). For reference, logarithmic state of freely growing cells is shown in the first column. The other columns present the effect of several concentration of α -factor during 2 or 4 hours.

As shown in Fig. 7.1.2, the logarithmic state of the MCM869 strain is notably and reproducibly (see annexe A.1) biases towards S and G2 phases (Fig. 7.1.2 top left) compared to WT BY4741 strain (Fig.7.1.2 bottom left) that contain a similar

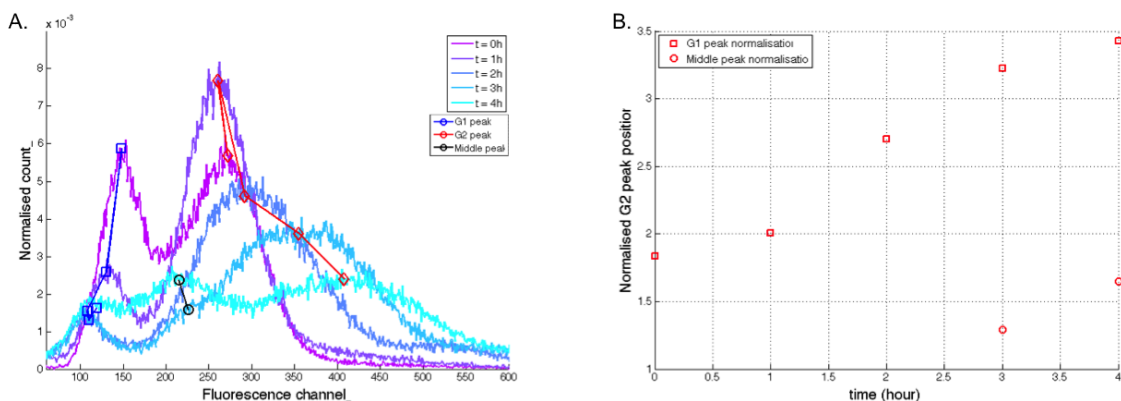
amount of cells in G1 and in G2. In both cases, cells in S phase are not directly visible and the quantification of the S-phase population is described in the next section and in material and methods.

Fig. 7.1.2 shows that most cells have a DNA content of 1 which means they are well synchronised with all concentrations of α -factor tested. Moreover, 2 hours are sufficient to obtain an optimal block. In addition after being too long in α -factor cells can escape the block for example BY4741 cells escape the block after 4 hours in presence of 10^{-7} M α -factor (Fig. 7.1.2 bottom right). Therefore, except described otherwise, MCM869 strain is blocked with 10^{-7} M α -factor during 2 to 4 hours while BY4741 strain is blocked with 10^{-6} M α -factor during 2 to 3 hours.

Effect of nocodazole treatment

The microtubule inhibitor nocodazole prevents nuclear migration and nuclear division in yeast (Jacobs and Szanislo 1982) therefore blocking the cells before the M phase of the cell cycle (Zieve et al. 1980).

Figure 7.1.3: Nocodazole effect on BY4741 cell



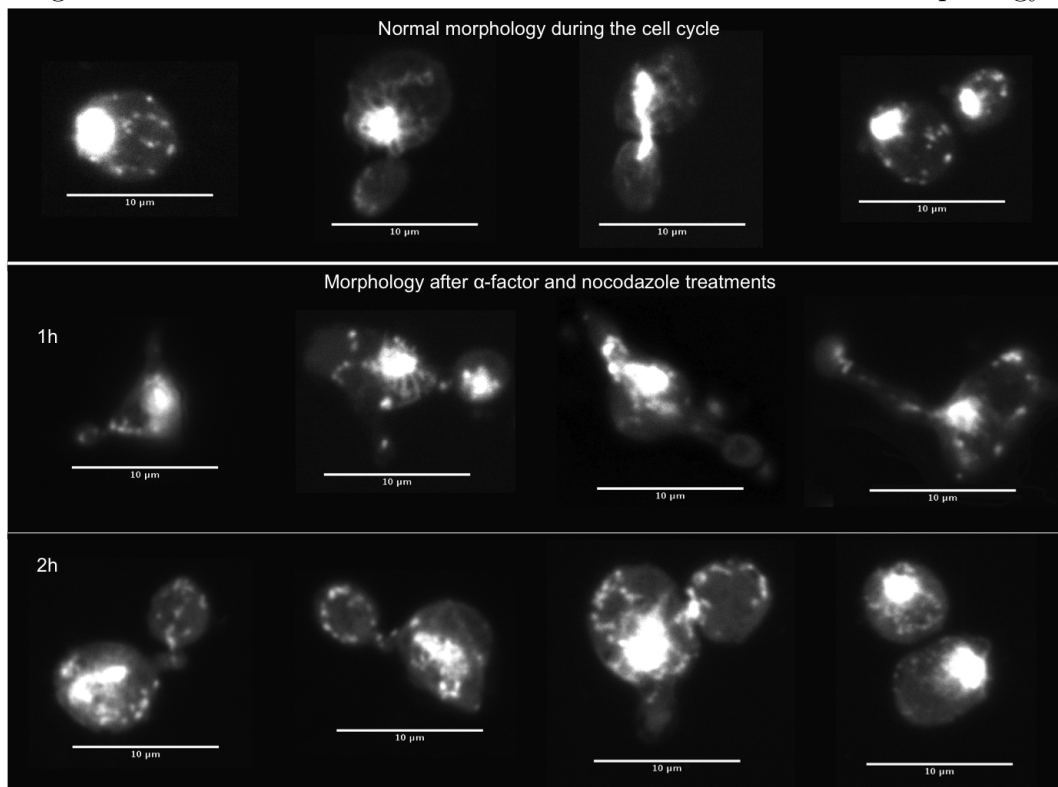
A. Normalised count of cell as a function of a the fluorescence level for several time after the addition of nocodazole. Half the volume of nocodazole is subsequently added every hour. The value of the fluorescence level can be arbitrary shifted during data acquisition however the ratio between the G1 and G2 peak should always be close to 2. Peak position is determined on each curve and represented as linked dots (blue = G1, red = G2, black = middle peak).

B. G2 peak position normalised by the position of G1 peak (square) or middle peak (round) are reported for each time point.

Logarithmic cell population of BY4741 strain are treated with 15mg/mL of nocodazole in YPD medium containing 1% DMSO to optimise the nocodazole solubility. In addition, a larger stock of nocodazole is unfrozen at the beginning of the experiment and half the volume of nocodazole initially added is added every hour to maintain its activity (Jacobs et al. 1988). Fig. 7.1.3 A. represents non normalised fluorescence intensity instead of DNA content. Indeed, the ratio

between the position of the G1 and G2 peak should be equal to ~ 2 (Van Dilla et al. 1968). However, Fig. 7.1.3 B. shows that the quantity of fluorescence in G2 peak increases with time. In logarithmic state, the ratio of 2 between the G1 and G2 peaks is respected (~ 1.8) but keeps increasing during the experiment (~ 3.5). Moreover after 3 hours of nocodazole treatment a second small peak appears and its ratio with the G2 peak is closer to 2 (~ 1.7).

Figure 7.1.4: Effect of nocodazole treatment on MCM869 cells morphology



Fluorescence induced by PI in MCM869 cells at logarithmic state (first row) or after 4 hours α -factor treatment followed by growth in nocodazole for 1 (second row) or 2 (last row) hours. Cell morphology is affected by the formation of the shmoo due to α -factor treatment and by the disorganisation of microtubules by nocodazole leading to abnormal morphology (second row). Isolated dots are probably emanating from mitochondria. Nocodazole treatment mainly prevent the migration of the nucleus. Moreover, DNA seems less localised.

Nocodazole is preventing the polymerisation of microtubule and has therefore an effect on the cell morphology (Fig. 7.1.4). Several possible explanations could account for an abnormal DNA quantity per event : for example, a higher propensity of these cells to form aggregates would lead several cells to be measured as one, and the disturbance of the microtubules could induce re-replication in cells that are not properly divided. However, another possible explanation is that the effect of nocodazole affect PI incorporation leading to a stronger signal from the cells

affected by nocodazole. Indeed, if the chromatin is less condensed, the accessibility to PI is increased. In this hypothesis, the apparition of the peak in the middle could come from cells that divided and retained this altered morphology.

If this middle peak correspond to nocodazole treated cells that have cycled, it means that some cells can escape the nocodazole block after 3 hours. For BY4742 cells, the block is the most efficient after one hour although numerous cells are still in G1.

7.1.2 Effect on S phase

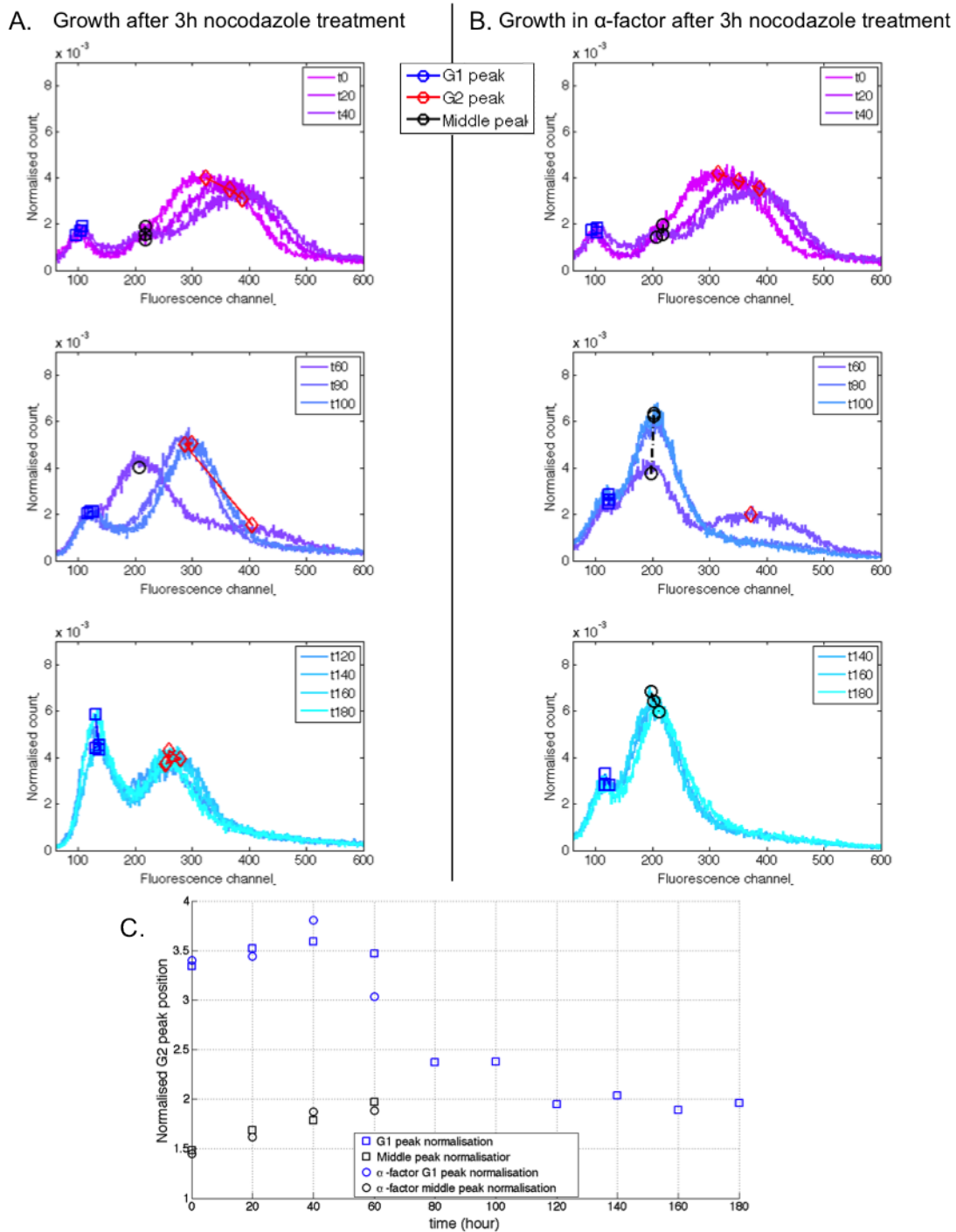
Replication after nocodazole treatment

To investigate further this effect, nocodazole is washed after 3 hours of treatment and cells are allowed to grow with (Fig. 7.1.5 B.) or without (Fig. 7.1.5 A.) α -factor.

The middle peak is still visible as well as the wide G2 peak. In addition, the position of the G2 peak keeps increasing during 40 min (top plot). At $t = 60$ min a drastic change occurs with a clear decrease of the G2 peak and an increase of the middle peak amplitudes (middle plot) which correspond to cells containing twice as less DNA. This suggests that a large fraction of the cell population divided between $t = 40$ min and $t = 60$ min. At $t = 80$ min another drastic change occurs in absence but not in presence of α -factor middle plot. Indeed, cells treated with α -factor remain at this position for the rest of the experiment suggesting that these cells are in a G1-phase state. However without α -factor, the fluorescence level increases and the some cells seem to undergo a second division at $t = 120$ min as the G1 peak increases. At these point, the obtain FACS profile resembles a logarithmic state (Fig. 7.1.2 bottom left). Despite the abnormal morphology, cells have a normal behaviour in the cell cycle.

The most plausible hypothesis is therefore that the PI incorporation is altered by nocodazole treatment leading to a wider G2 peak displaced toward high fluorescence levels. An easier incorporation of PI suggest that the DNA is more accessible and the chromatin less condensed (Rabinovitch 1994). Interestingly, this alteration of the chromatin condensation is conserved during one division but is mainly removed after DNA replication which underlies the role of DNA replication in chromatin reorganisation. Moreover, the replication of this population is well synchronised and rather fast as the totality of the population is replicated in 20 min ($t_{60} \rightarrow t_{80}$ min). This could be explained by a more open conformation of chromatin. Indeed, it was observed that the initiation of late origins is advanced in the case of acetylation (Mantiero et al. 2011, Yoshida et al. 2014).

Figure 7.1.5: DNA replication after nocodazole treatment on BY4741 cells

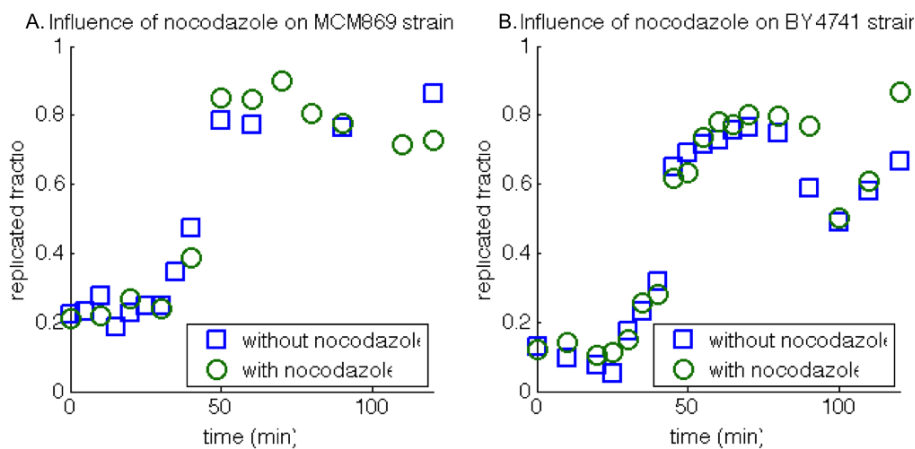


A. and B. Normalised count of cell as a function of a the fluorescence level for several time after the release from the 3 hours nocodazole treatment. For clarity, time points are split in three subplots. Peak position is determined on each curve and represented as linked dots (blue = G1, red = G2, black = middle peak).
 C. G2 peak positions at each time point is normalised by G1 peak position (blue) or middle peak position (black) with (round) or without (square) α -factor treatment.

Replication in nocodazole

MCM869 and BY4741 cells are allowed to grow after α -factor treatment for respectively 4 hours and 2 hours. α -factor is washed out and cells are resuspended in pre-warmed medium containing 15 mg/mL of nocodazole at $t = 0$ min. In addition, for BY4741 cells, half the volume of initially added nocodazole is added every hour. However in both case, the effect of nocodazole on the first cell cycle after the release is not visible (Fig. 7.1.6 A. and B.).

Figure 7.1.6: Influence of nocodazole on DNA replication



Replicated fraction extracted from normalised FACS data (description in material and methods) is plotted as a function of time for MCM869 (A.) and BY4741 (B.) strains in absence (blue square) or presence (green round) of nocodazole.

It was shown that in BY4741 strain nocodazole blocks the cells in G2 after one hour (Fig. 7.1.3) and induces alteration that lead to a stronger fluorescence level which is re-established to normal during DNA replication. Therefore it seems that the effect on cycling cells is negligible. Another possibility is that only very fresh nocodazole can block cells efficiently and that it is rapidly degraded at room temperature therefore adding new nocodazole from a stock unfrozen since the beginning of the experiment is pointless. In addition, nocodazole is soluble in DMSO that has a melting point at $\sim 19^{\circ}\text{C}$ which is about room temperature in air-conditioned growth chamber. Therefore, in routine protocol, nocodazole is solubilised at 30°C in growth medium for at least 30 min which grant a correct solubilisation of nocodazole in growth medium but probably lead to a rapid degradation or deactivation of nocodazole.

In conclusion, the block with nocodazole is not perfectly efficient to prevent cells to cycle after their release from α -factor block. Subsequent analysis will mainly focus on the ~ 100 min after the release from α -factor which corresponds to the

DNA replication and when nocodazole effect on the cell population is negligible but also when cells have an extremely low probability to cycle.

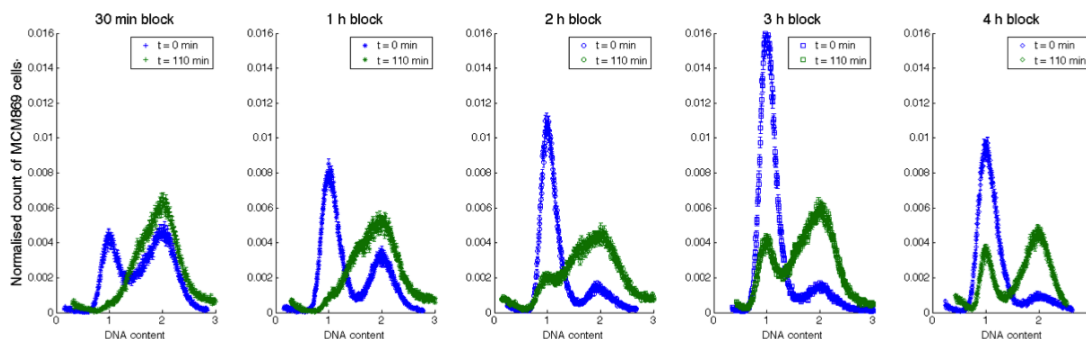
In addition, cells remaining in G2 after the α -factor treatment are exposed to fresh nocodazole. Their division time is thus increased and they can be considered to be retained in G2. Despite an overall negligible effect, nocodazole can be used to improve the population homogeneity so cells are treated with nocodazole in experiment where only the first S phase is to be analysed.

Synchrony of cells in α -factor

In Fig. 7.1.6, the maximum value of replicated fraction is inferior to one which means that some cells did not re-enter the cell cycle after the α -factor treatment. In order to improve the synchrony, the influence of the blocking duration is tested in MCM869 cells. After various duration of α -factor block, cells are allowed to grow.

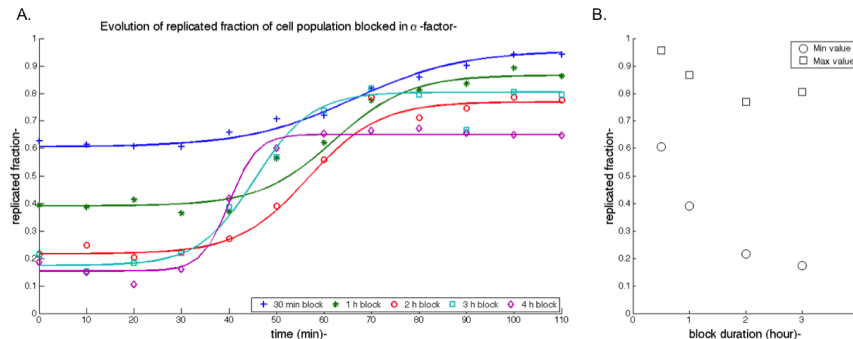
Blue curves of Fig. 7.1.7 confirms that block is efficient after more than 2 hours of treatment, and that cells do not escape the block after 4 hours. This effect is quantified in the Fig. 7.1.8 right plot where the minimum value of replicated fraction is decreasing with block hour duration.

Figure 7.1.7: Blocking duration by α -factor in MCM869 cells influences the synchronisation in G1



Normalised cell count as a function of DNA content directly after and 110 min after α -factor washing. The duration of α -factor block is variable 30min, 1h, 2h, 3h, 4h from left to right

In addition to the synchronisation in G1, the duration of α -factor treatment has an impact on the subsequent growth (Fig. 7.1.8 A.). Indeed, the maximal value of replicated fraction reached at $t = 110$ min also decrease with the duration of the treatment (Fig. 7.1.8 B.). Therefore, the duration treatment increases the synchronisation in G1 but reduce the number of cells that re-enter the cell cycle after the α -factor treatment. This reveals the existence of population subset which is differently affected by α -factor.

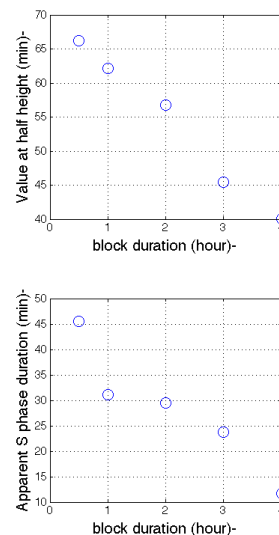
Figure 7.1.8: Cell cycle after various duration of treatment by α -factor in MCM869 cells

A. Replicated fraction extracted from the normalised FACS data (see material and methods and next section) as a function of time for each blocking duration (blue plus = 30 min, red star = 1h, red round = 2h, pigeon blue square = 3h, purple diamond = 4h). In addition, data points are fitted with a sigmoid curve.
 B. Minimal ($f(t = 0\text{min})$) and maximal ($f(t = 110\text{min})$) values of replicated fraction extracted from the sigmoid fit (see material and methods and next section)

However, the cause of this difference is still unknown. In particular, whether the cells that do not re-enter the cell cycle are the same that require more time to be synchronised in G1 or the one that are more sensitive to α -factor treatment remains an open question.

The parameters of S-phase extracted from fit also present an almost linear relation to the block duration (Fig. 7.1.9). After a 4h block duration, the S-phase lasts only $\sim 12\text{min}$. Therefore, the population that re-enters the cell cycle in this condition is smaller but more synchronised. This clearly points out that the S-phase duration extracted from the population analysis does not only depend on the actual duration of the S-phase in one cell but also on the population variability and therefore needs to be referred as 'population S-phase'.

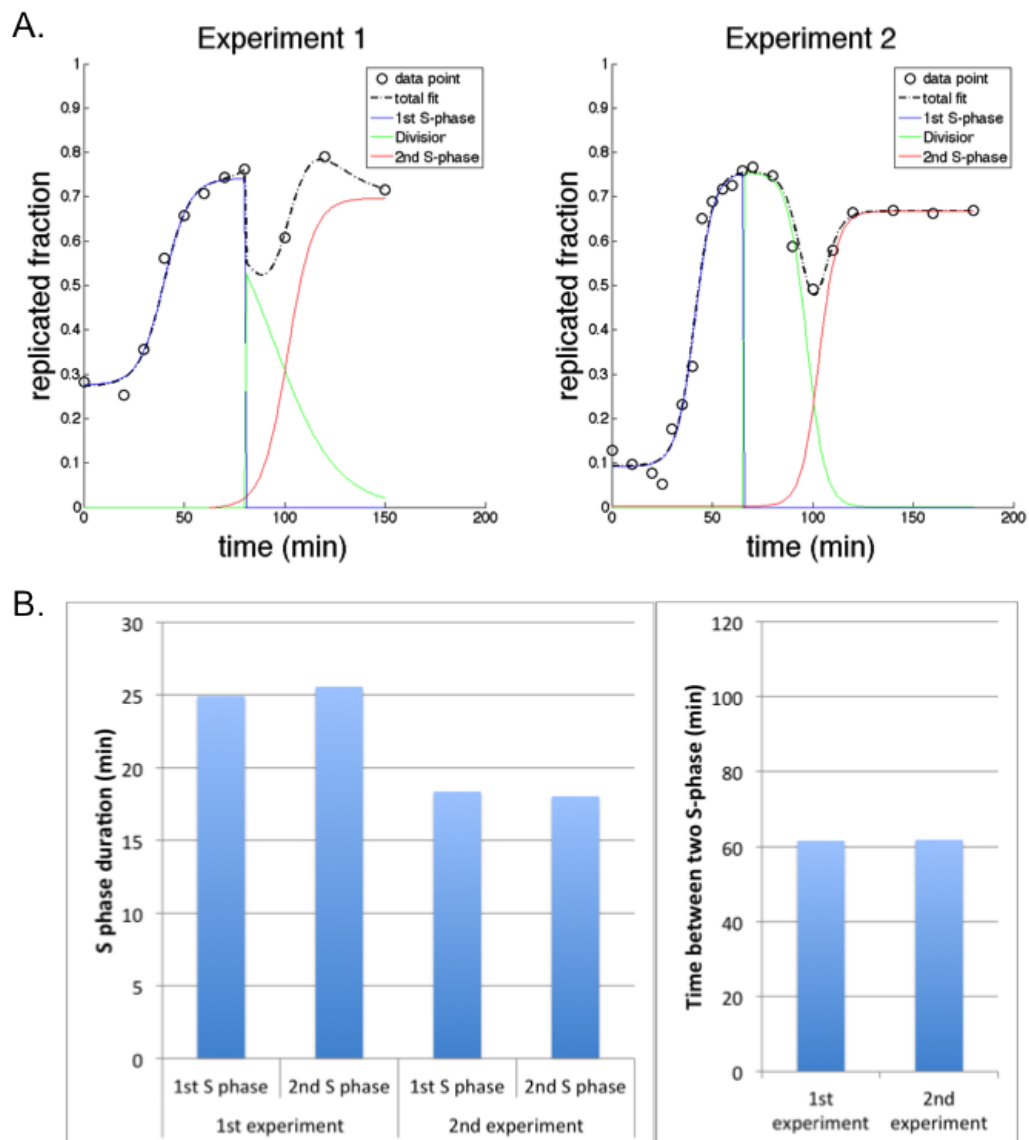
Once again, this also reveals the presence of a population subset with a population S-phase duration inferior compared to other conditions. This could mean that the other populations that re-enter the cell cycle in other conditions are less synchronised or have a longer individual S-phase or both.

Figure 7.1.9: S-phase parameters after various α -factor treatment duration

Value at half replicated fraction (top) and population S phase duration (bottom) extracted from the sigmoid fit

Multiple cell cycles

Figure 7.1.10: Evolution of replicated fraction during two S phases in BY4741 cells



BY4741 cells are allowed to grow in nocodazole free medium after 2 hours α -factor treatment.

A. Replicated fraction as a function of time extracted from normalised data (see material and methods and next section) for two experiments conducted in the same conditions. Data points are fitted with three sigmoids (black dotted line) to represent two consecutive S phase (blue and red) and cell division (green). In addition, in order to prevent the two first sigmoids to interfere, they are multiplied by a unit step function.

B. S phase duration and time between the middle point of the two S phases is extracted from the fit parameters.

To further test the strength of the synchrony obtained after α -factor block, BY4741 cells are grown without nocodazole after a 2 hours α -factor treatment (Fig. 7.1.10). Replicated fraction extracted from two experiments increases during ~ 80 minutes after the release from α -factor block, decreases until a minimum at $t \sim 100$ min then increases again and reaches a plateau corresponding to a DNA quantity in a cell population in logarithmic state (Fig. 7.1.10 A.). Therefore, the synchrony obtained by a two hours α -factor block is partially lost after the first S-phase and totally lost after the second. Fitting the data with three sigmoids to represent the 2 S-phases and the division allows to obtain the parameters of the two consecutive S-phases (Fig. 7.1.10 B.).

In the two experiments, the quality of the α -factor block is different with a high initial value for the replicated fraction at the first time point for the first experiment. This could cause a greater variability in S-phase entry which explains the slightly longer apparent S-phase duration (Fig. 7.1.10 B. left). Interestingly, the duration of consecutive S-phase is really similar. Moreover, the time between two consecutive S-phase is similar in the two experiments. The measure of the optical density of these cells in logarithmic state shows that the concentration of cells doubles in ~ 76 min (see annexe A.2). The value of ~ 61 min found with this analysis is plausible and shows that cells gained in synchronisation.

Synchrony requires that cell are at the same state at one time point but also that they progress together in the subsequent steps. α -factor allows to robustly and rapidly block most cells of a population in G1. However, cells do not re-enter the cell cycle in perfect synchronisation. In addition no optimal block duration can grant both a good accumulation in G1, an entering of the majority of cells in S phase and the synchrony of the cells progressing in the cell cycle. Nocodazole can block most cells in G2 after a short time but cells escape the block easily. In addition, nocodazole has an impact on the PI incorporation that is not removed by nocodazole washing or cell division but by DNA replication. This effect is probably due to an alteration of the cell morphology which lead to a chromatin decondensation that favours the PI incorporation. The influence of nocodazole on the first cell cycle after α -factor block is not visible in the replicated fraction of a cell population.

7.2 Progression in S phase

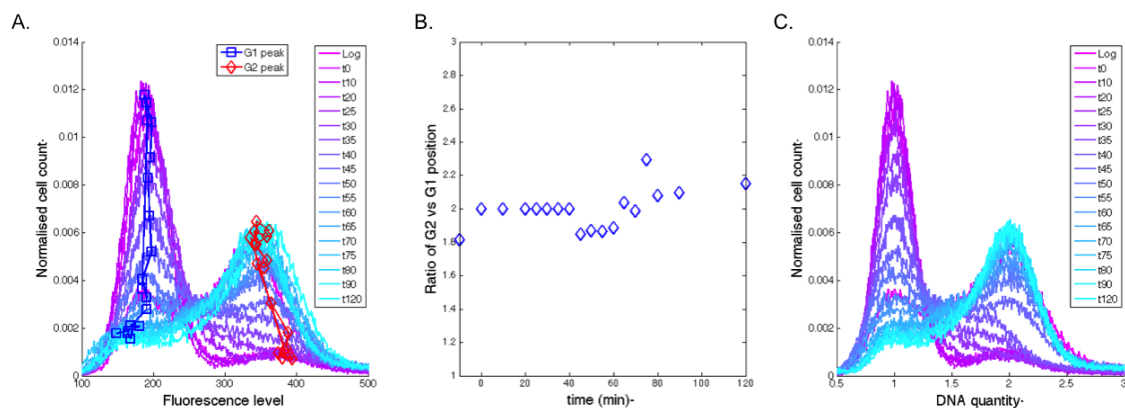
7.2.1 FACS profile contains information on three main populations of cells

Data normalisation

The value of fluorescence level is arbitrary set and can undergo small modifications during one experiment. Fitting the data requires to define initial values which can impact the fit. As fluorescence levels can be different from one experiment to another, the initial values need to be modified for each experiment in order to perform a correct data fitting. The internal normalisation of the data allows to keep the same set of initial values for all time points and all experiments.

To normalise the data G1 and G2 peaks are defined (see material and methods), and the data are normalised to obtain DNA content value of 1 for G1 peak and 2 for G2 peak.

Figure 7.2.1: Fluorescence level to DNA content



A. Normalised cell count from MCM869 cells as a function of the fluorescence level. peaks are not perfectly aligned as highlighted by the G1 (blue square) and G2 (red diamond) peaks position.

B. The defined peak position correspond to a ratio around 2 between the G2 and the G1 position C. Same data with fluorescence level normalised by peaks position

The normalised data is the proportion of cells with a given DNA content and therefore can be used to determine the replicated fraction. The obtained replicated fraction as a function of time can be fitted with a sigmoid function (Yabuki et al. 2002), providing information about the S phase duration (Fig. 7.1.9). However, as previously described, the precise evolution of cells is complex with some cells staying in the G1 phase while other rapidly reach G2 phase. Therefore a quantification of each of these cells populations is of great interest to determine cycling cells.

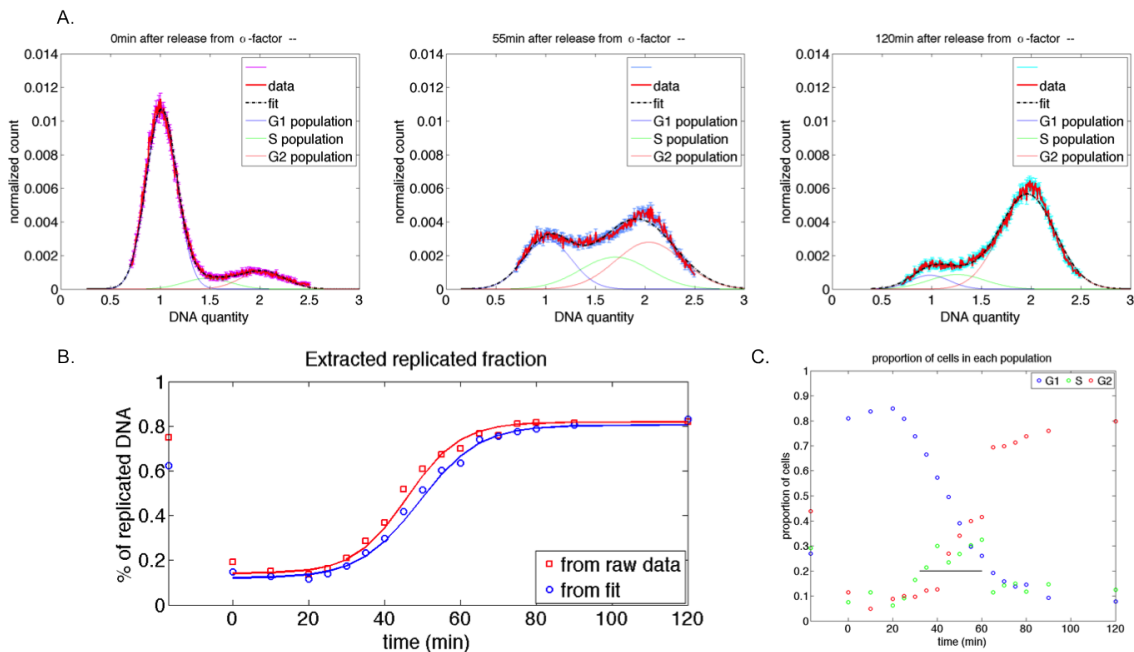
Normalised data can be fitted with three gaussian curves

Normalised data are fitted with three gaussians (see material and methods) (Fig. 7.2.2 A.) representing three populations : G1, S and G2. The accordance between the data and the fit can be assessed by the high R^2 and low χ^2 values (table in annexe ?).

In addition, the replicated fraction extracted from normalised data or from fit are similar (Fig. 7.2.2 B.) except that replicated fraction from fit values are lower. Indeed, normalised data can present a small trail that is due to the signal from aggregates that can contribute strongly to the replicated fraction but is not present in fitted data. Therefore, the gaussians fitting reproduces well the data and remove the noise from multiplets.

Proportion of cells in each population

Figure 7.2.2: Fit of normalised FACS data



A. Example of data fitted with three gaussians at three time points. Noisy data in red and their error bars in different colours are fitted by the black black dotted line and the gaussians corresponding to each population are represented as thin color lines (blue = G1 population, green = S population, red = G2 population)

B. Replicated fraction extracted from raw data (red squares) and from the fit (blue rounds) plotted as a function of time (negative time point is logarithmic state). Data points are fitted with a sigmoid curve (red and blue thick lines).

C. The proportion of cells in each population (G1 = blue, S = green, G2 = red) is plotted as function of time (negative time point is logarithmic state). Black horizontal bar represent the S phase duration computed from the sigmoid fit on raw data.

Splitting the data into three populations allows to quantify the proportion of each population at all time points (Fig. 7.2.2 C.). The G1 population is clearly decreasing during from $t = 20$ min until the end of experiment. G2 and S populations are more difficult to decipher as the G2 peak is significantly wider and often encompass the S peak (Fig. 7.2.2 middle). However, the G2 population is also increasing as expected.

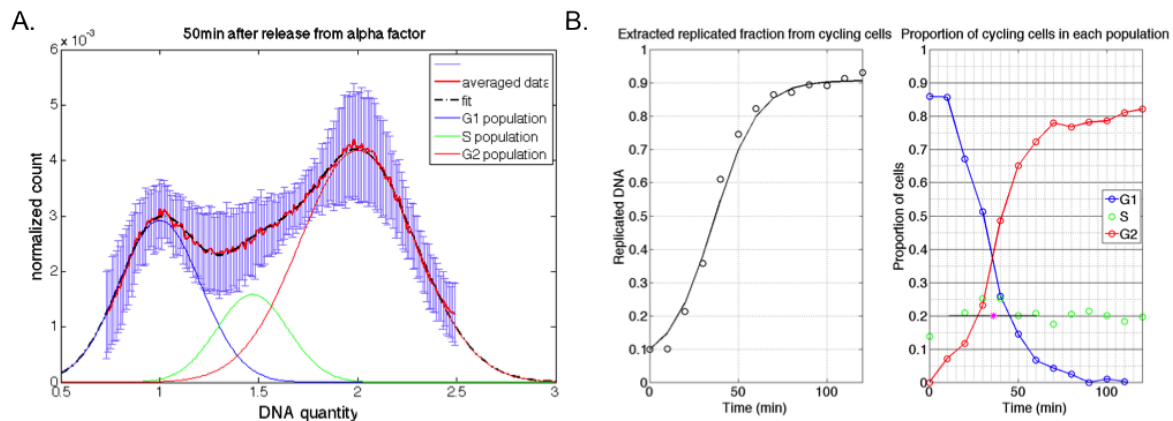
In addition, the evolution of S phase population is consistent with the S phase half value (~ 45 min) and duration (~ 30 min) extracted from the sigmoid fit of normalised data in Fig. 7.2.2 B. red. Therefore, data fitting provide consistent and more precise informations on the DNA replication progression.

7.2.2 Average S phase progression of MCM869 cells

Fit of averaged data

The data fitting can be improved by averaging several data sets. Ten normalised data set from MCM869 cells grown in the same conditions are averaged. The obtained data set with wider error bars is fitted with three gaussians with significantly better χ^2 values compared to single data sets (all parameters of the goodness of the fits are in annexe) (Fig. 7.2.3 A.).

Figure 7.2.3: High quality fit on averaged data allow to extract cycling cells behaviour



A. Example of fit of averaged normalised FACS data

B. Left : replicated fraction of cycling cells : data points and sigmoid fit. Right : proportion of population evolving during S phase from cycling cells (G1 = blue, S = green, G2 = red). The magenta star mark the middle of S phase and the black line mark the duration of S phase obtained by the sigmoid fit on the left figure.

Cycling cells

For the analysis of combing data described in the last chapter, the cell population of interest is the one capable of incorporating modified thymidine analogs referred here as cycling cells. As previously explained, some cells do not re-enter the cell cycle after α -factor block and the block is not perfect. In consequence, the replicated fraction varies only from ~ 0.2 to ~ 0.8 .

Thanks to the population analysis, cells blocked in G1 can be removed. Moreover, cells that are in G2 after the α -factor block will only start to incorporate modified thymidine analog after cell division that can take several tens of minute. Moreover, despite the conclusion that nocodazole show no effect on the replicated fraction variation, it is likely that fresh nocodazole delays the division of these cells with no visible impact on the global replicated fraction. Therefore, this small population of cells would most probably not incorporate thymidine analogs during at least the first hour and it can be removed as well from the analysis. The replicated fraction and proportion of cells in each population can therefore be computed for cycling cells only (see material and methods) (Fig. 7.2.3 B.).

The proportion of G1 and G2 cells in Fig. 7.2.3 B. right presents an intersection that correspond to the middle of the apparent S phase measured by sigmoid fitting of the replicated fraction. The S phase in 2h synchronised MCM869 cells lasts ~ 36 min.

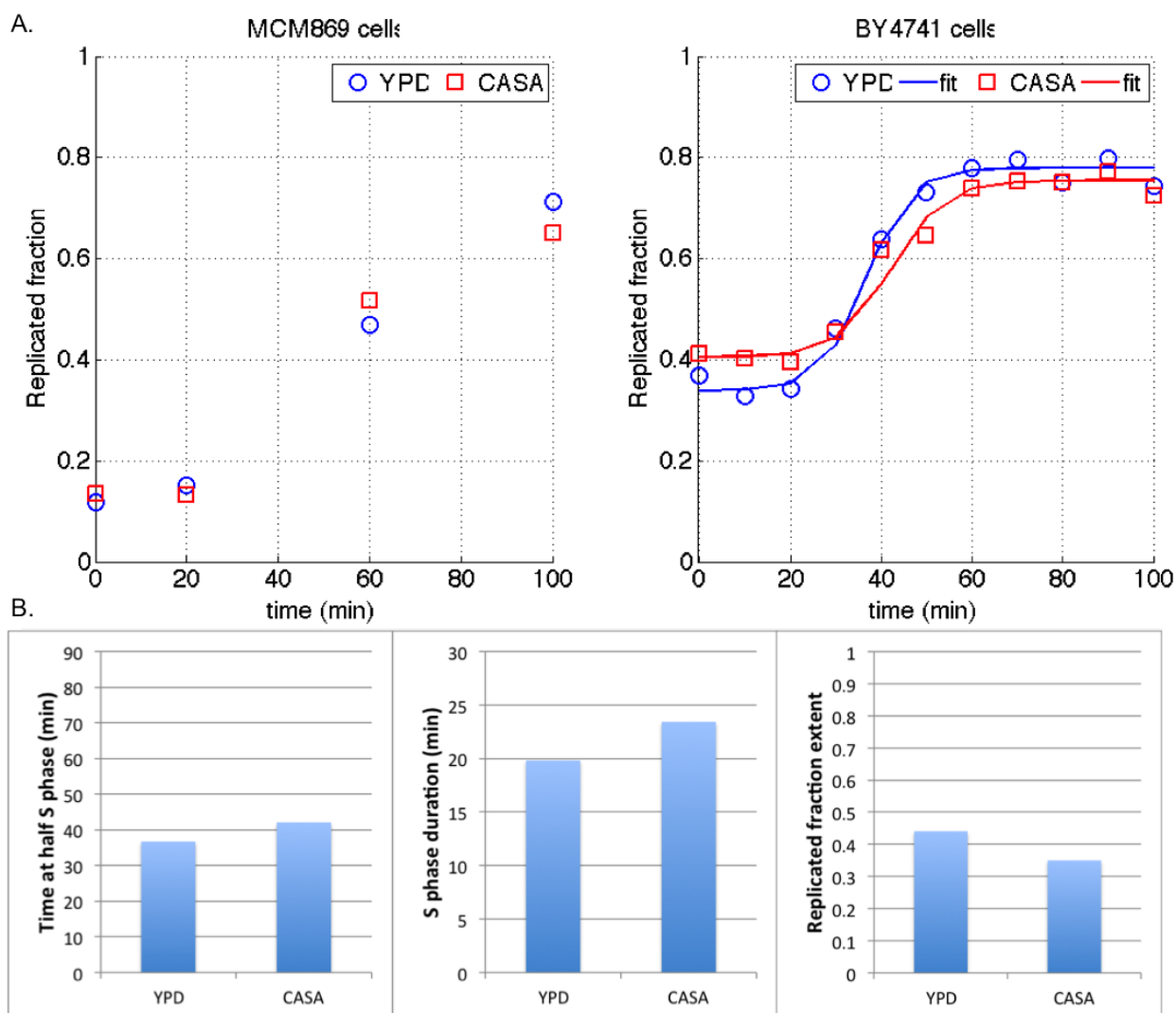
7.2.3 Effect of growth conditions

Growth medium

Most experiments were conducted in rich medium, YPD, which contains yeast extract (extraction of the cells content by removing the cell walls), peptone, distilled water and glucose. In addition, thymidine or thymidine analog was added to a concentration of 100 μ M in the medium for the growth of MCM869 cells that are not able to produce their own thymidine but can pump it from the growth medium (see material and methods).

Although the YPD is ideal for yeast growth, the exact concentration of its component is not precisely quantified and can vary. Therefore, synthetic medium CASA complemented with amino-acids can be used as replacement as it provides only the strict necessary for yeast growth.

Figure 7.2.4: Evolution of replicated fraction of cells growing in YPD or CASA



A. MCM869 (left) and BY4741 (right) cells are grown in YPD (blue rounds) or CASA (red squares) after α -factor block. Replicated fraction extracted from normalised data is plotted as a function of time. Sigmoid fit is performed on BY4741 cells data but not in MCM869 cells for which too few data points are available.

B. The parameters of S phase in BY4741 cells are extracted from fit.

Cells are grown in YPD or CASA after being blocked in G1 by α -factor (Fig. 7.2.4 A.). The two data sets are mostly similar showing that the change of medium does not have a strong impact on the growth of cells.

However, fit of BY4741 cells data reveals small differences on the two medium conditions with a slightly delayed and longer S-phase in CASA as well as a smaller replicated fraction extent (Fig. 7.2.4 B.).

Nature of thymidine analogs

Thymidine is a deoxyribonucleoside that is the association of the thymine and a sugar which is the form in which the bases are incorporated to DNA. Yeast strain MCM869 was modified and is not able to produce thymidine but needs to take it from the growth medium. Thymidine analogs are molecules that differ from thymidine by one chemical group (Fig. 7.2.5). This difference allows a specific recognition by antibodies in the case of BrdU (Dien and Srienc 1991) or a specific chemical reaction that links a fluorophore to the DNA in case of EdU (Salic and Mitchison 2008).

