Nucleosome positioning on large tandem DNA repeats of the '601' sequence engineered in Saccharomyces cerevisiae

Astrid Lancrey, Alexandra Joubert, Evelyne Duvernois-Berthet, Etienne Routhier, Saurabh Raj, Agnès Thierry, Marta Sigarteu, Loic Ponger, Vincent Croquette, Julien Mozziconacci, et al.

Nucleosome positioning on large tandem DNA repeats of the '601' sequence engineered in *Saccharomyces cerevisiae*

Astrid Lancrey[1], Alexandra Joubert[1], Evelyne Duvernois-Berthet[2], Etienne Routhier[3], Saurabh Raj[4,5], Agnès Thierry[6], Marta Sigarteu[1], Loic Ponger[1], Vincent Croquette[4,7], Julien Mozziconacci[1,3,8,*] and Jean-Baptiste Boulé[1,*]

[1]Structure et Instabilité des Génomes (StrInG), Museum National d'Histoire Naturelle, INSERM, CNRS, Alliance Sorbonne Université 75005 Paris, France.
[2] Physiologie Moléculaire et Adaptation (PhyMA), Museum National d'Histoire Naturelle, CNRS, Alliance Sorbonne Université, 75005 Paris, France.
[3] Laboratoire de Physique Théorique de la Matière Condensée (LPTMC), CNRS, Sorbonne Université, 75005 Paris, France.
[4] Laboratoire de Physique de L'École Normale Supérieure de Paris (LPENS), Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS, Ecole Normale Supérieure, Université PSL, Sorbonne Université, Université de Paris, 75005 Paris, France.
[5] Current address: Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, 110016, India.
[6] Institut Pasteur, Unité Régulation Spatiale des Génomes, UMR 3525, CNRS, 75015 Paris, France.
[7] ESPCI, Université PSL, 10 rue Vauquelin, 75005 Paris, France.
[8] Institut Universitaire de France, Paris 75005, France.

[*] correspondance: julien.mozziconacci@mnhn.fr, jean-baptiste.boule@mnhn.fr

Abstract

The artificial 601 DNA sequence is often used to constrain the position of nucleosomes on a DNA molecule in vitro. Although the ability of the 147 base pair sequence to precisely position a nucleosome in vitro is well documented, application of this property in vivo has been explored only in a few studies and yielded contradictory conclusions. Our goal in the present study was to test the ability of the 601 sequence to dictate nucleosome positioning in *Saccharomyces cerevisiae* in the context of a long tandem repeat array inserted in a yeast chromosome. We engineered such arrays with three different repeat size, namely 167, 197 and 237 base pairs. Although our arrays are able to position nucleosomes in vitro, analysis of nucleosome occupancy in vivo revealed that nucleosomes are not preferentially positioned as expected on the 601-core sequence along the repeats and that the measured nucleosome repeat length does not correspond to the one expected by design. Altogether our results demonstrate that the rules defining nucleosome positions on this DNA sequence in vitro are not valid in vivo, at least in this chromosomal context, questioning the relevance of using the 601 sequence in vivo to achieve precise nucleosome positioning on designer synthetic DNA sequences.

Introduction

The nucleosomes are more than structural proteins that enable folding of a meter-long molecule into a micrometer-diameter nucleus, and are now recognized as master players in the many facets of genome metabolism. The deposition of post-translational modifications on nucleosomes along genomes are tightly correlated with the regulation of gene expression [1]. The first experimental evidence for the existence of nucleosomes came from their regular, periodic assembly along the DNA molecule [2, 3]. In the following decades, the nucleosomal periodicity was measured in various contexts and the Nucleosomal Repeat Length (NRL) was shown to vary in different species, different cell types and even in different regions of chromosomes [4]. As soon as the first

eukaryotic DNA sequences were available, a link between the DNA helical pitch and the folding of the molecular into chromatin has been noted [5]. As more NRL measures became available, Jonathan Widom noticed that NRL took preferential values of $167 + n * 10$ base pairs (bp) and linked this observation to the value of the DNA helical pitch [6]. As the periodic position of nucleosomes became of high mechanistic interest to explain chromatin function, nucleosomal array reconstruction protocols for in vitro studies were developed. In order to test the physical and structural properties of DNA wrapping around an histone octamer and the role of the underlying DNA sequence, Widom and colleagues isolated, using a SELEX approach, a set of DNA sequences that could precisely position nucleosomes in vitro [7]. Among the DNA sequences obtained, the "601" sequence was retained to be the one forming the most stable nucleosomal complex in vitro, together with an histone octamere. This "601" sequence has been successfully used in many in vitro studies to reconstruct regular arrays of nucleosomes for electron microscopy [8, 9] and single molecule manipulations [10–12]. Structures of nucleosome core particles containing the 601 positioning sequence were also obtained by X-ray christallography, revealing in this case 145 bp of the 601 sequence wraped around the nucleosome [13]. Regardless of the exact size of the 601 nucleosome, these studies confirmed that the $167 + n * 10$ bp NRL rule was required to form regular arrays and revealed both their ambivalent ability to be highly resilient to supercoiling in their extended form [10, 14] and to fold into very dense and compact fibers under specific buffer conditions [15]. The detailed biophysical modeling of nucleosomal array structures showed that the observed quantized behavior of the NRL is primarily due to the combination of the nucleosome/nucleosome steric hindrance and the 10 bp helical pitch of the DNA linker [16, 17].

The potential use of the 601 sequence for in vivo studies has early been envisioned by Lowary and Widom themselves. However, only a handful of studies have been carried so far in this direction and the ability of one inserted sequence to position a nucleosome has led to divergent conclusions that pointed to the dependence on the context in which the 601 sequence was used (episomal vs genomically inserted) [18–21]. In yeast, single copies of the 601 sequence inserted genomically in an open reading frame (ORF) were shown to not promote strong positioning [20]. Regular arrays of repeats of various sizes have also been used in the context of extrachromosomal plasmids. Although the 601 repeats affected nearby gene expression, the positioning of nucleosomes on the repeated sequences was not directly assessed, leaving open the possibility for indirect effects [22]. Our goal here was to assess the ability of 601 sequence to position nucleosomes in vivo on long tandem arrays using three different linker lengths of 20, 50 and 90 bp. These lengths would respectively constrain the NRL respectively to 167, 197 or 237 bp. Using a genome editing approach that we recently developed to engineer long DNA tandem repeats in yeast chromosomal DNA [23], we assembled tens of 601 repeats and analyzed the statistical positioning of nucleosomes on these arrays. Our results show that, in vivo, the 601 sequence is not able to position the nucleosomes on the repeated 601 cores, yielding NRL which are different from the ones we designed, except for the control 167 bp tandem array, which approximates the naturally observed yeast NRL.

Results

In vivo assembly of 601 tandem DNA repeats in the yeast genome

We used a strategy that we developed recently [23] to assemble tandem DNA repeats containing the 601 sequence in the genome of *Saccharomyces cerevisiae*. Specifically, we engineered three designs of overlapping oligonucleotides to partially replace the non-essential *YMR262W* gene and its promoter, located in chromosome XIII, to give rise to arrays of tandem repeats of 167, 197 or 237 bp long monomers (Figure 1 and supplementary Figure S1). Integration at the *YMR262W* locus resulted in the removal of 129 bp upstream the start codon and 231 bp downstream the start codon [23]. Sizes of arrays obtained by this approach ranged from one repeat to arrays longer than 15 *kb*, as estimated by gel electrophoresis, corresponding to more than 50 repeats (Figure 2). Given the high efficiency of this assembly method, screening of only a few dozen of clones were needed to find clones containing large repeats. Correct assembly was verified by sequencing of the assembly junctions (Supplementary Figure S2) as well as the homogeneity of the tandem repeats (Supplementary Figure S3). For each of the three designs, the clone with the largest array was kept for chromatin studies. Finally, since tandem repeats may be susceptible to genomic instability, we measured the stability of the arrays upon mitotic division. After around 25 generations, the percentage of rearranged clones was respectively 1, 2 and 4%, demonstrating that large tandem repeats of the 601 sequences are easily maintained in the yeast genome (Supplementary Figure