The incorporation of these modified bases is not harmless for the cell as the replacing chemical groups cause a bigger sterical hindrance compared to the normal methyl group. For this reason cells are not able to perform consecutive cell cycles by incorporating BrdU only and the incorporation of EdU is believed to cause DNA breaks (Ligasová et al. 2012). In addition, difference of affinity could lead to a slower incorporation of analogs compared to natural thymidine. Therefore, after α -factor block, MCM869 cells were grown in YPD medium complemented with dT or EdU (Fig. 7.2.6). The nature of the analog do not impact the evolution of replicated fraction. Surprisingly, low concentration and even absence of analog do not seem to have no significant impact.

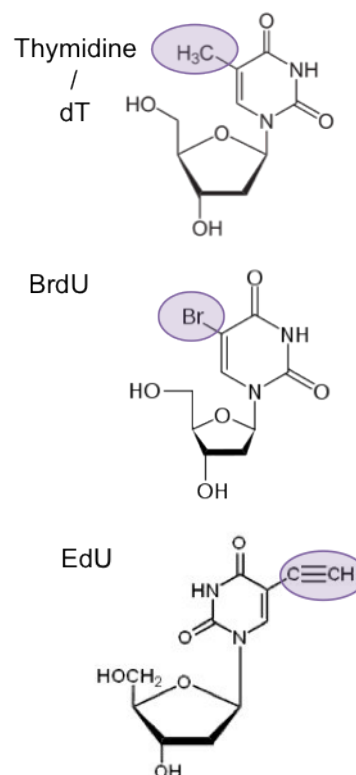
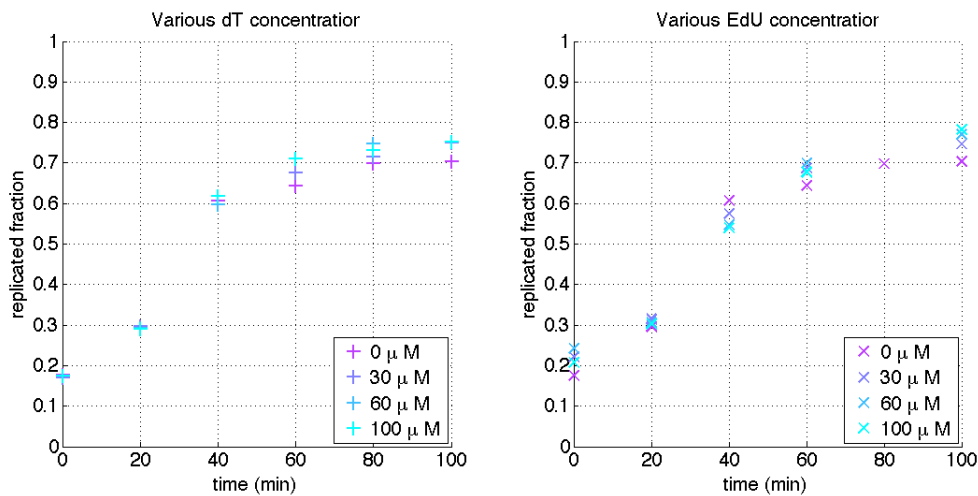


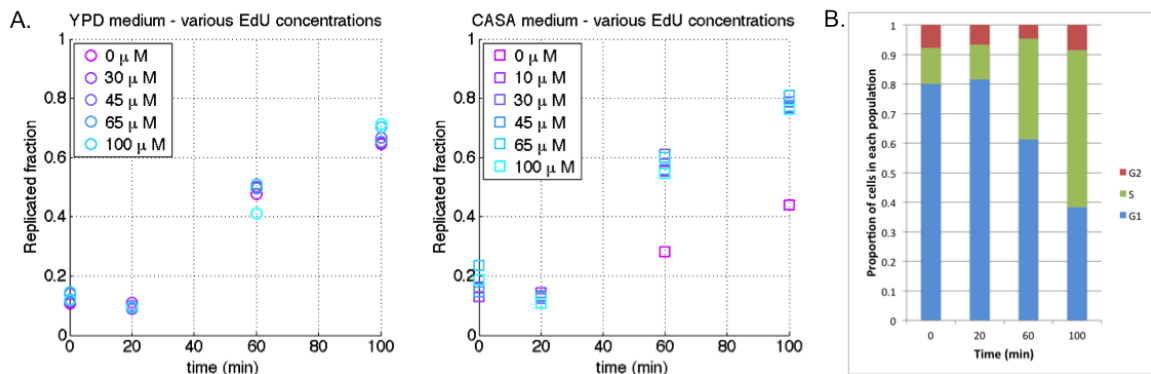
Figure 7.2.5: Thymidine and its analogs
Thymidine analogs are molecules that differ from thymidine by a chemical group (highlighted in violet).

Figure 7.2.6: Effect of the variation of concentration of thymidine analog in YPD



MCM869 cells are grown in YPD medium complemented with various concentrations ($0\mu\text{M}$ = purple, $30\mu\text{M}$ = violet, $60\mu\text{M}$ = blue, $100\mu\text{M}$ = cyan) of thymidine (dT, left) or Edu (right).

Figure 7.2.7: Effect of the concentration of thymidine analog in CASA vs YPD



A. Replicated fraction extracted from normalised FACS data plotted as a function of time for MCM869 cells grown in YPD (left) or CASA (right) complemented with various concentration of Edu (from $0\mu\text{M}$ = purple to $100\mu\text{M}$ = cyan.) after α -factor block.

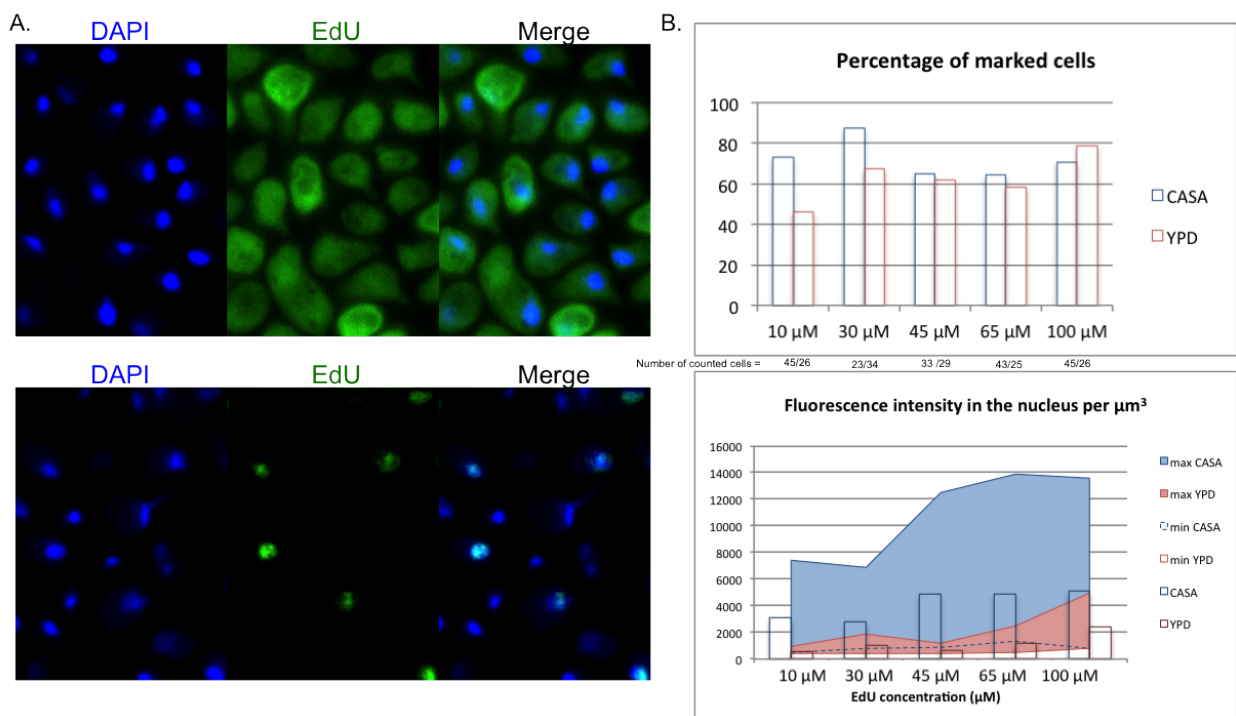
B. Normalised FACS data from the condition $0\mu\text{M}$ Edu in CASA medium are fitted with three gaussians (see material and methods and previous section) to extract the proportion of cells in G1 phase (blue bar), S phase (green bar) and G2 phase (red bar).

To confirm this effect, MCM869 cells were grown in YPD or CASA complemented with various concentration of Edu. The lack of dose dependant effect is confirmed in YPD medium (7.2.7 A. left). In CASA medium, the evolution of replication is normal with the tested concentrations of analog and only a total absence of analog causes a clear alteration (7.2.7 A. right). This result suggests that YPD indeed contains enough thymidine to support normal DNA replication

which is consistent with the fact that low analog concentrations are sufficient to maintain a normal evolution of the replicated fraction in CASA. In absence of any analog in the medium, ~40% of the cells nonetheless start replicating their DNA up to more than 50 % but without completion of the S-phase as no cell reaches G2-phase (Fig. 7.2.7 B.). The extend of the replicated fraction corresponds to ~30 min of a normal cell cycle.

Analog incorporation

Figure 7.2.8: Incorporation of EdU in CASA vs YPD medium



A. MCM869 cells grown in YPD complemented with 10 μM EdU at $t = 0$ min (top) and $t = 100$ min (bottom) after release from α -factor block. Fluorophore chemically linked to EdU is green and cell nuclei are marked by DAPI in blue. In absence of EdU the green fluorophore is distributed everywhere in the cell at low intensity level whereas in presence of EdU it is strongly localised at the nucleus.

B. The percentage of cells that incorporated EdU (top) and the fluorescence intensity by volume unit in the nucleus (bottom) are quantified in cells at $t = 100$ min after release from α -factor block. In the bottom plot, bars show the average intensity value while shade shows the space between minimum and maximum intensity values (blue = CASA, red = YPD)

Thymidine can seemingly be incorporated from non-complemented YPD medium. In addition, the affinity with the polymerisation machinery is greater for thymidine than for its analogs. Therefore, an important question for further work was to verify that thymidine analogs are correctly incorporated. Incorporation of EdU

is quantified in the nucleus of the cells from the previously described experiment at $t = 100$ min after release from α -factor block. A specific chemical reaction on the C-C triple bond of EdU allows to link a fluorophore to DNA at the position where EdU was incorporated. Therefore, the incorporation of EdU is quantified by the intensity of fluorescence from this fluorophore.

The proportion of cells with a EdU signal detected is ~60 to 80% (Fig. 7.2.8 B. top) which is consistent with the fact that a significant proportion of cells do not enter the cell cycle after the release from α -factor block as described in the first section and observed in this experiment by the maximum value of the replicated fraction of ~0.7 (Fig. 7.2.7 A.).

In CASA medium, the average fluorescence intensity is not dependant on the EdU concentration (Fig. 7.2.8 B. bottom) probably because a detection saturation value is reached at $45\mu\text{M}$ EdU as shown by a plateau in the maximum values. The incorporation of EdU is diminished in YPD medium (Fig. 7.2.8 B. bottom) despite similar progression in DNA replication (Fig. 7.2.7 A.) which confirms that cells are incorporating the thymidine contained in YPD. However, the incorporation is dose dependant with a correct level reached at $100\mu\text{M}$.

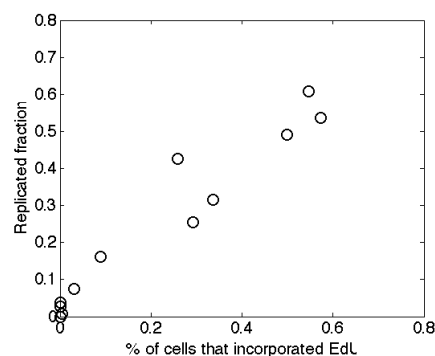


Figure 7.2.9: EdU incorporation during replication in YPD

Replicated fraction extracted from normalised FACS data is plotted as a function of the percentage of cells that incorporated EdU.

In addition, EdU incorporation in cells grown in YPD complemented with $100\mu\text{M}$ EdU was quantified during S phase. The fluorescence intensity in the nucleus was not measured because the nucleus was stained with PI that is less well localised in the cell compared to DAPI and did not allow to make a correct nucleus detection. Nonetheless, the number of cells marked with EdU could be counted. This is also a good indicator of the population progression in the cell cycle that is less precise but also less sensitive to background noise than fluorescence intensity in the nucleus. EdU incorporation follows S-phase progression in YPD medium (Fig. 7.2.9) which implies that a more precise analysis of analog incorporation by DNA combing is relevant to study the DNA replication progress.

7.3 Deciphering the replication program from FACS data

Quantitative analysis of FACS histogram.

Claire Panciatici¹, Agathe Bacquin¹, Anne Peyroche¹ and Arach Goldar^{1,**}

¹Commissariat à l'Énergie Atomique (CEA), iBiTec-S, 91191 Gif-sur-Yvette, France

Received xxxxxxxx; Revised xxxxxxxx; Accepted xxxxxxxx

ABSTRACT

Cell division is subject to a reproducible 4 stages cycle, called "cell cycle". The duplication of the cell's genome is restricted to the second stage of the cell cycle. The invention of flow active cytometry allows to follow the evolution of a cell population with regards to their DNA content. Here we develop a methodology that deciphers the population averaged temporal program of DNA replication from a simple FACS histogram of an exponentially growing cell population. By analysing the cell cycle of 3 *Saccharomyces cerevisiae* mutants, where the regulation of Ribonucleotide reductase is challenged, we show that the replication origin firing is slowed down if the speed of replication forks (v) is increased. This observation leads us to predict that cell control mechanisms (checkpoints) are not activated if $\frac{v_0}{4} < v < v_0$ where v_0 is the speed of replicative helicases

INTRODUCTION

Cell division was discovered more than 150 years ago (1) and studied in many model organisms including yeast and in-vitro culture of mammalian cells. The cell cycle is the general model describing the intracellular control of the cellular division in eukaryotes. It is composed of two main stages: the interphase (I) during which the chromosomes are replicated and the mitosis (M) during which the chromosomes are split between the two daughter cells. Interphase can be more accurately described as a succession of three phases: the **Gap1** phase (G1) where the genome is prepared for duplication, the **Synthesis** phase (S) where the genome is replicated and the **Gap 2** phase (G2) where the state of the genome is verified and the cell gets ready for mitosis.

These phases follow each other in a coordinated manner; none of them can start before the previous one being successfully completed. Moreover, the cell cycle is coupled with other cycles necessary for the functioning of the cell (cytoplasmic cycle, centrosome cycle, nucleolar cycle and Golgi apparatus cycle) and they are all interdependent (2). Numerous control mechanisms (checkpoints) can halt the progression of the cell cycle when an abnormality in the genome integrity is detected (DNA damage, DNA not completely replicated, untied chromosomes...). Since the discovery of such mechanisms in budding yeast (3), the scientific community put tremendous efforts into identifying the molecules involved in this quality control. These molecular interactions are (partly)

characterised showing that the general mechanisms that ensure the smooth progression of the cell cycle seem to be largely conserved among eukaryotes.

Distribution of a cell population around the cell cycle can be monitored by a microfluorometry method based on DNA content developed in 1969 by Van Dilla et al. (4). The principle is simple : cells are labelled with a fluorochrome regularly intercalating in DNA and are flowed in a narrow stream across a beam of exciting light. The resultant fluorescent light pulse is viewed by a photomultiplier and the DNA content is proportional to the fluorescence intensity. The DNA content distribution of an exponentially growing cell population shows two peaks corresponding to cells in G1 phase and cells in G2/M phase. The region between peaks represents cells synthesizing DNA (S phase). This flow cytometry measure is performed by a **F**luorescent-**A**ctivated **C**ell **S**orter (FACS) and is extensively used in a qualitative manner, to study eukaryotic cell cycle and detect their abnormalities.

Therefore, to study the cell cycle, two approaches can be undertaken. On one hand, the system biology is constantly looking for the cell cycle building blocks and relation between molecules identified by genetic and molecular biology. This approach have been successfully applied to the yeast *Saccharomyces cerevisiae* (S.c), where an important number of genetic studies have identified the majority of proteins and metabolic pathways that are involved in the yeast cell cycle (5). On the other hand, as did Van Dilla et al (4), one could wonder why and how the complex combination of actions of an important number of molecules create a reproducible cell cycle. In this approach, no molecule or protein is identified and no track is kept of their interactions, only important degrees of freedom and variables are considered. Choosing the latter is an art, because it involves guessing what is going on and there are many sensible choices, however by parsimony we look for parameters that give the simplest description of the system. In the case of the cell cycle, the method developed by Van Dilla et al. shows that the changes in the amount of DNA content gives the simplest description of progression in the cell cycle.

Therefore, the dynamic of the cell cycle is intimately related to the temporal program of DNA replication which dictates the progression of cells from G1 to G2. The replication of eukaryotic genomes is initiated from multiple chromosomal loci called replication origins (Oris)(6). During G1, Oris are licensed by protein complexes (pre-RCs) that load the

loci called replication origins (Oris)(6). During G1, Oris are licensed by protein complexes (pre-RCs) that load the core of the eukaryal replicative helicase (dh-MCM2-7) as a double hexamers (7). Inactive dh-MCM2-7 at the start of S phase correspond to potential origins (8). Multiple origins fire during S phase at different times by the action of trans-acting factors, but not every origin fires in each S phase. The firing of a replication origin creates two replication forks that progress in opposite directions and synthesize DNA by polymerizing the available deoxyribonucleotide triphosphates (dNTPs). The coordination of the multiple initiations, terminations and replication fork progression events is essential to ensure replication completion within a limited time. Here, we attempt to show how the quantitative study of the cell cycle can decipher the temporal program of DNA replication. First we describe the developed methodology. In a second part, by taking yeast S.c as an example, we discuss how our methodology allows to extract quantitative information about the temporal program of DNA replication from the FACS histogram of an exponentially growing cell population.

MATERIALS AND METHODS

Strains

The strain, MCM869 (*MATa ade2-1 trp1-1 can1-100 leu2-3, his3-11,15 URA3::GPD - TK_{7x} AuR1c::ADH - hENT1 bar1Δ::LEU2 cdc21Δ::kanMX*) is used to compare the inferred temporal program of DNA replication and the experimentally measured one. It is a derivative of the E1000 strain (9) in which the CDC21 gene (encoding thymidylate synthase) was disrupted to increase BrdU incorporation (10). To infer the temporal program of DNA replication in relation with the dNTP pool we used 4 strains:

1. BY4741, (wt): *MATa trp1Δ::hist5+ hist3Δ-1 leu2Δ0 met15Δ0 ura3Δ0*
2. BY4741.*dif1Δ*, (*dif1Δ*): *MATa trp1Δ::hist5+ hist3Δ-1 leu2Δ0 met15Δ0 ura3Δ0 dif1::KANMX4*
3. BY4741.*sml1Δ*, (*sml1Δ*): *MATa trp1Δ::hist5+ hist3Δ-1 leu2Δ0 met15Δ0 ura3Δ0 sml1::KANMX4*
4. BY4741.*hug1Δ*, (*hug1Δ*): *MATa trp1Δ::hist5+ hist3Δ-1 leu2Δ0 met15Δ0 ura3Δ0 hug1::KANMX6*

Unless indicated otherwise, cells were grown at 30°C in YPD.

Cell culture

Exponential population From stationary phase, cells are diluted to low optical density (OD) ($\approx 0.05-0.2$) and

their growth is monitored by the evolution of the OD with time. After a lag phase, cells enter the logarithmic state where the relation between OD and time is exponential. The curve of $\ln(OD)$ as a function of time was fitted by a linear curve in Excel. The slope of this curve (growth coefficient, α) was used to determine the OD of the cells at t_0 (OD_i) necessary to obtain cells in the logarithmic state ($OD_f \approx 0.5$) the next morning ($\Delta t = t_f - t_0$) by the formula: $OD_f = OD_i \exp(\alpha \Delta t)$. Cells in logarithmic state are diluted to OD_i in the volume needed and grown overnight.

Synchronized population Exponentially growing cells ($\approx 5 \times 10^6$ cells/mL) are arrested in G1 by incubation for 2 hours in the presence of $0.1 \mu M$ α -factor and $100 \mu M$ thymidine (Sigma), harvested to be released in S-phase by washing out the α -factor with three washes with fresh medium and transferring cells in pre-warmed medium containing $100 \mu M$ BrdU and 15 mg/mL nocodazole (Sigma). 1 mL samples ($\approx 10^7$ cells) are collected at relevant time points (time step 10 to 20 minutes) to monitor S phase progression by FACS according to Epstein and Cross (11)

Quantitative FACS analysis

Analysis of synchronized cell population To extract the distribution of cells in each phase and the fraction of DNA synthesis we used a home-made package developed using Matlab. The FACS histograms are extracted from raw FACS files by counting the number of events in each channel. Extracted signals are normalized by the total number of counted cells, and channel position (μ) of G1 and G2 peaks are detected. The position of the G1 peak corresponds to a cell DNA content of 1 and the G2 peak position to a cell DNA content of 2. As the values of channels are linearly proportional to the cell DNA content, the channels are converted to DNA content using $x = \frac{\mu - \mu_{G1}}{\mu_{G2} - \mu_{G1}} + 1$. FACS histograms contain 3 population of cells each in a different cell cycle phase: G1 cells are represented by a peak centered around $x=1$, G2 cell by a peak centered around $x=2$ and S phase cells form a peak around a x value that is between 1 and 2. We represent each of these populations by a Gaussian function centered around 1 for G1 with a standard deviation σ_{G1} ($g(\mu_1, \sigma_{G1})$), around 2 for G2 with a standard deviation σ_{G2} ($g(\mu_2, \sigma_{G2})$) and an arbitrary value $\mu \in]1, 2[$ for S phase with a standard deviation σ_S ($g(\mu, \sigma_S)$). The standard deviation of G2 phase is defined as $\sigma_{G2} = \sigma_{G1} \times \sqrt{2}$ (12). The normalized FACS signal is modeled as a linear combination of these three functions: $y = A_1 \times g(\mu_1, \sigma_{G1}) + A_2 \times g(\mu, \sigma_S) + A_3 \times g(\mu_2, \sigma_{G2}) + C$, where C is a constant.

Data are fitted using the constrained trust-region-reflective algorithm that is based on the lsqnonlin.m Matlab solver inside constrained boundaries. The boundaries are necessary to attribute population to a particular peak. The boundaries are defined to force the second gaussian representing S population to have its mean always included between the first and the last Gaussian. The temporal distribution of cells in each phase corresponds to the integral of each Gaussian function for $x \in [1, 2]$ divided by the sum of the integral of the three Gaussian function at each experimental time points. The fraction of

replicated DNA is calculated as :

$$f_{DNA}(t) = \frac{\int_1^2 A_1 g(\mu_1, \sigma_{G1}, x, t) dx \times \mu_1 + \int_1^2 A_2 g(\mu_2, \sigma_S, x, t) dx \times \mu_2 + \int_1^2 A_3 g(\mu_3, \sigma_{G2}, x, t) dx \times \mu_2}{\int_1^2 A_1 g(\mu_1, \sigma_{G1}, x, t) dx + \int_1^2 A_2 g(\mu_2, \sigma_S, x, t) dx + \int_1^2 A_3 g(\mu_3, \sigma_{G2}, x, t) dx}$$

RESULTS AND DISCUSSION

Extraction of replicated fraction from a G1 blocked-and-released *S. cerevisiae* cells

The progression through S phase of *S.c* cells released from α -factor mediated G1 arrest was monitored by FACS (Figure.1). The relative amount of DNA, x , was set to $x=1$ in G1 cells. A broad peak of fluorescence intensity was observed at all time points because the propidium iodide (PI) fluorescence follows a Gaussian dispersion whose standard deviation varies with DNA concentration as $\sigma = C\sqrt{x}$, where C is a constant (12). As the α -factor was washed out, the cells entered S phase and the DNA content increased to reach $x=2$ (G2/M phase). Following Van Dilla et al (4) the FACS histogram was considered as the combination of three populations. We quantified the proportion of each population at all time points (Figure.1) by representing them as Gaussian distributions. The G1 population is clearly reduced during the experiment (from $t=20$ min until the end of the experiment), whereas, the G2 and S populations are more difficult to characterise individually, because the G2 peak is significantly wider and often encompasses the S peak. Despite this difficulty, one could observe that the G2 population is also increasing as expected. Each time point was fitted with the combination of three Gaussians (Figure1) allowing us to determine the temporal distribution of cells in G1, S and G2 phases (Figure2.a and Figure2.b). As was observed earlier, cells do not enter and exit S phase synchronously (13), and their rate of S-phase entrance is slower than their rate of S phase exit. Demonstrating both the heterogeneity of the temporal starting point of replication program in a cell population after release from the α -factor treatment (13), and the variability of the rate of DNA synthesis in a cell population. Indeed, the transition time for the cell population between G1 and S phase is longer than the transition time between S and G2/M phase (Figure2.c), because some of the cells that entered S phase earlier are cached up with by cells that entered S phase later. In other words, cells that start their replication program first have a slower rate of DNA synthesis than cells that start the replication of their genome latter.

One might argue that as these experiments are performed in the presence of nocodazole (an inhibitor of microtubules), one would expect an accumulation of cells in G2/M phase by the end of the replication process which could influence the rate of S phase exist of cells and therefore, bias the apparent rate of DNA synthesis. To verify if the difference between the rate of S phase entrance and exit is an experimental artifact or not, we performed the same type of experiment in the absence of

nocodazole (Figure2.b). Interestingly we observed that the rate of S phase entrance is still longer than the rate of S phase exit and the observed values are similar to those obtained in the presence of nocodazole (Figure2.c). Thus, the variability in the rate of DNA synthesis in a replicating cell population is an inherent property of the replication process. In budding yeast, the speed of replication forks (v) is assumed to be independent of chromosomal location and a constant (14). As the rate of DNA synthesis depends on the speed of replication forks and the rate of origin firing ($I(t)$), the variability of its amplitude in a cell population implies that the amplitude of $I(t)$ and/ or the velocity of replication forks are not the same in all cells. By determining the distribution of cells in each cell cycle phase, one could calculate the population averaged kinetics of DNA replication ($f(t)$) as defined in the Material and Methods. The obtained curve reproduce in the limits of statistical errors ($\chi^2=0.25$, $P \leq 10^{-4}$) the previously experimentally measured kinetics of DNA replication in the same yeast strain (Figure3.a).

Modeling the kinetics of DNA replication

Our population based analysis of FACS histograms comfort the idea that the amplitude of the rate of origin firing is not the same in all individuals of a replicating cell population. This result confirms that in a cell population the number of fired replication origins, their positions and their times of firing is not the same in all cells. To link the extracted population averaged kinetics of DNA replication to molecular mechanisms in a consistent manner, it is necessary to take into account the cell to cell variability in the replication program. Recently, we proposed a nonlocal model for the temporal behavior of DNA replication that takes into account this variability (15). Based on concepts from multiple scattering, we have shown that the dynamics of DNA replication (kinetics of replicated fraction $f_{DNA}(t)$, rate of origin firing $I(t)$ and temporal profile of fork density $N_f(t)$) during the S phase can be completely characterized by the knowledge of 7 measurable parameters: the number of loaded dh-Mcm2-7 m_0 , the total number of fired origins O_{total} , fractal dimension of chromatin d_f , the dynamical fractal dimension d_w , the duration of S phase t_{end} , the speed of replication forks v and the size of the genome L .

It is assumed that the yeast chromatin adopts an equilibrium globule conformation (16) and that the diffusion of proteins is not hindered in the yeast nucleus. Therefore, it is reasonable to fix values of the geometrical fractal dimension to $d_f=3$ and the dynamical fractal dimension to $d_w=2$ (15). The size of the yeast genome is experimentally measured as $L=12 \times 10^3$ kb. Once the value of these three variables is defined, we use the nonlocal model to fit the extracted kinetics of replicated fraction $f_{DNA}(t)$ in order to extract the values of the four other parameters.

To reduce the size of variable space, we use experimental observations to define the boundaries of the region of interest. The number of chromatin bounded dh-Mcm2-7 double-hexamers has been measured by Chip (17) and Chip-seq (18) methodologies, $m_0 \approx 250$. A bias of these methodologies is that during the chromatin immuno-precipitation step, one might not collect all genomic fragments where the protein of interest is bound and also during the analysis of the

	m_0	O_{total}	v (kb.min ⁻¹)	t_{end} (min)
$f_{DNA}(t)$	311 ± 31	173 ± 17	1.46 ± 0.3	36.8 ± 5
FACS	299 ± 30	174 ± 20	1.41 ± 0.3	45 ± 4
ref (15)	322	168 ± 20	1.68	42

Table 1. Parameters values. The $f_{DNA}(t)$ row corresponds to parameters obtained by fitting the extracted $f_{DNA}(t)$. The FACS row corresponds to parameters obtained by fitting the FACS profile of the exponentially growing cell population.

genomic regions where the protein of interest is enriched, one might not detect all of them due to the normalization step of the data analysis protocol. Therefore, the number of chromatin-bound dh-Mcm2-7 double-hexamers estimated by these methods is an under estimation of the real number. The development of the nonlocal model allowed us to bound the possible values of fired replication with respect to the number of loaded dh-Mcm2-7 double-hexamers as $\frac{m_0}{2} < O_{total} < m_0$ (15). In the seminal work of Raghuraman al (19), the speed of replication forks was estimated to span between 0.5 kb.min⁻¹ to 3kb.min⁻¹. Alvino et al (20) have demonstrated that in the presence of 200mM hydroxyurea the speed of replication forks is reduced by a factor of 16. Furthermore, Sekedat et al (14) showed that the speed of replication forks in budding yeast under unchallenged conditions is a constant $v = 1.68 \text{ kb.min}^{-1}$. Therefore, in order to take into account all these observations, we set the lower bound for v to 0 and the higher bound to 3 kb.min⁻¹. Finally, we set the lower bound of t_{end} to 0 min.

Using these constraints, the extracted $f_{DNA}(t)$ is fitted using a simplex algorithm ($\chi^2 = 0.035$ $R^2 = 0.96$ $P \leq 10^{-4}$) (Figure3.a). The fit converges for parameter values that are in good agreement with the previous estimations (Table.1). This allows us to reproduce faithfully the experimentally measured $I(t)$ ($\chi^2 = 0.030$ $R^2 = 0.89$ $P \leq 10^{-4}$) and $N_f(t)$ ($\chi^2 = 0.037$ $R^2 = 0.90$ $P \leq 10^{-4}$) (Figure3.b-c). Encouraged by this result, we propose that one could infer the temporal program of DNA replication by analyzing only the experimentally measured $f_{DNA}(t)$.

However, the experimental extraction of $f_{DNA}(t)$ requires :i) the synchronization of the cell population in G1 phase, which might be a disturbance of the replication program and ii) either complex labeling and detection of newly synthesised DNA or complex numerical analysis of FACS histograms as we did in the previous section. To overcome these labour intensive procedures, we take advantage of the fact that in an exponentially growing cell population the FACS profile depends solely on the rate of DNA synthesis (21, 22), and therefore, if our fitting methodology is correct, one could extract the values of m_0 , O_{total} , v and t_{end} by modeling the FACS profile of an exponentially growing cell population.

Evaluating the temporal-program of DNA replication using FACS histograms

Bertuzzi et al (21) developed an equation linking the FACS profile and the total rate of DNA synthesis for an exponentially grown cell population:

$$\begin{aligned} \tilde{v}(\xi) &= 2K(\xi,1) - K(\xi,2) + (2 - \vartheta_1) \int_1^2 \frac{\delta K(\xi,x)}{\delta x} \exp\left(-\int_1^2 \frac{dz}{w(z)}\right) dx \\ K(\xi,x) &= \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(\xi-x)^2}{2\sigma(x)^2}\right) \\ \sigma(x) &= C\sqrt{x} \end{aligned} \quad (1)$$

where $\tilde{v}(\xi)$ is the normalised fluorescence density of the detection channel ξ normalised by the position of the G1 peak. $K(\xi,x)$ is the kernel function representing the dispersion of PI fluorescence over the detection channels for a cell whose DNA content is x . The form of $\sigma(x)$ was proposed by Bruni et al. (12) on the basis of a model of DNA-dye interaction. The parameter C corresponds to the width of the G1 peak ($x=1$) and is measured to be $C = 0.24 \pm 1.3 \times 10^{-3}$. The parameter ϑ_1 represents the fraction of cells in G1 and represents a fraction of 0.3 of the total population.

The function $w(z)$ is equal to $w(z) = \frac{1}{\alpha} \frac{df(z)}{dt}$, where $\frac{df(z)}{dt}$ represents the rate of total DNA synthesis in a cell with z DNA content. $\alpha = 5 \times 10^{-3} \pm 10^{-3} \text{ min}^{-1}$ is the Malthusien growth exponent of our strain. Once again, using the nonlocal model of DNA replication we calculate $\frac{df(z)}{dt}$ as a function of m_0 , O_{total} , v and t_{end} . Figure3.d represents the fit to the experimentally measured FACS profile of an exponentially growing cell population using a simplex algorithm ($\chi^2 = 4.1 \times 10^{-4}$ $R = 0.991$ $P \leq 10^{-4}$). The extracted values of m_0 , O_{total} , v and t_{end} , are in good accordance with previously published results (Table.1), and again the inferred $f_{DNA}(t)$ (comparison between inferred $f_{DNA}(t)$ and the one extracted from the synchronised cell population $\chi^2 = 0.0623$ $R^2 = 0.97$ $P \leq 10^{-4}$) $I(t)$ (comparison between inferred $I(t)$ and the one extracted from the synchronised cell population $\chi^2 = 8.8 \times 10^{-5}$ $R^2 = 0.999$ $P \leq 10^{-4}$) and $N_f(t)$ (comparison between inferred $N_f(t)$ and the one extracted from the synchronised cell population $\chi^2 = 6.3 \times 10^{-3}$ $R^2 = 0.98$ $P \leq$

10^{-4}) reproduce the profiles extracted in the previous section and experimentally measured kinetics (Figure3.a-c).

Application of the developed methodology.

To illustrate how this methodology could be used, we decided to investigate the impact of dNTP pools on DNA replication. The progression of replication forks and therefore the synthesis of DNA during S phase requires adequate and balanced dNTP pools. Ribonucleotide reductase (RNR) (23) catalyzes an essential step in the production of dNTPs. In S.c, RNR is subject to complex regulation during normal cell cycle progression and in the presence of genotoxic stress. Besides transcriptional regulation, several RNR inhibitors such as Sml1, Dif1 and Hug1 have been identified. The Facs histogram profile variations are subtle and blurred by the statistical errors during data acquisition. To convince ourselves that the Facs histograms are different, we enhance differences among profiles by removing the high frequency noise using a Fourier smoothing algorithm (Figure4.a). Once we visualised these differences, we analyse the raw flow cytometry profile of non-synchronised cell populations of Sml1, Dif1 and Hug1 yeast mutant strains and predict the kinetics of DNA replication (Figure4.b-e, Table.2). In good

	m_0	O_{total}	$v (kb.min^{-1})$	$t_{end} (min)$
<i>wt</i>	271±27	166±17	1.17±0.2	37.3±4
<i>dif1</i> Δ	263±30	155±16	1.26±0.3	28.2±7
<i>sml1</i> Δ	335±25	182±20	1.51±0.15	35.1±3
<i>hug1</i> Δ	277±20	160±18	1.32±0.2	33.7±5

Table 2. Parameters values for *wt*, *dif1*Δ, *sml1*Δ and *hug1*Δ strains.

accordance with others, we find that *sml1*Δ and *dif1*Δ strains replicate faster their genome than the *wt* strain (24), whereas no significant difference was found between the *wt* and the *hug1*Δ strain (25). However, the detailed inspection of $I(t)$ shows that while more origins fire earlier in S phase in *sml1*Δ and *dif1*Δ strains than in a *wt* strain, in the *hug1*Δ strain the rate of origin firing is similar to the *wt* strain at the beginning of S phase but $\approx 15min$ after the start of the latter, less origins are fired and the faster S phase can be attributed to an increase in the speed of replication forks. These observations are in good agreement with the mechanisms where Sml1 and Dif1 regulate the pool of dNTP by down-regulating the RNR from the start of G1 phase either by allosteric interaction (Sml1) (26) or by sequestration of RNR subunits in the nucleus (*dif1*) (27), whereas Hug1 regulates the rate of dNTP production by a competitive interaction and therefore its action becomes effective as it accumulates during S phase (28).

The ratio between the number of fired origins and the number of potential origins ($P = \frac{O_{total}}{m_0}$) defines the strength of origin usage. Figure4.h shows the variation of P as a function of the inferred speed of replication forks in each of the studied strains. Remarkably we find that the origin usage decreases as the speed of replication fork increases (29). Furthermore, following Gauthier et al (30), we calculate $P = \frac{L}{2m_0} \sqrt{\frac{I\pi}{v}} = \frac{1}{2} \sqrt{\frac{v_0}{v}}$, where I is the rate of origin firing averaged over the genome and cell cycle ($kb^{-1}min^{-1}$), and $v_0 = \frac{L^2 I \pi}{m_0^2}$ is a characteristic speed corresponding to the optimal usage of origins by cells to replicate their genome ($P = \frac{1}{2}$ (15)). By fitting the data ($\chi^2 = 4.3 \times 10^{-6}$ $R^2 = 0.995$ $P \leq 10^{-4}$) in Figure4.h to this expression we find that $v_0 = 1.76 \pm 6.3 \times 10^{-3} kb^{-1}min^{-1}$ a value close to the measured rate of replicative helicase progression in S.c (14). As, v_0 corresponds to a limiting speed of the DNA synthesis, we propose that the rate of origin firing is optimal when the speed of replicative polymerases matches the speed of replicative helicases and therefore, v_0 corresponds to the speed of replicative helicases. However, as genome duplication is possible solely for $0.5 < P < 1$ (15), we also propose that the checkpoint mechanism is activated if the rate of polymerisation of the newly synthesised DNA does not satisfy: $\frac{v_0}{4} < v < v_0$. This conclusion is in good agreement with our observation of a moderate activation of Rad53 in the presence of a constitutive replication stress, and an important Rad53 activation in the presence of hydroxyurea (31).

The inferred temporal behaviour of important kinetic parameters of the DNA replication program and the extracted values of the number of loaded Mcm2-7 double-hexamers, the total number of fired origins, the speed of replication forks and the duration of S phase are in good agreement with experimental observations. During the last 50 years, important experimental methodologies have been developed to study the temporal program of DNA replication in eukaryotes. While these approaches give us important insight into this fundamental process, they are demanding and time consuming. Over the last 15 years, it has been shown that mathematical modeling could be a very powerful approach to verify if an assumed molecular mechanism is able to reproduce the experimental observations. Here, we show that by analysing the experimental outcome of a common and simple methodology using a coarse grained model of DNA replication, one could infer the temporal program of the DNA replication with a quite good accuracy. Furthermore, accurate experimental determination of the rate of origin firing $I(t)$ and the density of replication forks $N_f(t)$ is only possible by using single molecule experiments that require the incorporation of dNTP analogues into genomic DNA during its replication which might challenge the latter. Particularly in the yeast S.c this requires genetic modification to give this organism the ability to import analogues from the culture medium. Here, the proposed methodology does not require any particular modification of the organism and the replication program is not challenged. Furthermore, to illustrate the feasibility of this approach, we studied the temporal program of DNA replication with respect of the level of dNTP pools. Our conclusions are in good accordance with experimental

observations in a quantitative manner and confirm proposed molecular mechanisms. Furthermore we propose that while the Sml1, Dif1 and Hug1 mutant have the same phenotype (25) their temporal program of DNA replication are different. In this work, we developed a methodology to study the dynamics of the replication process in *S.c.*, but because of its general character, we believe that it could be used in any eukaryotic organism in order to study the temporal program of DNA replication.

ACKNOWLEDGEMENTS

Authors would like to thanks XXXXXXXX for useful discussion and critical reading of this article. This work was supported by Commissariat à l’Énergie Atomique (CEA). CP is supported by the Soleil, LLB and the Idex.

Conflict of interest statement. None declared.

REFERENCES

1. Nurse, P., Masui, Y., Hartwell, L. H. (1998) Understanding the cell cycle.. *Nat Med*, **4**, 1103–1106.
2. Eigen, M. and Schuster, P. (1979) *The Hypercycle: A Principle of Natural Self-Organization*, Springer Berlin Heidelberg, .
3. Hartwell, L. H. and Weinert, T. A. (1998) Checkpoints: controls that ensure the order of cell cycle events.. *Science*, **246**, 629–634.
4. Van Dilla, M. A., Trujillo, T. T., Mullaney, P. F., and Coulter, J. R. (1969) Cell microfluorometry: a method for rapid fluorescence measurement.. *Science*, **163**, 1213–1214.
5. Chen, K. C., Csikasz-Nagy, A., Gyorfy, B., Val, J., Novak, B., and Tyson, J. J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle.. *Mol Biol Cell*, **11**, 369–391.
6. Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M. (2010) Eukaryotic chromosome DNA replication: where, when, and how?. *Annu Rev Biochem*, **79**, 89–130.
7. Labib, K., and Diffley, J. F. (2001) Is the MCM2-7 complex the eukaryotic DNA replication fork helicase?. *Curr Opin Genet Dev*, **11**, 64–70.
8. Hyrien, O., Marheineke, K., and Goldar, A. (Feb, 2003) Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem.. *Bioessays*, **25**(2), 116–125.
9. Lengronne, A., Pasero, P., Bensimon, A., and Schwob, E. (2001) Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains.. *Nucleic Acids Res*, **29**, 1433–1442.
10. Vernis, L., Piskur, J., and Diffley, J. F. X. (Oct, 2003) Reconstitution of an efficient thymidine salvage pathway in *Saccharomyces cerevisiae*.. *Nucleic Acids Res*, **31**(19), e120.
11. Epstein, C. B. and Cross, F. R. (Sep, 1992) CLB5: a novel B cyclin from budding yeast with a role in S phase.. *Genes Dev*, **6**(9), 1695–1706.
12. Bruni, C., Koch, G., and Rossi, C. (Mar, 1983) On the inverse problem in flow cytometry recovering DNA distribution from FMF data.. *Cell Biophys*, **5**(1), 5–19.
13. Ma, E., Hyrien, O., and Goldar, A. (Mar, 2012) Do replication forks control late origin firing in *Saccharomyces cerevisiae*?. *Nucleic Acids Res*, **40** (5), 2010–2019.
14. Sekedat, M. D., Fenyö, D., Rogers, and R. S., Tackett, A. J., Aitchison, J. D., and Chait, B. T. (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome.. *Mol Syst Biol*, **6**, 353.
15. Goldar, A., Arneodo, A., Audit, B., Argoul, F., Rappailles, A., Guilbaud, G., Petryk, N., Kahli, M., and Hyrien, O. (2016) Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations.. *Sci Rep*, **6**, 22469.
16. Mirny, L. A. (2011) The fractal globule as a model of chromatin architecture in the cell.. *Chromosome Res*, **19**, 37–51.
17. Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., Aparicio, O. M. (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins.. *Science*, **294**, 2357–2360.
18. Xu, W., Aparicio, J. G., Aparicio, O. M., Tavar, S. (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*.. *BMC Genomics*, **7**, 276.
19. Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., and Fangman, W. L. (Oct, 2001) Replication dynamics of the yeast genome.. *Science*, **294**(5540), 115–121.
20. Alvino, G. M., Collingwood, D., Murphy, J. M., Delrow, J., Brewer, B. J., and Raghuraman, M. K. (2007) Replication in hydroxyurea: it’s a matter of time.. *Mol Cell Biol*, **27**, 6396–6406.
21. Bertuzzi, A., Gandolfi, A., Germani, A., and Vitelli, R. (May, 1983) A general expression for sequential DNA-fluorescence histograms.. *J Theor Biol*, **102**(1), 55–67.
22. Bell, G. I., and Anderson, E. C. (1967) Cell growth and division. I. A mathematical model with applications to cell volume distributions in mammalian suspension cultures.. *Biophys J*, **7**, 329–351.
23. Kolberg, M., Strand, K. R., Graff, P., and Andersson, K. K. (2004) Structure, function, and mechanism of ribonucleotide reductases.. *Biochim Biophys Acta*, **1699**, 1–34.
24. Poli, J., Tsaponina, O., Crabb, L., Keszthelyi, A., Pantescio, V., Chabes, A., Lengronne, A., and Pasero, P. (2012) dNTP pools determine fork progression and origin usage under replication stress.. *EMBO J*, **31**,

- 883–894.
25. Kim, E., and Siede, W. (2011) Phenotypes associated with *Saccharomyces cerevisiae* Hug1 protein, a putative negative regulator of dNTP Levels, reveal similarities and differences with sequence-related Dif1.. *J Microbiol*, **49**, 78–85.
 26. Zhao, X., Chabes, A., Domkin, V., Thelander, L., and Rothstein, R. (2001) The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage.. *The EMBO Journal*, **20**, 3544–3553.
 27. Wu, X., and Huang, M. (2008) Dif1 controls subcellular localization of ribonucleotide reductase by mediating nuclear import of the R2 subunit.. *Mol Cell Biol*, **28**, 7156–7167.
 28. Meurisse, J., Bacquin, A., Richet, N., Charbonnier, J. B., Ochsenein, F., and Peyroche, A. (2014) Hug1 is an intrinsically disordered protein that inhibits ribonucleotide reductase activity by directly binding Rnr2 subunit.. *Nucleic Acids Res*, **42**, 13174–13185.
 29. Zhong, Y., Nellimoottil, T., Peace, J. M., Knott, S. R.V., Villwock, S. K., Yee, J. M., Jancuska, J. M., Rege, S., Tecklenburg, M., Sclafani, R. A., Tavar, S., and Aparicio, O. M. (2013) The level of origin firing inversely affects the rate of replication fork progression.. *The Journal of Cell Biology*, **201**, 373–383.
 30. Gauthier, M. G., Herrick, J., and Bechhoefer, J. (2010) Defects and DNA replication.. *Phys Rev Lett*, **104**, 218104.
 31. Ma, E., Goldar, A., Verbavatz, J. M., and Marsolier-Kergoat, M. C. (2011) Giant yeast cells with nonrecyclable ribonucleotide reductase.. *Mol Genet Genomics*, **285**, 415–425.

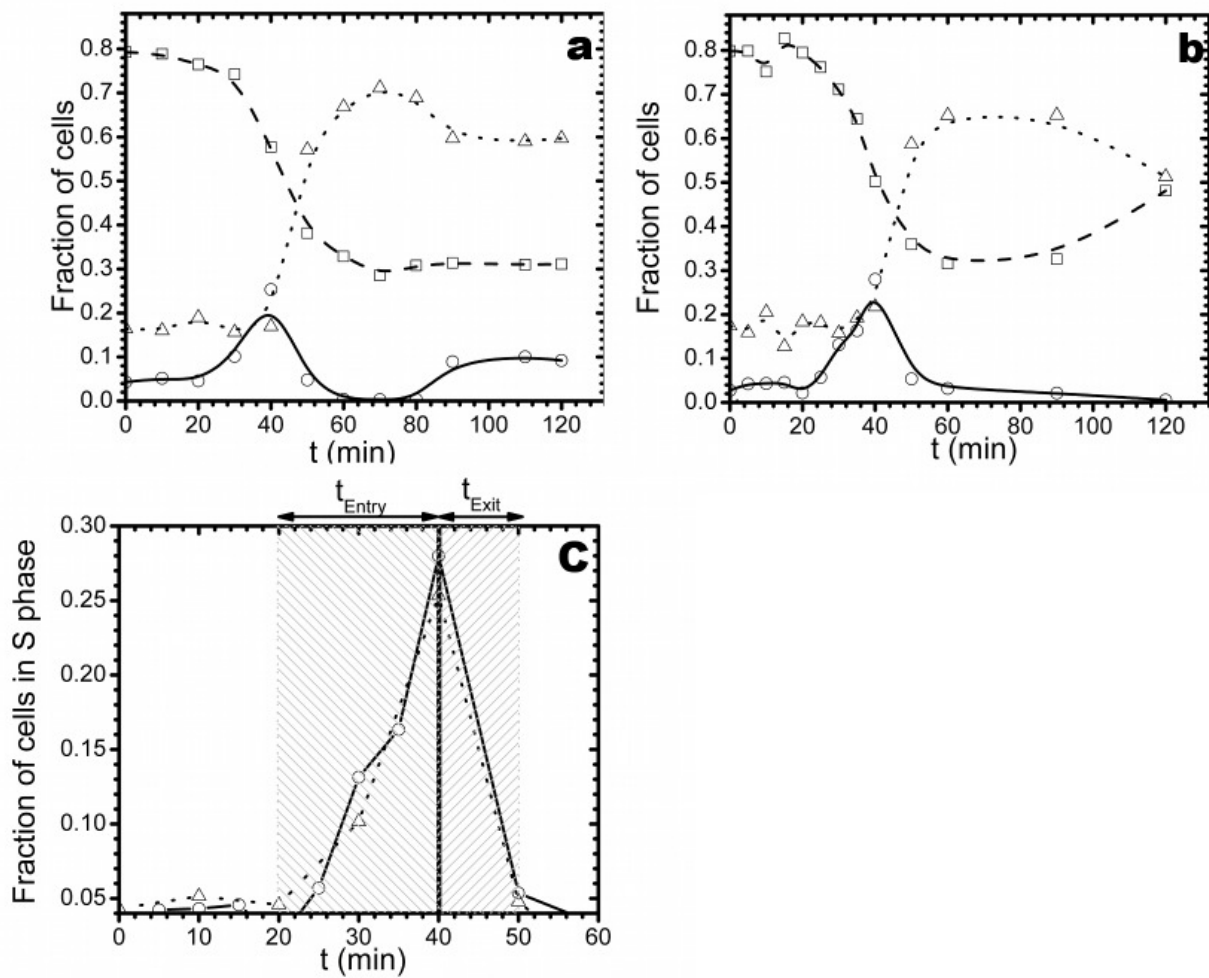


Figure 2. Fraction of cells in G1(□), S (○) and G2 (△) phases following α factor release in presence (a) and absence (b) of Nocodazole. The solid lines are visualisation guides. c. Fraction of cells in S phase in presence (△) and in absence of Nocodazole (○). The time of S phase entrance ($t_{Entry} = 20min$) is longer than the time of S phase exit ($t_{Exit} = 10min$) in accordance with Ma et al. (13).

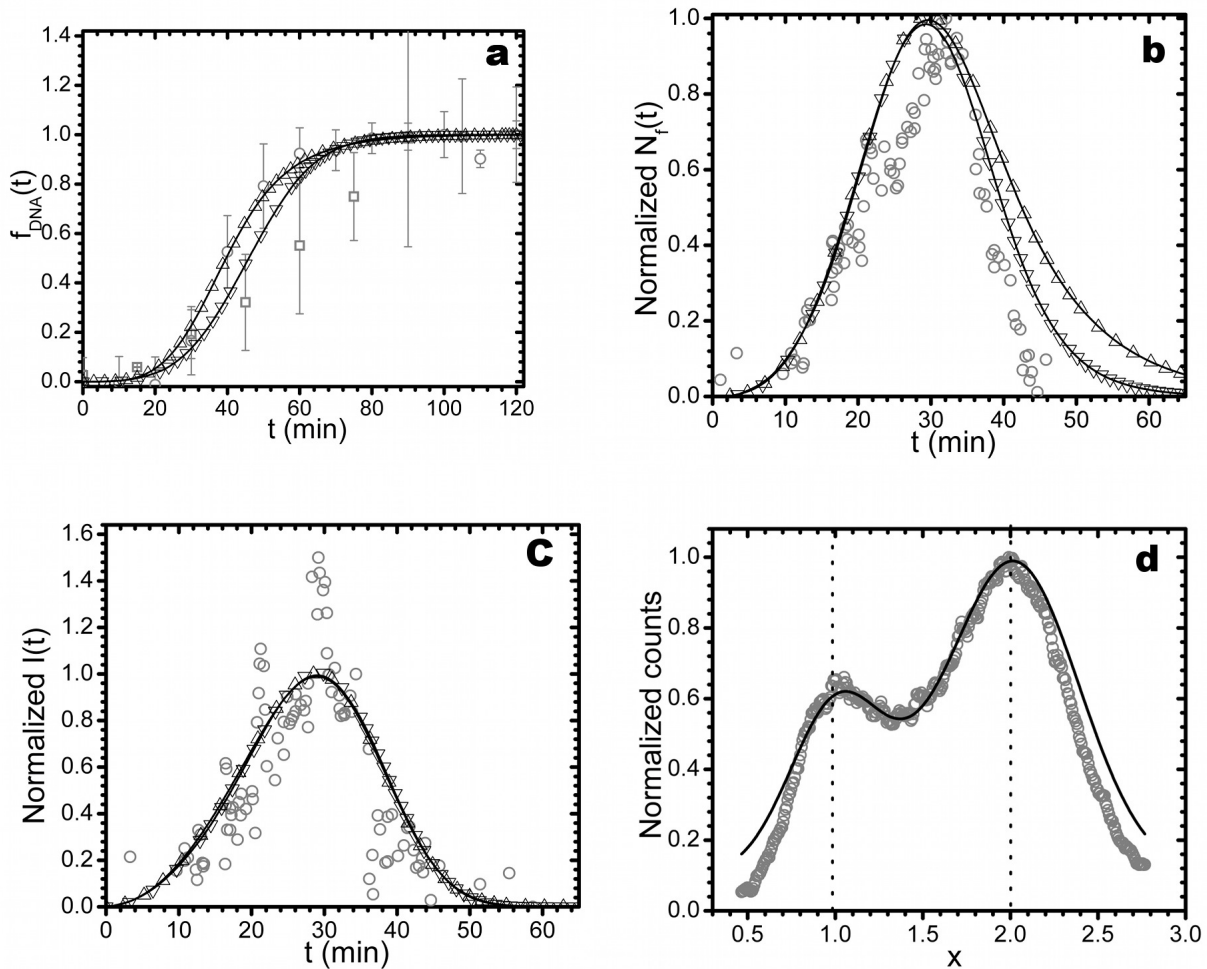


Figure 3. Comparison between inferred dynamic of DNA replication and experimental observations. **a.** Kinetics of fraction of replicated DNA, $f_{DNA}(t)$: experimental data (\square) (13), $f_{DNA}(t)$ extracted from FACS profiles of synchronized cells (\circ), fit of FACS extracted $f_{DNA}(t)$ (Δ) using the non local model, inferred $f_{DNA}(t)$ (∇) from the FACS profile of an exponentially grown cell population by using the non local model. **b.** Fork density: experimental data (\circ) (13), $N_f(t)$ inferred from the fit of $f_{DNA}(t)$ using the non local model (Δ), inferred from the fit of the FACS profile of an exponentially grown cell population by using the non local model (∇). **c.** Initiation rate: experimental data (\circ) (13), $I(t)$ inferred from the fit of $f_{DNA}(t)$ using the non local model (Δ), inferred from the fit of the FACS profile of an exponentially grown cell population by using the non local model (∇). **d.** FACS histogram of an exponentially growing cell population: experimental data (\circ), the solid black line is the fit obtained by using the methodology described in the main text.

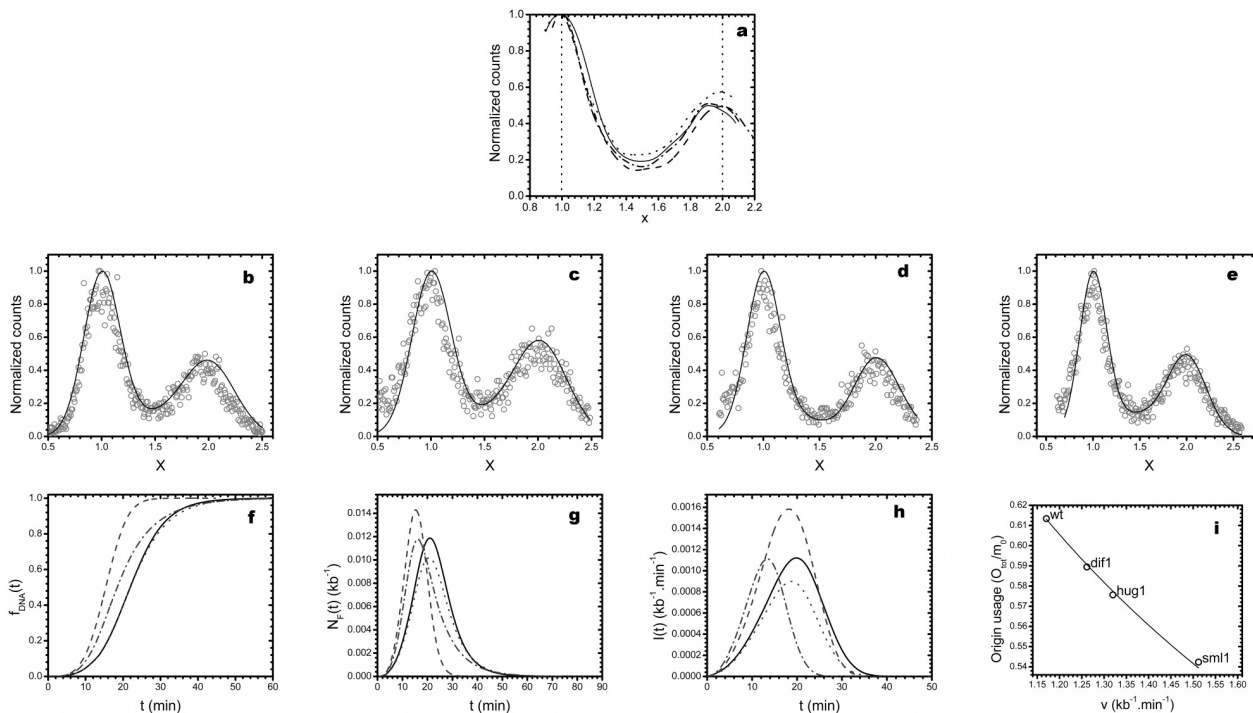


Figure 4. Temporal program of DNA replication in *wt*, *sml1Δ*, *dif1Δ* and *hug1Δ* yeast strains. **a**, Subtle changes of FACS histograms among *wt* (solid line), *sml1Δ* (dotted line), *dif1Δ* (dashed line) and *hug1Δ* (dashed-dotted line) yeast strains. To enhance small variations in FACS histograms, experimental data were smoothed using a Fourier methodology. Experimentally measured FACS histogram of exponentially growing cell population (\circ) in *wt* (**b**, $\chi^2 = 7.6 \times 10^{-3}$ $R = 0.96$ $P \leq 10^{-4}$), *sml1Δ* (**c**, $\chi^2 = 1.2 \times 10^{-2}$ $R = 0.92$ $P \leq 10^{-4}$), *dif1Δ* (**d**, $\chi^2 = 7.5 \times 10^{-3}$ $R = 0.96$ $P \leq 10^{-4}$) and *hug1Δ* (**e**, $\chi^2 = 7.3 \times 10^{-3}$ $R = 0.95$ $P \leq 10^{-4}$). The black solid line is the fit to the data obtained using the described methodology. Kinetics of DNA replication parameters: $f_{DNA}(t)$ (**f**), $N_f(t)$ (**g**) and $I(t)$ (**h**) in *wt* (black solid line), *sml1Δ* (gray dashed-dotted line), *dif1Δ* (gray dashed line) and *hug1Δ* (gray dotted line) yeast strains. **i**, Origin usage P versus the speed of replication forks v . Open circles (\circ) represent inferred values and the black solid line is the fit using $P = \frac{1}{2} \sqrt{\frac{v_0}{v}}$.

FACS analysis can be used to follow the progression of cells in the cell cycle by quantifying the distribution of levels of fluorescent agent inserted in DNA in a cell population. However, the detection of small effects requires a precise data analysis.

First, a proper normalisation of the data allows to extract the evolution of the replicated fraction with time. This evolution reveals the apparent duration of S phase. In addition, normalised data can be interpreted as three gaussians accounting for the populations of cells in G1, S and G2 which allows to remove a part of the noise from synchronisation defects. All these results allow to use FACS data as an indicator of cellular state and growth.

We show here that the growth of cells is not significantly altered by a change of the growth medium nor by the nature of the thymidine analog used (in the case of the MCM869 strain that requires the medium to be complemented with a thymidine analog). We also show that the incorporation of thymidine analogs in this strain follows DNA replication and can therefore be more precisely quantified to gather precise information about the DNA replication process.

Furthermore, we developed a methodology to extract the temporal program of DNA replication from FACS histogram of an exponentially growing cell population. By applying the procedure to yeasts dNTP mutants, we predict that the speed of local DNA synthesis is bounded by $\frac{v_0}{4} < v < v_0$, where v_0 is the speed of helicases.

Chapter 8

Using Small-Angle Scattering to study yeast nuclear organisation

As described in the introduction, SAS techniques have been already used to study chromatin conformation of nucleosomes assembly, isolated chromosomes, inside cells compartments and even inside entire cell for specific cell types. However, this study is to our knowledge the first attempt to study yeast nuclear organisation with SAS techniques. Most advances in structural biology using SAS were made on pure proteic samples that fulfil some conditions of homogeneity. Our samples are not homogenous, leading to a more difficult data interpretation. The first section set the condition to produce SAS profiles from yeast during the cell cycle, the second chapter focus on the extraction of information from these patterns that lead to the discussion in the last part about the origin of the signal.

8.1 Experimental conditions

8.1.1 Sample preparation

Usual requirements

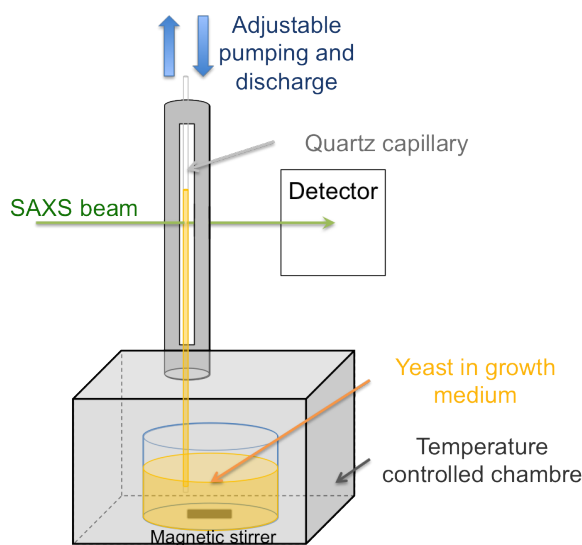
The main advantage of SAS techniques is that they require very few sample preparation. However, the difficulty of the technique resides in the data analysis and interpretation. The sample should be pure and monodisperse. In case of macromolecules in solution, multiple concentrations should be measured to differentiate unambiguously between the average signal from the macromolecule structure and the signal coming from the interaction of several macromolecules.

Sedimentation in capillaries

Concentration of cells in solution can be estimated. However, yeast sediment rapidly forming a pellet and the concentration of cells in the supernatant decreases very fast resulting in a very weak signal. To overcome this difficulties, samples are injected in a quartz capillary of 1.5 or 2 mm diameter and left overnight at 4°C to sediment with gravity. With this method a regular close packing is reached for cells and all samples are at a similar concentration. Spheroplasts or cells at several time points during replication can be sedimented in individual capillaries to measurement the effect of S phase progression. However, even with MCM869 strain where it is easier to control S phase progression by the thymidine supply, it is difficult to address how the cells evolve during the sedimentation process overnight.

Cell growth under stirring conditions

Figure 8.1.1: Schematic view of the stirrer

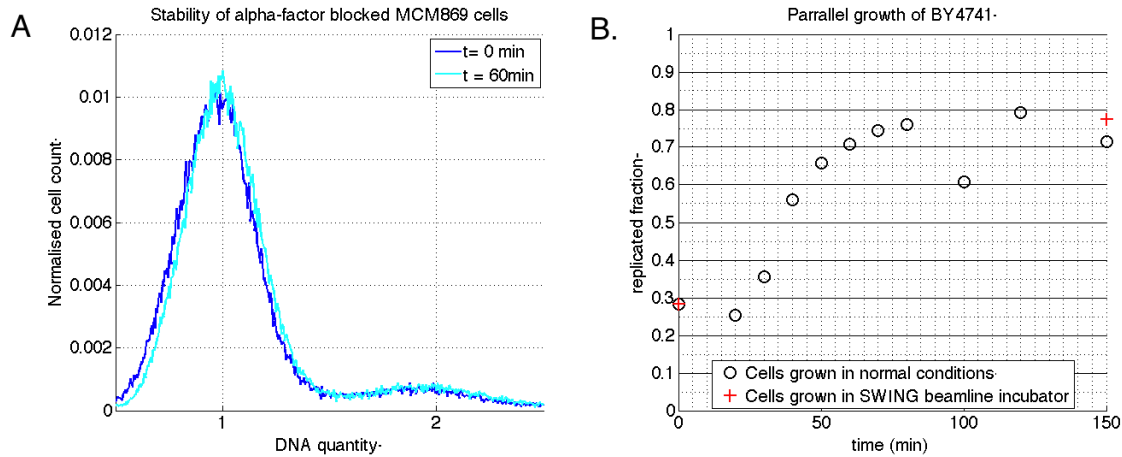


The other option used with SAXS takes advantage of a device developed by the team on SWING beam line which consists in a thermo-regulated incubation chamber with magnetic stirring. A quartz capillary is immersed in the incubator and is connected to an automated syringe that allows pumping a small volume of solution for measurement and rejecting it in the incubation chamber after the measurement. The sampled volume could also be thrown away however, re-injecting the latter allows to maintain the total

volume constant, and therefore ensure that the capillary stays immersed in the broth during the whole experiment. The agitation, temperature regulation and materials of this incubator are slightly different from normal growth condition. Therefore the behaviour of cells growing in this incubator is assessed by FACS and show that the cell cycle progression is not altered. Indeed, cells stay blocked in G1 in presence of α -factor (Fig. 8.1.2 A.) and present an increased replicated fraction at the end of the experiment consistent with the growth of the same batch of cells

grown in parallel in normal conditions (Fig. 8.1.2 B.).

Figure 8.1.2: Stability and evolution of yeast cells in SWING incubator



A. FACS profiles of MCM869 cells in complemented YPD containing α -factor (after 2 hours by α -factor treatment) after 0 min and 60 min in the SWING beam line incubator.

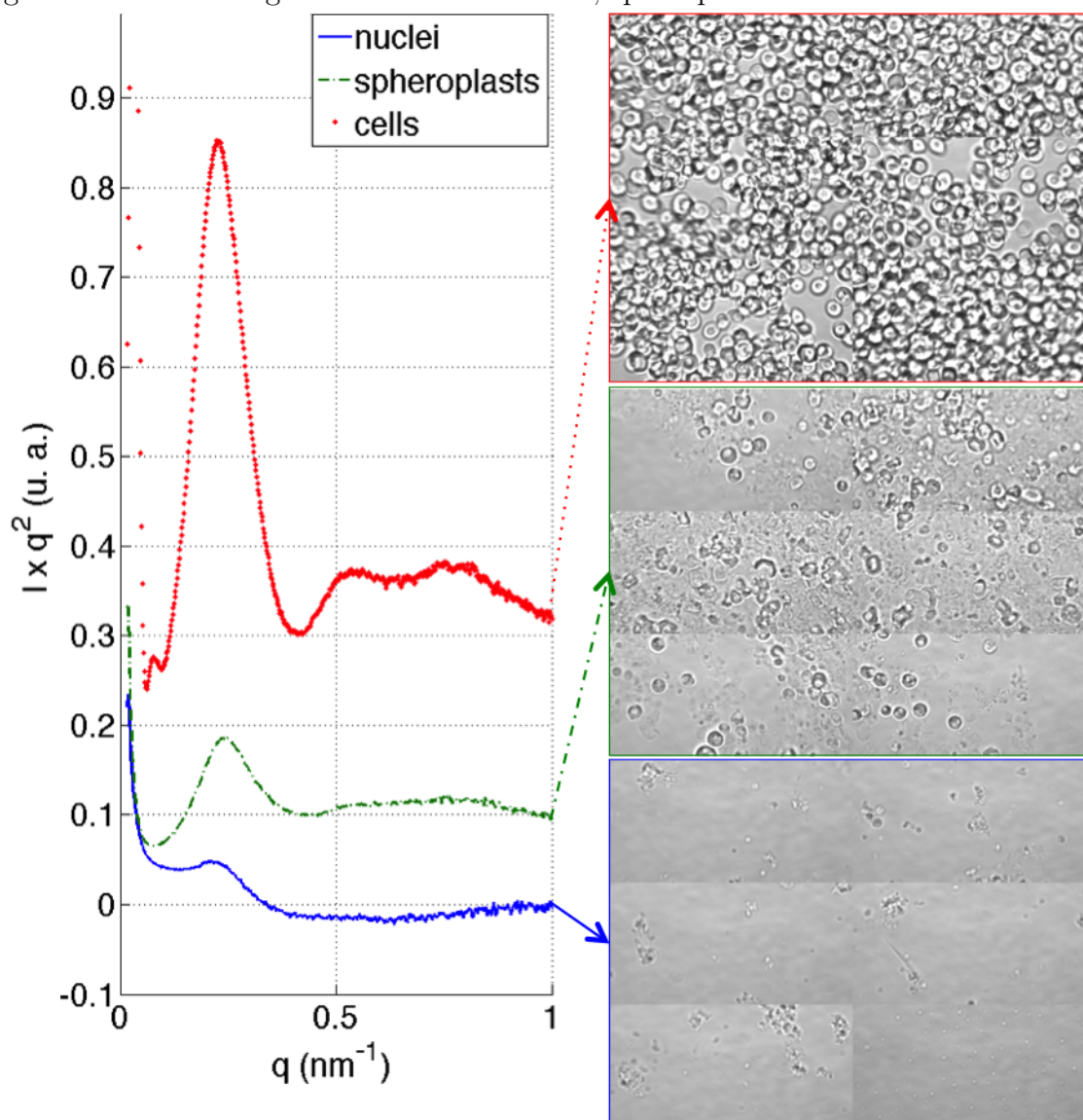
B. Growth of BY4741 cells in complemented CASA medium after washing from 2 hours α -factor treatment in normal incubator (black circles) and in SWING beamline incubation before and after experimental measures (red plus).

Usually, cells are grown in YPD, however this growth medium contains yeast extract which causes an important level of scattering. In addition, its composition is rather variable which can impede proper buffer subtraction. Therefore, it is preferable to grow cells in complemented CASA medium which causes no major change on the progression of the cell cycle as described in previous chapter.

Scattering by cell compartments

The few previous studies focused on isolated nuclei or entire cells with very small cytoplasm volume (Langmore and others 1983). Isolated nuclei, spheroplasts and entire MCM869 cells SAXS profiles are thus measured in capillaries and after the measurement, the samples are observed by phase contrast microscope (Fig. 8.1.3, normalised figure in annexe B. 1). Spheroplasts and nuclei are suspended in rather viscous buffer (1M Ficoll and 20% glycerol respectively) which are meant to control osmotic pressure to avoid blast of the spheroplasts or nuclei, but as a consequence, concentrations of the latter is diminished in the pellet after sedimentation. This effect can be clearly seen on pictures but also on the SAXS profiles as the intensity depends on the number of scatterers in the irradiated volume (see chapter 3 of introduction) and therefore on the cell concentration. The concentration of intact elements is moreover reduced by the presence of cells fragments in spheroplasts and nuclei samples that are completely absent in the entire cell sample.

Figure 8.1.3: SAXS signal from MCM869 cells, spheroplasts and isolated nuclei



Kratky plot of the SAXS subtracted data collected from MCM869 cells, spheroplasts and isolated nuclei pellets in quartz capillaries and corresponding images from the pellet after exposition.

Moreover, electron density of ficoll and glycerol is slightly higher than water decreasing the contrast of the structures of interest (Kilburn et al. 2010). Therefore, the signal over noise ratio is lower in spheroplasts and nuclei samples leading to a less visible peak.

One of the advantage of spheroplasts is that they have a round shape contrary to entire cells whose sugar cell wall can impose oval or more irregular shapes. Therefore, one could believe that working with spheroplasts would have improved

the population homogeneity. However, the spheroplasts are squeezed by the force of sedimentation and spheroplasts only appear rounds where some free space is available (Fig. 8.1.3 middle image).

The main peak visible in the cell sample ($q \sim 0.23 \text{ nm}^{-1}$, $d \sim 28 \text{ nm}$) is also present in spheroplasts and nuclei. The other peaks ($q \sim 0.53 \text{ nm}^{-1}$, $d \sim 12 \text{ nm}$ and $q \sim 0.78 \text{ nm}^{-1}$, $d \sim 8 \text{ nm}$) however are much less prominent and are only apparent in the entire cells sample where the signal over noise ratio is the best. A detailed investigation of the structure that can cause these peaks is present in the subsequent parts.

The signal scattered from entire cells is partly emanating from nuclear structure. Although removing the cell walls and cytoplasmic membrane theoretically allows to obtain samples more concentrated in nuclei, the harsh biochemical treatments and the restrictive buffer condition lead to a low concentration of nuclei in the irradiated volume. In addition, progression in the cell cycle is much more difficult to assess in spheroplasts or isolated nuclei and can be disturbed by the isolation protocol. Therefore, unless specified otherwise, entire cells are used in subsequent experiments.

8.1.2 Radiation damages

Absorption and scattering

A part of the radiation passing through the sample is scattered and another part is absorbed by the sample. The absorbed radiation can be harmful for biological samples.

In particular, X-rays are ionising radiation that can induce DNA double-strand breaks and other damages to the cell due to the creation of free radicals or oxydated species in the buffer. Moreover, ionisation by X-rays can cause the aggregation of free macromolecules in solution. This could potentially happen in our buffer.

In water 1.5mm of water 35% of the X-ray radiation is absorbed. Denser elements can absorb radiation even further. Overall density of yeast cells have been estimated slightly higher than water ($\sim 1.1 \text{ g/mL}$ Bryan et al. 2010) but particular organelles such as lipid bodies absorb more X-rays than other (Uchida et al. 2011).

Neutrons are less harmful than X-rays because they are less absorbed than X-rays (Svergun et al. 2013). In addition, the irradiated surface is 10 to 100 times larger and the dose being dependant on the mass of the sample is therefore reduced.

For this reason the rest of this part will focus on the potential damages created by X-rays exposure.

Computation of irradiation dose

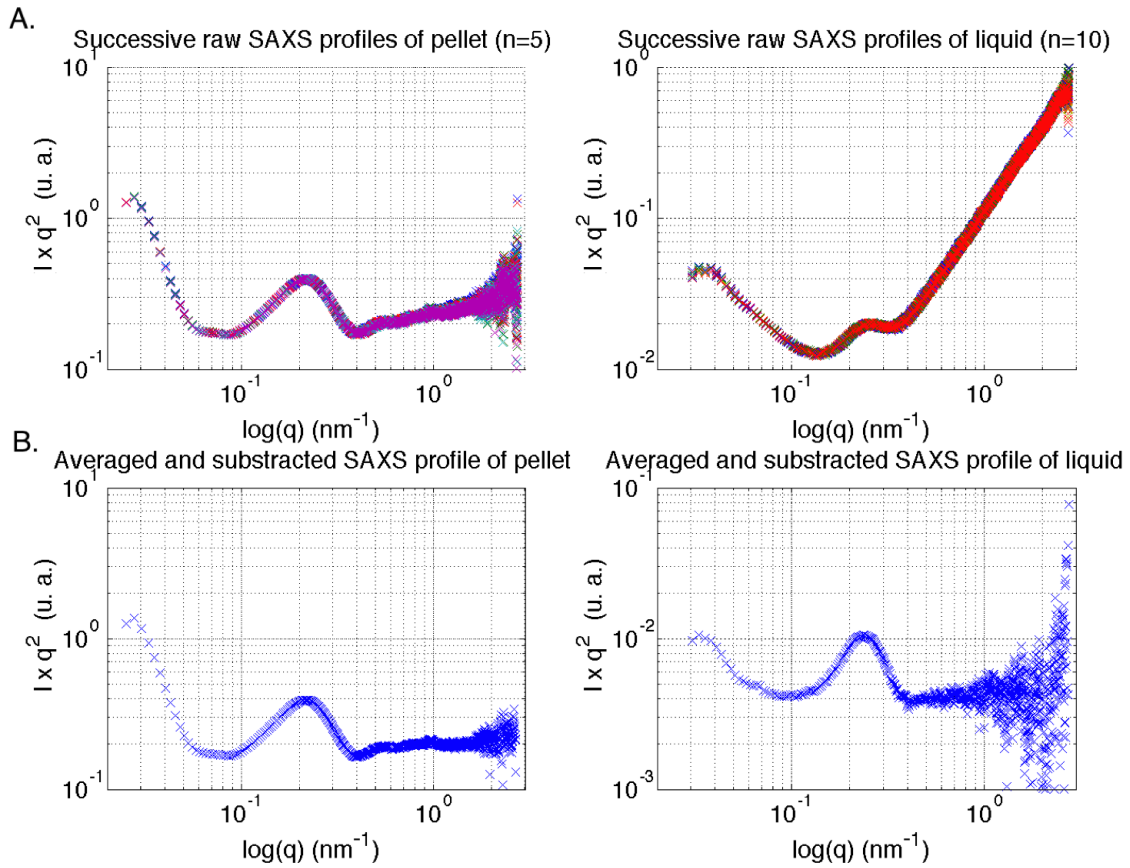
Despite an important number of studies, the radiation dose leading to cell damages or lethality is not easy to define because the energy of the photons produced by irradiators is not directly given. Instead, the energy of the electrons used to produce the photons is given without an efficiency of the transmission of this energy to photons. In addition, the doses of radiation are given in Gy or Gy/s ($\text{Gy} = \text{J/kg}$) but the sample volume or weight is not taken into account by the device. As a reference, it was shown that half the yeast cells do not grow after an irradiation of 75Gy with X-rays of energy ~ 30 to 100keV (Cao et al. 2015). On the other hand, doses of 10^8 Gy deposited by 7.8keV photons caused no damages on cryo-immobilized diploid yeast cells (Giewekemeyer et al. 2015).

Synchrotron radiation that we use have a high flux of photon giving high deposited dose (~ 11 kGy/s) with ~ 12 keV photons. In pellet samples the exposed mass is higher than water therefore the dose per kg is reduced however a small fraction of the cell receives all the radiation. On the contrary, in liquid samples, the dose is higher but cells are mobiles and have a short residency time in the beam. Potential damages on DNA as well as lethality due to the ionisation of cell components is not possible to exclude. In the case of liquid samples only 0.01 μL is exposed at each frame and this volume is re-injected in 8.5mL therefore even if a significant proportion of irradiated cells were damaged this effect would be negligible on the population. This is supported by the evolution of the replicated fraction after 150min incubation that suggest a normal growth of the population (Fig. 8.1.2 B.).

Frame invariance

Potential damage accumulation is investigated by taking a succession of images on the same position in the sample with the experimental exposure time (respectively 40 and 250 ms for pellet and liquid conditions). In both pellet and liquid conditions, curves superimpose even after 5 to 10 expositions approximately every second (Fig. 8.1.4 A. top). Therefore if damages are caused to the biological sample there are not in the observation scale and do not disturb the structure that cause the observed features of the curve. Moreover, curves from successive exposition can be averaged to obtain higher quality data and the solvent and capillary contribution subtracted giving one curve (Fig. 8.1.4 B. bottom).

Figure 8.1.4: SAXS curves from yeast cells after successive exposures



A. Yeast cells as a pellet (left) or liquid sample (right) (see first subsection of this chapter and material and methods) are submitted to a succession of irradiation from X-rays (10^{12} ph/s at ~ 12 keV, exposure time 5×40 ms and 10×250 ms respectively).

B. Curves are averaged and the background (solvent and capillary) are subtracted.

Very few technical restriction directly apply to the use of SAS for biological samples. One requirement is that the sample causes a sufficiently intense scattering compare to its buffer. Entire cells as well as spheroplasts respect this condition but not nuclei that are difficult to concentrate in a pellet because of their light weight and the high viscosity of their buffer.

On the other hand, the beam should not alter the sample in a way that the acquired pattern is influenced by the exposure time. This was verified for all types of samples therefore even if damages are induced by the absorption of the radiation they are not at the scale observed here and do not disturb the structure creating the observable pattern.

Yeast cells can provide reproducible SAXS patterns as it was done with other cell types. However the interpretation of these profiles is ambiguous.

8.2 Data analysis

8.2.1 Qualitative analysis

Profiles observation

SAS data are usually presented as $\log(I) = f(\log(q))$ called log-log plots (Jacques and Trehwella 2010) (Fig. 8.2.1 B.) because in this representation, a plateau at small q gives a first approximation of the size of the macromolecule. In our case however, a similar plateau could only be obtained for the size of the nucleus, spheroplasts or cell which is out of our observed range. A common representation for polymers is the Kratky plot $Iq^2 = f(q)$ (Fig. 8.2.1 A.). This representation divides the natural decay of scattering with distance in disordered sampled and therefore allows to compare peaks weight.

SAS profiles obtained from yeast spheroplasts are compared to data published for chicken erythrocyte nuclei (Fig. 8.2.1) where the existence of a chromatin fibre is supported by multiple techniques (Langmore and Schutt 1980, Scheffer et al. 2011, Nishino et al. 2012). The SAXS profiles obtained are comparable to the profile obtained in chicken erythrocyte nuclei (Fig. 8.2.1 A.). The SANS profiles tend to differ with a notably steeper slope measured in yeast spheroplasts compare to chicken erythrocyte nuclei. Both however have in common that there is no peak similar to the one observed in SAXS. This observation is not dependant on the mode of plotting as Kratky plots of SANS data do not show any peak either (see annexe B. 2). Older data on chicken erythrocyte nuclei reveal a shoulder at a characteristic distance of ~ 40 nm with SANS which can become clearer with an increase of ionic strength (Notbohm 1986).

In addition to the main peak, two smaller peaks are visible at high Q for concen-

trated samples. Despite being only visible with high concentration these peaks are not likely to be concentration artefacts corresponding to interaction between the individual scatterer that mostly contribute to the signal at low q . Similar peaks are also visible in chicken erythrocyte nuclei (Fig. 8.2.1 A.) as well as in HeLa cell nuclei SAXS profiles (Nishino et al. 2012) where the main peak is removed by washing the cell nuclei but not these two small peaks.

Current conclusions

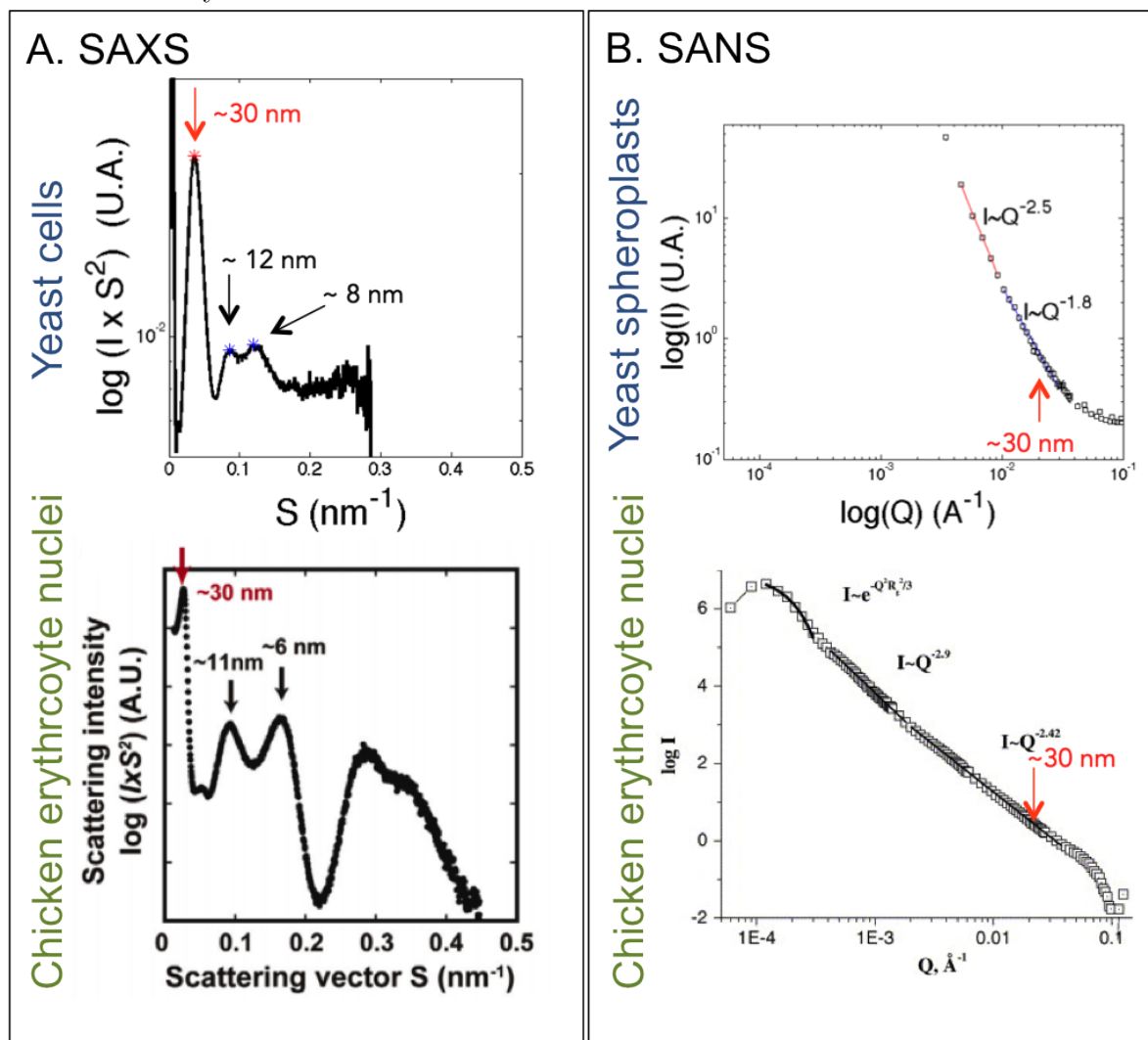
As already mentioned in the introduction, the slope visible on chicken erythrocyte nuclei SANS data has been interpreted as the scattering from a fractal structure (Lebedev et al. 2005), supporting the fractal model for eukaryotes genome organisation (Bancaud et al. 2012). However, the discrepancies between SANS and SAXS data were never addressed. Indeed, SAXS data on the chicken erythrocytes nuclei reveals a series of peaks corresponding to dense repeated structures with a characteristic interaction distance that are not visible in SANS profiles.

The main peak corresponding to a distance of ~ 30 to ~ 40 nm was attributed to chromatin fibre folded into itself while smaller peaks at 11 and 6 nm were associated with nucleosomes face to face and side to side interactions inside this compact organisation. Moreover, an increase of the ionic strength in the buffer do change the peak position corresponding to closer distance between the fibres and isolated chromatin do form similar peaks in presence of an increasing concentration of $MgCl_2$ (Jim Allan and Nick Gilbert personal communication). A similar peak around ~ 30 nm in HeLa cells was however attributed to ribosomes stacking on the surface of isolated nuclei (Nishino et al. 2012, Kiseleva et al. 2007) and disappeared when properly washed. In addition, the relative height of the 11 and 6nm peaks is different whereas they are similar in chicken erythrocyte nuclei which also plead in favour of the absence of a regularly folded structure in this cell type. Indeed, if these peaks are caused by the interaction of nucleosomes inside compact organisation, the face-to-face and side-to-side interactions should occur with the same frequency.

By extension, one hypothesis is that cell types that have a low genome activity such as chicken erythrocytes form a robust chromatin fibre organisation while active genomes do not. This is however not consistent with data on sperm nuclei that do not present signal from folded structure despite an absent transcriptional activity (Langmore and others 1983).

SAS data currently available therefore lead to propose several interpretation of the chromatin conformation but are not directly transposable from one cell type to another.

Figure 8.2.1: SAXS and SANS compared profiles from yeast and chicken erythrocytes



A. Plot of SAXS data for pellet yeast cells (top) and chicken erythrocyte nuclei (bottom) reused from Nishino et al. 2012 with permission. Data presentation has been adapted to match the published data (note that the scattering vector $S = \frac{Q}{2\pi}$ where q is the commonly used scattering vector), arrow indicate the peaks and their corresponding characteristic distance $d = \frac{1}{S}$.

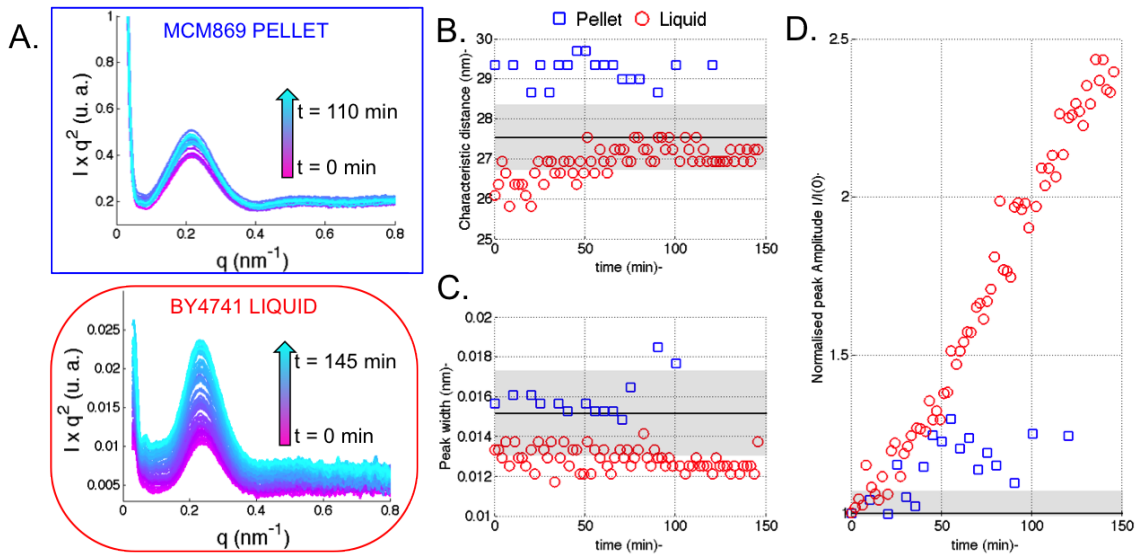
B. Plot of SANS data for yeast spheroplasts sedimented in 1% agarose (top) and chicken erythrocyte nuclei (bottom) reused from Lebedev et al. 2005 with permission. Data presentation has been adapted to match the published data (note that in spite of the legend, the figure is a loglog plot). Arrows point at the characteristic distance of ~ 30 nm where no maximum is visible.

8.2.2 Evolution with time

Peak parameters

In the hypothesis where the peak observed at $\sim 30\text{nm}$ is due to chromatin, one could expect that it is affected by DNA replication. Therefore it would be interesting to measurement its evolution during S phase. The peak amplitude, position and width is evaluated by automatic peak detection at each experimental time point in capillaries or liquid conditions (Fig. 8.2.2).