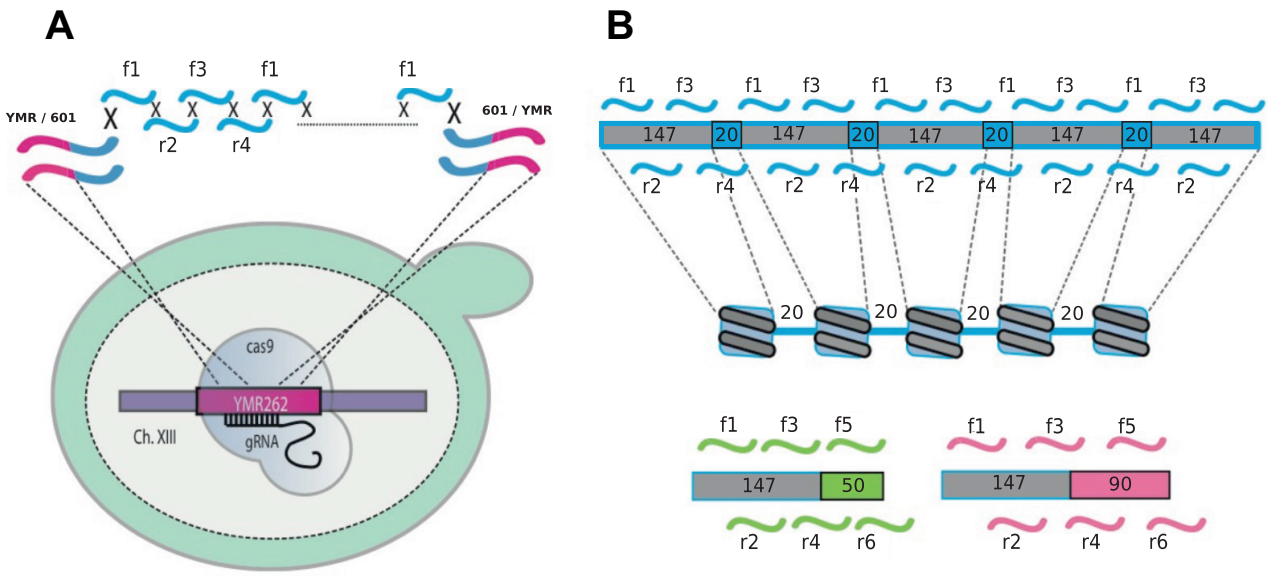

Figure 1. In vivo assembly of '601' tandem DNA repeats. (A) Schematic representation of the CRISPR/Cas9-assisted genomic integration of overlapping oligonucleotides. The method is represented for integration of 601-167 repeats in the *YMR262W* gene of *S. cerevisiae*. (B) Design of the 601-167 repeats: four 50 nt oligonucleotides f1-r2-f3-r4 overlapping on 20 nt were used to construct the 601-167 array. The 147 bp core sequence (grey) corresponds to the DNA portion theoretically wrapped around one nucleosome, followed by a 20 bp linker (blue). The f4 oligo overlaps with f1 so that several monomers can assemble into a repeated array. (C) Design of the 601-197 and 601-237 arrays, using six overlapping oligonucleotides.

S4).