Figure 8.2.2: Evolution of the SAXS peak with time after release from α -factor in pellet or liquid



A. Kratky plots of SAXS data from MCM869 cells pelleted in quartz capillaries (top) or BY4741 cells growing in CASA medium in SWING incubator collected over time after release from α -factor. B. Position of the peak reported as a characteristic distance ($d_0 = \frac{2\pi}{q_0}$), C. Peak width and D. Normalised amplitude at the peak position for cells in pellet (blue square) or liquid (red circles). As a reference, average value obtained for stable α -factor blocked MCM869 cells in liquid condition is plotted as a black line and the grey shade represent the standard deviation to the mean.

The average values obtained for stable α -factor blocked MCM869 cells in liquid condition serve as reference for the measurement of variation (black line and grey shade). Thus measured position and width of the peak do not present any significant change during the time of the experiment. Even accounting for the experimental dispersion, the characteristic distance extracted from the peak position is different for the two conditions. However, this difference between liquid and pellet conditions and between the strains is not the reason for this difference as proven by additional measurements made in both conditions (see annexe B.3). This difference is due to the variability from one cell population to another and possibly to

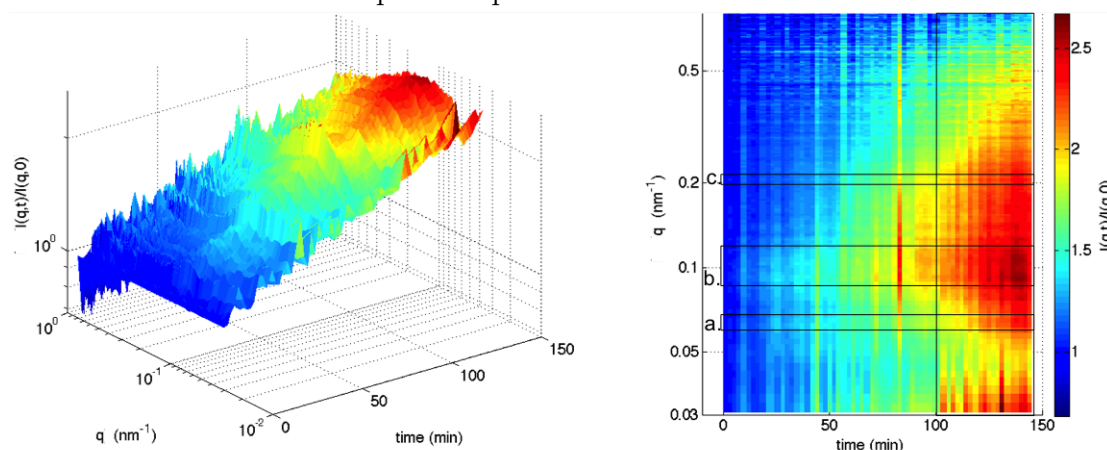
measurements artifacts, such as the level of noise, that can change slightly the estimation of the peak position.

The amplitude increases in both cases until 60 min, although it presents a higher dispersion for pellet sample that is probably due to a higher heterogeneity of concentration. Although the slope is equivalent at the beginning, a plateau seems to be reached after 60min for the pellet sample whereas the amplitude is continuously increasing for the liquid condition.

Amplitude domains

The stacking of the curves as a function of time allow to see that the whole curve is undergoing a rather monotonous increase in amplitude. To visualise better the relative increase of each part of the curve, each $I(q)$ values are normalised by $I(q)$ at $t=0$) and plotted as a surface coloured plot (Fig. 8.2.3).

Figure 8.2.3: Global evolution of SAXS amplitude with time after release from α -factor in liquid samples



The amplitude for each q and time, obtained for liquid condition samples, is normalised by the amplitude at the first time ($\frac{I(q,t)}{I(q,t_0)}$) and plotted as a function of time and q and represented as a colour. For visibility, q and ($\frac{I(q,t)}{I(q,t_0)}$) are represented in log scales and the plot is repeated in top view in the right panel.

Box a delimits the region around $q \sim 0.065 \text{ nm}^{-1}$, box b $q \sim 0.1 \text{ nm}^{-1}$ around and box c around $q \sim 0.22 \text{ nm}^{-1}$, which is the main peak area.

Despite few outliers time points, the evolution of the amplitude is rather smooth. A clear change of behaviour appears at $t \sim 100$ min which corresponds to an inflexion points in the replication fraction of the same batch of cells grown in parallel and analysed by FACS (Fig. 8.1.2 B.). Notably, the amplitude at very small q is increased which could corresponds to intracellular interaction due to the doubling of the cell concentration is solution. The amplitude around $q \sim 0.1\text{-}0.2 \text{ nm}^{-1}$ is also

increasing strongly after $t = 100$ min but the effect is less impressive as the increase start strongly around $t \sim 90$ min. In particular, a zone around $q \sim 0.1 \text{ nm}^{-1}$ (zone b) is increasing strongly up to ~ 2.5 times its initial value and seems to impose a local increase of the amplitude. Some vertical lines also highlight particular regions for example the region around $q \sim 0.065 \text{ nm}^{-1}$ (zone a) corresponds to the small peak visible in the Kratky plot at small q (Fig. 8.2.2) while the region around $q \sim 0.22 \text{ nm}^{-1}$ (zone c) corresponds at the summit of the main peak.

Overall, this representation shows that the amplitude is increasing all along the curve which makes interpretation of the increase of the amplitude in one position (such as shown in Fig. 8.2.2 D difficult. However, the evolution of the amplitude increase with time is not exactly the same in all points, and the characterisation of these small variations call for data fitting.

8.2.3 Data fitting

SAXS and SANS fitting

Data obtained in sedimented spheroplasts in SAXS and SANS need to be fitted with a similar equation but with different multiplicative factors for each contribution. These data present a slope at small q that can be the combination of one or several contributions. In addition, an equation will be proposed for the peak with a multiplication value of zero for SANS data. Instead, a constant will be added for SANS data to account for incoherent scattering which is stronger and appears at smaller q in SANS data and will therefore be ignored in SAXS.

Equation of models

Two approaches allow to give model equations to fit the data. First, a non directive model is proposed based on the SANS profile data. Indeed, the apparent change of slope can suggest an organisation of two phases of different densities.

$$I_{ph}(q) = A.q^{-4} + B.q^{-2}$$

Another approach consist in hypothesising that the signal is due to chromatin as it was done in literature (Lebedev et al. 2005) and equates proposed models for chromatin organisation.

First, the scattering by a fractal structure (Mirny 2011) has a known equation (Kjems and Freltoft 1985, Sinha 1989) :

$$I_{sfr}(q) = 1 + A \frac{\sin((d_f-1)\arctan(q.\xi))}{(d_f-1)q.\xi(1+\xi^2.q^2)^{\frac{d_f-1}{2}}}$$

and can be simplified for $q\xi \gg 1$ in :

$$I_{fr}(q) = A.q^{-d_f}$$

This model actually described a polymer organisation whose density can vary from an ideal gaussian chain ($d_f = 2$) to a clustered network ($d_f = 3$) into domains of characteristic size ξ .

In addition, the notion of globule includes the notion of a surface that delimits this organisation which corresponds to a contribution in q^{-4} .

$$I_{gfr}(q) = A.q^{-4} + B.q^{-d_f}$$

Moreover, it is believed that in yeast the chromatin has enough space to reach an equilibrium and therefore occupies all the available space which suppose that d_f reach its maximum value of 3.

$$I_{geq}(q) = A.q^{-4} + B.q^{-3}$$

Another model for chromatin organisation is the colloidal gel (Marko and Poirier 2003). Scattering by colloidal gel has also been described (Yeh et al. 1998) :

$$I_{coll}(q) = A.exp(-q^s \xi_1^s) + \frac{B}{1+q^2 \xi_2^2}$$

Where ξ_1 and s are the parameters of the network structure while ξ_2 is the size of the colloid.

Peak equation

Following the hypothesis that this peak which corresponds to a characteristic distance of $\sim 30\text{nm}$ could be due to chromatin, a lorentzian function was chosen because it can be used to describe liquid crystals which is a state proposed for nucleosomes and chromosomes organisation (Livolant et al. 2006, Chow et al. 2010).

$$I_{peak}(q) = A \times \frac{1}{(1+\xi^2.|q-q_0|^2)}$$

Where ξ is the correlation length and $d_0 = \frac{2\pi}{q_0}$ is the characteristic spacing between the inhomogeneities.

Few other equations have be tested such as gaussian peak which does not allow to fit an asymmetrical peak or a cardinal sinus which aimed at taking into accounts the small variation at high q but the two peaks at high q are too strong and too similar in intensity to be simple harmonics of the first main peak.

Goodness of the fit

The best fit for both SAXS and SANS data is obtained with the simplified fractal model and the globule fractal model (see annexe B. 4). However, the amplitude in front of the term in q^{-4} is almost equal to 0. These models are good to fit SANS data, however, the fit is not in good agreement with the SAXS data, particularly around the peak position. Overall, even with an attempt to reduce the complexity, tested models present many parameters to fit and often lead to unsuccessful fit, particularly for SAXS data where the error bars are very small. Failing to fit the entirety of the signal means that cell samples present an additional complexity induced by other biological structures of the sample or by multiple scattering.

Despite failing to fit the entire curve or even large section with the models proposed, the peak equation fits locally well, in the area where the signal from the peak is preponderant and the other contributions can be neglected. This section covers 45 to 70 points depending on the noise at large q and yields good values for χ^2 (Fig. 8.3.4 A. and annexe B.5) and can therefore be exploited to extract the parameters of the peak observed in SAXS.

Qualitative analysis of SAXS data reveals an evolution of the signal with time. In particular, the amplitude is increasing at the peak position as well as in several q domains. A number of equations describing potential organisation of the chromatin as well as non directive model have been tested but none lead to a reliable fit of the data. The main reason is probably that many elements of the spheroplasts or entire cells are contributing to the SAS signal in addition to chromatin. Therefore, adding other contributions to the equation could eventually lead to a correct fit but requires too many assumptions and many free parameters. For further insight on these data sets, *in silico* models of the cell organisation need to be built in order to compare the scattering signal of the numerically simulated structure to the data.

Data fitting efforts have failed to select a simple model to represent the whole curve. SAXS signal of yeast spheroplasts, and furthermore entire cells, presents a complexity that is the addition or interaction of scattering from multiple components and is not possible to describe with a model with few parameters. Therefore, the only way to accurately link the evolution of the total curve to physical parameters of the structures observed would be to make *in silico* models of the structure and compare their simulated SAS profiles with the experimental data.

However, the peak area is fitted with a Lorentzian equation which represents an organisation as a liquid crystal consistent with the described behaviour of nucleosomes arrangement in physiological conditions.

8.3 Origin of the signal

8.3.1 SAXS and SANS comparison

Difference of contrast in SANS and SAXS

Scattering by SAXS and SANS by the same structure should provide the same profile. However, spheroplasts samples similarly prepared consistently result in different scattering profiles with the absence of the peak in SANS.

As hydrogen causes important level of incoherent scattering (scattering independent of q only linked to the material density) in SANS, a plausible hypothesis was that this information could be masked by a poor signal to noise ratio. However, as seen in Fig. 8.2.1, the approximate position of the peak average ($\sim 0.02 \text{ \AA}^{-1}$) is in the well defined range ($\sim 0.003\text{-}0.4 \text{ \AA}^{-1}$ for yeast spheroplasts), thus this argument is invalid.

Therefore the main hypothesis is that difference of contrast from the different cell parts cause this absence.

Application of SANS contrast ratio to SAXS data

Table 8.1: SANS and SAXS density and contrast

	SAXS ($e^{-} \cdot \text{\AA}^{-3}$)		SANS (100% D2O) ($10^{-14} \text{ cm} \cdot \text{\AA}^{-3}$)	
	ρ	$\Delta\rho^2$	ρ	$\Delta\rho^2$
Solvent	0.33	0	6.4	0
Lipids	0.38	0.0025	0.15	39.06
Proteins	0.43	0.01	3.1	10.89
DNA	0.48	0.0225	3.9	6.25

The main components of the cell are lipids, proteins and DNA/RNA. Their approximate scattering power value in both SANS and SAXS are tabulated (8.1). The contrast $\Delta\rho^2 = (\rho_{solvent} - \rho)^2$ values are very different in SANS and SAXS. Lipids contrast the most in SANS (3.6 times more than proteins and 6.25 times more than DNA) while DNA contrast the most in SAXS (2.25 times more than proteins and 9 times more than lipids).

From this observation, a reasonable hypothesis is that signal from lipids is masking the contribution of DNA and proteins in the SANS profile. Contrast absolute values cannot be directly compared between SANS and SAXS as they are corres-

ponding to completely different type of interaction. However, the ratio between two contribution is unitless and can be moderated.

SAXS data best fitting equation is composed of a part describing the loglog slope and a part describing the peak.

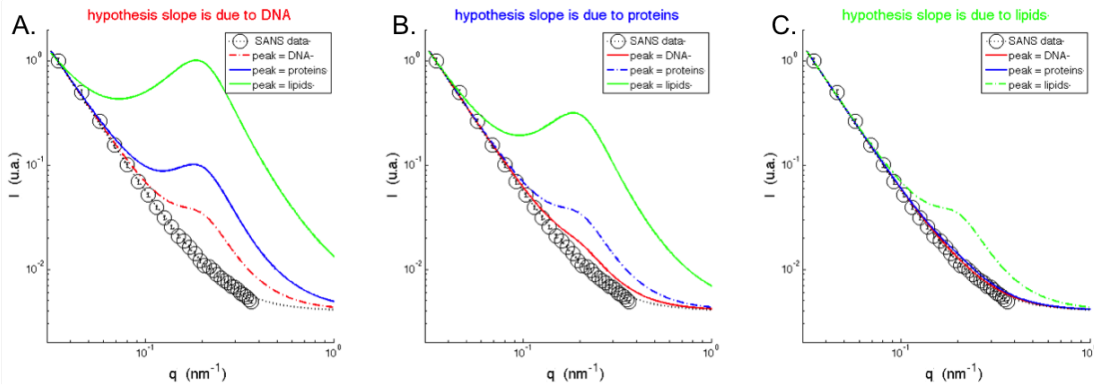
$$I_{SAXS}(q) = A.q^{-df} + B.\frac{1}{(1+\xi^2.(q-q_0)^2)}$$

This equation can be used to apply contrast ratio $\frac{SANS}{SAXS}$ to the SAXS values A and B and try to recapitulates the SANS data. Indeed, one can assume that the peak is due to lipids, proteins or DNA and with these hypothesis, balance the weight that the slope should have in SANS profile if it was from lipids, proteins or DNA. All hypothesis are made and plotted with SANS data point (black dots) as a reference and the fit (thin black dotted line) corresponding to the following equation.

$$I_{SANS}(q) = C.q^{-df} + D$$

The hypothesis for the slope contribution correspond to different plots A,B,C while the hypothesis for the peak contribution correspond to the several curves in each plot.

Figure 8.3.1: Application of SAXS/SANS contrast difference to SANS fit curve



Loglog plots of normalised SANS data points and fit (black circle and thin black dotted line) of yeast spheroplasts are compared with the equation from SAXS data with corrected slope contribution. The hypothesis where peak is from DNA (red), proteins (blue) or lipids (green) are combined with the hypothesis where slope is from DNA (A.), proteins (B.) or lipids (C.).

Even if the chosen equation is not fitting the data perfectly, few hypothesis were obviously wrong. For example, in the hypothesis where the slope comes from a structure made by DNA in SAXS (Fig. 8.3.1 A.), the peak is be at least as much prominent in SANS whatever its origin. Same observation is true for the hypothesis where the slope is due to protein and the peak to lipids (Fig. 8.3.1 B.

green) as well as for the cases where the slope and peak are from the same element (Fig. 8.3.1 dotted lines). The case where slope is due to proteins and the peak to DNA (Fig. 8.3.1 B. red) leads to a peak in SANS, smaller than in SAXS but still much more prominent than the actual SANS data.

Therefore, the slope cannot be due to DNA and is unlikely to be due proteins. However, if the slope is due to lipid it totally masks the peak whether it is due to protein (Fig. 8.3.1 C. blue) or DNA (Fig. 8.3.1 C. red) and reduce it to an extremely small bump included in the SANS data error bars. Interestingly, the slope at small q is the same for SANS and SAXS which suggest that in both sample, the low q behaviour is dictated by lipids in yeast spheroplasts samples.

Case of chicken erythrocyte nuclei

Interestingly, the divergence between SANS and SAXS observed here in yeast spheroplasts was also observed in chicken erythrocytes isolated nuclei (Fig. 8.2.1). Although chicken erythrocyte nuclei possess one layer less of membrane compared to yeast spheroplasts, the contribution from lipids might be the cause of the SANS scattering as well. In any case, it is not possible to interpret both SAXS and SANS data as scattering signal from the chromatin while they manifestly differ. The peak in chicken erythrocyte SAXS data have been convincingly attributed to a spacing between fibres. First, the existence of a fibres inside the nuclei has been confirmed by electron microscopy (Woodcock 1994, Scheffer et al. 2011). Then this value diminishes when the salt concentration of the buffer is increased which is consistent with a diminution of the characteristic spacing due to compaction by salts. Finally, after treatment with DNase, an enzyme that cut DNA is small fragments, the SAXS profile do not present this peak anymore. Moreover, the artifactual layer of ribosomes observable in HeLa isolated nuclei that has a similar characteristic spacing (Nishino et al. 2012) is not present after chicken erythrocyte nuclei isolation (K. Maeshima personal communication).

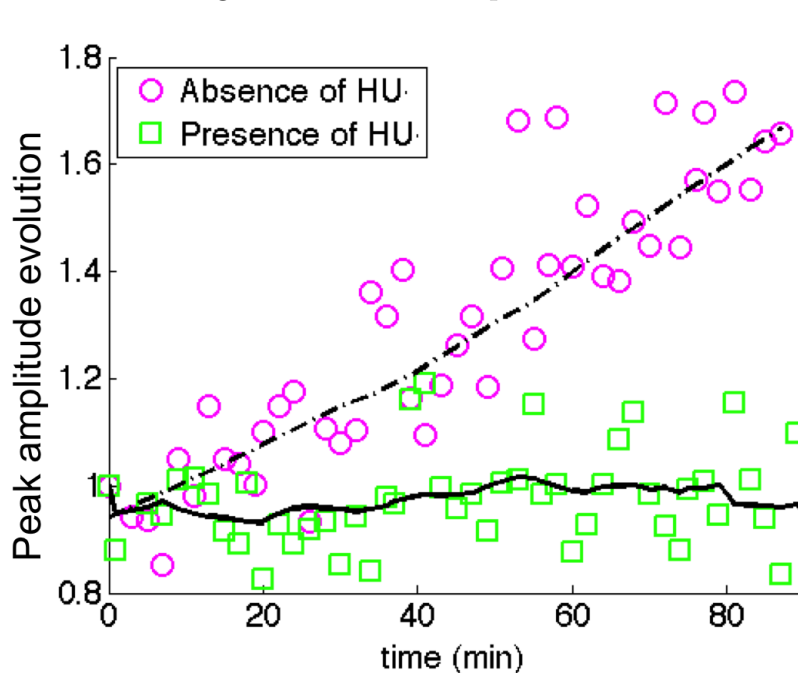
In SANS, increasing the percentage of hydrogen in water allows to change the scattering power of the solvent to resemble the scattering power of one of the component to mask it. In the DNA match condition (no scattering from DNA), the scattering profile remains substantially unchanged (Lebedev et al. 2005) which also suggests that the total signal is from lipids which scatter 7.6 times more than proteins in these conditions. Therefore we propose that the SANS data obtained in chicken erythrocyte nuclei cannot be interpreted by invoking the fractal organisation of chromatin. Nevertheless, this does not imply that the chromatin is not organised as a fractal, only that the SANS method is not adapted to detect it.

8.3.2 Link to replication

Effect of HU

MCM869 cells are grown in the SWING incubator in YPD medium in presence or absence of hydroxyurea. Hydroxyurea is a ribonucleotide reductase inhibitor that therefore blocks the synthesis of dNTP and slow down drastically the DNA replication program (Koc et al. 2004, Alvino et al. 2007). The peak amplitudes for both conditions are plotted as a function of time and smoothed (black lines).

Figure 8.3.2: Peak amplitude in HU



MCM869 cells grown in YPD in SWING incubator with or without Hydroxyurea (HU). Normalised peak amplitude of SAXS data are plotted as a function of time (pink circle : no HU, green square HU) and smoothed (dotted line : no HU, solid line : HU).

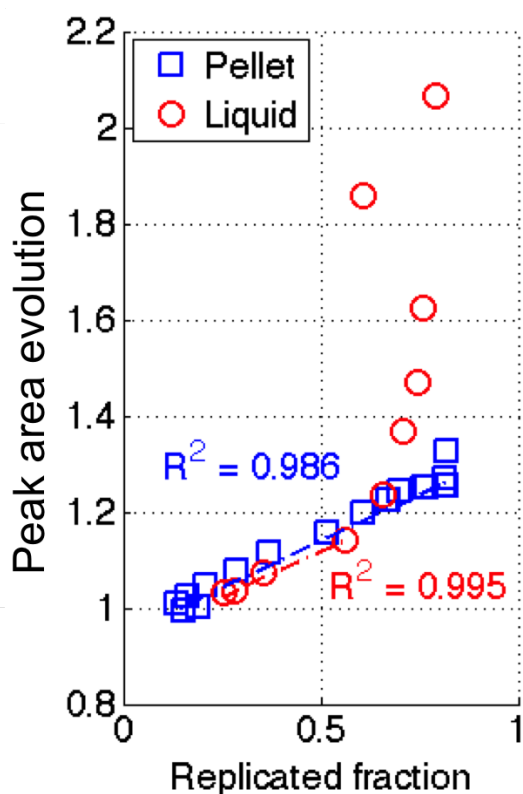
The increasing of the amplitude of MCM869 cells is consistent with the one observed for BY4741 cells grown in complemented CASA (Fig. 8.2.2 D.) with a slope of ~ 0.01 unit/min between 10 and 90 min. In presence of HU however, the amplitude increase slowly (~ 0.0005 unit/min between 20 and 300min) which corresponds to a rate approximately 20 time slower which is consistent with the relative evolution of replicated fraction already published (most cells in G2 at $t=30$ min vs $t=510$ min in HU) (Alvino et al. 2007). The peak amplitude is proportional to the number of scatterers. Therefore the number of scatterers is increasing with time at a rate which is proportional to the rate of DNA synthesis. This sug-

gests that the structure giving the SAXS curve, and notably the observed peak, is formed during DNA replication.

Link to replicated fraction

As seen previously, the amplitude of the signal is not increasing only at the peak position. To quantify more precisely the peak evolution the area under the peak is computed for each time. However, the data from cells in presence of HU as well as the control data have been collected in YPD medium which induce a lot of noise after subtraction and only allow to measurement the peak position but no other parameters.

Figure 8.3.3: Correlation of peak area with replicated fraction



The area under the peak is quantified, smoothed and plotted as a function of replicated fraction for MCM869 cells in pellet and BY4741 cells in liquid condition. Linear regression is showed as a dotted line with their respective regression coefficient (pellet/blue : $R^2 = 0.9886$ over all points, liquid/red : $R^2 = 0.995$ for replicated fraction < 0.6).

The area under the SAXS peak of MCM869 cells in pellet and BY4741 cells in liquid is smoothed and plotted as a function of the replicated fraction (see annexe B. 6). For pelleted cells, sample for FACS and SAXS are collected together while replicated fraction for liquid cell is measured from the same batch of cells growing in parallel of the SAXS measurements (Fig. 8.1.2 B.).

The peak area in pellet cells is correlated to replicated fraction for all the experiment while this relation is not linear for the liquid sample except for the early beginning of the curve. The area is proportional to the number of scatterers which is therefore increasing with DNA synthesis in pellet cells. This is in agreement with the observation in the previous paragraph. The fact that this relation is not true for high replicated fraction of cells in liquid condition is due to a coupled increase of the number

of scatterers per cell as well as the number of cells. Therefore, the structure behind this peak can be made by DNA or any component of the cell that is linearly produced during DNA synthesis. As we discussed in the introduction, the nucleosomes are very rapidly deposited on DNA after replication therefore chromatin is linearly produced during DNA synthesis.

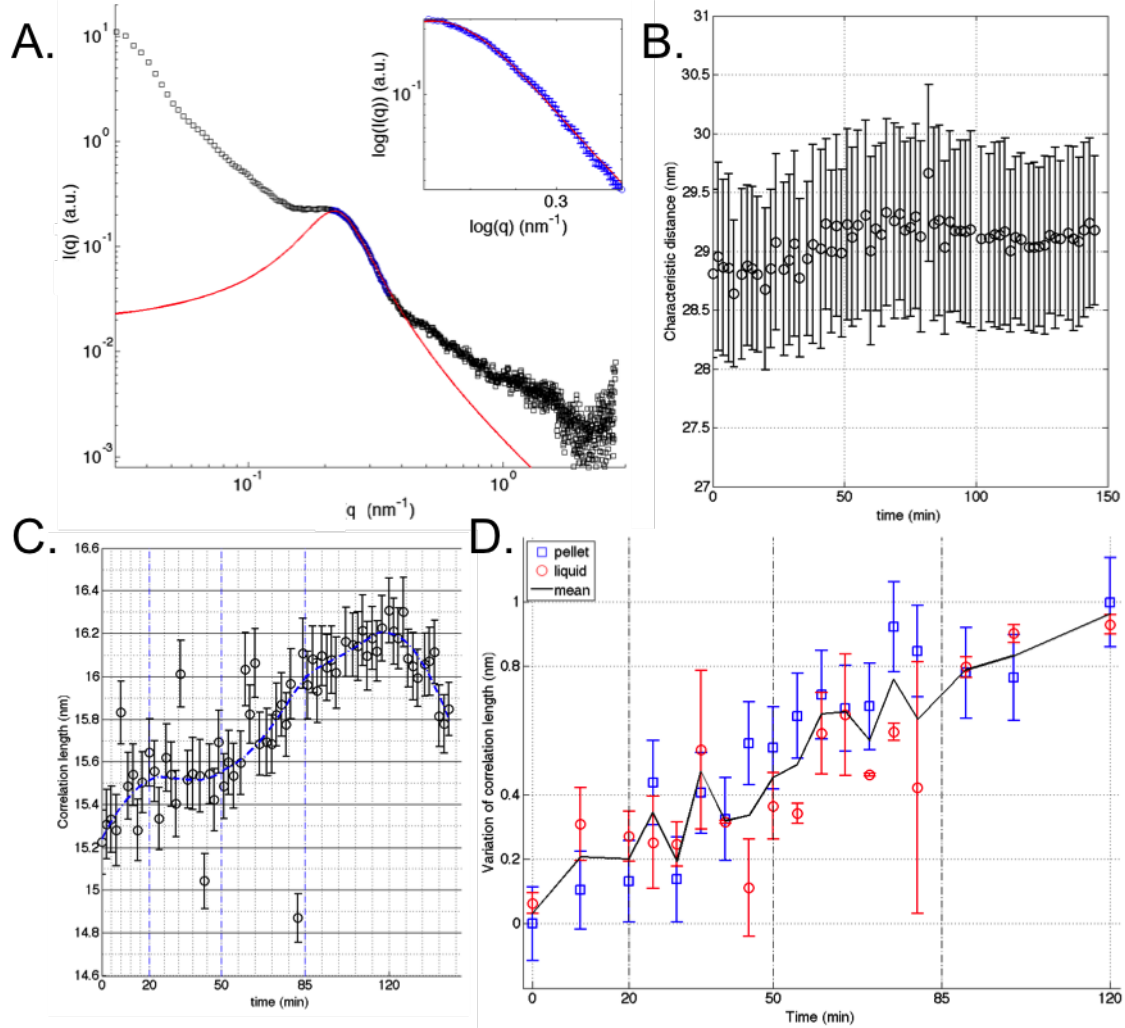
Concerning the hypothesis that the peak is created by ribosomes aggregates, the conclusion is more difficult. Indeed, ribosomes biogenesis is well documented (Woolford and Baserga 2013) but its kinetics during the cell cycle is less known. The production of ribosomes is increasing with the size of the cell and therefore the quantity of ribosomes globally increases with the growth of the population (Ju and Warner 1994). A peak of ribosome synthesis is expected in G1 where the cell need to grow to reach a critical size before entering S phase (Rudra and Warner 2004). Ribosomes need to be transmitted in daughter cell to insure early homeostasis and are divided equally between the two daughter cells in HeLa (Tsai et al. 2012) and presumably also in yeast although budding yeast have a size asymmetry. Therefore, even if what we know about the synthesis of ribosomes does not indicate a linear correlation with DNA synthesis, as the relation was never measured it is not yet possible to discriminate it. In order to investigate the characteristic of the structure leading to a peak in SAXS, more precise informations need to be gathered by data fitting.

Correlation length evolution during cell cycle

The SAXS data for BY4741 cells grown in complemented CASA are fitted with the lorentzian function on the peak range. First observation is that, due to the general slope of the curve, the value of q_0 derived from lorentzian fit does not exactly coincide with the experimental local maximum position. In addition, $d_0 = \frac{2\pi}{q_0}$ is a characteristic distance that does not require a chromatin fibre organisation. The value of this distance is close to 29 nm and do not significantly changed during the time of the experiment (Fig. 8.3.4 B.). The correlation length is the distance beyond which the positions of objects are independent. This value is globally increasing with time which suggests that the structure is becoming more and more organised (Fig. 8.3.4 C.). Once again, this evolution is not linearly correlated to replicated fraction but the smoothed data (dashed blue line in Fig. 8.3.4 C.) allows to observe an interesting variations of behaviour with time.

During the first 20 minutes, DNA replication is almost absent as the cells are re-organising after the treatment with α -factor which also corresponds to an increase of the correlation length. From 20 to 50 min most cells are in S phase and therefore undergoing DNA replication. Interestingly, the correlation length does not evolves with time during this period suggesting that the organisation is unchanged. The organisation increases strongly between 50 to 85 min which is consistent with a

Figure 8.3.4: Fit of the peak zone of the liquid sample SAXS curve



A. SAXS data from BY4741 cells grown in complemented CASA in SWING incubator are plotted as log-log. Points in blue are the fitted range and the red curve the fitting curve. The small inset present a zoom of the fitted data with error bars. B. Characteristic distance $d_0 = \frac{2\pi}{q_0}$ is plotted as a function of time after release from α -factor treatment C. Correlation distance ξ is plotted as a function of time after α -factor and smoothed (blue dashed line). D. $\xi - \xi_0$ is plot for the MCM869 cells in pellet (blue square) and BY4741 cells in liquid (red circles). For BY4741, the points are average in bins centred on the MCM869 points to obtain the same number of points. As the two data sets are equivalent ($\chi^2=0.2442$) the mean is also plot (black line).

higher compaction preceding division. Because the cells are mainly desynchronised after one cell cycle (see first chapter of the results) the following part of the curve is difficult to interpret as it is composed of a combination of different states. Moreover, the initial state corresponds to cells just released from α -factor treatment which might be different from a normal G1 state. However, a decrease is expected in order to return to the organisation before DNA replication. Indeed, local chromatin compaction is not inherited (Dickerson et al. 2016). The decrease at the end of the experiment is encouraging in this direction. An interesting experiment would be to add α -factor at some point of the experiment to re-block the cells in their initial state to observe more clearly the variation of the correlation length.

The evolution of the correlation length $\xi - \xi_0$ is plotted (Fig. 8.3.4 C.) for MCM869 cells in pellet as well as BY4741 cells in liquid (time points are averaged in bins centred on the MCM869 time points for comparison, see material and methods). The two data sets are significantly similar ($\chi^2 = 0.2442$) and the average is increasing during G1, S and G2 phase. Despite being at a higher time resolution, liquid data are also more noisy which cast a doubt on the plateau observed during S phase. The global evolution however is consistent with a maximal variation of the correlation length of 1 nm. This small variation does not support a view of a major rearrangement during DNA replication. However, as the correlation length is averaged over the genome, it does not exclude local drastic changes.

The short value of the correlation length (respectively ~ 15 nm and ~ 16 nm for MCM869 and BY4741) in regard of the nucleosome size (6 nm of height and 11 nm of width) argues against the presence of a well defined DNA fibre but is consistent with the description of nucleosome organisation as small packs (Ricci et al. 2015) or loose chromatin fibre (Dekker 2008) with a small persistence length (Hajjoul et al. 2013, Lassadi et al. 2015). This small packs of nucleosomes could be organised locally as column which would induce the direction of the nematic liquid crystal. Columnar state was obtained from free nucleosomes *in vitro* in physiological conditions (Mangenot et al. 2003) and was recently supported by an analysis of the nucleosome repeat length (Trifonov 2016). Overall, this view is consistent with recent cryo-electron microscopy data in yeast cells (Chen et al. 2016) that show no 30nm chromatin fibre but an organisation as 10 nm structure with sparse oligonucleosomes assembly.

A short correlation length value also discredits the hypothesis of a ribosome origin of the peak. The stacking of ribosomes on the nucleus surface results in a 30 nm peak in SAXS as observed in ribosome fraction and in on HeLa cell nuclei (Nishino et al. 2012). The size of a ribosome being close to 30nm, an aggregation could result in a characteristic distance of 30nm. Even if this aggregation effect is believed to be an artefact due to the protocol of nuclei isolation and should not

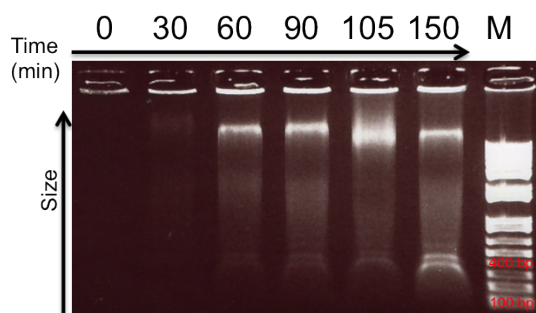
happen in-vivo. Indeed, recent cryo-electron microscopy study of yeast cells show that important number of ribosomes are visible in the cytoplasm but do not form any aggregate (Chen et al. 2016). Moreover, the correlation length expected for a ribosome stacking would be at least equal to the size of a ribosome. Consistently with the available data, this would imply a sharper peak in SAXS profile which seems to be the case for HeLa cells although a proper fit would be needed to confirm this affirmation.

8.3.3 MNase digestion

Chromatin digestion by MNase enzyme

A convincing argument that supports that the SAXS peak is due to chromatin in chicken erythrocyte nuclei is that it is not present when the nuclei are treated with DNase enzyme. This experiment could not be reproduced simply in yeast where the cell wall and membrane are less permeable.

Figure 8.3.5: DNA digestion by MNase after calcium induction



BY4741 with chromosomally integrated MNase enzyme (BYMNase) cells are sampled in parallel of SAXS experiment after induction addition of calcium. The DNA is extracted and separated in 1% agarose gel containing BET, a DNA intercalating agent (see material and methods).

Several protocols exist to permeabilise yeast cells or yeast spheroplasts in order to provide the access to the nucleus to enzymes however they generally do not need morphologically intact cells at the end of the protocol. After several tests, no good balance between enzyme access to the cell and integrity of the cell could be found.

Therefore, MNase gene from bacteria was cloned in BY4741 (see material and methods for BYMNase strain). Upon calcium activation, the enzyme cuts DNA between nucleosomes. To allow the calcium to enter the cell, a mild permeabilisation with digitonin 0.1% is necessary which increases the osmotic pressure but do not lead to visible dam-

ages to the cell.

Digestion is performed at 30°C and at several time point after the addition of calcium, cells are sampled and injected in an horizontal quartz capillary for SAXS measurement. Cells are also sampled at the same time and mixed with stop buffer to stop the enzymatic reaction and analyse the DNA cuts later on (Fig. 8.3.5, see material and methods). In the gel, DNA is marked and the apparition of short fragments shows that the digestion occurred. The smaller fragments, clearly visible

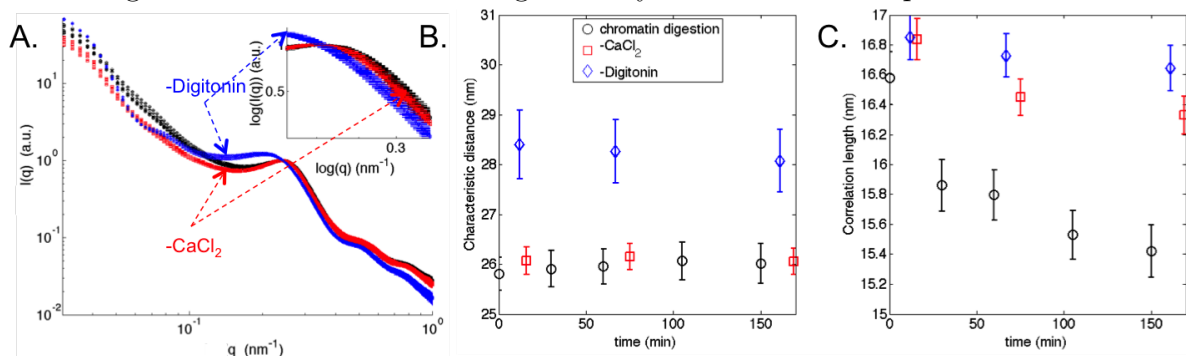
from $t = 60$ min are characteristic of mono-, di- and multinucleosomes assembly. Note that because no proteinase treatment was performed on DNA, it is not perfectly linear which explain why the smaller band is at 300-400bp whereas they are expected at ~ 150 bp. This analysis is not quantitative, in particular because of the low signal on the early time points but clearly shows that the enzyme operated during the experiment from 30 min.

As controls, two batches from the same initial pool of cells were treated in parallel respectively with digitonin or calcium.

Effect on SAXS curves

The SAXS data obtained from the digestion and the control samples are plot as loglog (Fig. 8.3.6 A.). The first observation is that both digitonin and calcium have an effect on the curve. The calcium (blue, - digitonin) seems to have an impact on the small q values while digitonin (red, $-\text{CaCl}_2$) clearly affects the peak position. The specific behaviour of the experimental curves (black) at small q is therefore difficult to interpret because it is probably an intermediate of the two effects. The second observation is that the evolution with time is extremely small and the curves are mostly superposable over time. The peak region (insert of Fig. 8.3.6 A.) is fitted with the lorentzian function. Confirming the visible effect, the characteristic distance is significantly shorter in presence of digitonin (Fig. 8.3.6 B.) which is consistent with an increase of an osmotic pressure. In addition, the correlation length decreases in presence of both digitonin and calcium (Fig. 8.3.6 C.). The diminution of correlation length is consistent with a loss of organisation that could be caused by breaks between nucleosomes.

Figure 8.3.6: Effect of DNA digestion by MNase on SAXS profiles



A. Loglog plot of SAXS data from BY MNase cells in the buffer activating digestion that contains digitonin and calcium (black), in the same buffer lacking calcium (red) or digitonin (blue) at multiple times. The inset present the region fitted with the peak that give the characteristic distance (B.) and the correlation length(C.) as a function of time for cells in the digestion buffer (black circle), lacking calcium (red squares) or digitonin (blue diamond).

The variability of the correlation length presented in Fig. 8.3.4 C. is rather high compared to this effect. However in this experiment the cell concentration is much higher than in a liquid experiment where cells must be at a low concentration to grow therefore the noise is reduced leading to a better fit and a consistent evolution with time. Moreover, the DNA is not entirely cut and forces other than covalent link such as crowding or electrostatic forces are also regulating the chromatin organisation. Therefore this effect is small but reliable and significant compared with the control conditions.

Multiple experiments were performed to investigate the origin of the signal obtained with SAS techniques. First, the comparison of SANS and SAXS data from yeast spheroplasts shows that the differences observed in the two profiles can be explained by differences of contrast in the case where the SANS signal is mostly imposed by lipids. Contrary to the published interpretation, the regular slope obtained cannot be interpreted as a fractal dimension of chromatin in this organism as well as in other organisms where this difference was observed such as chicken erythrocyte nuclei. In addition, the presence or absence of a peak corresponding to $\sim 30\text{nm}$ does not imply or deny the existence of a nucleosome fibre. Indeed, this value is a characteristic distance between dense objects and does not define the size of these objects except that their width cannot be bigger than this distance.

In yeast cells, the amplitude of the peak observed in SAXS which is proportional to the number of objects that are giving this structure is linked to replication. Indeed, in presence of HU which slows considerably DNA replication, the increase of amplitude is significantly reduced compared to normal replication condition. Moreover, the prominence of the peak is strongly correlated to the replicated fraction which implies that the number of structures giving the SAXS peak is linearly increasing with replicated fraction. In addition, cuts in chromatin by MNase reduce the organisation of this structure

Together, these results support the hypothesis that the peak is caused by chromatin. The short correlation length reflects a low level of organisation which is in accordance with loose chromatin fibre with a small persistence length and not with an organised 30-nm fibre. During replication, the compaction remains constant which is consistent with a rapid formation of the chromatin after the passage of the fork and increases in G2 which is consistent with a higher compaction before mitosis. On the other hand, the characteristic distance remains constant which implies that it corresponds to an equilibrium value depending on the crowding and electrostatic forces such as a change in the osmotic pressure of the buffer but not on the local modifications.

Chapter 9

Temporal and spatial replication program

The rules leading to the establishment of a spatio-temporal program of replication in eucaryotes are still elusive. The first part of this chapter is dedicated to describe the precaution to analyse DNA combing experiment. Then the kinetic parameters of DNA replication are extracted from DNA combing data and analysed to obtain informations on the spatio-temporal program of DNA replication.

9.1 Doing the spadework for DNA combing

DNA combing experiment

For DNA combing, MCM869 cells are synchronised in G1 by 2 hours treatment in 10^{-7} M α -factor and grown in YPD medium containing 100 μ M BrdU. Every 10 minutes, cells are sampled for FACS and 10ml of culture is transferred to YPD containing 100 μ M thymidine. Therefore, each cell batch is exposed to BrdU during a given time and the replication is then resumed in thymidine until $t = 120$ min. The purpose of this method is to obtain mostly entirely replicated cells with a low amount of single-strand DNA which can be easily broken during the experiment. After two hours, samples are treated to isolate DNA. Isolated DNA is stretched on silanized glass coverslips and stained with antibodies specific of total DNA signal and of BrdU. Fluorochromes are conjugated to specific antibodies and image of the fibres are acquired under fluorescent microscopy (see material and methods).

Obtention of the replicated tracks

Data analysis is detailed in the material and methods chapters however, the careful analysis of combing data requires caution and the main key steps are sum-

marised here. Images are cut to isolate a fibre and the pixel intensity of BrdU fluorescence is averaged on the thickness of the window to obtain the fluorescence intensity along the fibre. To convert this intensity into replicated tracks, several thresholds are defined :

- Intensity threshold : The intensity value above which a pixel is considered as replicated. Indeed, even after background subtraction, the fluorescence from total DNA is spreading causing a residual signal. To prevent this effect, this threshold value is set to obtain the higher averaged replicated fraction difference between the first time ($t = 10$ min) and the last time ($t = 120$ min). This thresholding transforms the fibre to an array of 0 and 1 where ones represent replicated parts and 0 the unreplicated parts.
- Size of false negatives : False negative are caused by an irregular BrdU staining (Kaykov et al. 2016). The threshold is defined at 50 pixels ($\sim 2\mu\text{m}$) because at this value, the total replicated fraction at $t = 120$ min from combing reaches the value from FACS data (Fig. 9.1.1). Given that FACS and DNA combing do not have the same sensitivity, it might be abusive to use FACS as an external reference. Moreover, other combing studies have used smaller values (Platel et al. 2015). Therefore, even if this value is set as the reference, several values have been tested for this threshold and the conclusions that are exposed in the subsequent parts are not modified in a large threshold range (see annexe).
- Size of false positives : False positive are defined as tracks that are below twice the optical limit ($1\mu\text{m} \sim 23$ pixels). This threshold also removes very small tracks that could be due to incorporation of BrdU for DNA repair.

This result in fibres containing replicated parts (also called tracks or eyes) and unreplicated parts (gaps). Replicated fraction at a given time can be computed by dividing the total length of all tracks by the total length of the all fibres at this time.

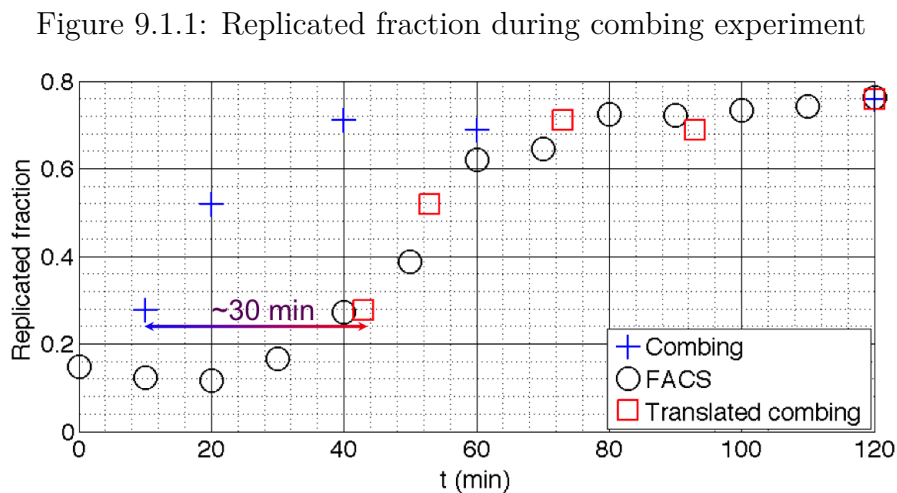
Conversion to genomic distance

The distance are measured in pixel and are converted to μm by the calibration rate of $0.0434 \mu\text{m}$ per pixel. To convert this combed distance into genomic distance, dsDNA from the bacteriophage lambda of 48.5kb is combed on the same cover slips (see material and methods). A important number of the fibres are smaller than the expected size of lambda DNA ($\sim 16\mu\text{m}$) which is mainly due to breakage of the DNA molecules during combing. Indeed, the force exerted on the fibres by the surface tension of water is $\sim 150\text{pN}$ (Bensimon et al. 1995) which is sufficient to break

DNA molecules (Cluzel et al. 1996). Fibres could also appear shorter because of an inhomogenous stretching. However, the variation of fluorescence along the fibres is similar for long and short fibres which suggest that short fibre do not present more combing defect than long fibres. The mean fluorescence per length unit is almost equivalent although it is slightly smaller for short fibres which suggest a $\sim 10\%$ better stretching. The stretching is assumed to be overall homogenous as it was done in previous studies (Michalet et al. 1997). The long fibres are $21.2 \pm 2.7 \mu\text{m}$ which corresponds to an elongation of ~ 1.3 and a conversion of $2.3 \pm 0.3 \text{ kb}/\mu\text{m}$ consistent with previous combing studies (Michalet et al. 1997).

Function of time vs function of f

The replicated fraction inferred from combing data is significantly shifted compared to FACS data. In the first chapter of results we saw that cells stock a substantial amount of dNTP which allows to replicated up to a replicated fraction that would normally be reached in 33min in a normal cell cycle (Fig. 7.2.7 B.). In our experimental conditions, cells are able to stock BrdU and incorporate it even when then are released in thymidine medium. Therefore the experimental time which corresponds to the medium change is not the time where the cells actually stop incorporating BrdU but it is substantially shifted. By translating combing data points by 33min (except for the last time point that was sampled at the end of the experiment and not resuspended in growth medium), the replicated fraction from combing matches the replicated fraction from FACS (Fig. 9.1.1).



Replicated fraction from combing (blue plus) and from parallel FACS measurement (black circles) as a function of time. The time values from combing can be corrected by 33min due to incorporation of BrdU stock (red squares)

In addition to this uncertainty on the absolute time, the lack of synchrony which

was already visible by FACS is reinforced in this single cell technique which is more sensitive and has a poorer statistics. DNA replication can be considered spatially homogenous (Jun and Bechhoefer 2005) and all segments with a similar replicated fraction f are at roughly the same point in S phase. Therefore, fibres are sorted by their replicated fraction and their replicated fraction can be related to the time when they start DNA replication by normalised $F(t)$ obtained from FACS (Ma et al. 2012). Because the apparent S phase duration on a population obtained from replicated fraction curves depends on the variability of S phase entry as well as on the S phase duration of single cells, the measured S phase duration is longer than the length of S phase in each cell.

9.2 Kinetics of DNA replication

9.2.1 Replication dynamic

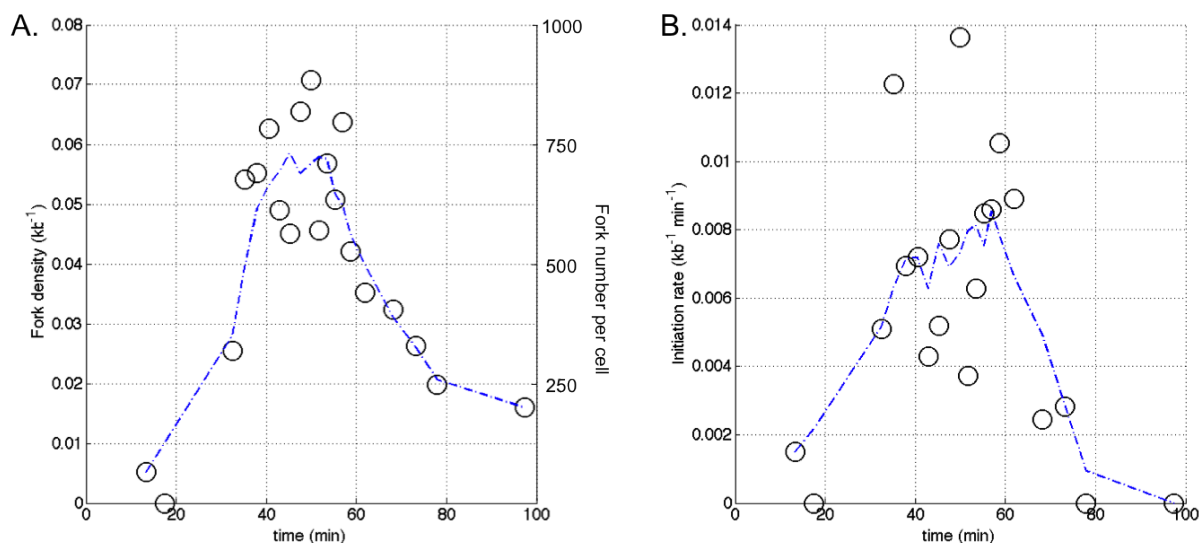
Fork number and initiation rate

The number of active forks is directly obtained by DNA combing by counting the tracks borders. The density of forks $N(t)$ is the number of fork per length unit (see material and methods) (Fig. 9.2.1 A.). The fork density increases at the beginning of S phase, reaches a maximum value and diminishes. The ascending slope is due to the creation of new forks while the descending slope is due to fork fusion or merge. Indeed, after $t = 60$ min initiation events are rare (Fig. 9.2.1 B.). These general shape of the fork density is in accordance with already published data obtained by DNA combing or population methods (Ma et al. 2012). The fact that single molecule and population average behaviours are similar support the idea that the replication process is spatially homogenous above the size of combed fibres. Therefore, the fork density multiplied by the genome size should represent the number of active forks in a cell (Fig. 9.2.1 A. right axis).

Initiation rate is the number of initiation per time unit and unreplicated DNA length (see material and methods). In our DNA combing data, all tracks <5 kb were considered to be a replicon (created by a single origin and not by the merging of forks from several origins). Initiation rate gives the probability of firing on the non replicated DNA and is independent of time and space therefore giving valuable information of the dynamics of the replication dynamics.

The initiation rate inferred from the DNA combing data (Fig. 9.2.1 B.) is in good agreement with previously published results (Goldar et al. 2009, Ma et al. 2012). As expected, the initiation rate is increasing during S phase to promote the entire replication of the genome. The integration of initiation rate over time gives the number of 0.18 initiation per kb which corresponds to 2200 initiations

Figure 9.2.1: Replication parameters extracted from DNA combing



A. Fork density as a function of time (black circles) smoothed with a mobile mean on five points (blue dashed line)
 B. Initiation rate as a function of time (black circles) smoothed with a mobile mean on five points (blue dashed line)

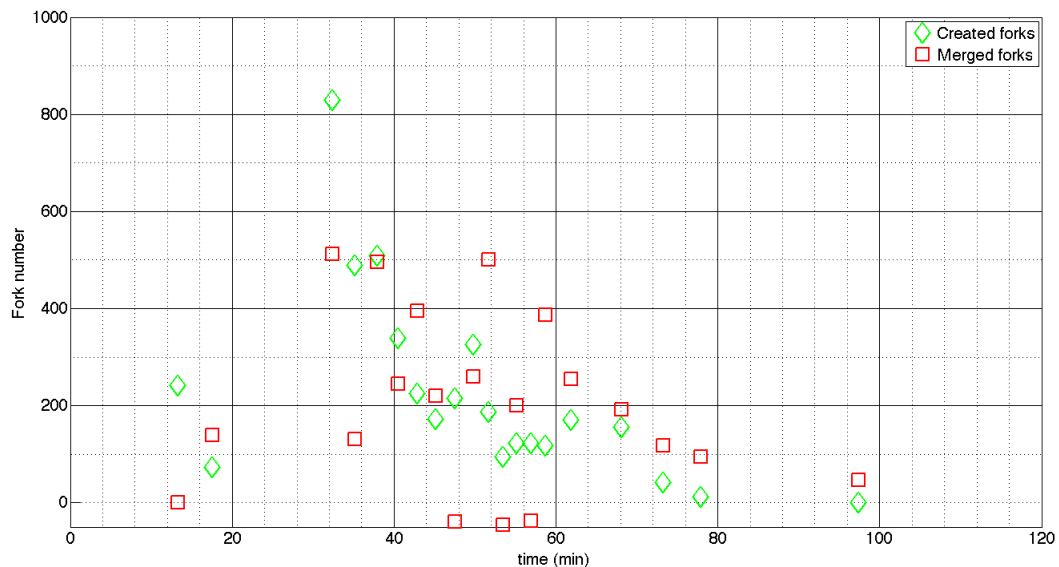
over the genome in the analysed population. The number of cells analysed is not possible to obtain directly. The total quantity of analysed DNA corresponds to the genomes of 1.8 cells which, always in the hypothesis of spatial homogeneity, corresponds to ~ 1200 initiations per cell and one initiation every $\sim 10\text{kb}$.

Forks creation and merging

The number of created forks can be obtained thanks to the initiation rate and thus the number of merged forks can be deduced from the number of active forks (see material and methods) (Fig. 9.2.2). One could imagine that first many forks are created, and only start to merge after a significant delay. In this case, the maximum value of the active forks number would coincide with the number of created forks.

The number of initiation reveals that many forks are created and therefore they start to merge very early. For example at time $\sim 32\text{min}$, ~ 800 forks are created while no active forks were remaining from the previous time point which implies that the ~ 300 forks that have merged between time $\sim 18\text{min}$ and time $\sim 32\text{min}$ were also created during this gap time. In $\sim 14\text{min}$, $\sim 60\%$ of the created forks have merged. This means that they were less than 50kb apart (distance travelled by 2 forks during the gap time considering a constant speed of $1.68\text{kb}/\text{min}$).

Figure 9.2.2: Creation and merge of forks



Number of created (green diamond) and merged (red square) forks during the interval between two time points as a function of time

The subsequent time points have a higher resolution showing that up to 40% of initiations occurred at less than 10kb apart.

9.2.2 Origin distribution

Eye-to-eye distance

An approximation of spatial origin distribution can be obtained by plotting the eye-to-eye distance (ETED) which is the distance between the centres of two adjacent tracks (Lengronne et al. 2001, see material and methods). If most tracks are the result of merges of forks from several origins, therefore, the ETED is an overestimation of the Inter-Origin distance (IOD) (Fig. 9.2.3 A.).

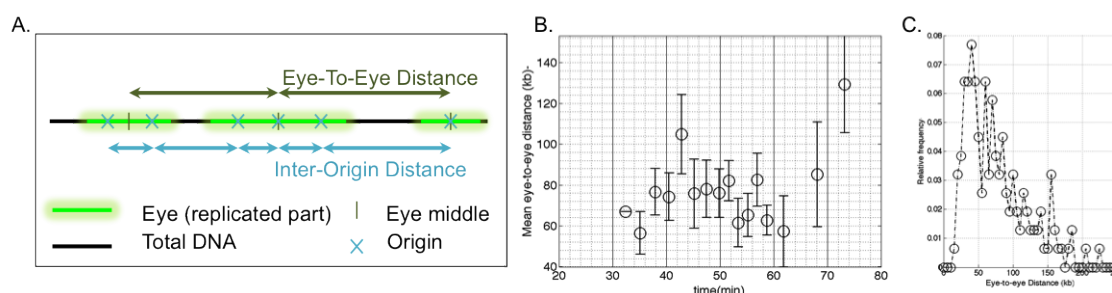
The average ETED presents a non linear behaviour as a function of time (Fig. 9.2.3 B.). Indeed, ETED increase until ~42min, then is stabilised around an average value of ~70kb and is increasing at the end of the S phase. The increase of the ETED implies that close eyes merged while the decrease means that new initiations occurred. In *Xenopus*, the ETED decrease rapidly at the beginning, reaches a plateau at a length of ~20kb and also increase at the end of S phase (Herrick et al. 2002). The initial decrease and the low value of the plateau suggests that the origins are located close to each other in cluster. Our data does not present an

initial decrease that could be due to the fact that we lack very early time points. In this case, the increase around $t \sim 40$ min could be the sign of the end of a first wave of initiation. The plateau value is also higher which suggest an other mode of clustering.

The average value at early time point (56.5 ± 10.3 kb at $t \sim 35$ min) is consistent with the value of 57.8 ± 38.1 kb obtained by DNA combing on cells treated with HU (Tourrière et al. 2005). However, as the distances are not normally distributed, the average is not a reliable indicator.

The distribution of the distances on pooled time points (Fig.9.2.3 C.) clearly shows that an important proportion of fibres are closer (40% of distances lower than 50kb) with a peak around 40kb.

Figure 9.2.3: Eye-to-eye distance



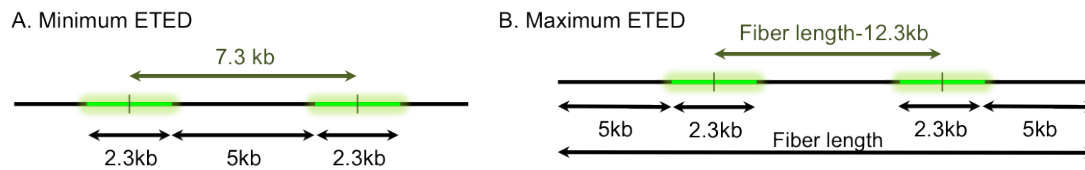
- A. Schematic representation of Eye-To-Eye Distance (ETED) and Inter-Origin Distance (IOD).
 B. Mean ETED as a function of time
 C. Distribution of ETED for pooled time points in 5 kb bins

These values are higher than the initiation distance inferred from the dynamic behaviour of replication. Several biases in the measure of ETED could explain this discrepancy.

Possible biases

The different thresholds chosen in the first subsection to analyse the data are determinant for a reliable analysis. Indeed, the false negative (≤ 5 kb) and false positive (≥ 2.3 kb) threshold values impose the minimum length of measurable ETED (Fig. 9.2.4 A.). In addition, to be identified as an eye with a determined centre position, tracks must not be on the borders of the fibre, therefore the maximum measurable length is also dependant on the threshold values and on the length of the fibre (Fig. 9.2.4 B.). In our data set, the fibres are not entire chromosomes but are rather short (101.77 ± 3.72 kb). However, the subsequent analysis suggests that the replication is spatially homogenous over the size of the fibres because the replication parameters measured are consistent with those obtained from population global behaviour.

Figure 9.2.4: Biases impose minimum and maximum measurable ETED (Fig.)



In addition, ETED is intrinsically an overestimation of the IOD. Indeed, one eye is not necessarily a replicon but can be composed of the fusion of tracks from several origins. The apparent initiation time of an eye t_{eye} depends only on the initiation time of its most distant origins t_1 and t_n , their distances $d_{1,n}$ and the fork velocity (see annexe):

$$t_{eye} = \frac{t_1 + t_n}{2} - \frac{d_{1,n}}{2 \times v}$$

Therefore, for small tracks, $d_{1,n}$ is small and the apparent initiation time is the averaged of the the initiation time of the extreme origins of the eye. In addition, if the distance between two origins is small, their sizes are also small and their initiation times are comparable which implies that their average value is close to their actual initiation time. However, for bigger eyes, initiation time is an overestimation of the first initiation time. Considering ETED instead of actual IOD lead to overestimation of both the IOD and the initiation time.

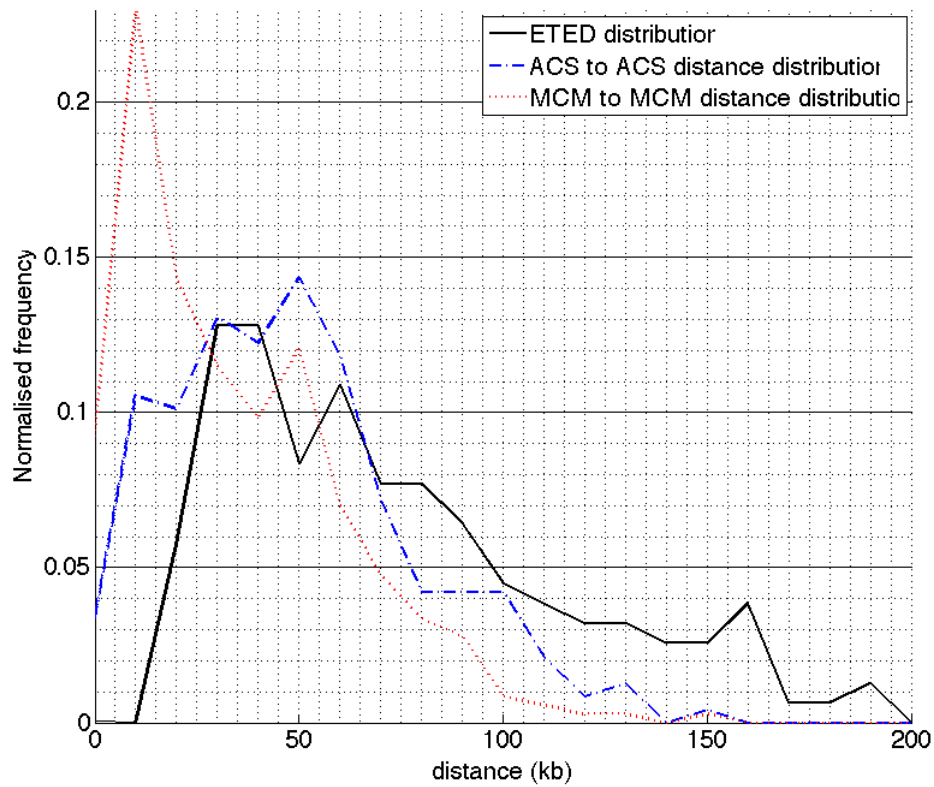
MCM are potential origins

ACS (consensus sequence for ORC binding) positions that link ORC were mapped by Eaton et al. (Eaton et al. 2010). These positions are correlated to peaks of initiation in replication profiles. Distance between adjacent ACS were computed from this data set and their distribution presented in Fig. 9.2.5 (blue dashed line). The main distance is comparable to ETED distribution ~ 40 - 50 kb. The distributions are not totally similar because the small ETED values cannot be detected by our method.

MCM positions have been mapped by ChIP-seq (Xu et al. 2006) and distance between adjacent MCM is also presented in Fig. 9.2.5 (red dotted line). As expected, MCM are closer between them than ORC loading positions. More than 20% of MCM are closer than 10kb which is corresponds to the distance of ~ 10 kb between origins that was inferred from the dynamic analysis of the DNA combing data. MCM are loaded by ORC and can slide along DNA therefore the concentration of MCM is higher around ACS.

Detection by ChIP-seq only allows to detect well positioned MCM. The total number of MCM detected ~ 400 is far below the number of initiation measured

Figure 9.2.5: ACS and MCM distances distributions



ETED distribution in 10kb bins (black line) presented in Fig. 9.2.3 in 5 kb bins C. with an increased binning for clarity is compared to the distribution of ACS to ACS distances distribution (blue dashed line) and MCM to MCM distances distribution (red pointed line). ACS positions values were obtained from Eaton et al. 2010 (ACS-ORC) and MCM positions from Xu et al. 2006

here ~1200 and MCMs should be in excess compared to the number of initiations. However, more local analysis lead to conclude that one ORC could load more than 5 to 10 MCMs (Donovan et al. 1997). Given that ORCs are well positioned and therefore more accurately detected by Chip-seq and that ~250 ORCs are detected, our value of 1200 initiations is plausible. Therefore, MCM are not perfectly detected by Chip-seq, however, the bias is the same around each ORC therefore one can assume that the distribution is still valid.

One possibility to explain both the similarity of ETED and ACS distances distribution and the discrepancy between ETED and IOD is that the initiation locally occurs in a short time range which implies that all MCM (that are potential origins) initiates roughly around the same time and rapidly merge into one eye detected by DNA combing which is more or less centred in the ACS position.

9.2.3 Timing

Timing cluster

According to the the extracted parameters of the replication kinetics, the S phase lasts ~40min (Fig. 9.2.1 C.). However, potential origins that are at a distance of 10kb need to fire less than 6min apart to be both activated and not passively replicated by one another. Therefore spatially close origins fire at close time.

Initiation is executed by limiting factors that can be recycled after the merging of forks. Timing clusters result from an increased probability of initiation around already fired origins. Recycling of limiting factor that would be therefore locally available cannot account for the clustering because initiation occurs before neighbouring forks have merged. Therefore, spatial clustering suggests an effect of active forks on the initiation of neighbouring origins. Accessibility of chromatin is increased around active forks in a range of several kbs (Rodriguez and Tsukiyama 2013) which could participate in the increased probability of initiation.

To be efficient this increase in local initiation probability should occur in an area concentrated in potential origins.

MCM distribution and timing

All ARS do not have the same probability of initiation and the MCM concentration around ARS has been proposed as one of the regulator of the probability of initiation (Yang et al. 2010, Das and Rhind 2016, Hyrien 2016). Indeed, the time at which an ARS is replicated is not globally correlated to the local number of MCM (Belsky et al. 2015, McGuffee et al. 2013). Indeed, this correlation is only true for the ARS that are not dependant on another identified mode of regulation

(Das et al. 2015). This observation supports the fact that in the absence of specific regulators, the initiation is randomly caused by the encounter of initiation factor and a potential origin.

Instead of considering the replication time which is the time at which one position is replicated in half a population of cells, one can consider the window Δt during which a position is replicated in a cell population (Fig.2.1.1). These two values have been shown to present a correlation (Yang et al. 2010). However, the correlation is not perfect. Moreover, contrary to absolute timing, Δt is inversely proportional to the probability of initiation independently of the beginning of the S phase and of the contribution of other modes of regulation that would delay the replication of given regions.

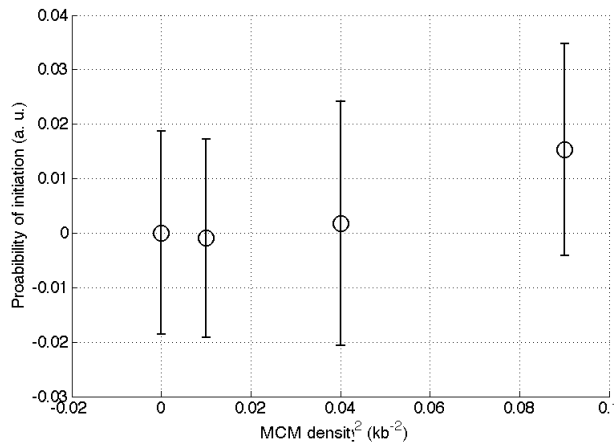
MCM positions are obtained from the same data set previously used (Xu et al. 2006) and Δt is obtained by fitting the quantity of DNA as a function of time by a sigmoid curve for each position measured published by Hawkins et al. (Hawkins et al. 2013). To guarantee that the fit gives accurate value despite the small number of points, the value of Δt are considered for $\chi^2 < 1$. Average value of Δt and the MCM density are computed inside 10kb bins.

The probability of initiation P is linked to the number of MCM m_0 (Gauthier et al. 2010) by:

$$P = \frac{v_0}{\pi \times L^2} \times m_0^2$$

where v_0 is the speed corresponding to an optimal origin usage and L the genome length.

Figure 9.2.6: Probability of initiation linked to square MCM density



$\frac{1}{\Delta t}$ as a function of the square MCM density. Note that $\frac{1}{\Delta t}$ is proportional to the probability of initiation when the probability of passive replication can be neglected

Δt depends on the probability of initiation and the probability of passive replication. The probability of initiation is expected to be inversely proportional to Δt if the probability of passive replication is neglected. $1/\Delta t$ is plotted as a function of the square MCM density (Fig. 9.2.6).

The probability of initiation increases with the MCM density. The relation seems linear for MCM density > 0 but the number of points is low because of the underdetection of the number of MCM by ChIP-seq.

Therefore MCM density is linked to the genome wide variability of replication but not directly to the replication time of many ARS. Moreover, the dependance between MCM density and probability of initiation is not valid for large bins. This suggest that general regulation factor define timing domains that are further locally regulated by the grouped initiation of MCMs.

This conclusion is a first approximation, because the probability of passive replication is assumed to be constant. However, it is likely that the probability of passive replication also evolves with the MCM density in a non linear manner.

DNA combing allows a direct detection of the replication parameters such as eye-to-eye distance (ETED). However, even with the single-cell resolution, numerous biases makes these observations difficult to link directly to the origin behaviour. In particular because eyes are not necessarily replicons but can result from multiple initiation and fork merges. Moreover, focusing on small eyes that are more likely to be single replicons drastically reduce the statistic. Efforts are made in the direction of the improvement of the technique throughput (De Carli et al. 2015, Kaykov et al. 2016).

Another approach consist in obtaining the initiation rate from small tracks which describe the dynamic of the system. Under an assumption of large scale spatial homogeneity, one can compute the replication parameters freed from several biases. This approach reveals that the number of initiation and active forks is greater than anticipated and that to respect this dynamic, origins must be closer than described from the ETED only.

MCM distribution on the genome is consistent with this short inter-origin distance. MCM are distributed around ORC-binding sites and form timing cluster which result in the ETED distance measured at low time resolution. Moreover, the density of MCMs is linked to the local probability of initiation.

Conclusions and perspectives

Measure of nuclear organisation by SAS techniques

One of the objective of the present work, was to analyse the higher order organisation of chromatin inside yeast cell nuclei and its evolution during DNA replication. Based on the work of Lebedev and co-workers (Lebedev et al. 2005), small-angle scattering techniques were investigated to assess the nuclear structure in a wide scale range (5-500 nm). However a deeper investigation of the literature revealed inconsistency between the published data produced by SANS (using neutrons) and SAXS (using X-rays). Using both techniques to study yeast spheroplasts, we find the same inconsistency.

X-rays and neutrons are not scattered equally by the main components of the cells, as lipids scatter neutrons much more than X-rays. We show here that SAXS and SANS results can be reconciled if the signal obtained with SANS is mainly due to scattering by lipids. Lipids are also probably the main contributor of the SAXS curve of yeast spheroplasts at small q where SAXS and SANS data are in agreement. Therefore, we argue that the use of SANS signal to deduce the behaviour of chromatin inside nuclei or entire cells is abusive. In addition, the SAXS signal at small q is the addition or the interaction of signals from different structures.

In conclusion, SAS methods are not currently suited to study the large scale chromatin organisation *in vivo*. Development of a high resolution *in silico* model of nuclear organisation could allow deciphering the different contributions of lipids, proteins and DNA and understand their evolution with time.

Interestingly, SAXS data present a peak which is not present in SANS and is therefore due to proteins or DNA or a combination of the two. This peak corresponds to a characteristic distance of ~ 30 nm, which has been interpreted as the distance between compact 30 nm chromatin fibres in chicken erythrocyte nuclei (Langmore and others 1983) or aggregates of ribosomes which are artefactually stuck on isolated nuclei of HeLa cells (Nishino et al. 2012).

On the ribosome hypothesis

In entire yeast cells, recent electron microscopy images show a dense presence of ribosomes (Chen et al. 2016) which have a diameter of ~25-30 nm. Therefore, aggregates of ribosomes on close packing could lead to a correlation peak of a characteristic distance of ~25-30 nm. However, several lines of evidence described in the result section tend to discredit this hypothesis :

- the characteristic distance can diminish with buffer condition down to ~25 nm which corresponds to a perfect close packing,
- the correlation length, which is the distance beyond which the scatterers are not in interaction, is smaller than the ribosome size (~15-17 nm),
- the number of scatterers responsible for the peak is linearly evolving with DNA replication and depends on the DNA replication rate,
- the correlation length is affected by MNase treatment.

In conclusion, the ribosomes may be contributing to create a peak around 30nm but we argue here that it is not responsible of the measured evolution. Therefore in the subsequent development chromatin will be considered as the main contributor of this peak and responsible of the modification of its parameters.

Future work can be performed to further reinforce this conclusion. First, the presence of ribosomes on the surface of yeast isolated nuclei can be checked with specific antibodies. New SAXS measurement could be done on washed isolated nuclei devoid of these potentially found ribosome aggregates. In such nuclei, DNase treatment or MNase activation could be performed to measure the effect on the peak parameters. In addition, the ribosome quantity during the cell cycle could be precisely measured to confront to the time evolution of the peak parameters in entire cells. Finally, the *in silico* model of entire cell would need to simulate the dense presence of ribosomes in the cytoplasm.

Chromatin local structure

The peak corresponding to a characteristic distance of 30 nm do not imply that the chromatin forms a 30nm fibre. Indeed, only 30 nm fibres in very close packing would give rise to a repeated distance of 30 nm as in chicken erythrocyte (characteristic distance has a range of ~30-40 nm opposed to the characteristic distance in yeast of ~25-30 nm). On the contrary, the short correlation length do not support the existence of a well defined a 30 nm fibre organisation. Instead our data depicts chromatin as a nematic liquid crystal. In this picture, nucleosomes

are organised as a bead-on-a-string fibres of ~10nm which are distant of ~25 to 30 nm.

This distance is strikingly constant during the cell cycle but is affected by the osmotic pressure of the buffer. This suggests that the distance between objects corresponds to an equilibrium dictated by the physicochemical properties of the environment.

The value of the correlation length is an average over the genome and over the cell population therefore it does not exclude particular arrangement whose contribution would be reduced by the averaging. However value of ~16 nm suggests that nucleosomes are mostly influenced only by their direct neighbours and less often form small aggregates of few nucleosomes, in accordance to the recent electron microscopy images (Chen et al. 2016).

This vision of chromatin is close to the model of chromatin organisation in human cells, however no characteristic distance is observable by SAXS in HeLa cells after ribosomes removal. This suggests that, in human, chromatin has a completely liquid behaviour while in yeast, the particular Rabl organisation leads to a preferential orientation that give rise to a nematic order. This orientation is potentially more localised in centromeres regions were the crowding forces a local linear orientation.

More information on the local structure of the chromatin in the nucleus will be obtained by the X-ray imaging of single cells.

Evolution of the correlation length with time

The correlation length is a length over which scatterers are in interaction, and can be understood as an energy of interaction over an entropic force that tends to disorganise the system. Interactions between nucleosomes are made via linkers that can be modelled as springs with a given stiffness k . This gives :

$$\xi = \frac{k \times x^2}{2F}$$

where ξ is the correlation length, x is the elongation of the linker relative to its equilibrium position and F is the entropic force that tends to disorganise the system.

The difference between the correlation length at any time $\xi(t)$ and the initial correlation length ξ_0 depends on their respective stiffness and elongation :

$$\xi(t) - \xi_0 = \frac{k \times x(t)^2 - k_0 \times x_0^2}{2F}$$

Digestion of nucleosomes linker by MNase enzyme leads to a decrease of the correlation length. As MNase is cutting some linkers, their stiffness becomes zero, therefore diminishing the whole genome average stiffness.

During interphase however, the linker stiffness can be considered not affected and $k = k_0$ which gives :

$$\left| \frac{x(t)}{x_0} \right| = \sqrt{\frac{\xi(t)}{\xi_0}}$$

As ξ increases with time during interphase, the absolute value of x also increases which means that the linker is moving away from its equilibrium size. Whether this corresponds to a shorter or longer linker size is more difficult to predict. Chromatin is expected to become denser before cell division to allow the correct distribution of chromosomes between the two cells. However, the mechanism of condensation could involve a shortening of the linker to induce closer nucleosomes or a lengthening, in order to allow more complex nucleosome interactions. Furthermore, this effect is rather small in accordance with the observation of an absence of major reorganisation in yeast metaphase cells by electron microscopy (Chen et al. 2016).

The increase of the correlation length starts at the beginning of the experiment where cells are reorganising after α -factor release. It is not clear whether the correlation length stays constant or increases during S phase. Indeed, while the data set with the higher time resolution is almost constant during S phase, the data set in pellet that is in best accordance with the model suggests a linear increase during all interphase. In both cases, no drastic modification is observed at the start of DNA replication that could be linked to the peak of initiation (see annex B.4). Therefore, it seems that the potential modifications during S phase are not caused by a specific reorganisation that would allow DNA replication, for example a local chromatin opening, but mostly by the creation of matter in the nucleus that causes rearrangement in the cell.

Overall, these interpretations require more data sets with a precise knowledge of the state of the measured cells so the small effects detected here are confirmed. However, one can already say from this first data sets that the reorganisation during S phase is not as drastic as it was pictured. Moreover, as the modifications during S phase are small, it seems that the G1 conformation is suitable to perform S phase. This suggests that if any particular organisation is required to perform DNA replication, it is achieved during the G1 phase.

Timing decision point and temporal program of replication

Interestingly, the replication program is suggested to be imposed in G1 (Wu and Gilbert 1996). Although the exact biological mechanism behind the concept of timing decision point is still unknown, multiple non-exclusive hypotheses have been proposed.

The residency time of ORC in particular positions as well as its surrounding nucleosome environment leads to a specific distribution of MCMs along the gen-

ome. This repartition of potential origins induces an inhomogeneous probability of initiation along the genome and is set during G1-phase.

On the other hand, *fkh1* is associated to early firing and is associated to origins in G1. The exact mechanism by which Fkh induce early firing is still unknown but it is known to act on the clustering of early origins (Knott et al. 2012) where it could promote the recruitment of *cdc45* at the end of G1 to grant early replication. In addition, *Rif1* sequesters late origins through physical clustering at the nuclear periphery and prevents the initiation by a molecular inactivation of the replicative kinase (Peace et al. 2014).

It has been shown that the number of MCM is correlated to the initiation timing of origins but only if they are not regulated by other known elements such as *Fkh1/2* and *Rif1*. However, we showed that the probability of initiation is correlated genome-wide to the MCM density, which implies that the MCM distribution is the first level of regulation of the timing program.

Contrary to eukaryotes with larger and more compact genomes, the global chromatin structure do not impede the diffusion of limiting factors to focus the initiation in defined areas. However, in the hypothesis that limiting factors are asymmetrically imported, with a major input in the centromeric region, the sequestration of limiting factor to early origins slows down their diffusion in the nuclear volume.

Therefore, the temporal program is regulated on three scales set during G1-phase:

- nuclear layout guarantees the centromeres clustering and the positioning of telomeres as well as a free diffusion,
- scaffolding factors such as *Fkh1/2* and *Rif1* act locally on the initiation factors (respectively by recruiting *cdc45* or deactivation DDK), creating respectively preferential or disadvantageous areas for initiation and physically tethering potential origins in these areas,
- more locally, a concentration of potential origin increases the probability of initiation.

Spatio-temporal program and clusters

During S phase, the program is executed by the encounter between initiation factors and potential origins. To explain the measured behaviour of fork density and initiation rate during S-phase, the initiation factors must be present in limiting quantity that increases through S phase (Rhind et al. 2010). This, however, does not explain the spatial distribution of origins. The distribution of origins is difficult to detect with cell populations methods where rare events are erased by

averaging. Therefore, the distribution of origins was assessed by DNA combing. This method provides insight about single cells as the newly synthesised DNA is directly visualised on DNA. The distance between two zones of incorporation corresponds to about 50kb at the beginning of the S-phase. However, in this work, we perform a more in-depth analysis of DNA combing data based on the extract temporal behaviour of the initiation rate and the fork number. This reveals that 20 to 40% of initiations occur at less than 10kb of distance in a short period of time. Forks from this origins rapidly merge and mostly result in longer tracks with a distance between them comparable to the distance between ORC-ACS. Therefore, we suggest that rather synchronous initiation of close MCMs loaded by the same ORC rapidly results in regions of replication more or less centred in the ORC-ACS position.

MCM are therefore forming timing clusters. Such initiation clusters were observed in many eukaryotes which suggests that, in addition of a common temporal dynamic, the spatial program of replication is also mainly similar in all eukaryotes. Although larger genomes could require additional regulatory mechanisms, the spatio-temporal program, on adapted scales, is similar.

Timing clusters are created by a local increase of the initiation probability. The local recycling of limiting factors such as *cdc45* is expected to occur after fork fusion and therefore cannot explain the observed initiation that need to happen before merging of neighbouring forks. It was observed that the chromatin is more accessible in a 5 to 7kb range around active forks (Rodriguez and Tsukiyama 2013) which could explain an increase of the local initiation induced by active forks. In this hypothesis, the tension due to DNA unwinding or the action of chromatin remodellers associated with active forks would deplete nucleosomes upstream of active forks.

This hypothesis is difficult to reconcile with the fact that the peak of initiation does not induce local reorganisation. The number of nucleosomes displaced by a 7kb range influence of fork can be computed from the density of mapped nucleosomes. Moreover, in the hypothesis where the chromatin is linearly recreated after fork passage, we can also estimate the number of newly deposited nucleosomes to compute the proportion of nucleosomes displaced in the cell. Therefore, we estimate that up to 30% of nucleosomes are affected by the fork progression at the peak of initiation. It is possible that the exact fork number is overestimated by our analysis, as we consider an homogenous distribution on the genome. However, it is also possible that this effect is not visible on the correlation length because it is an average value over the population which is not perfectly synchronised and whose initiation peaks can be asynchronous.

Population versus single cell

Overall, this work highlighted that the variability in a cell population can strongly affect the interpretation of the data. The work on single cells gives precious insight on the precise mechanisms of DNA replication. However, as the throughput is low, it is difficult to obtain statistical significance and data have to be carefully interpreted. Working with a cell population is often necessary, not only because it allows to gather information easily, but also because sometimes the global behaviour and the noise already contain information on the process.

Two approaches can be undertaken to obtain more homogeneous population. On the one hand, experimental settings we used lead to a synchronisation of most cells in G1 but with a variability in the S-phase entry. In the future, cells could be sorted to obtain a more homogeneous population. On the other hand, the variability in the cell population can be estimated by analysing our population as composed of population subsets. Such an approach was already applied to the analysis of FACS data, where considering three populations allows to measure the entry of S phase of the global population. Moreover, instead of trying to analyse the population behaviour, one can use a single cell behaviour and a noise function to model the population data.

The present work investigated several directions to study the evolution of cells during DNA replication in order to get insight into the DNA replication program. Measurement of the structure inside the cell during DNA replication *in vivo* revealed a striking stability even during period of the S phase with high-frequency of initiation which was recently observed *in vitro* as well. Moreover, the high frequency of initiation can be interpreted in terms of fork creation and merge which suggests that the previous interpretation of combing data underestimated the number of origins and overestimated their distance. These new data set confirms the role of MCMs as potential origin and supports an initiation in cluster comparable to what is observed in other eukaryotes.

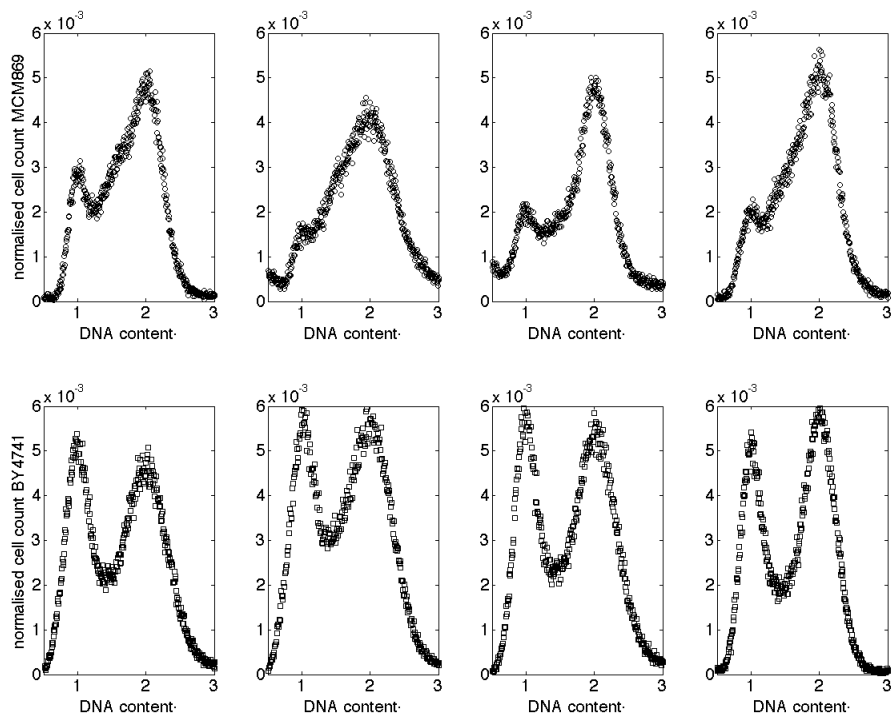
Overall, this work proposed new technical and analytical methods to study DNA replication that opened paths for future investigations. Furthermore, this work intended to review and take advantage of the extensive literature and data sets concerning budding yeast replication. The effort in building a comprehensive model of eukaryotes DNA replication process should focus on integrating all these data.

Annexes

Annexe A

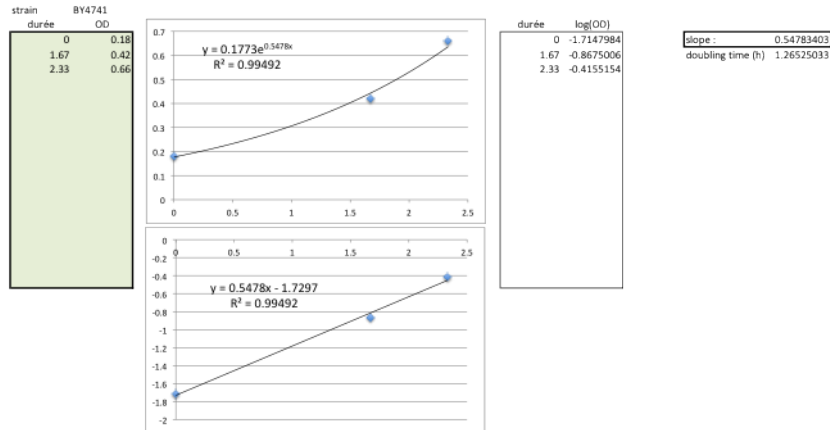
Evolution of the cell cycle - supplementary informations

FIGURE A.1 : Logarithmic state of MCM869 and BY4741 cells



FACS histogram of logarithmic state of MCM869 strain (top) and BY4741 strain (bottom) shows that that logarithmic state of MCM869 cells is notably biased towards late S-G2 phase. This effect is probably caused by the longer S phase of MCM869 strain that need to incorporate thymidine analog from the growth medium.

FIGURE A.2 : Logarithmic growth of BY4741



Optical density is monitored with time during BY4741 growth. The linear evolution of $\ln(OD)$ with time is characteristic of logarithmic growth. The doubling time is $\frac{\ln(2)}{\text{slope}}$ with here gives 1.26h or 76min.

FIGURE A.3 : Goodness of the fit for FACS histograms and replicated fraction

	χ^2	R2	p<1e-4		χ^2	R2	p<1e-4		
Fig. 7.1.8	30min	6.106	0.991	Y	0	0.0812	0.9958	Y	
	1h	3.5028	0.9876	Y	10	0.053	0.9981	Y	
	2h	3.1276	0.9897	Y	20	0.0468	0.9964	Y	
	3h	2.5342	0.9883	Y	30	0.0426	0.9942	Y	
	4h	8.6551	0.9956	Y	40	0.0192	0.9917	Y	
Fig. 7.1.10	experiment1	0.7533	0.9938	Y	50	0.0121	0.9967	Y	
	experiment2	3.7481	0.9908	Y	60	0.0146	0.9978	Y	
Fig. 7.2.2	log	10.4395	0.9635	Y	70	0.0176	0.9979	Y	
	0	4.9387	0.9965	Y	80	0.0222	0.9978	Y	
	10	2.5524	0.9973	Y	90	0.0243	0.9977	Y	
	20	2.4049	0.9978	Y	100	0.0281	0.9968	Y	
	25	4.1017	0.9974	Y	110	0.0394	0.9953	Y	
	30	4.9573	0.9956	Y	120	0.0382	0.9968	Y	
	35	6.2421	0.9929	Y	fraction fit	4.4761	0.9964	Y	
	40	7.0566	0.9837	Y	Fig. 7.2.4	fraction ypd	2.3524	0.9937	Y
	45	7.4235	0.9556	Y	fraction casa	2.8698	0.9851	Y	
	50	8.5861	0.9286	Y	Fig. 7.2.7	0	4.5983	0.9967	Y
	55	8.4475	0.9437	Y	20	3.4239	0.9969	Y	
	60	7.9049	0.9534	Y	60	4.3739	0.9815	Y	
	65	8.1537	0.9672	Y	100	6.2914	0.9724	Y	
	70	7.9294	0.9869	Y					
	75	8.0848	0.9902	Y					
	80	9.7887	0.9801	Y					
	90	8.4947	0.9811	Y					
120	7.6346	0.987	Y						
replicated fraction raw	4.3212	0.8677	Y						
replicated fraction fit	2.807	0.9087	Y						

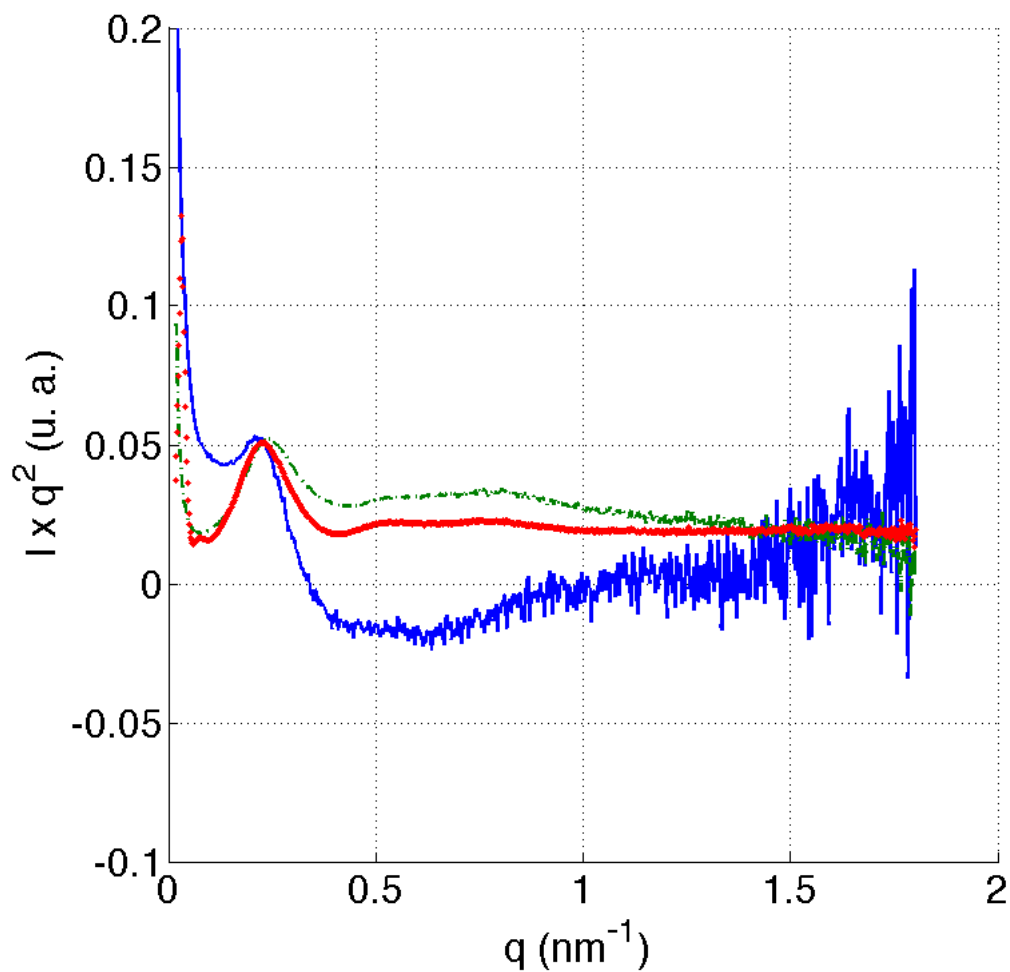
R² and p value of each fit performed in the chapter 7.