The NRL measured on 601 repeats are different in vitro and in vivo

In order to verify that our design of the DNA linkers 20, 50 and 90 bp did not affect the ability of the 601 sequence to position nucleosomes in vitro, we assembled chromatin on plasmids containing respectively 20, 14 and 14 repeats of the 167, 197 and 237 bp monomers (supplementary Figure S5A-C). In vitro assembled chromatin was digested by Micrococcal Nuclease (MNase), separated on gel and probed by Southern Blotting using a 601 probe. The results confirmed that nucleosomes were positionned by the 601 core on all three 601 repeats (supplementary Figure S5D, E). However with the 601-237 design we observed two additional shorter ladders reflecting that some nucleosomes are positioned according to a second NRL when the designed NRL is equal to 237 bp.

To get insights into nucleosome organization in the 601 repeats in vivo we first performed MNase-gel experiments on the wild type YPH499 strain and the three yeast strains harboring the repeats. MNase-gel experiments consisted on partially digesting the chromatin of cells harboring 601 repeats with MNase, followed by gel analysis of NRL over the whole genome or specifically within the repeats (Figure 3, see Material and Methods for details). We measured a NRL of 163 $\pm$1 bp for the whole genome, similar to the values expected for *S. cerevisiae* [24]. The values measured after Southern blotting within the 601 arrays revealed that nucleosomes are spaced in average according to a 170$\pm$4 bp NRL, a 176$\pm$7 bp NRL and a 171$\pm$9 bp NRL respectively for the 167 bp, 197 bp and 237 bp repeats (Figure 3C,D). The NRL measured for the 601-167 repeats is close to the one expected, however nucleosomes did not appear preferentially spaced according to the pre-defined 197 and 237 bp NRL and are therefore not phased with the repeats.

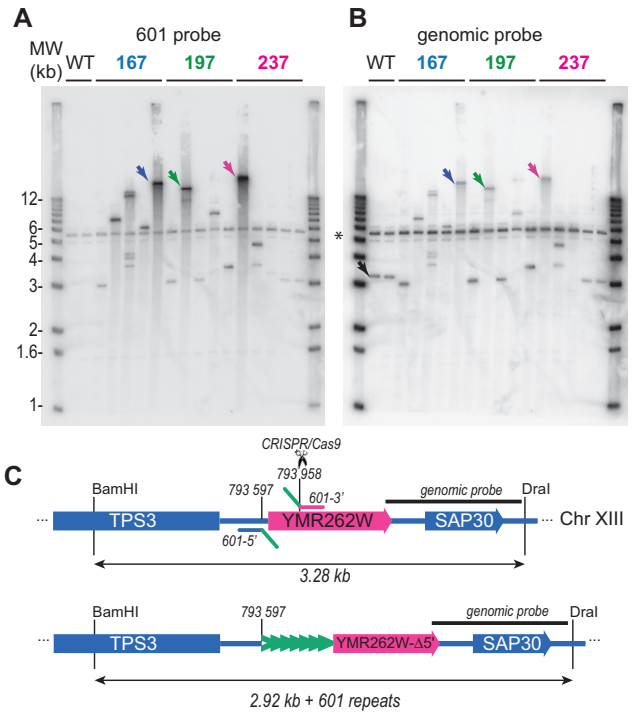

Figure 2. Overlapping oligonucleotides assemble to form long repeated arrays at the *YMR262W* locus. (A) Diagram of the assembly region before and after CAS9-assisted recombinational assembly. Position of the genomic probe is represented by a dark line. Localization of the genomic-601 junction primers are also indicated. (B) Southern Blot analysis of yeast recombinant strains. Genomic DNA from strains containing 601 repeats were digested with BamhI and DraI. Samples were electrophoresed, blotted and hybridized with a 601 probe containing a 601-147 nt repeat unit. The first two lanes (WT) correspond respectively to the wild-type and CAS9 expressing strains. Five clones are shown for respectively the 167, the 197 and the 237 design; the strains containing arrays largers than 15 kb (marked with color arrows) were retained for further manipulations. (C) The membrane shown in B was stripped and rehybridized using the genomic probe shown in A. The star marks a non specific band present on both Southern blots and present in the WT controls

Nucleosomes are not positioned on the 601 core sequence in vivo

To have a more precise assessment of nucleosome occupancy in the repeats in vivo, we performed a MNase-seq analysis of mononucleosomal DNA from the 601 strains. Every midpoint of MNase resistant fragments, (which correspond to the central nucleotide of each fragment), was positioned to a reference genome allowing to aggregate on a single 601 monomer sequence all fragments containing the 601 sequence (see Material and Methods and Supplementary Figure S6, S7 for details on the building of the reference genome and our MNase-seq analysis pipeline). We used two levels of MNase digestion: moderate and over-digested, with two replicates for each conditions. Nucleosomal DNA midpoint density on the three array types are presented on Figure 4A-C. The length distribution of moderately digested fragments was found to be centered on 147 bp (coral and purple distributions on Figure 4D-F) whereas the length distribution of over-digested fragments was closer to 125 bp, exhibiting two to three peaks corresponding to various levels of digestion of the DNA wrapped around the nucleosome (green and cyan distributions on Figure 4D-F). For over-digested chromatin, we observed a main peak in the distribution of fragments midpoints, at position 110 on the 167 bp repeats, and at position 150 and 160 on respectively the 197 and 237 bp repeats (Figure 4 green and cyan). Density patterns of fragments midpoints gave similar peak positions, although patterns obtained with overdigested fragments gave a stronger main peak for the 167 and 197 repeats (Figure 4 coral and purple curves). As a control, the distribution of genome-wide nucleosome position aligned to the transcriptional start site provided the expected nucleosome

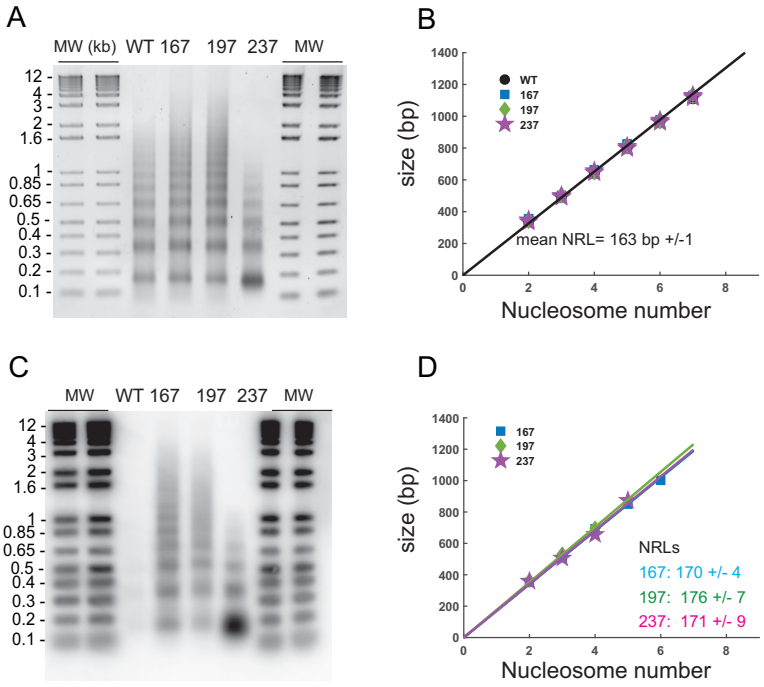

Figure 3. NRL estimation in '601' tandem DNA repeats in vivo. (A) MNase gel on which were migrated the purified DNA from partial MNase digestions of the wt and the three 601 strains chromatin. Each band from one nucleosomal ladder corresponds respectively, from bottom to top, to the mononucleosomal DNA, di- tri-, etc.. Conversion of migration distance inside the gel into DNA size enabled to determine the DNA size of each nucleosomal band of the ladders and to calculate the expected whole genome average 163 bp NRL of the WT strain (B). (C) The EtBr gel in (A) was transferred onto a membrane and hybridized with the 601 radio-labeled probe to specifically determine the NRL in the 601 repeats. (D) Results of NRL calculation in the 601 repeats for respectively the 167, 197 and 237 strain.

organization in all libraries [25], regardless of the two extent of digestion by MNase (Supplementary Figure S8).