χ^2 ,

Annexe B

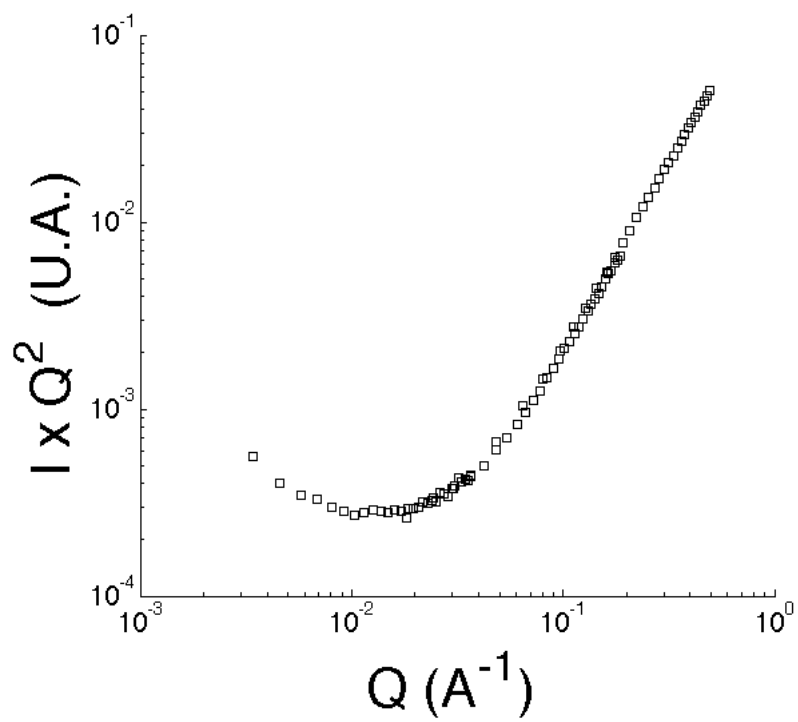
Yeast nuclear organisation - supplementary informations

FIGURE B.1 : Cell, spheroplasts and nuclei normalised SAXS data



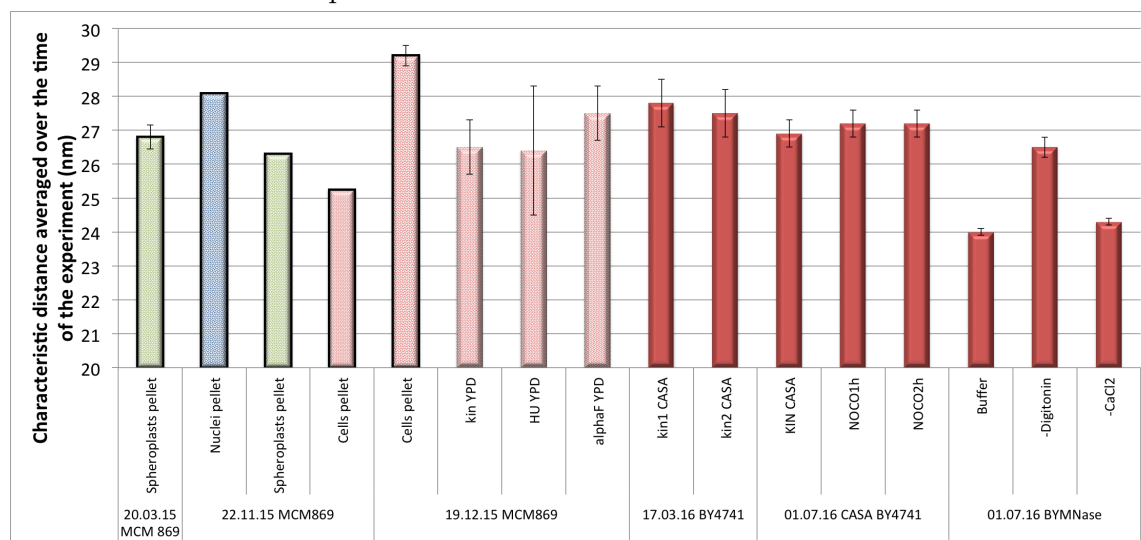
Cell curve (in red), spheroplast curve (in green) and nuclei curve (in blue) represented in Fig. 8.1.3 are shown here normalised by the amplitude at the maximum position.

FIGURE B.2 : Kratky plot of SANS data



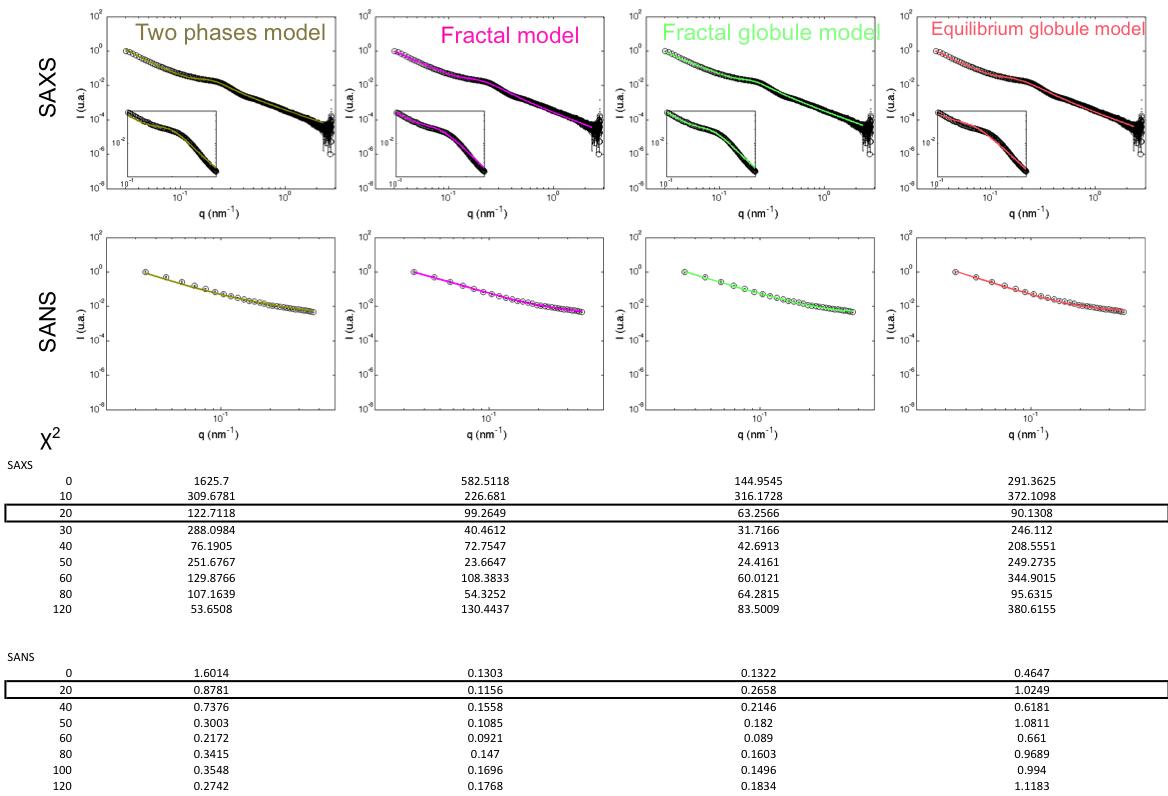
SANS data plotted in Fig. 8.2.1 B is represented in a Kratky plot.

FIGURE B.3 : Maximum position converted into characteristic distances for all SAXS experiments



The sample is precised on the x-axis and by the format of the bar : full color is BY4741 strain while dotted color is MCM869 ctrain, red appoint cells, green spheroplasts and blue nuclei. The black framing designate pellets. The value plotted is the mean over all time points of the experiment and the standard deviation of the mean.

FIGURE B.4 : SAXS and SANS curves of yeasts spheroplasts fit with the described models



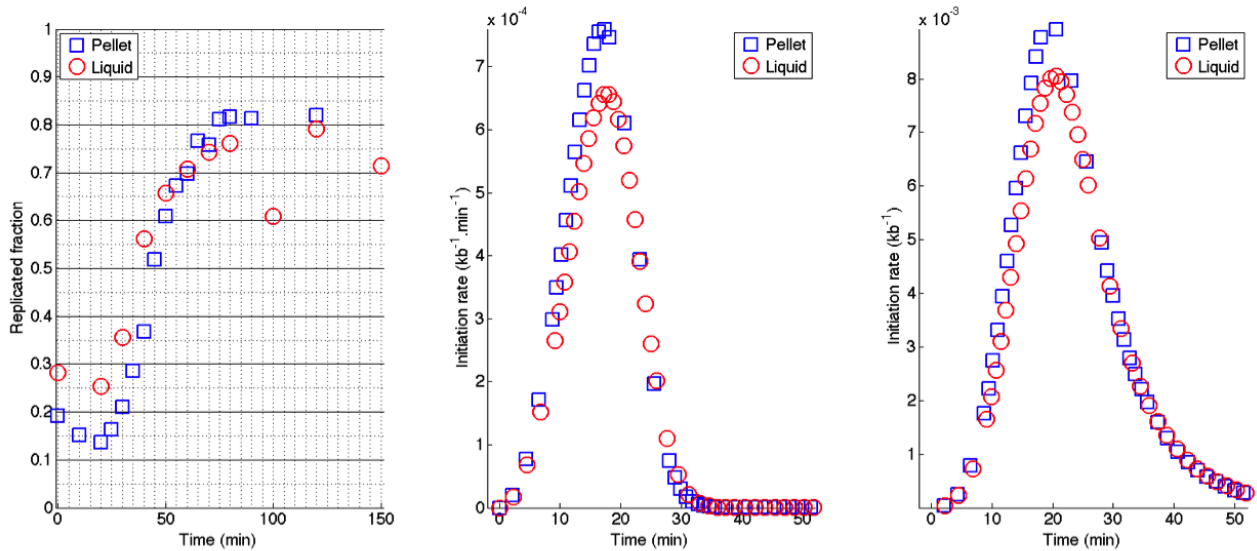
The curves plotted are from the $t = 20$ min time point. The table compiles the χ^2 values.

FIGURE B.5 : Goodness of the fit of the peak area in SAXS data

MCM869 Pellet fit range : 0.1991-0.3881nm ⁻¹ (75 points)		BY4741 Liquid fit range : 0.2130-0.3525nm ⁻¹ (56 points)				BYMNase fit range : 0.23-0.33nm ⁻¹ (39 points)							
time (min)	χ^2	time (min)	χ^2	time (min)	χ^2	Digestion buffer		-CaCl2		-Digitonin			
						time (min)	χ^2	time (min)	χ^2	time (min)	χ^2		
0	0.1159	0	1.7348	51	2.8561	102	6.3586	0	1.7221	16	0.7106	12	1.1644
10	0.1399	2	2.1871	53	2.9276	105	5.3911	30	1.1728	75	0.6677	67	1.202
20	0.1686	4	1.9978	55	3.729	107	6.3596	60	1.5415	169	0.7373	161	1.3067
25	0.2052	6	2.3197	58	3.1851	109	7.3105	105	1.4738				
30	0.2112	8	2.3041	60	4.3974	111	6.5447	150	2.0388				
35	0.1638	11	2.2273	62	3.1215	113	5.7103						
40	0.199	13	1.8632	64	4.0539	115	5.8845						
45	0.2096	15	1.8398	66	3.4608	117	6.3328						
50	0.2001	17	2.0532	69	4.3431	120	6.9472						
55	0.2422	20	2.3211	71	3.29	122	7.0753						
60	0.2472	22	2.1629	73	3.7451	124	6.3926						
65	0.2349	24	2.3082	75	3.977	126	7.818						
70	0.2205	27	2.4418	77	3.905	128	5.4675						
75	0.2688	29	2.3222	79	5.3054	130	7.7486						
80	0.267	31	2.5662	82	3.4145	132	5.6326						
90	0.2555	33	2.6459	84	5.2049	135	7.3307						
100	0.2336	36	2.1504	86	4.9424	137	8.5985						
120	0.2663	38	3.2368	88	4.3889	139	7.3856						
		41	3.3593	90	5.0478	141	6.8896						
		43	2.3619	92	4.3362	143	6.0174						
		45	2.5493	94	5.5391	145	4.7327						
		47	2.6201	96	5.5232								
		49	2.6421	98	5.61								

χ^2 value of goodness of the fit for MCM869 cells in pellet and BY4741 in liquid condition and BYMNase cells in presence of the digestion buffer or the same buffer without calcium or without digitonin.

FIGURE B.6 : F(t) of MCM869 and BY4741 cells



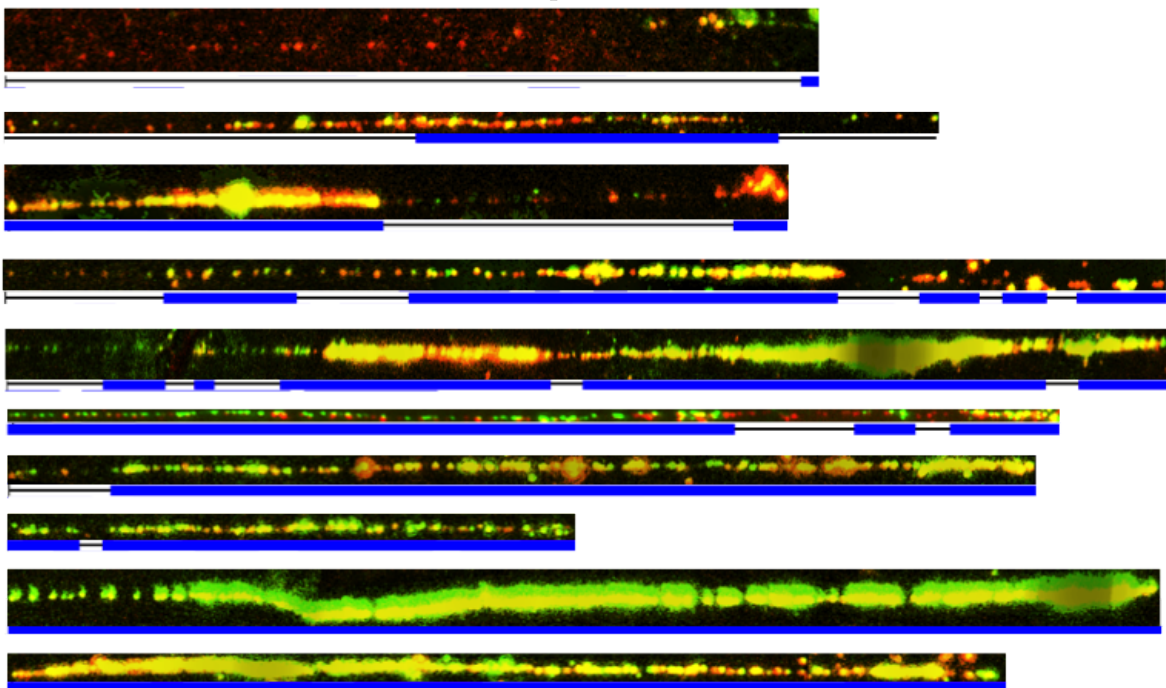
Replicated fraction measured by FACS of MCM869 cells (pellet) and BY4741 cells (Liquid) used for the SAXS experiments. Initiation rate and fork density computed from the cells in logarithmic state by the method described in the paper.

Annexe C

DNA combing - supplementary informations

Combing report	Unit	1px 1 μ m	0.0434 μ m 2.3kb
	Tresholds	false positive false negative	1 μ m 50px
		Value	sdom
Number of fibers		220	
Average size of fibers		101.77	3.72
Number of tracks		280	
Average size of tracks		19.13	1.46
Number of gaps		364	
Average gap length		13.7	0.64
Number of ETED		156	
Average ETED (kb)		75.8	3.6
Average I(t) (kb-1.min-1)		0.0057	0.0006
Average N(t) (kb-1)		0.0369	0.002

FIGURE C.1 : Example of combed fibres

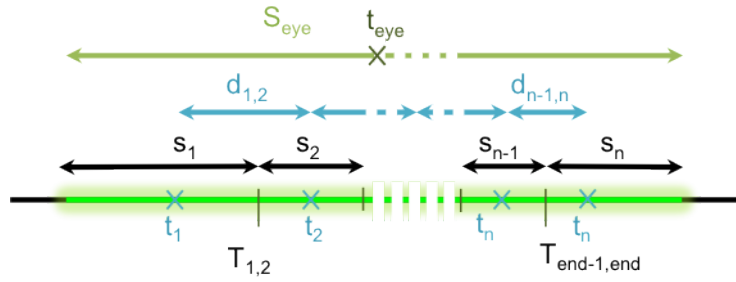


Example of several combed fibers images (red total DNA, green replicated parts, intensity and contrast were modified for visualisation) and representation after treatment (black total DNA, blue replicated parts).

Annexe D

Replication time of an eye

FIGURE D.1 : Schematic composition of an eye composed of n replicons



The size of an eye is the sum of the size of the n replicons that compose it :

$$S_{eye} = \sum_1^n s_i$$

For any middle replicon the size depend on the time of initiation of the origin and the time of fusion of the left and right forks assuming they travel at a constant speed v :

$$s_i = v \times (T_{i-1,i} - t_i) + v \times (T_{i,i+1} - t_i) = v \times (T_{i-1,i} + T_{i,i+1} - 2 \times t_i)$$

At a time t , the border replicons have the respectively the sizes :

$$s_1 = 2 \times v \times (T_{1,2} - t_1) + v \times (t - T_{1,2}) = v \times (t + T_{1,2} - t_1)$$

$$s_n = 2 \times v \times (T_{n-1,n} - t_n) + v \times (t - T_{n-1,n}) = v \times (t + T_{n-1,n} - t_n)$$

Therefore the size of eye can be expressed as a function of the initiation and fusion time of its replicons :

$$S_{eye} = 2 \times v \times \left(\sum_1^{n-1} T_{i,i+1} - \sum_1^n t_i + t \right)$$

The size of the eye at a time t is also linked to its apparent initiation time :

$$S_{eye} = 2 \times v \times (t - t_{eye})$$

Therefore the apparent replication time of the eye becomes :

$$t_{eye} = \sum_1^n t_i - \sum_1^{n-1} T_{i,i+1}$$

Moreover, the distance between two adjacent origins is :

$$d_{i,i+1} = v \times (T_{i,i+1} - t_i) + v \times (T_{i,i+1} - t_{i+1})$$

Therefore the fusion time of two adjacent replicon is :

$$T_{i,i+1} = \frac{t_i + t_{i+1}}{2} + \frac{d_{i,i+1}}{2 \times v}$$

Therefore :

$$\sum_1^{n-1} T_{i,i+1} = \frac{t_1}{2} + \sum_2^{n-1} t_i + \frac{t_n}{2} + \frac{1}{2 \times v} \sum_1^{n-1} d_{i,i+1}$$

The apparent time of initiation of an eye is the average initiation time of its extreme (1 and n) origins minus their distance divide by twice the fork speed v :

$$t_{eye} = \frac{t_1 + t_n}{2} - \frac{d_{1,n}}{2 \times v}$$

Note that if the eye is a replicon $t_1 = t_n$ and $d_{1,n} = 0$.

Bibliography

- Abdel-Monem, M., Dürwald, H., Hoffman-Berling, H., 1976. Enzymic unwinding of DNA. *European Journal of Biochemistry* 65, 441–449.
- Alabert, C., Barth, T. K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O. N., Imhof, A., Groth, A., 2015. Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes & development* 29(6), 585–590.
- Alabert, C., Groth, A., 2012. Chromatin replication and epigenome maintenance. *Nature Reviews Molecular Cell Biology* 13(3), 153–167.
- Aladjem, M. I., Falaschi, A., Kowalski, D., 2006. 2 Eukaryotic DNA Replication Origins. *Cold Spring Harbor Monograph Archive* 47, 31–61.
- Albert, B., Mathon, J., Shukla, A., Saad, H., Normand, C., Leger-Silvestre, I., Villa, D., Kamgoue, A., Mozziconacci, J., Wong, H., Zimmer, C., Bhargava, P., Bancaud, A., Gadal, O., 2013. Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. *The Journal of Cell Biology* 202(2), 201–210.
- Allis, C. D., Chicoine, L. G., Richman, R., Schulman, I. G., 1985. Deposition-related histone acetylation in micronuclei of conjugating Tetrahymena. *Proceedings of the National Academy of Sciences* 82(23), 8048–8052.
- Alvino, G. M., Collingwood, D., Murphy, J. M., Delrow, J., Brewer, B. J., Raghuraman, M. K., 2007. Replication in Hydroxyurea: It's a Matter of Time. *Molecular and Cellular Biology* 27(18), 6396–6406.
- Amitai, A., Holcman, D., 2013. Diffusing Polymers in Confined Microdomains and Estimation of Chromosomal Territory Sizes from Chromosome Capture Data. *Physical Review Letters* 110(24).
- Amitai, A., Toulouze, M., Dubrana, K., Holcman, D., 2015. Analysis of Single Locus Trajectories for Extracting In Vivo Chromatin Tethering Interactions. *PLOS Computational Biology* 11(8), e1004433.

- Annunziato, A., 2015. The Fork in the Road: Histone Partitioning During DNA Replication. *Genes* 6(2), 353–371.
- Aparicio, O. M., 2013. Location, location, location: it's all in the timing for replication origins. *Genes & Development* 27(2), 117–128.
- Aparicio, O. M., Weinstein, D. M., Bell, S. P., 1997. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* 91(1), 59–69.
- Arias, E. E., Walter, J. C., 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes & Development* 21(5), 497–518.
- Arya, G., Maitra, A., Grigoryev, S. A., 2010. A Structural Perspective on the Where, How, Why, and What of Nucleosome Positioning. *Journal of Biomolecular Structure and Dynamics* 27(6), 803–820.
- Aves, S. J., Liu, Y., Richards, T. A., 2012. Evolutionary Diversification of Eukaryotic DNA Replication Machinery. In: MacNeill, S. (Ed.), *The Eukaryotic Replisome : a Guide to Protein Structure and Function*, number 62 in *Subcellular Biochemistry*, pp. 19–35, Springer Netherlands, doi: 10.1007/978-94-007-4572-8_2.
- Avşaroğlu, B., Bronk, G., Gordon-Messer, S., Ham, J., Bressan, D. A., Haber, J. E., Kondev, J., 2014. Effect of Chromosome Tethering on Nuclear Organization in Yeast. *PLoS ONE* 9(7), e102474.
- Ay, F., Noble, W. S., 2015. Analysis methods for studying the 3d architecture of the genome. *Genome Biology* 16(1).
- Azvolinsky, A., Giresi, P. G., Lieb, J. D., Zakian, V. A., 2009. Highly Transcribed RNA Polymerase II Genes Are Impediments to Replication Fork Progression in *Saccharomyces cerevisiae*. *Molecular Cell* 34(6), 722–734.
- Backlund, M. P., Joyner, R., Weis, K., Moerner, W. E., 2014. Correlations of three-dimensional motion of chromosomal loci in yeast revealed by the double-helix point spread function microscope. *Molecular biology of the cell* 25(22), 3619–3629.
- Bancaud, A., Lavelle, C., Huet, S., Ellenberg, J., 2012. A fractal model for nuclear organization: current evidence and biological implications. *Nucleic Acids Research* 40(18), 8783–8792.
- Bar-Ziv, R., Voickek, Y., Barkai, N., 2016. Chromatin dynamics during DNA replication. *Genome Research* p. gr.201244.115.

- Barutcu, A. R., Fritz, A. J., Zaidi, S. K., van Wijnen, A. J., Lian, J. B., Stein, J. L., Nickerson, J. A., Imbalzano, A. N., Stein, G. S., 2016. C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization: CHROMOSOME CONFORMATION CAPTURE METHODS. *Journal of Cellular Physiology* 231(1), 31–35.
- Bass, H. W., Hoffman, G. G., Lee, T.-J., Wear, E. E., Joseph, S. R., Allen, G. C., Hanley-Bowdoin, L., Thompson, W. F., 2015. Defining multiple, distinct, and shared spatiotemporal patterns of DNA replication and endoreduplication from 3d image analysis of developing maize (*Zea mays* L.) root tip nuclei. *Plant Molecular Biology* .
- Baxter, J., Diffley, J. F., 2008. Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. *Molecular Cell* 30(6), 790–802.
- Bechhoefer, J., Rhind, N., 2012. Replication timing and its emergence from stochastic processes. *Trends in Genetics* 28(8), 374–381.
- Belsky, J. A., MacAlpine, H. K., Lubelsky, Y., Hartemink, A. J., MacAlpine, D. M., 2015. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes & Development* 29(2), 212–224.
- Belton, J.-M., Lajoie, B., Audibert, S., Cantaloube, S., Lassadi, I., Goiffon, I., Baù, D., Marti-Renom, M., Bystricky, K., Dekker, J., 2015. The Conformation of Yeast Chromosome III Is Mating Type Dependent and Controlled by the Recombination Enhancer. *Cell Reports* 13(9), 1855–1867.
- Bensimon, D., Simon, A. J., Croquette, V., Bensimon, A., 1995. Stretching DNA with a receding meniscus: experiments and models. *Physical review letters* 74(23), 4754.
- Berbenetz, N. M., Nislow, C., Brown, G. W., 2010. Diversity of Eukaryotic DNA Replication Origins Revealed by Genome-Wide Analysis of Chromatin Structure. *PLoS Genetics* 6(9), e1001092.
- Berezney, R., Dubey, D. D., Huberman, J. A., 2000. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* 108(8), 471–484.
- Berger, A. B., Cabal, G. G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., Olivio-Marin, J.-C., Gadal, O., Zimmer, C., 2008. High-resolution statistical mapping reveals gene territories in live yeast. *Nature Methods* 5(12), 1031–1037.

- Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., Lemaitre, J.-M., 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Structural & Molecular Biology* 19(8), 837–844.
- Bi, X., 2014. Heterochromatin structure: Lessons from the budding yeast: Structural Aspects of Heterochromatin in the Budding Yeast. *IUBMB Life* 66(10), 657–666.
- Bianchi, A., Shore, D., 2007. Early Replication of Short Telomeres in Budding Yeast. *Cell* 128(6), 1051–1062.
- Bloom, J., Cross, F. R., 2007. Multiple levels of cyclin specificity in cell-cycle control. *Nature Reviews Molecular Cell Biology* 8(2), 149–160.
- Bochman, M. L., Paeschke, K., Zakian, V. A., 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* 13(11), 770–780.
- Bochman, M. L., Schwacha, A., 2009. The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. *Microbiology and Molecular Biology Reviews* 73(4), 652–683.
- Bordas, J., Perez-Grau, L., Koch, M. H. J., Vega, M. C., Nave, C., 1986. The superstructure of chromatin and its condensation mechanism. *European Biophysics Journal* 13(3), 157–173.
- Bowers, J. L., Randell, J. C., Chen, S., Bell, S. P., 2004. ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Molecular cell* 16(6), 967–978.
- Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., Boeke, J. D., 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288c: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *YEAST-CHICHESTER* 14, 115–132.
- Brambati, A., Colosio, A., Zardoni, L., Galanti, L., Liberi, G., 2015. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Frontiers in Genetics* 6.
- Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M., Broach, J. R., 1996. Efficient transcriptional silencing in *Saccharomyces cerevisiae* requires a heterochromatin histone acetylation pattern. *Molecular and Cellular Biology* 16(8), 4349–4356.

- Bryan, A. K., Goranov, A., Amon, A., Manalis, S. R., 2010. Measurement of mass, density, and volume during the cell cycle of yeast. *Proceedings of the National Academy of Sciences* 107(3), 999–1004.
- Buckland, R. J., Watt, D. L., Chittoor, B., Nilsson, A. K., Kunkel, T. A., Chabes, A., 2014. Increased and Imbalanced dNTP Pools Symmetrically Promote Both Leading and Lagging Strand Replication Infidelity. *PLoS Genetics* 10(12), e1004846.
- Burgers, P. M. J., 2009. Polymerase Dynamics at the Eukaryotic DNA Replication Fork. *Journal of Biological Chemistry* 284(7), 4041–4045.
- Bystricky, K., 2015. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Letters* 589(20PartA), 3014–3022.
- Bystricky, K., Heun, P., Gehlen, L., Langowski, J., Gasser, S. M., 2004. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proceedings of the National Academy of Sciences of the United States of America* 101(47), 16495–16500.
- Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M., Gasser, S. M., 2005. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *The Journal of Cell Biology* 168(3), 375–387.
- Bücking-Throm, E., Duntze, W., Hartwell, L. H., Manney, T. R., 1973. Reversible arrest of haploid yeast cells at the initiation of DNA synthesis by a diffusible sex factor. *Experimental Cell Research* 76(1), 99–110.
- Cabal, G. G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., Feuerbach-Fournier, F., Olivo-Marin, J.-C., Hurt, E. C., Nehrbass, U., 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* 441(7094), 770–773.
- Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., Labib, K., 2005. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes & Development* 19, 1905–1919.
- Cao, G., Zhang, M., Miao, J., Li, W., Wang, J., Lu, D., Xia, J., 2015. Effects of X-ray and carbon ion beam irradiation on membrane permeability and integrity in *Saccharomyces cerevisiae* cells. *Journal of Radiation Research* 56(2), 294–304.
- Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., Méchali, M., 2012. New insights into replication origin characteristics in metazoans. *Cell Cycle* 11(4), 658–667.

- Cayrou, C., Coulombe, P., Vigneron, A., Stanojčić, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., Mechali, M., 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Research* 21(9), 1438–1449.
- Chabes, A., Stillman, B., 2007. Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 104(4), 1183–1188.
- Chagin, V. O., Casas-Delucchi, C. S., Reinhart, M., Schermelleh, L., Markaki, Y., Maiser, A., Bolius, J. J., Bensimon, A., Fillies, M., Domaing, P., Rozanov, Y. M., Leonhardt, H., Cardoso, M. C., 2016. 4d Visualization of replication foci in mammalian cells corresponding to individual replicons. *Nature Communications* 7, 11231.
- Champoux, J. J., 2001. DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry* 70(1), 369–413.
- Chang, F., Riera, A., Evrin, C., Sun, J., Li, H., Speck, C., Weinreich, M., 2015. Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. *eLife* 4, e05795.
- Chargaff, E., 1989. In retrospect: a commentary on ‘Studies on the structure of ribonucleic acids’ by B. Magasanik and E. Chargaff *Biochim. Biophys. Acta* 7 (1951) 396–416. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1000, 15–33.
- Chen, C., Lim, H. H., Shi, J., Tamura, S., Maeshima, K., Surana, U., Gan, L., 2016. Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Molecular Biology of the Cell* .
- Chen, C.-L., Duquenne, L., Audit, B., Guilbaud, G., Rappailles, A., Baker, A., Huvet, M., d’Aubenton Carafa, Y., Hyrien, O., Arneodo, A., Thermes, C., 2011. Replication-Associated Mutational Asymmetry in the Human Genome. *Molecular Biology and Evolution* 28(8), 2327–2337.
- Chereji, R. V., Morozov, A. V., 2014. Ubiquitous nucleosome crowding in the yeast genome. *Proceedings of the National Academy of Sciences* 111(14), 5236–5241.
- Chilkova, O., Stenlund, P., Isoz, I., Stith, C. M., Grabowski, P., Lundstrom, E.-B., Burgers, P. M., Johansson, E., 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but

- have comparable processivity in the presence of PCNA. *Nucleic Acids Research* 35(19), 6588–6597.
- Chow, M. H., Yan, K. T. H., Bennett, M. J., Wong, J. T. Y., 2010. Birefringence and DNA Condensation of Liquid Crystalline Chromosomes. *Eukaryotic Cell* 9(10), 1577–1587.
- Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J. L., Chatenay, D., Caron, F., 1996. DNA an extensible molecule. *Science (New York, N.Y.)* 271(5250), 792–794.
- Clément, C., Almouzni, G., 2015. MCM2 binding to histones H3-H4 and ASF1 supports a tetramer-to-dimer model for histone inheritance at the replication fork. *Nature Structural & Molecular Biology* 22(8), 587–589.
- Cole, H. A., Nagarajavel, V., Clark, D. J., 2012. Perfect and imperfect nucleosome positioning in yeast. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819(7), 639–643.
- Cook, J. G., Chasse, D. A. D., Nevins, J. R., 2004. The Regulated Association of Cdt1 with Minichromosome Maintenance Proteins and Cdc6 in Mammalian Cells. *Journal of Biological Chemistry* 279(10), 9625–9633.
- Cooley, C., Dave, A., Garg, M., Bianchi, A., 2014. Tel1atm dictates the replication timing of short yeast telomeres. *EMBO reports* 15(10), 1093–1101.
- Cortini, R., Barbi, M., Caré, B. R., Lavelle, C., Lesne, A., Mozziconacci, J., Victor, J.-M., 2015. The physics of epigenetics. *arXiv preprint arXiv:1509.04145* .
- Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M. R., Berger, J. M., 2011. The structural basis for MCM2–7 helicase activation by GINS and Cdc45. *Nature Structural & Molecular Biology* 18(4), 471–477.
- Costas, C., de la Paz Sanchez, M., Stroud, H., Yu, Y., Oliveros, J. C., Feng, S., Benguria, A., López-Vidriero, I., Zhang, X., Solano, R., Jacobsen, S. E., Gutierrez, C., 2011. Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nature Structural & Molecular Biology* 18(3), 395–400.
- Coster, G., Frigola, J., Beuron, F., Morris, E., Diffley, J., 2014. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. *Molecular Cell* 55(5), 666–677.
- Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R., Mozziconacci, J., 2012. Normalization of a chromosomal contact map. *BMC genomics* 13(1), 1.

- Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., Fakan, S., 2006. Chromosome territories – a functional nuclear landscape. *Current Opinion in Cell Biology* 18(3), 307–316.
- Czajkowsky, D. M., Liu, J., Hamlin, J. L., Shao, Z., 2008. DNA Combing Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI. *Journal of Molecular Biology* 375(1), 12–19.
- Das, M., Singh, S., Pradhan, S., Narayan, G., 2014. MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. *Molecular Biology International* 2014, 1–11.
- Das, S. P., Borrman, T., Liu, V. W., Yang, S. C.-H., Bechhoefer, J., Rhind, N., 2015. Replication timing is regulated by the number of MCMs loaded at origins. *Genome Research* 25(12), 1886–1892.
- Das, S. P., Rhind, N., 2016. How and why multiple MCMs are loaded at origins of DNA replication. *BioEssays* .
- Davidson, I. F., Li, A., Blow, J. J., 2006. Deregulated Replication Licensing Causes DNA Fragmentation Consistent with Head-to-Tail Fork Collision. *Molecular Cell* 24(3), 433–443.
- Davé, A., Cooley, C., Garg, M., Bianchi, A., 2014. Protein Phosphatase 1 Recruitment by Rif1 Regulates DNA Replication Origin Firing by Counteracting DDK Activity. *Cell Reports* 7(1), 53–61.
- De Carli, F., Gaggioli, V., Millot, G. A., Hyrien, O., 2015. Single-molecule, antibody-free fluorescent visualisation of replication tracts along barcoded DNA molecules. *The International Journal of Developmental Biology* .
- de Moura, A. P. S., Retkute, R., Hawkins, M., Nieduszynski, C. A., 2010. Mathematical modelling of whole chromosome replication. *Nucleic Acids Research* 38(17), 5623–5633.
- de Wit, E., de Laat, W., 2012. A decade of 3c technologies: insights into nuclear organization. *Genes & Development* 26(1), 11–24.
- Deegan, T. D., Diffley, J. F., 2016. MCM: one ring to rule them all. *Current Opinion in Structural Biology* 37, 145–151.
- Dekker, J., 2006. The three ‘C’ s of chromosome conformation capture: controls, controls, controls. *nature publishing group* 3(1), 17–21.

- Dekker, J., 2008. Mapping in Vivo Chromatin Interactions in Yeast Suggests an Extended Chromatin Fiber with Regional Variation in Compaction. *Journal of Biological Chemistry* 283(50), 34532–34540.
- Dekker, J., Marti-Renom, M. A., Mirny, L. A., 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics* 14(6), 390–403.
- Dekker, J., Rippe, K., Dekker, M., Kleckner, N., 2002. Capturing Chromosome Conformation. *Science* 295(5558), 1306–1311.
- Delbrück, M., 1954. On the replication of desoxyribonucleic acid (DNA). *Proceedings of the National Academy of Sciences* 40(9), 783–788.
- Dellino, G. I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., Pelicci, P. G., 2013. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Research* 23(1), 1–11.
- Deniz, O., 2014. Nucleosome positioning in budding yeast.
- Deniz, O., Flores, O., Aldea, M., Soler-Lopez, M., Orozco, M., 2016. Nucleosome architecture throughout the cell cycle. *Scientific Reports* 6, 19729.
- Denker, A., de Laat, W., 2016. The second decade of 3c technologies: detailed insights into nuclear organization. *Genes & Development* 30(12), 1357–1382.
- DePamphilis, M. L., de Renty, C. M., Ullah, Z., Lee, C. Y., 2012. “The Octet”: Eight Protein Kinases that Control Mammalian DNA Replication. *Frontiers in Physiology* 3.
- Descorps-Declère, S., Saguez, C., Cournac, A., Marbouty, M., Rolland, T., Ma, L., Bouchier, C., Moszer, I., Dujon, B., Koszul, R., Richard, G.-F., 2015. Genome-wide replication landscape of *Candida glabrata*. *BMC Biology* 13(1).
- Desprat, R., Thierry-Mieg, D., Lailier, N., Lajugie, J., Schildkraut, C., Thierry-Mieg, J., Bouhassira, E. E., 2009. Predictable dynamic program of timing of DNA replication in human cells. *Genome Research* 19(12), 2288–2299.
- Dewar, J. M., Budzowska, M., Walter, J. C., 2015. The mechanism of DNA replication termination in vertebrates. *Nature* 525(7569), 345–350.

- Dickerson, D., Gierliński, M., Singh, V., Kitamura, E., Ball, G., Tanaka, T. U., Owen-Hughes, T., 2016. High resolution imaging reveals heterogeneity in chromatin states between cells that is not inherited through cell division. *BMC Cell Biology* 17(1).
- Dien, B. S., Srienc, F., 1991. Bromodeoxyuridine labeling and flow cytometric identification of replicating *Saccharomyces cerevisiae* cells: lengths of cell cycle phases and population variability at specific cell cycle positions. *Biotechnology Progress* 7(4), 291–298.
- Dileep, V., Ay, F., Sima, J., Vera, D. L., Noble, W. S., Gilbert, D. M., 2015. Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. *Genome Research* 25(8), 1104–1113.
- Dimitrova, D. S., 2002. The spatio-temporal organization of DNA replication sites is identical in primary, immortalized and transformed mammalian cells. *Journal of Cell Science* 115(21), 4037–4051.
- Dimitrova, D. S., Gilbert, D. M., 1999. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Molecular Cell* 4(6), 983–993.
- Ding, Q., MacAlpine, D. M., 2011. Defining the replication program through the chromatin landscape. *Critical Reviews in Biochemistry and Molecular Biology* 46(2), 165–179.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398), 376–380.
- Donovan, S., Harwood, J., Drury, L. S., Diffley, J. F., 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proceedings of the National Academy of Sciences* 94(11), 5611–5616.
- Dorn, E. S., Cook, J. G., 2011. Nucleosomes in the neighborhood: New roles for chromatin modifications in replication origin control. *Epigenetics* 6(5), 552–559.
- Drissi, R., Dubois, M.-L., Douziech, M., Boisvert, F.-M., 2015. Quantitative Proteomics Reveals Dynamic Interactions of the Minichromosome Maintenance Complex (MCM) in the Cellular Response to Etoposide Induced DNA Damage. *Molecular & Cellular Proteomics* 14(7), 2002–2013.

- Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., Noble, W. S., 2010. A three-dimensional model of the yeast genome. *Nature* 465(7296), 363–367.
- Duzdevich, D., Warner, M., Ticau, S., Ivica, N., Bell, S., Greene, E., 2015. The Dynamics of Eukaryotic Replication Initiation: Origin Specificity, Licensing, and Firing at the Single-Molecule Level. *Molecular Cell* 58(3), 483–494.
- Eaton, M. L., Galani, K., Kang, S., Bell, S. P., MacAlpine, D. M., 2010. Conserved nucleosome positioning defines replication origins. *Genes & Development* 24(8), 748–753.
- Ebrahimi, H., Robertson, E. D., Taddei, A., Gasser, S. M., Donaldson, A. D., Hiraga, S.-i., 2010. Early initiation of a replication origin tethered at the nuclear periphery. *Journal of Cell Science* 123(7), 1015–1019.
- Edwards, M. C., 2002. MCM2-7 Complexes Bind Chromatin in a Distributed Pattern Surrounding the Origin Recognition Complex in *Xenopus* Egg Extracts. *Journal of Biological Chemistry* 277(36), 33049–33057.
- Eltsov, M., MacLellan, K. M., Maeshima, K., Frangakis, A. S., Dubochet, J., 2008. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences* 105(50), 19732–19737.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shores, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., Bernstein, B. E., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345), 43–49.
- Errico, A., Costanzo, V., 2010. Differences in the DNA replication of unicellular eukaryotes and metazoans: known unknowns. *EMBO reports* 11(4), 270–278.
- Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., Speck, C., 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences* 106(48), 20240–20245.
- Fangman, W. L., Brewer, B. J., 1992. A question of time: replication origins of eukaryotic chromosomes. *Cell* 71(3), 363–366.
- Fanning, E., 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Research* 34(15), 4126–4137.

- Feng, W., Bachant, J., Collingwood, D., Raghuraman, M. K., Brewer, B. J., 2009. Centromere Replication Timing Determines Different Forms of Genomic Instability in *Saccharomyces cerevisiae* Checkpoint Mutants During Replication Stress. *Genetics* 183(4), 1249–1260.
- Ferguson, B. M., Brewer, B. J., Reynolds, A. E., Fangman, W. L., 1991. A yeast origin of replication is activated late in S phase. *Cell* 65(3), 507–515.
- Filion, G. J., van Bommel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., van Steensel, B., 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143(2), 212–224.
- Finch, J. T., Klug, A., 1976. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences* 73(6), 1897–1901.
- Forsburg, S. L., 2004. Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiology and Molecular Biology Reviews* 68(1), 109–131.
- Forterre, P., Filée, J., Myllykallio, H., 2013. Origin and Evolution of DNA and DNA Replication Machineries. Landes Bioscience.
- Foti, R., Gnan, S., Cornacchia, D., Dileep, V., Bulut-Karslioglu, A., Diehl, S., Buness, A., Klein, F., Huber, W., Johnstone, E., Loos, R., Bertone, P., Gilbert, D., Manke, T., Jenuwein, T., Bonomo, S., 2016. Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Molecular Cell* 61(2), 260–273.
- Foulik, M. S., Urban, J. M., Casella, C., Gerbi, S. A., 2015. Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Research* 25(5), 725–735.
- Fraser, H. B., 2013. Cell-cycle regulated transcription associates with DNA replication timing in yeast and human. *Genome biology* 14(10), 1.
- Friedman, K. L., Brewer, B. J., Fangman, W. L., 1997. Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes to Cells* 2(11), 667–678.
- Fudenberg, G., Belton, J.-M., Goloborodko, A., Imakaev, M., Dekker, J., Mirny, L., 2013. Polymer models of yeast *S. cerevisiae* genome organization. *Epigenetics & Chromatin* 6(Suppl 1), P128.

- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., Mirny, L., 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* 15(9), 2038–2049.
- Fudenberg, G., Mirny, L. A., 2012. Higher-order chromatin structure: bridging physics and biology. *Current Opinion in Genetics & Development* 22(2), 115–124.
- Fussner, E., Ching, R. W., Bazett-Jones, D. P., 2011. Living without 30nm chromatin fibers. *Trends in Biochemical Sciences* 36(1), 1–6.
- Gauthier, M. G., Herrick, J., Bechhoefer, J., 2010. Defects and DNA Replication. *Physical Review Letters* 104(21).
- Gauthier, M. G., Norio, P., Bechhoefer, J., 2012. Modeling inhomogeneous DNA replication kinetics. *PLoS one* 7(3), e32053.
- Ge, X. Q., Jackson, D. A., Blow, J. J., 2007. Dormant origins licensed by excess Mcm2 7 are required for human cells to survive replicative stress. *Genes & Development* 21(24), 3331–3341.
- Gehlen, L. R., Gruenert, G., Jones, M. B., Rodley, C. D., Langowski, J., O’Sullivan, J., 2012. Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus* 3(4), 370–383.
- Gibson, D. G., Bell, S. P., Aparicio, O. M., 2006. Cell cycle execution point analysis of ORC function and characterization of the checkpoint response to ORC inactivation in *Saccharomyces cerevisiae*. *Genes to Cells* 11(6), 557–573.
- Giewekemeyer, K., Hackenberg, C., Aquila, A., Wilke, R., Groves, M., Jordanova, R., Lamzin, V., Borchers, G., Saksl, K., Zozulya, A., Sprung, M., Mancuso, A., 2015. Tomography of a Cryo-immobilized Yeast Cell Using Ptychographic Coherent X-Ray Diffractive Imaging. *Biophysical Journal* 109(9), 1986–1995.
- Gilbert, D. M., 2004. In search of the holy replicator. *Nature Reviews Molecular Cell Biology* 5(10), 848–855.
- Gillespie, P. J., Blow, J., 2010. Clusters, factories and domains: The complex structure of S phase comes into focus. *Cell Cycle* 9(16), 3238–3246.
- Gillespie, P. J., Gambus, A., Blow, J. J., 2012. Preparation and use of *Xenopus* egg extracts to study DNA replication and chromatin associated proteins. *Methods* 57(2), 203–213.

- Gindin, Y., Valenzuela, M. S., Aladjem, M. I., Meltzer, P. S., Bilke, S., 2014. A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Molecular Systems Biology* 10(3), 722–722.
- Goldar, A., Arneodo, A., Audit, B., Argoul, F., Rappailles, A., Guilbaud, G., Petryk, N., Kahli, M., Hyrien, O., 2016. Deciphering DNA replication dynamics in eukaryotic cell populations in relation with their averaged chromatin conformations. *Scientific Reports* 6, 22469.
- Goldar, A., Marsolier-Kergoat, M.-C., Hyrien, O., 2009. Universal Temporal Profile of Replication Origin Activation in Eukaryotes. *PLoS ONE* 4(6), e5899.
- Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A., Nag, A., 1984. Replication timing of genes and middle repetitive sequences. *Science (New York, N.Y.)* 224(4650), 686–692.
- Goloborodko, A., Imakaev, M. V., Marko, J. F., Mirny, L., 2016a. Compaction and segregation of sister chromatids via active loop extrusion. *eLife* 5, e14864.
- Goloborodko, A., Marko, J., Mirny, L., 2016b. Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal* 110(10), 2162–2168.
- Gong, K., Tjong, H., Zhou, X. J., Alber, F., 2015. Comparative 3d Genome Structure Analysis of the Fission and the Budding Yeast. *PLOS ONE* 10(3), e0119672.
- Gosling, R. G., Franklin, R. E., 1953. Molecular configuration of sodium thymonucleate. *Nature* 171(4356), 740–741.
- Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., Gasser, S. M., 1996. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *The Journal of cell biology* 134(6), 1349–1363.
- Goulian, M., Richards, S. H., Heard, C. J., Bigsby, B. M., 1990. Discontinuous DNA synthesis by purified mammalian proteins. *Journal of Biological Chemistry* 265(30), 18461–18471.
- Gray, J. V., Petsko, G. A., Johnston, G. C., Ringe, D., Singer, R. A., Werner-Washburne, M., 2004. "Sleeping Beauty": Quiescence in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* 68(2), 187–206.
- Grigoryev, S. A., 2012. Nucleosome spacing and chromatin higher-order folding. *Nucleus* 3(6), 493–499.

- Grigoryev, S. A., Bascom, G., Buckwalter, J. M., Schubert, M. B., Woodcock, C. L., Schlick, T., 2016. Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proceedings of the National Academy of Sciences* p. 201518280.
- Gros, J., Devbhandari, S., Remus, D., 2014. Origin plasticity during budding yeast DNA replication in vitro. *The EMBO Journal* 33(6), 621–636.
- Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., Remus, D., 2015. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Molecular Cell* 60(5), 797–807.
- Grosberg, A., 2016. Extruding Loops to Make Loopy Globules? *Biophysical Journal* 110(10), 2133–2135.
- Grosberg, A., Rabin, Y., Havlin, S., Neer, A., 1993. Crumpled globule model of the three-dimensional structure of DNA. *EPL (Europhysics Letters)* 23(5), 373.
- Guidi, M., Ruault, M., Marbouty, M., Loïdice, I., Cournac, A., Billaudeau, C., Hocher, A., Mozziconacci, J., Koszul, R., Taddei, A., 2015. Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biology* 16(1).
- Guilbaud, G., Rappailles, A., Baker, A., Chen, C.-L., Arneodo, A., Goldar, A., d'Aubenton Carafa, Y., Thermes, C., Audit, B., Hyrien, O., 2011. Evidence for Sequential and Increasing Activation of Replication Origins along Replication Timing Gradients in the Human Genome. *PLoS Computational Biology* 7(12), e1002322.
- Gürsoy, G., Xu, Y., Liang, J., 2014. Computational predictions of structures of multichromosomes of budding yeast. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3945–3948, IEEE.
- Haber, J. E., Leung, W.-Y., 1996. Lack of chromosome territoriality in yeast: promiscuous rejoining of broken chromosome ends. *Proceedings of the National Academy of Sciences* 93(24), 13949–13954.
- Hahn, S., Kim, D., 2013. Physical Origin of the Contact Frequency in Chromosome Conformation Capture Data. *Biophysical Journal* 105(8), 1786–1795.
- Hahn, S., Kim, D., 2015. Identifying and Reducing Systematic Errors in Chromosome Conformation Capture Data. *PLOS ONE* 10(12), e0146007.

- Hajjoul, H., Mathon, J., Ranchon, H., Goiffon, I., Mozziconacci, J., Albert, B., Carrivain, P., Victor, J.-M., Gadai, O., Bystricky, K., Bancaud, A., 2013. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Research* 23(11), 1829–1838.
- Hamlin, J., Mesner, L., Lar, O., Torres, R., Chodaparambil, S., Wang, L., 2008. A revisionist replicon model for higher eukaryotic genomes. *Journal of Cellular Biochemistry* 105(2), 321–329.
- Hammouda, B., 2010. The SANS toolbox.
- Hansen, J. C., 2012. Human mitotic chromosome structure: what happened to the 30-nm fibre? *The EMBO journal* 31(7), 1621–1623.
- Harari, Y., Kupiec, M., 2014. Genome-wide studies of telomere biology in budding yeast. *Microbial Cell* 1(3), 70–80.
- Hatton, K. S., Dhar, V., Brown, E. H., Iqbal, M. A., Stuart, S., Didamo, V. T., Schildkraut, C. L., 1988. Replication program of active and inactive multigene families in mammalian cells. *Molecular and cellular biology* 8(5), 2149–2158.
- Hawkins, M., Retkute, R., Müller, C. A., Saner, N., Tanaka, T. U., de Moura, A. P. S., Nieduszynski, C. A., 2013. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Reports* 5(4), 1132–1141.
- Haye, J. E., Gammie, A. E., 2015. The Eukaryotic Mismatch Recognition Complexes Track with the Replisome during DNA Synthesis. *PLoS Genet* 11(12), e1005719.
- Hediger, F., Neumann, F. R., Van Houwe, G., Dubrana, K., Gasser, S. M., 2002. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Current biology* 12(24), 2076–2089.
- Heinzel, S. S., Krysan, P. J., Tran, C. T., Calos, M. P., 1991. Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Molecular and cellular biology* 11(4), 2263–2272.
- Herrick, J., Jun, S., Bechhoefer, J., Bensimon, A., 2002. Kinetic Model of DNA Replication in Eukaryotic Organisms. *Journal of Molecular Biology* 320(4), 741–750.
- Herrick, J., Stanislawski, P., Hyrien, O., Bensimon, A., 2000. Replication fork density increases during DNA synthesis in *X. laevis* egg extracts. *Journal of Molecular Biology* 300(5), 1133–1142.

- Heun, P., Laroche, T., Raghuraman, M. K., Gasser, S. M., 2001. The positioning and dynamics of origins of replication in the budding yeast nucleus. *The Journal of cell biology* 152(2), 385–400.
- Hewish, D. R., Burgoyne, L. A., 1973. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochemical and biophysical research communications* 52(2), 504–510.
- Higgins, J. S., Benoît, H., 1994. *Polymers and neutron scattering*. Clarendon Press, google-Books-ID: tQ0vAQAAIAAJ.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T. M., Schübeler, D., Gilbert, D. M., 2008. Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. *PLoS Biology* 6(10), e245.
- Hizume, K., Yagura, M., Araki, H., 2013. Concerted interaction between origin recognition complex (ORC), nucleosomes and replication origin DNA ensures stable ORC-origin binding. *Genes to Cells* 18(9), 764–779.
- Hoggard, T., Shor, E., Müller, C. A., Nieduszynski, C. A., Fox, C. A., 2013. A Link between ORC-Origin Binding Mechanisms and Origin Activation Time Revealed in Budding Yeast. *PLoS Genetics* 9(9), e1003798.
- Hsieh, T.-H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., Rando, O., 2015. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162(1), 108–119.
- Huang, H., Strømme, C. B., Saredi, G., Hödl, M., Strandsby, A., González-Aguilera, C., Chen, S., Groth, A., Patel, D. J., 2015. A unique binding mode enables MCM2 to chaperone histones H3–H4 at replication forks. *Nature Structural & Molecular Biology* 22(8), 618–626.
- Huet, S., Lavelle, C., Ranchon, H., Carrivain, P., Victor, J.-M., Bancaud, A., 2014. Relevance and Limitations of Crowding, Fractal, and Polymer Models to Describe Nuclear Architecture. In: *International Review of Cell and Molecular Biology*, volume 307, pp. 443–479, Elsevier.
- Hutchison, C. J., Bridger, J. M., Cox, L. S., Kill, I. R., 1994. Weaving a pattern from disparate threads: lamin function in nuclear assembly and DNA replication. *Journal of cell science* 107(12), 3259–3269.
- Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton Carafa, Y., Arneodo, A., Thermes, C., 2007. Human gene organization driven by the coordination of replication and transcription. *Genome Research* 17(9), 1278–1285.

- Hyrien, O., 2015. Peaks cloaked in the mist: The landscape of mammalian replication origins. *The Journal of Cell Biology* 208(2), 147–160.
- Hyrien, O., 2016. How MCM loading and spreading specify eukaryotic DNA replication initiation sites. *F1000Research* 5.
- Hyrien, O., Goldar, A., 2010. Mathematical modelling of eukaryotic DNA replication. *Chromosome Research* 18(1), 147–161.
- Hyrien, O., Marheineke, K., Goldar, A., 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays* 25(2), 116–125.
- Hyrien, O., Rappailles, A., Guilbaud, G., Baker, A., Chen, C.-L., Goldar, A., Petryk, N., Kahli, M., Ma, E., d'Aubenton Carafa, Y., Audit, B., Thermes, C., Arneodo, A., 2013. From Simple Bacterial and Archaeal Replicons to Replication N/U-Domains. *Journal of Molecular Biology* 425(23), 4673–4689.
- Ibarra, A., Schwob, E., Méndez, J., 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proceedings of the National Academy of Sciences* 105(26), 8956–8961.
- Ibel, K., Klingholz, R., Strätling, W. H., Bogenberger, J., Fittler, F., 1983. Neutron diffraction of chromatin in interphase nuclei and metaphase chromosomes. *European Journal of Biochemistry* 133(2), 315–319.
- Imakaev, M. V., Fudenberg, G., Mirny, L. A., 2015. Modeling chromosomes: Beyond pretty pictures. *FEBS Letters* 589(20PartA), 3031–3036.
- Indiani, C., O'Donnell, M., 2006. The replication clamp-loading machine at work in the three domains of life. *Nature Reviews Molecular Cell Biology* 7(10), 751–761.
- Iyer, B. V., Kenward, M., Arya, G., 2011. Hierarchies in eukaryotic genome organization: insights from polymer theory and simulations. *BMC biophysics* 4(1), 1.
- Jacobs, C. W., Adams, A. E., Szaniszló, P. J., Pringle, J. R., 1988. Functions of microtubules in the *Saccharomyces cerevisiae* cell cycle. *The Journal of Cell Biology* 107(4), 1409–1426.
- Jacobs, C. W., Szaniszló, P. J., 1982. Microtubule function and its relation to cellular development and the yeast cell cycle in *Wangiella dermatitidis*. *Archives of Microbiology* 133(2), 155–161.

- Jacques, D. A., Trehwella, J., 2010. Small-angle scattering for structural biology- Expanding the frontier while avoiding the pitfalls. *Protein Science* 19(4), 642–657.
- Jansen, A., Verstrepen, K. J., 2011. Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* 75(2), 301–320.
- Jaspersen, S. L., Winey, M., 2004. THE BUDDING YEAST SPINDLE POLE BODY: Structure, Duplication, and Function. *Annual Review of Cell and Developmental Biology* 20(1), 1–28.
- Jeon, Y., Bekiranov, S., Karnani, N., Kapranov, P., Ghosh, S., MacAlpine, D., Lee, C., Hwang, D. S., Gingeras, T. R., Dutta, A., 2005. Temporal profile of replication of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 102(18), 6419–6424.
- Jin, Q.-W., Fuchs, J., Loidl, J., 2000. Centromere clustering is a major determinant of yeast interphase nuclear organization. *Journal of cell science* 113(11), 1903–1912.
- Jin, Q.-w., Trelles-Sticken, E., Scherthan, H., Loidl, J., 1998. Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *The Journal of cell biology* 141(1), 21–29.
- Johnson, A., O'Donnell, M., 2005. CELLULAR DNA REPLICASES: Components and Dynamics at the Replication Fork. *Annual Review of Biochemistry* 74(1), 283–315.
- Joti, Y., Hikima, T., Nishino, Y., Kamada, F., Hihara, S., Takata, H., Ishikawa, T., Maeshima, K., 2012. Chromosomes without a 30-nm chromatin fiber. *Nucleus* 3(5), 404–410.
- Ju, Q., Warner, J. R., 1994. Ribosome synthesis during the growth cycle of *Saccharomyces cerevisiae*. *Yeast (Chichester, England)* 10(2), 151–157.
- Jun, S., Bechhoefer, J., 2005. Nucleation and growth in one dimension, part II: Application to DNA replication kinetics. *Physical Review E* 71.
- Jun, S., Zhang, H., Bechhoefer, J., 2005. Nucleation and growth in one dimension. I. The generalized Kolmogorov-Johnson-Mehl-Avrami model. *Physical Review E* 71(1).
- Kadauke, S., Blobel, G. A., 2009. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1789(1), 17–25.

- Kalverda, B., Röling, M. D., Fornerod, M., 2008. Chromatin organization in relation to the nuclear periphery. *FEBS Letters* 582(14), 2017–2022.
- Kamimura, Y., Tak, Y.-S., Sugino, A., Araki, H., 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *The EMBO journal* 20(8), 2097–2107.
- Kanemaki, M., Labib, K., 2006. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *The EMBO Journal* 25(8), 1753–1763.
- Kanoh, Y., Matsumoto, S., Fukatsu, R., Kakusho, N., Kono, N., Renard-Guillet, C., Masuda, K., Iida, K., Nagasawa, K., Shirahige, K., Masai, H., 2015. Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nature Structural & Molecular Biology* .
- Kara, N., Hossain, M., Prasanth, S. G., Stillman, B., 2015. Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. *Journal of Biological Chemistry* 290(19), 12355–12369.
- Kawakami, H., Ohashi, E., Kanamoto, S., Tsurimoto, T., Katayama, T., 2015. Specific binding of eukaryotic ORC to DNA replication origins depends on highly conserved basic residues. *Scientific Reports* 5, 14929.
- Kaykov, A., Taillefumier, T., Bensimon, A., Nurse, P., 2016. Molecular Combing of Single DNA Molecules on the 10 Megabase Scale. *Scientific Reports* 6, 19636.
- Kilburn, D., Roh, J. H., Guo, L., Briber, R. M., Woodson, S. A., 2010. Molecular Crowding Stabilizes Folded RNA Structure by the Excluded Volume Effect. *Journal of the American Chemical Society* 132(25), 8690–8696.
- Kimura, H., Cook, P. R., 2001. Kinetics of core histones in living human cells little exchange of H3 and H4 and some rapid exchange of H2b. *The Journal of cell biology* 153(7), 1341–1354.
- Kirkegaard, K., Wang, J. C., 1985. Bacterial Topoisomerase I can relax positively supercoiled DNA containing single-strand loop. *Journal of Molecular Biology* 185, 625–637.
- Kiseleva, E., Allen, T. D., Rutherford, S. A., Murray, S., Morozova, K., Gardiner, F., Goldberg, M. W., Drummond, S. P., 2007. A protocol for isolation and visualization of yeast nuclei by scanning electron microscopy (SEM). *Nature Protocols* 2(8), 1943–1953.

- Kitamura, E., Blow, J. J., Tanaka, T. U., 2006. Live-Cell Imaging Reveals Replication of Individual Replicons in Eukaryotic Replication Factories. *Cell* 125(7), 1297–1308.
- Kjems, J., Freltoft, T., 1985. Fractal structures studied by small-angle X-ray and neutron scattering. In: *Festkörperprobleme* 25, pp. 669–675, Springer.
- Knott, S., Peace, J., Ostrow, A., Gan, Y., Rex, A., Viggiani, C., Tavaré, S., Aparicio, O., 2012. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in *S. cerevisiae*. *Cell* 148(1-2), 99–111.
- Knott, S. R., Viggiani, C. J., Tavaré, S., Aparicio, O. M., 2009. Genome-wide replication profiles indicate an expansive role for Rpd3l in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. *Genes & Development* 23(9), 1077–1090.
- Koc, A., Wheeler, L. J., Mathews, C. K., Merrill, G. F., 2004. Hydroxyurea Arrests DNA Replication by a Mechanism That Preserves Basal dNTP Pools. *Journal of Biological Chemistry* 279(1), 223–230.
- Koren, A., Soifer, I., Barkai, N., 2010. MRC1-dependent scaling of the budding yeast DNA replication timing program. *Genome Research* 20(6), 781–790.
- Kornberg, R. D., 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* 184(4139), 868–871.
- Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell* 128(4), 693–705.
- Kumar, D., Viberg, J., Nilsson, A. K., Chabes, A., 2010. Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint. *Nucleic Acids Research* 38(12), 3975–3983.
- Kurth, I., O'Donnell, M., 2013. New insights into replisome fluidity during chromosome replication. *Trends in Biochemical Sciences* 38(4), 195–203.
- Kylie, K., Romero, J., Lindamulage, I. K., Knockleby, J., Lee, H., 2016. Dynamic regulation of histone H3k9 is linked to the switch between replication and transcription at the *Dbf4* origin-promoter locus. *Cell Cycle* pp. 00–00.
- Köhler, C., Koalick, D., Fabricius, A., Parplys, A. C., Borgmann, K., Pospiech, H., Grosse, F., 2016. *Cdc45* is limiting for replication initiation in humans. *Cell Cycle* 15(7), 974–985.

- Labib, K., Hodgson, B., 2007. Replication fork barriers: pausing for a break or stalling for time? *EMBO reports* 8(4), 346–353.
- Lande-Diner, L., Zhang, J., Cedar, H., 2009. Shifts in Replication Timing Actively Affect Histone Acetylation during Nucleosome Reassembly. *Molecular Cell* 34(6), 767–774.
- Langmore, J. P., others, 1983. Low angle x-ray diffraction studies of chromatin structure in vivo and in isolated nuclei and metaphase chromosomes. *The Journal of cell biology* 96(4), 1120–1131.
- Langmore, J. P., Schutt, C., 1980. The higher order structure of chicken erythrocyte chromosomes in vivo. *Nature* 288(5791), 620–622.
- Laporte, D., Courtout, F., Salin, B., Ceschin, J., Sagot, I., 2013. An array of nuclear microtubules reorganizes the budding yeast nucleus during quiescence. *The Journal of Cell Biology* 203(4), 585–594.
- Lassadi, I., Kamgoué, A., Goiffon, I., Tanguy-le Gac, N., Bystricky, K., 2015. Differential chromosome conformations as hallmarks of cellular identity revealed by mathematical polymer modeling. *PLoS Comput Biol* 11(6), e1004306.
- Lebedev, D., Filatov, M., Kuklin, A., Islamov, A., Kentzinger, E., Pantina, R., Toperverg, B., Isaev-Ivanov, V., 2005. Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties. *FEBS Letters* 579(6), 1465–1468.
- Lebofsky, R., Takahashi, T., Walter, J., 2009. DNA Replication in Nucleus-Free *Xenopus* Egg Extracts. In: Vengrova, S., Dalgaard, J. Z. (Eds.), *DNA Replication*, number 521 in *Methods in Molecular Biology*, pp. 229–252, Humana Press, doi: 10.1007/978-1-60327-815-7_13.
- Lee, J., Chastain, P. D., Kusakabe, T., Griffith, J. D., Richardson, C. C., 1998. Coordinated leading and lagging strand DNA synthesis on a minicircular template. *Molecular cell* 1(7), 1001–1010.
- Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., Nislow, C., 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics* 39(10), 1235–1244.
- Leipe, D. D., Aravind, L., Koonin, E. V., 1999. Did DNA replication evolve twice independantly ? *Nucleic Acids Research* 27(17), 3389–3401.
- Leman, A. R., Noguchi, E., 2012. Local and global functions of Timeless and Tipin in replication fork protection. *Cell Cycle* 11(21), 3945–3955.

- Lengronne, A., Pasero, P., Bensimon, A., Schwob, E., 2001. Monitoring S phase progression globally and locally using BrdU incorporation in TK⁺ yeast strains. *Nucleic Acids Research* 29(7), 1433–1442.
- Lesne, A., Riposo, J., Roger, P., Cournac, A., Mozziconacci, J., 2014. 3d genome reconstruction from chromosomal contacts. *Nature Methods* 11(11), 1141–1143.
- León-Ortiz, A. M., Svendsen, J., Boulton, S. J., 2014. Metabolism of DNA secondary structures at the eukaryotic replication fork. *DNA Repair* 19, 152–162.
- Li, G., Reinberg, D., 2011. Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development* 21(2), 175–186.
- Li, X., Feng, H., Zhang, J., Sun, L., Zhu, P., 2015. Analysis of chromatin fibers in HeLa cells with electron tomography. *Biophysics Reports* 1(1), 51–60.
- Li, Z., Pearlman, A. H., Hsieh, P., 2016. DNA mismatch repair and the DNA damage response. *DNA Repair* 38, 94–101.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326(5950), 289–293.
- Ligasová, A., Strunin, D., Liboska, R., Rosenberg, I., Koberna, K., 2012. Atomic Scissors A New Method of Tracking the 5 Bromo 2 Deoxyuridine Labeled DNA In Situ. *PLoS ONE* 7(12), e52584.
- Lin, Y.-L., Pasero, P., 2012. Interference between DNA replication and transcription as a cause of genomic instability. *Current genomics* 13(1), 65–73.
- Lipford, J. R., Bell, S. P., 2001. Nucleosome positioned by ORC facilitate the initiation of DNA replication. *Molecular Cell* 7, 21–30.
- Liu, J., McConnell, K., Dixon, M., Calvi, B. R., 2012. Analysis of model replication origins in *Drosophila* reveals new aspects of the chromatin landscape and its relationship to origin activity and the prereplicative complex. *Molecular biology of the cell* 23(1), 200–212.
- Livolant, F., Mangenot, S., Leforestier, A., Bertin, A., Frutos, M. d., Raspaud, E., Durand, D., 2006. Are liquid crystalline properties of nucleosomes involved in chromosome structure and dynamics? *Philosophical Transactions of the Royal*

- Society A: Mathematical, Physical and Engineering Sciences 364(1847), 2615–2633.
- Loeb, L. A., Monnat, R. J., 2008. DNA polymerases and human disease. *Nature Reviews Genetics* 9(8), 594–604.
- Lohman, T. M., 1993. Helicase-catalyzed DNA unwinding. *Journal of Biological Chemistry* 268(4), 2269–2272.
- Lowary, P. T., Widom, J., 1989. Higher-order structure of *Saccharomyces cerevisiae* chromatin. *Proceedings of the National Academy of Sciences* 86(21), 8266–8270.
- Loyola, A., Bonaldi, T., Roche, D., Imhof, A., Almouzni, G., 2006. PTMs on H3 Variants before Chromatin Assembly Potentiate Their Final Epigenetic State. *Molecular Cell* 24(2), 309–316.
- Lubelsky, Y., Prinz, J. A., DeNapoli, L., Li, Y., Belsky, J. A., MacAlpine, D. M., 2014. DNA replication and transcription programs respond to the same chromatin cues. *Genome Research* 24(7), 1102–1114.
- Lucchini, R., Wellinger, R. E., Sogo, J. M., 2001. Nucleosome positioning at the replication fork. *The EMBO Journal* 20(24), 7294–7302.
- Luger, K., Dechassa, M. L., Tremethick, D. J., 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology* 13(7), 436–447.
- Luger, K., Richmond, T. J., 1998. DNA binding within the nucleosome core. *Current Opinion in Structural Biology* 8, 33–40.
- Lygeros, J., Koutroumpas, K., Dimopoulos, S., Legouras, I., Kouretas, P., Heichinger, C., Nurse, P., Lygerou, Z., 2008. Stochastic hybrid modeling of DNA replication across a complete genome. *Proceedings of the National Academy of Sciences of the United States of America* 105(34), 12295–12300.
- Löb, D., Lengert, N., Chagin, V. O., Reinhart, M., Casas-Delucchi, C. S., Cardoso, M. C., Drossel, B., 2016. 3d replicon distributions arise from stochastic initiation and domino-like DNA replication progression. *Nature Communications* 7, 11207.
- Ma, E., Hyrien, O., Goldar, A., 2012. Do replication forks control late origin firing in *Saccharomyces cerevisiae*? *Nucleic Acids Research* 40(5), 2010–2019.
- Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P.-c., Meng, C., Berezney, R., 1998. Spatial and temporal dynamics of DNA replication sites in mammalian cells. *The Journal of cell biology* 143(6), 1415–1425.

- MacAlpine, D. M., Almouzni, G., 2013. Chromatin and DNA Replication. *Cold Spring Harbor Perspectives in Biology* 5(8), a010207–a010207.
- Maeshima, K., Hihara, S., Eltsov, M., 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology* 22(3), 291–297.
- Maeshima, K., Imai, R., Hikima, T., Joti, Y., 2014. Chromatin structure revealed by X-ray scattering analysis and computational modeling. *Methods* 70(2-3), 154–161.
- Mangenot, S., Leforestier, A., Durand, D., Livolant, F., 2003. Phase Diagram of Nucleosome Core Particles. *Journal of Molecular Biology* 333(5), 907–916.
- Mantiero, D., Mackenzie, A., Donaldson, A., Zegerman, P., 2011. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO journal* 30(23), 4805–4814.
- Margueron, R., Reinberg, D., 2010. Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics* 11(4), 285–296.
- Maric, C., Prioleau, M.-N., 2010. Interplay between DNA replication and gene expression: a harmonious coexistence. *Current Opinion in Cell Biology* 22(3), 277–283.
- Marko, J. F., Poirier, M. G., 2003. Micromechanics of chromatin and chromosomes. *Biochemistry and Cell Biology* 81(3), 209–220.
- Mattarocci, S., Shyian, M., Lemmens, L., Damay, P., Altintas, D., Shi, T., Bartholomew, C., Thomä, N. H., Hardy, C., Shore, D., 2014. Rif1 Controls DNA Replication Timing in Yeast through the PP1 Phosphatase Glc7. *Cell Reports* 7(1), 62–69.
- Maundrell, K., Hutchison, A., Shall, S., 1988. Sequence analysis of ARS elements in fission yeast. *The EMBO journal* 7(7), 2203.
- Maya-Mendoza, A., Olivares-Chauvet, P., Shaw, A., Jackson, D. A., 2010. S Phase Progression in Human Cells Is Dictated by the Genetic Continuity of DNA Foci. *PLoS Genetics* 6(4), e1000900.
- McCarroll, R. M., Fangman, W. L., 1988. Time of replication of yeast centromeres and telomeres. *Cell* 54(4), 505–513.
- McCulloch, S. D., Kunkel, T. A., 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell research* 18(1), 148–161.

- McCune, H. J., Danielson, L. S., Alvino, G. M., Collingwood, D., Delrow, J. J., Fangman, W. L., Brewer, B. J., Raghuraman, M. K., 2008. The Temporal Program of Chromosome Replication: Genomewide Replication in *clb5* *Saccharomyces cerevisiae*. *Genetics* 180(4), 1833–1847.
- McGinty, R. K., Tan, S., 2015. Nucleosome Structure and Function. *Chemical Reviews* 115(6), 2255–2273.
- McGuffee, S., Smith, D., Whitehouse, I., 2013. Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. *Molecular Cell* 50(1), 123–135.
- Mehanna, A., Diffley, J. F., 2012. Pre-replicative complex assembly with purified proteins. *Methods* 57(2), 222–226.
- Meister, P., Taddei, A., Gasser, S. M., 2006. In and out of the Replication Factory. *Cell* 125(7), 1233–1235.
- Mello, J. A., Silljé, H. H., Roche, D. M., Kirschner, D. B., Nigg, E. A., Almouzni, G., 2002. Human Asf1 and CAF-1 interact and synergize in a repair-coupled nucleosome assembly pathway. *EMBO reports* 3(4), 329–334.
- Menderlker, M., John G. Elias, Corthers, D. M., 1981. The dimension of DNA in solution. *Journal of Molecular Biology* 152, 153–161.
- Meryet-Figuere, M., Alaei-Mahabadi, B., Ali, M. M., Mitra, S., Subhash, S., Pandey, G. K., Larsson, E., Kanduri, C., 2014. Temporal separation of replication and transcription during S-phase progression. *Cell Cycle* 13(20), 3241–3248.
- Meselson, M., Stahl, F. W., 1958. The replication of DNA in *Escherichia coli*. *Proceedings of the national academy of sciences* 44(7), 671–682.
- Mesner, L. D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J. L., Bekiranov, S., 2013. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Research* 23(11), 1774–1788.
- Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J. S., Bensimon, A., 1997. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science (New York, N.Y.)* 277(5331), 1518–1523.
- Mirny, L. A., 2011. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research* 19(1), 37–51.

- Mitra, S., Kornberg, A., 1966. Enzymatic mechanisms of DNA replication. *The Journal of general physiology* 49(6), 59–79.
- Moggs, J. G., Grandi, P., Quivy, J.-P., Jónsson, Z. O., Hübscher, U., Becker, P. B., Almouzni, G., 2000. A CAF-1–PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Molecular and cellular biology* 20(4), 1206–1218.
- Moldovan, G.-L., Pfander, B., Jentsch, S., 2007. PCNA, the Maestro of the Replication Fork. *Cell* 129(4), 665–679.
- Moreno, S., Gambus, A., 2015. Regulation of Unperturbed DNA Replication by Ubiquitylation. *Genes* 6(3), 451–468.
- Mozziconacci, J., Koszul, R., 2015. Filling the gap: Micro-C accesses the nucleosomal fiber at 100–1000 bp resolution. *Genome Biology* 16(1).
- Méchali, M., Yoshida, K., Coulombe, P., Pasero, P., 2013. Genetic and epigenetic determinants of DNA replication origins, position and activation. *Current Opinion in Genetics & Development* 23(2), 124–131.
- Müller, C. A., Nieduszynski, C. A., 2012. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Research* 22(10), 1953–1962.
- Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., Fraser, P., 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502(7469), 59–64.
- Nannas, N. J., O’Toole, E. T., Winey, M., Murray, A. W., 2014. Chromosomal attachments set length and microtubule number in the *Saccharomyces cerevisiae* mitotic spindle. *Molecular biology of the cell* 25(25), 4034–4048.
- Necsulea, A., Guillet, C., Cadoret, J.-C., Prioleau, M.-N., Duret, L., 2009. The Relationship between DNA Replication and Human Genome Organization. *Molecular Biology and Evolution* 26(4), 729–741.
- Newman, T. J., Mamun, M. A., Nieduszynski, C. A., Blow, J., 2013. Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic Acids Research* 41(21), 9705–9718.
- Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A. S., Imamoto, N., Ishikawa, T., others, 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal* 31(7), 1644–1653.

- Norio, P., Kosiyatrakul, S., Yang, Q., Guan, Z., Brown, N. M., Thomas, S., Riblet, R., Schildkraut, C. L., 2005. Progressive Activation of DNA Replication Initiation in Large Domains of the Immunoglobulin Heavy Chain Locus during B Cell Development. *Molecular Cell* 20(4), 575–587.
- Notbohm, H., 1986. Small angle scattering of cell nuclei. *European Biophysics Journal* 13(6), 367–372.
- Ohya, T., Kawasaki, Y., Hiraga, S.-I., Kanbara, S., Nakajo, K., Nakashima, N., Suzuki, A., Sugino, A., 2002. The DNA polymerase domain of pol(epsilon) is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry* 277(31), 28099–28108.
- Olivera, B. M., Hall, Z. W., Anraku, Y., Chien, J. R., Lehman, I. R., 1968. On the Mechanism of the Polynucleotide Joining Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 33(0), 27–34.
- Olson, M. O., Dundr, M., 2015. Nucleolus Structure and Function. eLS pp. 1–9.
- On, K. F., Beuron, F., Frith, D., Snijders, A. P., Morris, E. P., Diffley, J. F. X., 2014. Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *The EMBO Journal* 33(6), 605–620.
- Oppikofer, M., Kueng, S., Gasser, S. M., 2013. SIR–nucleosome interactions: Structure–function relationships in yeast silent chromatin. *Gene* 527(1), 10–25.
- Ostrow, A. Z., Nellimoottil, T., Knott, S. R. V., Fox, C. A., Tavaré, S., Aparicio, O. M., 2014. Fkh1 and Fkh2 Bind Multiple Chromosomal Elements in the *S. cerevisiae* Genome with Distinct Specificities and Cell Cycle Dynamics. *PLoS ONE* 9(2), e87647.
- Oudet, P., Gross-Beliard, M., Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4, 281–300.
- O’Donnell, M., Li, H., 2016. The Eukaryotic Replisome Goes Under the Microscope. *Current Biology* 26(6), R247–R256.
- O’Toole, E. T., Winey, M., McIntosh, J. R., 1999. High-voltage electron tomography of spindle pole bodies and early mitotic spindles in the yeast *Saccharomyces cerevisiae*. *Molecular biology of the cell* 10(6), 2017–2031.
- Panday, A., Grove, A., 2016. The high mobility group protein HMO1 functions as a linker histone in yeast. *Epigenetics & Chromatin* 9(1).

- Pasero, P., Braguglia, D., Gasser, S. M., 1997. ORC-dependent and origin-specific initiation of DNA replication at defined foci in isolated yeast nuclei. *Genes & development* 11(12), 1504–1518.
- Patel, P. K., Arcangioli, B., Baker, S. P., Bensimon, A., Rhind, N., 2006. DNA replication origins fire stochastically in fission yeast. *Molecular biology of the cell* 17(1), 308–316.
- Patel, S. S., Donmez, I., 2006. Mechanisms of Helicases. *Journal of Biological Chemistry* 281(27), 18265–18268.
- Patterton, H. G., Landel, C. C., Landsman, D., Peterson, C. L., Simpson, R. T., 1998. The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 273(13), 7268–7276.
- Pavlov, Y., Frahm, C., McElhinny, S. N., Niimi, A., Suzuki, M., Kunkel, T., 2006. Evidence that Errors Made by DNA Polymerase alpha are Corrected by DNA Polymerase delta. *Current Biology* 16(2), 202–207.
- Peace, J. M., Ter-Zakarian, A., Aparicio, O. M., 2014. Rif1 Regulates Initiation Timing of Late Replication Origins throughout the *S. cerevisiae* Genome. *PLoS ONE* 9(5), e98501.
- Peace, J. M., Villwock, S. K., Zeytounian, J. L., Gan, Y., Aparicio, O. M., 2016. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Research* 26(3), 365–375.
- Pegoraro, G., Misteli, T., 2016. High-throughput Imaging as a versatile and unbiased discovery tool. *Methods* 96, 1–2.
- Perez-Arnaiz, P., Bruck, I., Kaplan, D. L., 2016. Mcm10 coordinates the timely assembly and activation of the replication fork helicase. *Nucleic Acids Research* 44(1), 315–329.
- Petryk, N., Kahli, M., d'Aubenton Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.-L., Hyrien, O., 2016. Replication landscape of the human genome. *Nature Communications* 7, 10208.
- Platel, M., Goldar, A., Wiggins, J. M., Barbosa, P., Libeau, P., Priam, P., Narassimprakash, H., Grodzinski, X., Marheineke, K., 2015. Tight Chk1 levels control replication cluster activation in *Xenopus*. *PloS one* 10(6), e0129090.

- Pohl, T. J., Brewer, B. J., Raghuraman, M. K., 2012. Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in *Saccharomyces cerevisiae*. *PLoS Genetics* 8(5), e1002677.
- Poli, J., Tsaponina, O., Crabbe, L., Keszthelyi, A., Pantesco, V., Chabes, A., Lengronne, A., Pasero, P., 2012. dNTP pools determine fork progression and origin usage under replication stress. *The EMBO journal* 31(4), 883–894.
- Pommier, Y., 2006. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Reviews Cancer* 6(10), 789–802.
- Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., Ren, B., Gilbert, D. M., 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature* 515(7527), 402–405.
- Powell, S. K., MacAlpine, H. K., Prinz, J. A., Li, Y., Belsky, J. A., MacAlpine, D. M., 2015. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *The EMBO Journal* 34(4), 531–543.
- Rabinovitch, P., 1994. *Introduction to cell cycle analysis*. Phoenix Flow Systems, Inc .
- Raghuraman, M. K., Brewer, B. J., 2010. Molecular analysis of the replication program in unicellular model organisms. *Chromosome Research* 18(1), 19–34.
- Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., Fangman, W. L., 2001. Replication dynamics of the yeast genome. *Science (New York, N.Y.)* 294(5540), 115–121.
- Randell, J. C., Bowers, J. L., Rodríguez, H. K., Bell, S. P., 2006. Sequential ATP Hydrolysis by Cdc6 and ORC Directs Loading of the Mcm2-7 Helicase. *Molecular Cell* 21(1), 29–39.
- Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., Lander, E., Aiden, E., 2014. A 3d Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159(7), 1665–1680.
- Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., Diffley, J. F., 2009. Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* 139(4), 719–730.

- Retkute, R., Nieduszynski, C. A., de Moura, A., 2011. Dynamics of DNA Replication in Yeast. *Physical Review Letters* 107(6).
- Retkute, R., Nieduszynski, C. A., de Moura, A., 2012. Mathematical modeling of genome replication. *Physical Review E* 86(3).
- Rhind, N., 2014. The three most important things about origins: location, location, location. *Molecular Systems Biology* 10(4), 723–723.
- Rhind, N., Gilbert, D. M., 2013. DNA Replication Timing. *Cold Spring Harbor Perspectives in Biology* 5(8), a010132–a010132.
- Rhind, N., Yang, S. C.-H., Bechhoefer, J., 2010. Reconciling stochastic origin firing with defined replication timing. *Chromosome Research* 18(1), 35–43.
- Ricci, M., Manzo, C., García-Parajo, M. F., Lakadamyali, M., Cosma, M., 2015. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* 160(6), 1145–1158.
- Rice, J. C., Allis, C. D., 2001. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Current opinion in cell biology* 13(3), 263–273.
- Richet, N., Liu, D., Legrand, P., Velours, C., Corpet, A., Gaubert, A., Bakail, M., Moal-Raisin, G., Guerois, R., Compper, C., Besle, A., Guichard, B., Almouzni, G., Ochsenein, F., 2015. Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork. *Nucleic Acids Research* 43(3), 1905–1917.
- Riera, A., Speck, C., 2016. Licensing of Replication Origins. In: Kaplan, D. L. (Ed.), *The Initiation of DNA Replication in Eukaryotes*, pp. 189–211, Springer International Publishing, Cham.
- Riera, A., Tognetti, S., Speck, C., 2014. Helicase loading: How to build a MCM2-7 double-hexamer. *Seminars in Cell & Developmental Biology* 30, 104–109.
- Riley, D., Weintraub, H., 1979. Conservative segregation of parental histones during replication in the presence of cycloheximide. *Proceedings of the National Academy of Sciences* 76(1), 328–332.
- Rivera-Mulia, J., Gilbert, D., 2016a. Replicating Large Genomes: Divide and Conquer. *Molecular Cell* 62(5), 756–765.

- Rivera-Mulia, J. C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R. A., Nazor, K., Loring, J. F., Lian, Z., Weissman, S., Robins, A. J., Schulz, T. C., Menendez, L., Kulik, M. J., Dalton, S., Gabr, H., Kahveci, T., Gilbert, D. M., 2015. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Research* 25(8), 1091–1103.
- Rivera-Mulia, J. C., Gilbert, D. M., 2016b. Replication timing and transcriptional control: beyond cause and effect—part III. *Current Opinion in Cell Biology* 40, 168–178.
- Rodley, C., Bertels, F., Jones, B., O’Sullivan, J., 2009. Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genetics and Biology* 46(11), 879–886.
- Rodriguez, J., Tsukiyama, T., 2013. ATR-like kinase Mec1 facilitates both chromatin accessibility at DNA replication forks and replication fork progression during replication stress. *Genes & Development* 27(1), 74–86.
- Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., others, 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO journal* 30(10), 1928–1938.
- Rudra, D., Warner, J. R., 2004. What better measure than ribosome synthesis? *Genes & development* 18(20), 2431–2436.
- Rutledge, M. T., Russo, M., Belton, J.-M., Dekker, J., Broach, J. R., 2015. The yeast genome undergoes significant topological reorganization in quiescence. *Nucleic Acids Research* 43(17), 8299–8313.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., Gilbert, D. M., 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Research* 20(6), 761–770.
- Ryu, H.-Y., Ahn, S. H., 2014. Yeast histone H3 lysine 4 demethylase Jhd2 regulates mitotic ribosomal DNA condensation. *BMC biology* 12(1), 1.
- Récamier, V., 2013. Single particle imaging in the cell nucleus: a quantitative approach. Ph.D. thesis, Université René Descartes-Paris V.
- Sakabe, K., Okazaki, R., 1966. A unique property of the replicating region of chromosomal DNA. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* 129(3), 651–654.

- Salic, A., Mitchison, T. J., 2008. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 105(7), 2415–2420.
- Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., Aiden, E. L., 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences* p. 201518552.
- Saner, N., Karschau, J., Natsume, T., Gierlinski, M., Retkute, R., Hawkins, M., Nieduszynski, C. A., Blow, J. J., de Moura, A. P. S., Tanaka, T. U., 2013. Stochastic association of neighboring replicons creates replication factories in budding yeast. *The Journal of Cell Biology* 202(7), 1001–1012.
- Schaaper, R. M., Radman, M., 1989. The extreme mutator effect of *Escherichia coli* mutD5 results from saturation of mismatch repair by excessive DNA replication errors. *The EMBO journal* 8(11), 3511.
- Scheffer, M. P., Eltsov, M., Frangakis, A. S., 2011. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences of the United States of America* 108(41), 16992–16997.
- Schober, H., Kalck, V., Vega-Palas, M. A., Van Houwe, G., Sage, D., Unser, M., Gartenberg, M. R., Gasser, S. M., 2008. Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Research* 18(2), 261–271.
- Schübeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., Groudine, M., 2002. Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nature Genetics* 32(3), 438–442.
- Sekedat, M. D., Fenyö, D., Rogers, R. S., Tackett, A. J., Aitchison, J. D., Chait, B. T., 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular Systems Biology* 6.
- Sekelja, M., Paulsen, J., Collas, P., 2016. 4d nucleomes in single cells: what can computational modeling reveal about spatial chromatin conformation? *Genome Biology* 17(1).

- Shachar, S., Pegoraro, G., Misteli, T., 2015. HIPMap: A High-Throughput Imaging Method for Mapping Spatial Gene Positions. *Cold Spring Harbor Symposia on Quantitative Biology* 80, 73–81.
- Siddiqui, K., On, K. F., Diffley, J. F. X., 2013. Regulating DNA Replication in Eukarya. *Cold Spring Harbor Perspectives in Biology* 5(9), a012930–a012930.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., de Laat, W., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4c). *Nature Genetics* 38(11), 1348–1354.
- Sinha, S., 1989. SCATTERING FROM FRACTAL STRUCTURES. *Physica* 38, 310–314.
- Smith, D. J., Whitehouse, I., 2012. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* 483(7390), 434–438.
- Smith, O. K., Aladjem, M. I., 2014. Chromatin Structure and Replication Origins: Determinants of Chromosome Replication and Nuclear Organization. *Journal of Molecular Biology* 426(20), 3330–3341.
- Smith, O. K., Kim, R., Fu, H., Martin, M. M., Lin, C. M., Utani, K., Zhang, Y., Marks, A. B., Lalande, M., Chamberlain, S., Libbrecht, M. W., Bouhassira, E. E., Ryan, M. C., Noble, W. S., Aladjem, M. I., 2016. Distinct epigenetic features of differentiation-regulated replication origins. *Epigenetics & Chromatin* 9(1).
- Smith, S., Stillman, B., 1989. Purification and characterization of CAF-1, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell* 58, 15–25.
- Sogo, J. M., Stahl, H., Koller, T., Knippers, R., 1986. Structure of replicating simian virus 40 minichromosomes. *Journal of Molecular Biology* 189, 189–204.
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., Li, G., 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 344(6182), 376–380.
- Speck, C., Stillman, B., 2007. Cdc6 ATPase Activity Regulates ORC.Cdc6 Stability and the Selection of Specific DNA Sequences as Origins of DNA Replication. *Journal of Biological Chemistry* 282(16), 11705–11714.
- Sridhar, A., Kedziora, S., Donaldson, A. D., 2014. At Short Telomeres Tell Directs Early Replication and Phosphorylates Rif1. *PLoS Genetics* 10(10), e1004691.