For each the three 601 arrays tested, the pattern of positioning of the nucleosome on the 601 repeat is very reproducible between libraries and does not depend strongly on their digestion level. For the 167 bp repeats most of the nucleosomes are positioned around the middle of the second half of the 601 core sequence (Figure 4A). For the 197 bp repeats, the positioning is more uniform, with a higher density at the boundary between the 601-core end and the linker (Figure 4B). For the 237 bp repeats, we detected three preferential positions centered respectively around 20, 90 and 160 bp (Figure 4C). In summary, nucleosomes are periodically positioned according to the pre-defined NRL only for the 601-167 repeats. However they are not preferentially found on the 601 core sequence. In the two other 601 designs we did not observe an exclusive nucleosome spacing according to the designed NRL, a results which confirms our MNase-gel results (Figure 3C, D).

In vivo assembled 601 repeats are weakly transcribed

Since the transcription machinery is strongly associated with chromatin remodeling complexes that could have a predominant role in nucleosome occupancy, we then checked the transcription activity of our synthetic 601 arrays. To see whether deleting the *YMR262W* promoter upon repeat assembly was sufficient to abolish polymerase II transcription through the assembled array and the remaining of the *YMR262W* gene, we measured genome wide transcription in exponentially growing cells by RNA-seq of polyadenylated or total RNAs, and assessed the transcription level of every gene in the three 601 strains and in the wild-type strain. To quantify how many transcripts carry the 601 sequence, we compared the expression level in the 601 array with the rest of the coding genes by ordering every transcript according to its counts-per-million (cpm) value.