- Stamatoyannopoulos, J. A., Adzhubei, I., Thurman, R. E., Kryukov, G. V., Mirkin, S. M., Sunyaev, S. R., 2009. Human mutation rate associated with DNA replication timing. *Nature Genetics* 41(4), 393–395.
- Stanojcic, S., Sollelis, L., Kuk, N., Crobu, L., Balard, Y., Schwob, E., Bastien, P., Pagès, M., Sterkers, Y., 2016. Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells. *Scientific Reports* 6, 23142.
- Stodola, J. L., Burgers, P. M., 2016. Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nature Structural & Molecular Biology* 23(5), 402–408.
- Struhl, K., Segal, E., 2013. Determinants of nucleosome positioning. *Nature Structural & Molecular Biology* 20(3), 267–273.
- Struhl, K., Stinchcomb, D. T., Scherer, S., Davis, R. W., 1979. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proceedings of the National Academy of Sciences* 76(3), 1035–1039.
- Sugino, A., Hirose, S., Okazaki, R., 1972. RNA-linked nascent DNA fragments in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 69(7), 1863–1867.
- Sun, J., Shi, Y., Georgescu, R. E., Yuan, Z., Chait, B. T., Li, H., O’Donnell, M. E., 2015. The architecture of a eukaryotic replisome. *Nature Structural & Molecular Biology* 22(12), 976–982.
- Svergun, D. I., Koch, M. H. J., Timmins, P. A., May, R. P., 2013. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Oxford University Press.
- Symeonidou, I.-E., Kotsantis, P., Roukos, V., Rapsomaniki, M.-A., Grecco, H. E., Bastiaens, P., Taraviras, S., Lygerou, Z., 2013. Multi-step Loading of Human Minichromosome Maintenance Proteins in Live Human Cells. *Journal of Biological Chemistry* 288(50), 35852–35867.
- Szerlong, H. J., Hansen, J. C., 2011. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure This paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal’s usual peer review process. *Biochemistry and Cell Biology* 89(1), 24–34.

- Taddei, A., Gasser, S. M., 2012. Structure and Function in the Budding Yeast Nucleus. *Genetics* 192(1), 107–129.
- Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., Araki, H., 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes & development* 17(9), 1153–1165.
- Tanaka, S., Diffley, J. F., 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2–7 during G1 phase. *Nature Cell Biology* 4(3), 198–207.
- Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., Araki, H., 2011. Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing. *Current Biology* 21(24), 2055–2063.
- Taylor, J. H., 1960. Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. *The Journal of biophysical and biochemical cytology* 7(3), 455–463.
- Theis, J. F., Newlon, C. S., 1997. The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proceedings of the National Academy of Sciences* 94(20), 10786–10791.
- Therizols, P., Duong, T., Dujon, B., Zimmer, C., Fabre, E., 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences* 107(5), 2025–2030.
- Thoma, F., Koller, T., 1977. Influence of histone H1 on chromatin structure. *Cell* 12(1), 101–107.
- Ticau, S., Friedman, L., Ivica, N., Gelles, J., Bell, S., 2015. Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. *Cell* 161(3), 513–525.
- Tjong, H., Gong, K., Chen, L., Alber, F., 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research* 22(7), 1295–1305.
- Tokuda, N., Terada, T., Sasai, M., 2012. Dynamical Modeling of Three-Dimensional Genome Organization in Interphase Budding Yeast. *Biophysical Journal* 102(2), 296–304.
- Toledo, L., Altmeyer, M., Rask, M.-B., Lukas, C., Larsen, D., Povlsen, L., Bekker-Jensen, S., Mailand, N., Bartek, J., Lukas, J., 2013. ATR Prohibits Replication Catastrophe by Preventing Global Exhaustion of RPA. *Cell* 155(5), 1088–1103.

- Touchon, M., Nicolay, S., Audit, B., d'Aubenton Carafa, Y., Arneodo, A., Thermes, C., others, 2005. Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proceedings of the National Academy of Sciences of the United States of America* 102(28), 9836–9841.
- Tourrière, H., Versini, G., Cerdón-Preciado, V., Alabert, C., Pasero, P., 2005. Mrc1 and Tof1 Promote Replication Fork Progression and Recovery Independently of Rad53. *Molecular Cell* 19(5), 699–706.
- Trask, B., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G., Wu, M., 1993. Studies of Metaphase and Interphase Chromosomes Using Fluorescence In Situ Hybridization. *Cold Spring Harbor Symposia on Quantitative Biology* 58(0), 767–775.
- Trifonov, E. N., 2016. Nucleosome repeat lengths and columnar chromatin structure. *Journal of Biomolecular Structure and Dynamics* 34(6), 1156–1158.
- Truong, L. N., Wu, X., 2011. Prevention of DNA re-replication in eukaryotic cells. *Journal of Molecular Cell Biology* 3(1), 13–22.
- Tsai, H.-J., Baller, J. A., Liachko, I., Koren, A., Burrack, L. S., Hickman, M. A., Thevandavakkam, M. A., Rusche, L. N., Berman, J., 2014. Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres. *MBio* 5(5), e01703–14.
- Tsai, Y.-J., Lee, H.-I., Lin, A., 2012. Ribosome Distribution in HeLa Cells during the Cell Cycle. *PLoS ONE* 7(3), e32820.
- Tuduri, S., Tourrière, H., Pasero, P., 2010. Defining replication origin efficiency using DNA fiber assays. *Chromosome Research* 18(1), 91–102.
- Tuteja, N., Tuteja, R., 2004. Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. *European Journal of Biochemistry* 271(10), 1835–1848.
- Tyler, J. T., Adams, C. R., Chen, S.-R., Kobayashi, R., Kamakaka, R. T., Kadonaga, J. T., 1999. The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature* 402, 555–560.
- Uchida, M., Sun, Y., McDermott, G., Knoechel, C., Le Gros, M. A., Parkinson, D., Drubin, D. G., Larabell, C. A., 2011. Quantitative analysis of yeast internal architecture using soft X-ray tomography. *Yeast* 28(3), 227–236.

- Urban, J. M., Foulk, M. S., Casella, C., Gerbi, S. A., 2015. The hunt for origins of DNA replication in multicellular eukaryotes. *F1000prime reports* 7.
- Vaillant, C., Audit, B., Arneodo, A., 2007. Experiments Confirm the Influence of Genome Long-Range Correlations on Nucleosome Positioning. *Physical Review Letters* 99(21).
- Valton, A.-L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.-F., Prioleau, M.-N., 2014. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *The EMBO Journal* 33(7), 732–746.
- van Brabant, A. J., Hunt, S. Y., Fangman, W. L., Brewer, B. J., 1998. Identifying sites of replication initiation in yeast chromosomes : looking for origins in all the right places. *electrophoresis* 19, 1239–1246.
- Van Dilla, M. A., Trujillo, T., Mullaney, P., Coulter, J., 1968. Cell Microfluorometry: A Method for Rapid Fluorescence Measurement. *Science* pp. 1213–1214.
- Van Holde, K. E., Zlatanova, J., 1995. Chromatin Higher Order Structure : Chasing a Mirage ? *The Journal of Biological Chemistry* 270(15), 8373–8376.
- van Leeuwen, F., Gafken, P. R., Gottschling, D. E., 2002. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* 109(6), 745–756.
- Vasquez, P. A., Hult, C., Adalsteinsson, D., Lawrimore, J., Forest, M. G., Bloom, K., 2016. Entropy gives rise to topologically associating domains. *Nucleic Acids Research* 44(12), 5540–5549.
- Verdaasdonk, J., Vasquez, P., Barry, R., Barry, T., Goodwin, S., Forest, M., Bloom, K., 2013. Centromere Tethering Confines Chromosome Domains. *Molecular Cell* 52(6), 819–831.
- Vernis, L., 2003. Reconstitution of an efficient thymidine salvage pathway in *Saccharomyces cerevisiae*. *Nucleic Acids Research* 31(19), 120e–120.
- Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B. J., Grunstein, M., 2002. Histone acetylation regulates the time of replication origin firing. *Molecular cell* 10(5), 1223–1233.
- Wang, R., Mozziconacci, J., Bancaud, A., Gadal, O., 2015. Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. *Current Opinion in Cell Biology* 34, 54–60.
- Wasson, T., Hartemink, A. J., 2009. An ensemble model of competitive multi-factor binding of the genome. *Genome Research* 19(11), 2101–2112.

- Watson, J. D., Crick, F., 1953a. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171(4361), 964–967.
- Watson, J. D., Crick, F. H. C., 1953b. THE STRUCTURE OF DNA. *Cold Spring Harbor Symposia on Quantitative Biology* 18(0), 123–131.
- Weinreich, M., Palacios DeBeer, M. A., Fox, C. A., 2004. The activities of eukaryotic replication origins in chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1677(1-3), 142–157.
- Wells, R. D., 2007. Non-B DNA conformations, mutagenesis and disease. *Trends in Biochemical Sciences* 32(6), 271–278.
- Wilkins, M. H. F., Zubay, G., Wilson, H. R., 1959. X-Ray diffraction studies of the molecular structure of nucleohistone and chromosomes. *Journal of Molecular Biology* 1, 179–185.
- Williams, S. P., Athey, B. D., Muglia, L. J., Schappe, R. S., Gough, A. H., Langmore, J. P., 1986. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophysical journal* 49(1), 233.
- Winston, F., Dollard, C., Ricupero-Hovasse, S. L., 1995. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288c. *Yeast* 11(1), 53–55.
- Wong, H., Arbona, J.-M., Zimmer, C., 2013. How to build a yeast nucleus. *Nucleus* 4(5), 361–366.
- Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., Fabre, E., Zimmer, C., 2012. A Predictive Computational Model of the Dynamic 3d Interphase Yeast Nucleus. *Current Biology* 22(20), 1881–1890.
- Woodcock, C. L., 1994. Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *The Journal of cell biology* 125(1), 11–19.
- Woodcock, C. L., Ghosh, R. P., 2010. Chromatin Higher-order Structure and Dynamics. *Cold Spring Harbor Perspectives in Biology* 2(5), a000596–a000596.
- Woodfine, K., 2003. Replication timing of the human genome. *Human Molecular Genetics* 13(2), 191–202.

- Woodward, A. M., Göhler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D. A., Blow, J. J., 2006. Excess Mcm2–7 license dormant origins of replication that can be used under conditions of replicative stress. *The Journal of Cell Biology* 173(5), 673–683.
- Woolford, J. L., Baserga, S. J., 2013. Ribosome Biogenesis in the Yeast *Saccharomyces cerevisiae*. *Genetics* 195(3), 643–681.
- Wu, C., McGeehan, J. E., Travers, A., 2016. A metastable structure for the compact 30-nm chromatin fibre. *FEBS Letters* 590(7), 935–942.
- Wu, J.-R., Gilbert, D. M., 1996. A distinct G1 step required to specify the Chinese hamster DHFR replication origin. *Science* 271(5253), 1270–1272.
- Wu, L., Liu, Y., Kong, D., 2014. Mechanism of chromosomal DNA replication initiation and replication fork stabilization in eukaryotes. *Science China Life Sciences* 57(5), 482–487.
- Wu, P.-Y. J., Nurse, P., 2009. Establishing the Program of Origin Firing during S Phase in Fission Yeast. *Cell* 136(5), 852–864.
- Xu, W., Aparicio, J. G., Aparicio, O. M., Tavaré, S., 2006. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC genomics* 7(1), 276.
- Yabuki, N., Terashima, H., Kitada, K., 2002. Mapping of early firing origins on a replication profile of budding yeast. *Genes to Cells* 7(8), 781–789.
- Yabuuchi, H., Yamada, Y., Uchida, T., Sunathvanichkul, T., Nakagawa, T., Masukata, H., 2006. Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins. *The EMBO journal* 25(19), 4663–4674.
- Yadav, T., Whitehouse, I., 2016. Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. *Cell Reports* 15(4), 715–723.
- Yaffe, E., Farkash-Amar, S., Polten, A., Yakhini, Z., Tanay, A., Simon, I., 2010. Comparative Analysis of DNA Replication Timing Reveals Conserved Large-Scale Chromosomal Architecture. *PLoS Genetics* 6(7), e1001011.
- Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., Shirahige, K., 1997. The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 2(11), 655–665.

- Yamazaki, S., Hayano, M., Masai, H., 2013. Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing. *Trends in Genetics* 29(8), 449–460.
- Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y., Masai, H., 2012. Rif1 regulates the replication timing domains on the human genome. *The EMBO journal* 31(18), 3667–3677.
- Yang, S. C.-H., Rhind, N., Bechhoefer, J., 2010. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Molecular Systems Biology* 6.
- Yardimci, H., Loveland, A. B., Habuchi, S., van Oijen, A. M., Walter, J. C., 2010. Uncoupling of Sister Replisomes during Eukaryotic DNA Replication. *Molecular Cell* 40(5), 834–840.
- Yardımcı, G. G., Noble, W. S., 2015. Predictive model of 3d domain formation via CTCF-mediated extrusion. *Proceedings of the National Academy of Sciences* 112(47), 14404–14405.
- Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A., Diffley, J. F. X., 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519(7544), 431–435.
- Yeh, F., Sokolov, E. L., Walter, T., Chu, B., 1998. Structure Studies of Poly(diallyldimethylammonium chloride-co-acrylamide) Gels/Sodium Dodecyl Sulfate Complex. *Langmuir* 14(16), 4350–4358.
- Yekezare, M., Gomez-Gonzalez, B., Diffley, J. F. X., 2013. Controlling DNA replication origins in response to DNA damage - inhibit globally, activate locally. *Journal of Cell Science* 126(6), 1297–1306.
- Yokota, H., Van Den Engh, G., Hearst, J. E., Sachs, R. K., Trask, B. J., 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *The Journal of cell biology* 130(6), 1239–1249.
- Yoshida, K., Bacal, J., Desmarais, D., Padioleau, I., Tsaponina, O., Chabes, A., Pantesco, V., Dubois, E., Parrinello, H., Skrzypczak, M., Ginalski, K., Lengronne, A., Pasero, P., 2014. The Histone Deacetylases Sir2 and Rpd3 Act on Ribosomal DNA to Control the Replication Program in Budding Yeast. *Molecular Cell* 54(4), 691–697.

- Zaghloul, L., Baker, A., Audit, B., Arneodo, A., 2012. Gene organization inside replication domains in mammalian genomes. *Comptes Rendus Mécanique* 340(11-12), 745–757.
- Zegerman, P., 2015. Evolutionary conservation of the CDK targets in eukaryotic DNA replication initiation. *Chromosoma* 124(3), 309–321.
- Zegerman, P., Diffley, J. F. X., 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445(7125), 281–285.
- Zegerman, P., Diffley, J. F. X., 2010. Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature* 467(7314), 474–478.
- Zeman, M. K., Cimprich, K. A., 2013. Causes and consequences of replication stress. *Nature Cell Biology* 16(1), 2–9.
- Zentner, G. E., Kasinathan, S., Xin, B., Rohs, R., Henikoff, S., 2015. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nature Communications* 6, 8733.
- Zhang, J., Xu, F., Hashimshony, T., Keshet, I., Cedar, H., 2002. Establishment of transcriptional competence in early and late S phase. *Nature* 420(6912), 198–202.
- Zhang, K., Gao, Y., Li, J., Burgess, R., Han, J., Liang, H., Zhang, Z., Liu, Y., 2016. A DNA binding winged helix domain in CAF-1 functions with PCNA to stabilize CAF-1 at replication forks. *Nucleic Acids Research* p. gkw106.
- Zieve, G. W., Turnbull, D., Mullins, J. M., McIntosh, J. R., 1980. Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor Nocodazole. *Experimental Cell Research* 126(2), 397–405.
- Zimmer, C., Fabre, E., 2011. Principles of chromosomal organization: lessons from yeast. *The Journal of Cell Biology* 192(5), 723–733.
- Zlatanova, J., Seebart, C., Tomschik, M., 2007. Nap1: taking a closer look at a juggler protein of extraordinary skills. *The FASEB Journal* 21(7), 1294–1310.

Résumé français

L'information génétique contenue dans le noyau de la cellule donne les instructions pour construire, maintenir et reproduire la cellule. Il est donc crucial de dupliquer ces instructions sans erreurs afin d'obtenir deux cellules avec une information génétique correcte après la division cellulaire. L'ADN est le porteur moléculaire de l'information génétique et le processus de sa duplication est nommé réplication.

L'histoire de l'étude de la réplication de l'ADN a débuté il y a plus de cinquante ans. La découverte de la structure de l'ADN a permis d'imaginer la façon dont il était dupliqué, a débloqué la compréhension du fonctionnement de nombreuses enzymes et soulevé de nombreuses nouvelles questions. A plusieurs échelles, le processus de réplication de l'ADN est lié à sa structure. Ainsi la réplication linéaire est indubitablement organisée par la structure en double hélice. L'ADN subit plusieurs niveaux de repliement afin de tenir dans le noyau de la cellule. Cependant, l'hyperstructure de la chromatine est un obstacle à la propagation des fourches de réplication et nécessite une prise en charge par de nombreuses protéines partenaires. La position des nucléosomes est aussi d'une importance majeure pour la réplication de l'ADN, en particulier pour choisir les positions d'initiation. Enfin, l'organisation tri-dimensionnelle du noyau et le repliement de l'ADN dans le noyau est l'un des facteurs influençant l'exécution temporelle de la réplication. La réplication est donc un processus multi-échelle et l'organisation de l'ADN à plusieurs échelles est cruciale pour comprendre son déroulement. De plus, chez les eucaryotes, la réplication de l'ADN débute en plusieurs positions à des moments différents, faisant du temps la quatrième dimension du processus de réplication.

Ces quatre dimensions ont été étudiées au cours de ce travail.

Mesure de l'organisation nucléaire par les techniques de diffusion de rayonnement aux petits angles

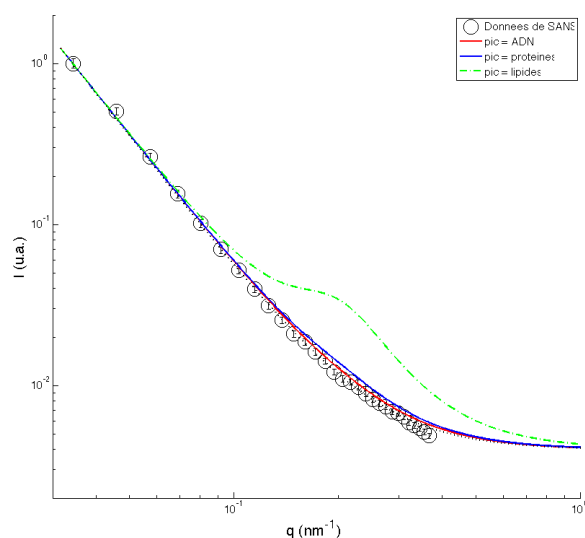
Un des objectifs du travail présenté ici, consiste à analyser l'organisation à grande échelle de la chromatine dans le noyau de la cellule de levure et l'évolution de cette structure au cours de la réplication d'ADN. En se basant sur les travaux de Lebedev et collègues, l'option d'utiliser les techniques de diffusion de rayonnement aux petits angles a été envisagée puisqu'elles devaient permettre d'explorer la structure dans le noyau sur une large gamme de distances (5-500nm). Cependant, une analyse de la littérature plus approfondie, a montré qu'ils existait une incohérence entre les données publiées obtenues par la diffusion de neutrons aux petits angles (SANS) et la diffusion de rayon X (SAXS). En utilisant les deux méthodes pour étudier des sphéroplastes de levures, nous obtenons la même incohérence.

Les neutrons et les rayons X ne sont pas diffusés de la même façon par les principaux composants de la cellule. En effet, les lipides diffusent les neutrons bien plus que les rayons X. Nous montrons ici que les résultats de SANS et de SAXS peuvent être réconciliés si le signal obtenu par SANS est principalement dû à la diffusion par les lipides (Figure 1). Les lipides sont aussi probablement le contributeur majoritaire de la courbe de SAXS de sphéroplastes de levures à petit q où les données de SANS et de SAXS sont en accord. En conséquence, nous soutenons que l'usage de données de SANS pour déduire le comportement de la chromatine dans le noyau est abusive. De plus, le signal de SAXS à petit q (correspondant aux grandes distances) est la somme ou bien l'interaction de signaux provenant de plusieurs structures qui ne sont pas toutes nécessairement de la chromatine.

Pour conclure, les méthodes de diffusion de rayonnement aux petits angles ne sont pas actuellement appropriées pour étudier l'organisation à grande échelle de la chromatine *in vivo*. Le développement d'un modèle *in silico* hautement résolu pourrait cependant permettre d'identifier les contributions des lipides, des protéines et de l'ADN ainsi que de comprendre leurs évolutions au cours du temps.

De façon intéressante, les données de SAXS présentent un pic qui n'est pas visible en SANS (Figure 1, courbe verte) et est donc dû aux protéines, à l'ADN ou à une combinaison des deux. Ce pic correspond à une distance caractéristique d'environ 30 nm, qui a été interprétée comme la distance entre des fibres de 30nm compactes dans les noyaux d'érythrocytes de poulet ou à des ribosomes agglomérés sur la surface des noyaux de cellules HeLa.

Figure 1 : Réconciliation des données de SANS et de SAXS



Les données de SAXS (courbe verte) sont ajustées en tenant compte de la différence du poids de diffusion entre SANS et SAXS dans différentes hypothèses (courbes colorées) et comparées au données de SANS (ronds noirs). Seule l'hypothèse présentée ici où la pente est due au lipides permet de retrouver les données de SANS si le pic est du aux protéines (courbe bleue) ou à l'ADN (courbe rouge).

L'hypothèse des ribosomes

Dans les cellules entières de levures, de récentes images de microscopie électronique montrent une dense présence de ribosomes dans le cytoplasme ayant un diamètre de ~ 25 à 30 nm. Des agrégats de ribosomes pourraient donc créer un pic à 25 - 30 nm. Cependant plusieurs indices développés dans ce travail discréditent cette hypothèse :

- la distance caractéristique peut diminuer avec les conditions de tampon jusqu'à ~ 25 nm qui correspondrait à un empilement parfaitement compact peu crédible,
- la longueur de corrélation qui est la distance au delà de laquelle les diffuseurs ne sont plus en interaction est inférieure à la taille d'un ribosome (~ 14 - 17 nm),
- le nombre de diffuseurs qui créent le pic augmente linéairement avec la répllication d'ADN et dépend de la vitesse de répllication,
- la longueur de corrélation est affectée par un traitement par l'enzyme MNase qui coupe les liaisons entre nucléosomes.

En conclusion, les ribosomes peuvent contribuer à créer un pic à ~ 30 nm mais pas aux évolutions mesurées ici. En conséquence, dans la suite du raisonnement, la chromatine sera considérée comme le principal contributeur de ce pic et responsable de la modifications de ses paramètres.

De futures travaux peuvent être entrepris pour renforcer encore cette conclusion. D'abord, la présence de ribosomes en surface des noyaux isolés de levures peut être vérifiée à l'aide d'anticorps spécifiques. De nouvelles mesures de SAXS pourront être effectuées sur ces noyaux isolés desquels on aura enlevé les potentiels agrégats de ribosomes. Dans de tels noyaux, un traitement par la DNase ou l'activation de la MNase permettront de mesurer l'effet de ces enzymes sur les paramètres du pic. De plus, la quantité de ribosomes pendant le cycle cellulaire pourra être mesuré, pour être confronté à l'évolution des paramètres du pic pendant le cycle cellulaire dans des cellules entières. Enfin, le modèle *in silico* des cellules entières devra prendre en compte la présence des ribosomes dans le cytoplasme.

Structure locale de la chromatine

Le pic correspondant à une distance caractéristique de 30 nm n'implique pas nécessairement que la chromatine forme une fibre dense de 30 nm de large. En effet, seules des fibres formant un empilement très proche donnent lieu à une distance répétée de 30 nm comme c'est le cas dans les erythrocytes de poulet (distance caractéristique de ~ 30 -40 nm comparée à la distance caractéristique mesurée dans la levure de ~ 25 -30). Au contraire, la faible longueur de corrélation mesurée n'est pas compatible avec l'existence d'une organisation bien définie composée de fibres de 30 nm. Nos données dépeignent plutôt la chromatine comme un cristal liquide avec un ordre nématique. Dans cette optique, les nucléosomes seraient organisés selon le modèle de collier de perle formant des fibre de ~ 10 nm distantes de ~ 25 à 30 nm.

Cette distance est remarquablement constante pendant le cycle cellulaire mais est affecté par la pression osmotique du tampon. Cela suggère que cette distance entre objets correspond à un équilibre dicté par les propriétés physico-chimiques de l'environnement nucléaire.

La valeur de la longueur de corrélation est une moyenne sur le génome donc cela n'exclut pas la présence d'arrangements particuliers dont la contribution serait réduite par le moyennage. Cependant, la valeur de ~ 16 nm suggère que les nucléosomes sont principalement influencés par leurs voisins directs et forment de rares agrégats de quelques nucléosomes, en accord avec les récentes images de

microscopie électronique.

Cette vision de la chromatine est proche du modèle de l'organisation de la chromatine chez l'homme. Cependant, la distance caractéristique n'est pas observable par SAXS dans les cellules HeLa après le retrait des ribosomes. Cela suggère que, dans les cellules humaines, la chromatine adopte un comportement totalement liquide tandis que chez la levure, l'organisation particulière en Rabl conduit à une orientation préférentielle qui engendre un ordre nématique. Cette orientation est potentiellement mieux définie dans la région des centromères où l'encombrement force une orientation linéaire locale. Des informations supplémentaires sur cette structure locale de la chromatine dans le noyau pourront être obtenue par des images aux rayons X de cellules uniques.

Evolution de la longueur de corrélation au cours du temps

La longueur de corrélation est la longueur sur laquelle les diffuseurs sont en interaction and peut être compris comme une énergie d'interaction divisée par une force entropique qui tend à désorganiser le système. Les interactions entre les nucléosomes sont permises par un segment d'ADN leur qui peut être considéré comme un ressort de raideur k . On obtient :

$$\xi = \frac{k \times x^2}{2F}$$

où ξ est la longueur de corrélation, x l'élongation du leur par rapport à sa position d'équilibre et F la force entropique qui tend à désorganiser le système.

La différence entre la corrélation à tout temps $\xi(t)$ et la longueur de corrélation initiale ξ_0 dépend de leurs raideurs et élongations respectives :

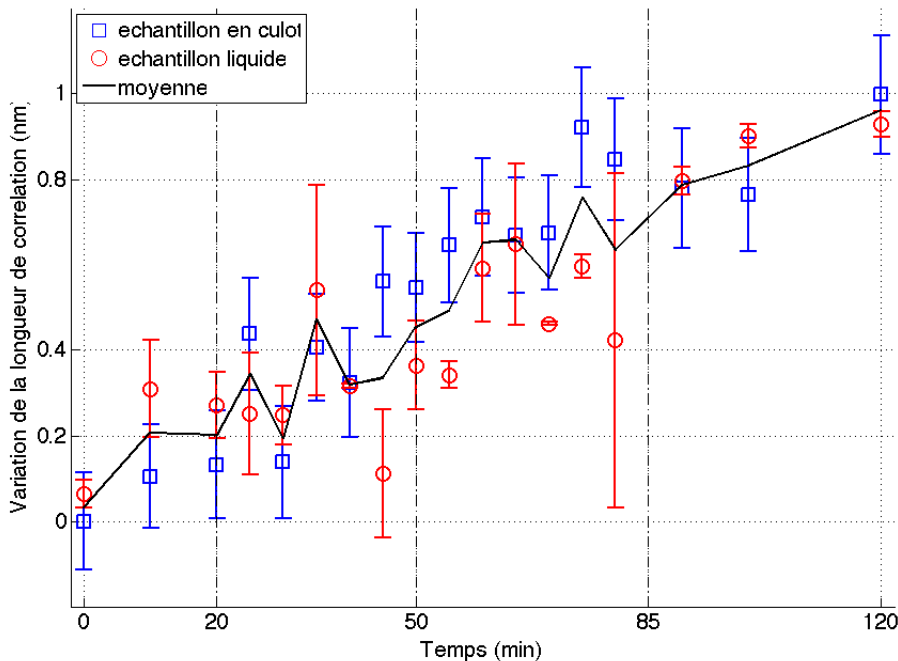
$$\xi(t) - \xi_0 = \frac{k \times x(t)^2 - k_0 \times x_0^2}{2F}$$

La digestion par la MNase tend à diminuer la longueur de corrélation. Comme la MNase coupe les leurs, leur raideur devient zéro, diminuant ainsi la raideur moyenne sur le génome.

En revanche, au cours de l'interphase, la raideur du leur peut être considérée constante d'où $k = k_0$, ce qui donne :

$$\left| \frac{x(t)}{x_0} \right| = \sqrt{\frac{\xi(t)}{\xi_0}}$$

Figure 2 : Evolution de la longueur de corrélation au cours du temps



La longueur de corrélation mesurée dans des échantillons en culot et en condition liquide au cours du temps après le lavage de l' α -facteur qui synchronise la population de cellules en phase G1 du cycle cellulaire. La plupart des cellules effectuent leur phase S intervient environ entre les temps 20 et 50min. Plus de points temporels ont été mesurés pour la condition liquide mais pour une meilleure visualisation des données, elles ont été interpolées sur le même intervalle temporel que les données en condition culot.

Comme ξ augmente au cours du temps pendant l'interphase, la valeur absolue de x augmente également ce qui signifie que le lieu s'éloigne de sa position d'équilibre. Il est difficile de prédire si cela correspond à un allongement ou un raccourcissement. On s'attend à ce que la chromatine se condense avant la division cellulaire pour permettre la formation de chromosomes qui pourront être séparés dans les deux cellules. Cependant, le mécanisme de condensation pourrait impliquer un raccourcissement du lieu qui entrainerait un rapprochement des nucléosomes, ou bien un allongement qui autoriserait des interactions entre nucléosomes plus complexes. De plus, cet effet est assez faible, en accord avec l'observation de l'absence de réorganisation majeure dans les cellules en métaphase vu par microscopie électronique.

L'augmentation de la longueur de corrélation commence dès le début de l'expérience où les cellules se réorganisent après le traitement par α -facteur. L'évolution de la longueur de corrélation au cours de la phase S n'est pas claire. En effet, le jeu de données possédant la meilleure résolution temporelle (Figure 2, ronds rouges) suggère que cette longueur pourrait être constante au cours de la phase S mais le

jeu de données obtenus sur des culots de cellules (Figure 2, carrés bleus) , qui sont mieux représentées par l'équation, suggère une augmentation linéaire au cours de toute l'interphase. Dans les deux cas, aucune modification drastique n'est observée au cours de la phase S qui pourrait être corrélée avec le pic d'initiation. Il semble donc que les potentielles modifications au cours de la phase S ne sont pas causées par une réorganisation spécifique qui permettrait la réplication de l'ADN, comme par exemple une ouverture locale de la chromatine, mais plutôt par une création de matière dans le noyau qui entraîne un réarrangement de la cellule.

Ces interprétations demandent d'être soutenues par des jeux de données supplémentaires, associés à une connaissance de l'état des cellules lors de la mesure afin que les petites variations mesurées ici puissent être confirmées. Cependant, ces premiers jeux de données laissent penser que la réorganisation au cours de la phase S n'est pas aussi drastique qu'il a pu être envisagé. De plus, les réorganisations au cours de la phase S étant faibles, il semble que la configuration atteinte en phase G1 soit adaptée pour effectuer la phase S. Cela suggère que si une organisation particulière est requise pour procéder à la réplication de l'ADN, elle est mise en place au cours de la phase G1.

Point de décision du timing et programme temporel de la réplication

Il est donc intéressant de noter qu'il est proposé que le programme de réplication soit établi au cours de la phase G1. Bien que le mécanisme biologique exact derrière le concept de point de décision du timing de la réplication soit toujours inconnu, plusieurs hypothèses non-exclusives ont été formulées.

Le temps de résidence du complexe ORC en des positions particulières et l'environnement nucléosomal conduisent à une distribution spécifique des doubles hexamères de MCM le long du génome. Cette répartition des origines potentielles le long du génome est effectuée durant la phase G1.

Par ailleurs, fkh1 est lié au déclenchement précoce des origines par son association avec les origines lors de la phase G1. Le mécanisme exact par lequel Fkh induit un déclenchement précoce des origines au cours de la phase S est inconnu mais semble être lié au regroupement physique d'origines précoces où le recrutement de cdc45 pourrait être important dès la fin de la phase G1, ce qui conduirait à une réplication précoce. De plus, Rif1 séquestre les origines tardives par un regroupement physique en périphérie du noyau où il empêche l'initiation par une inactivation des kinases de la réplication.

Il a été montré que le nombre de MCM est corrélé au moment de l'initiation des origines, mais seulement lorsque ces dernières ne sont pas régulées par un autre élément tel que Fkh1/2 ou Rif1. Cependant, nous montrons ici que la probabilité d'initiation serait bien corrélée à la densité de MCM sur le génome entier, ce qui implique que la distribution de MCM est le premier niveau de régulation du programme temporel.

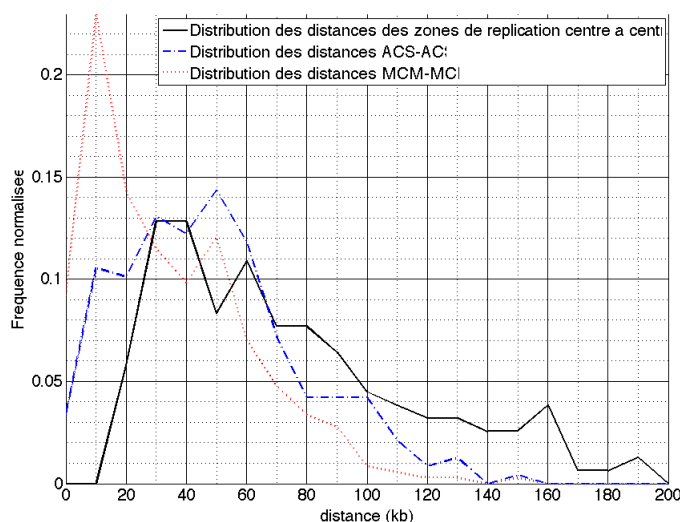
Contrairement aux eucaryotes ayant des génomes grands et compacts, la structure globale de la chromatine n'empêche pas la diffusion de facteurs limitants pour contraindre la réplication à certaines zones. Cependant, dans l'hypothèse où les facteurs limitant sont importés de façon asymétrique dans le noyau, avec un apport majeur au niveau du pôle centromérique, la séquestration des facteurs limitants au niveau des origines précoces ralentit leur diffusion dans le volume nucléaire. Dans ce cas, le programme temporel est régulé à trois échelles définies pendant la phase G1 :

- l'organisation nucléaire garantit le regroupement des centromères et le positionnement des télomères en périphérie ainsi que la libre diffusion dans le volume nucléaire,
- des facteurs de structuration tels que Fkh1/2 et Rif1 ont un rôle local sur la distribution des facteurs d'initiation (respectivement en recrutant cdc45 et en désactivant DDK), créant respectivement des zones préférentielles ou défavorables pour l'initiation et maintenant les origines potentielles dans ces régions,
- de façon plus locale, la concentration en origines potentielles influe sur la probabilité d'initiation.

Programme spatio-temporel et regroupement

Au cours de la phase S, le programme de réplication est exécuté par la rencontre entre un facteur d'initiation et une origine potentielle. Afin de rendre compte de l'évolution mesurée de la densité de fourche et du taux d'initiation au cours de la phase S, les facteurs d'initiation doivent être présents en quantité limitante et augmenter au cours de la phase S. Cela n'explique pas en revanche, la distribution spatiale des origines. La distribution des origines est difficile à mesurer dans une population de cellules où les événements rares sont effacés par la moyenne faite sur l'ensemble de la population, et doit donc être mesurée par la technique du

Figure 3 : Distribution de distances entre origines, entre ACS et entre MCM



Distribution des distances entre zones de répliation centre à centre en bins de 10kb (ligne noire) comparé à la distribution de distances entre positions de fixation d'ORC (ACS) (ligne bleue pointillée) et entre positions de MCM (ligne rouge pointillée). A cause d'un manque de sensibilité, les distances très proches ne peuvent pas être mesurées, ce qui explique les valeurs négatives pour des distances courtes obtenues pour notre courbe expérimentale (ligne noire).

peignage d'ADN. Par cette méthode, l'ADN nouvellement synthétisé est directement visualisé sur des molécules d'ADN, ce qui donne de l'information sur des cellules uniques. La distance entre deux zones d'initiation est d'environ 50 kb au début de la phase S. Cependant, au cours de ce travail, nous avons mené une analyse plus poussée des données de peignage d'ADN basée sur le comportement temporel extrait du taux d'initiation et du nombre de fourches. Cette analyse montre que 20 à 40% des initiations ont lieu à moins de 10 kb de distance dans un court laps de temps. Les fourches émanant de ces origines fusionnent rapidement, donnant lieu à des zones répliquées plus grandes, avec une distance entre elles comparable à la distance entre deux ORC consécutifs. Cela suggère donc qu'une initiation presque synchrone de MCM proches, mis en place par le même ORC, résulte très rapidement en des régions de répliation plus ou moins centrées sur la position d'ORC.

Les MCMs forment donc des regroupements de timing. De tels regroupements d'initiation ont été observés chez de nombreux eucaryotes, ce qui suggère qu'en plus d'un programme temporel commun, la dimension spatiale du programme est aussi largement similaire chez tous les eucaryotes. Bien que les génomes plus grands nécessitent des mécanismes de régulations additionnels, le programme spatio-temporel, en adaptant l'échelle, est équivalent.

Les regroupements de timing sont créés par une augmentation locale de la pro-

babilité d'initiation. Un recyclage local de facteurs d'initiation limitant, comme cdc45, est supposé avoir lieu après la fusion des fourches et ne peut donc pas expliquer l'initiation observée qui doit nécessairement intervenir avant la fusion des fourches. Il a été observé que la chromatine est plus accessible dans une zone de 5 à 7 kb autour d'une fourche active ce qui pourrait expliquer une augmentation de la probabilité d'initiation locale due à des fourches actives. Dans cette hypothèse, la tension créée par le déroulement de l'ADN, ou bien l'action de remodeleurs de la chromatine associés aux fourches, retirerait les nucléosomes en amont des fourches actives.

Cette hypothèse est difficile à réconcilier avec notre observation que le pic d'initiation n'induit pas de réorganisation locale de la chromatine. Le nombre de nucléosomes déplacés dans une zone de 7 kb peut être calculé à partir de la densité de nucléosomes qui ont été cartographiés, et dans l'hypothèse où la chromatine est reconstruite linéairement après le passage de la fourche, le nombre de nucléosomes nouvellement déposés peut aussi être estimé afin de calculer le nombre de nucléosomes déplacés. Ainsi, nous pouvons estimer que 30% des nucléosomes sont affectés par la réplication au moment du pic d'initiation. Il est possible que le nombre de fourche soit surestimé car dans notre analyse la densité de fourche a été présumée homogène sur le génome. Cependant, il est aussi possible que l'impact de cet effet n'est pas visible sur la longueur de corrélation à cause de la moyenne faite sur des cellules qui ne sont pas parfaitement synchronisées et n'ont pas un pic d'initiation exactement au même moment.

Population contre cellule unique

Globalement, ce travail a mis en lumière le fait que la variabilité dans une population de cellule peut grandement affecter l'interprétation des résultats obtenus. Le travail en molécules uniques a donné un regard précieux sur le mécanisme précis de la réplication de l'ADN. Cependant, son rendement est faible, ce qui rend l'obtention de résultats statistiquement significatifs difficile et l'interprétation des données, là encore, précautionneuse. Travailler avec une population de cellules est souvent nécessaire, non seulement car cela permet de collecter plus facilement de l'information mais aussi car le comportement global et la déviation par rapport à ce comportement contiennent tous deux des informations sur le processus.

Deux approches peuvent être entreprises pour obtenir l'information voulue. D'un côté, les paramètres expérimentaux permettant de synchroniser la population de cellules en phase S peuvent être modifiés pour tenter d'obtenir une population

plus homogène. D'un autre côté, la variabilité dans une population de cellules peut être estimée par un découpage de la population globale en sous populations. Cette approche a déjà été appliquée pour l'analyse des données de FACS, où le découpage en trois populations a permis de mesurer l'entrée en phase S de la population globale. De plus, au lieu d'analyser le comportement de la population, le comportement d'une cellule unique et une fonction définissant le bruit peuvent être utilisés pour modéliser les données de population.

Ce travail a pris plusieurs directions pour étudier l'évolution des cellules durant la réplication de l'ADN afin d'obtenir un aperçu du programme de réplication dans son ensemble. La mesure de la structure dans la cellule pendant la réplication *in vivo* a montré une stabilité étonnante, y compris pendant la période de la phase S où l'initiation est très fréquente, ce qui corrobore des données *in vitro*. La haute fréquence d'initiation peut aussi être interprétée en termes de création et fusion de fourches qui suggère que les précédents résultats obtenus par l'interprétation des données de peignage ont sous estimé le nombre d'origines et surestimé leurs distances. Ces nouvelles données confirment le rôle de MCM en tant qu'origines potentielles et vont dans le sens de l'existence d'initiations regroupées, comparables à ce qui a été observé chez les autres eucaryotes.

Globalement, ce travail a proposé de nouvelles méthodes techniques et analytiques pour étudier la réplication de l'ADN, qui ouvre la voie à de nouvelles recherches. De plus, ce travail a cherché à passer en revue et à tirer parti de l'importante littérature scientifique ainsi que des jeux de données disponibles sur la réplication de l'ADN chez la levure. L'effort pour construire un modèle complet de la réplication de l'ADN chez les eucaryote devra se concentrer sur l'intégration de toutes ces données.

Titre : Réplication de l'ADN chez la levure de boulanger : lien entre la conformation de la chromatine et la cinétique de réplication

Mots clés : Réplication de l'ADN, Chromatine, Saccharomyces Cerevisiae, Timing, SAXS

Résumé : L'information génétique contenue dans le noyau de la cellule doit être dupliquée fidèlement afin d'être transmise aux cellules filles pendant la division cellulaire. Pour organiser leur division, les cellules suivent un cycle reproductible composé de quatre étapes appelé cycle cellulaire. La préparation et l'exécution du programme de réplication de l'ADN ont lieu pendant des phases spécifiques du cycle grâce à l'intervention de multiples partenaires protéiques et de régulateurs structuraux. En particulier, la réplication de l'ADN s'effectue sur une matrice complexe constituée d'ADN associé à des protéines appelée chromatine. Cette dernière influence et est influencée par la réplication de l'ADN. Le travail présenté ici a pour objectif de faire le lien entre la conformation de la chromatine et la cinétique de réplication de l'ADN. Pour ce faire, nous combinons plusieurs techniques. La cytométrie de flux nous permet de suivre la quantité d'ADN présent dans une population de cellules et,

à l'aide d'une méthode développée dans notre laboratoire, d'extraire le programme de réplication moyen d'une population de cellules. La technique de SAXS fournit des informations sur l'organisation locale des protéines et de l'ADN *in vivo*. Nos données peuvent être interprétées comme un cristal liquide avec un ordre nématique et une faible longueur de corrélation, ce qui suggère que la chromatine de la levure est majoritairement dépourvue d'une organisation en fibre de 30nm *in vivo*. Par ailleurs, par la méthode de peignage d'ADN, nous reproduisons les résultats précédemment obtenus montrant que la distance entre zones répliquées est d'environ ~60kb qui correspond à la distance entre des origines de réplication identifiées. Cependant, d'après l'étude du comportement dynamique de l'initiation, nous proposons que les initiations sont plus fréquentes que ce qui a été mesuré précédemment et correspondent à la distance entre les protéines MCM disposées sur le génome.

Title : DNA replication in budding yeast : link between chromatine conformation and kinetics of replication

Keywords : DNA replication, Chromatine, Saccharomyces Cerevisiae, Timing, SAXS

Abstract : Genetic information carried in the cell nucleus must be faithfully duplicated to be transmitted to daughter cells during cell division. In order to orchestrate their division, cells go through a reproducible 4 stages cycle called «cell cycle». The preparation and execution of the DNA replication program is restricted to specific phases and implies many proteic and structural regulators. In particular, DNA replication occurs on a complex template of DNA associated with proteins. The latter is both influencing and influenced by DNA replication. This work aims at investigating the link between chromatin conformation and the kinetics of DNA replication. In order to do so, we combine several techniques. Using flow cytometry, we follow the evolution of a cell population with regards to their DNA content and,

with a method developed in our laboratory, decipher the population averaged temporal program of DNA replication. SAXS data provide information on the local organisation of protein and DNA *in vivo*. Our data can be interpreted as a liquid crystal with a nematic order and a short correlation length, which suggest that yeast chromatin *in vivo* is predominantly devoid of 30 nm fibres organisation. On the other hand, we performed DNA combing to study the replication program in single cells. We reproduce previously obtained result showing that distance between replicated tracks is of ~60kb which corresponds to the distance between known origins of replication. However, studying the behaviour of initiation, we propose that the initiation events are more frequent than previously measured and correspond to distances between MCMs proteins loaded on the genome.