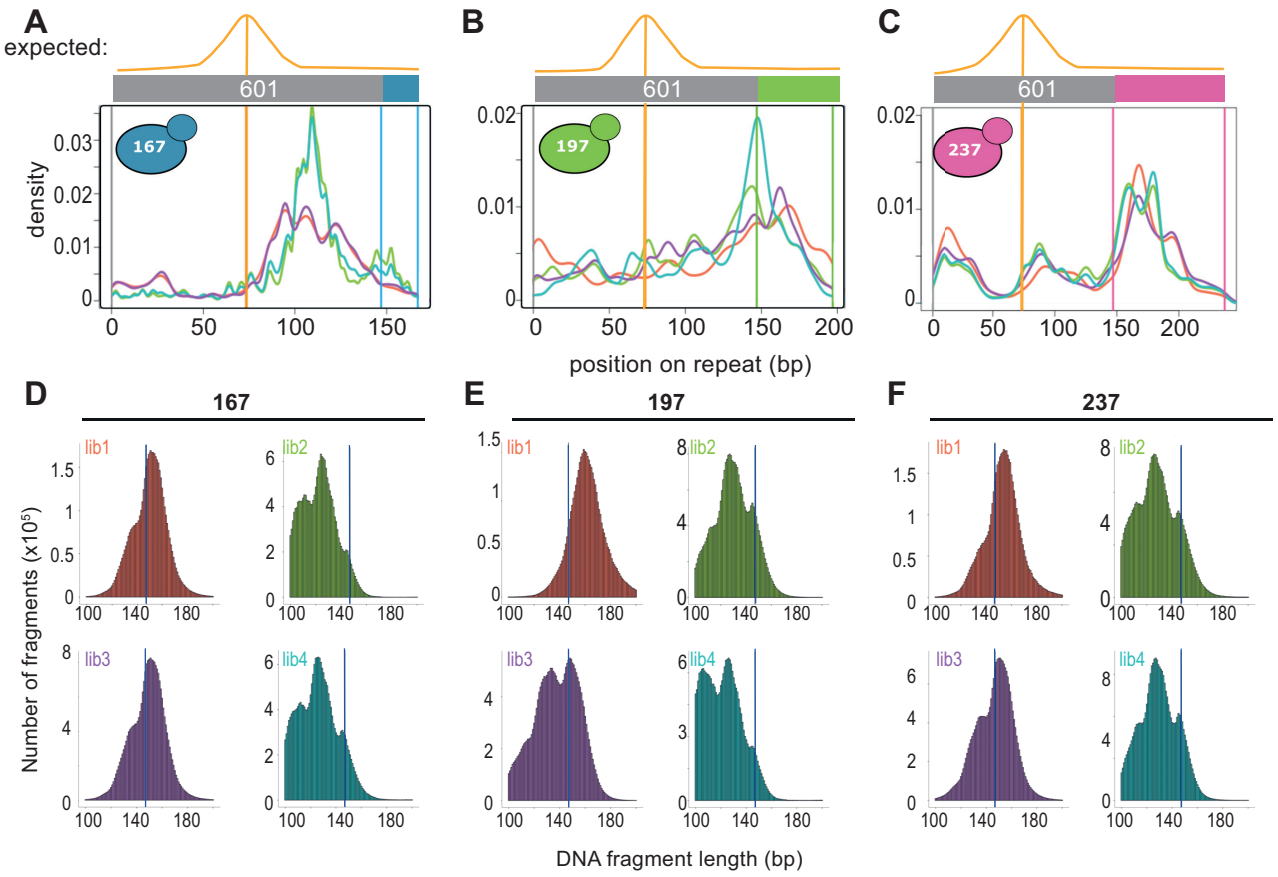

Figure 4. Nucleosomal fragment midpoints are not localized at the middle of 601 cores in the repeats *in vivo*. Midpoints Kernel density curves of each individual library are displayed on the same graphic for each 601-strain: the 601-167 strain (A), the 601-197 strain (B) and the 601-237 strain (C). The orange vertical line marks the 601-core middle and the two other colored vertical lines delimit the linker portion of the corresponding 601 design. The expected distribution is schematized on top of each diagram. The fragment length distribution histograms of the four corresponding libraries are displayed for the 601-167 strain (D), the 601-197 strain (E) and the 601-237 strain (F).

When considering poly adenylated transcripts only, respectively 4.28, 1.94 and 12.7 per million transcripts were identified in the 167, 197 and 237 bp NRL strains. When ordered from high to low expression, transcripts containing the 601 sequence ranked respectively at position 5325, 5522 and 4789, meaning that in each strain more than 80% of the genes expressed are detected at a higher level. If we take into consideration the repeated nature of the array, the number of transcripts should also be divided by the approximate number of repeat within each array (∼50). For total RNAs, more 601 sequences were detected in the 167 and 237 bp arrays, hinting at some transcription in the arrays of non polyadenylated RNA species. Once normalized by the number of repeats, however, this still amounted to a low steady state level compared to most genes (Table 1), supplementary Tables S4, S5).

Table 1. 601 repeats inserted at the *YMR262W* locus are weakly transcribed.

RNA sample	601 array type	cpm	rank / total genes which cpm is > 0	rank 601 monomer/ total genes which cpm is > 0
polyA	167	4.28	5325 / 6068	5915 / 6068
polyA	197	1.94	5522 / 6031	5928 / 6031
polyA	237	12.7	4789 / 6014	5711 / 6014
total	167	19.56	1316 / 6612	5558 / 6612
total	197	2.1	4830 / 6612	5555 / 6612
total	237	38	637 / 6612	5534 / 6612

Discussion

The in vivo spacing between nucleosomes has been proposed to play an important role in many chromosome functions ranging from the definition of chromatin states and gene activity [26] to homologous pairing [27]. In this study we built synthetic DNA arrays in the genome of *S. cerevisiae*, as a first step to address experimentally the biological effects of unnatural nucleosome spacing. Based on the extensive published work on nucleosome positioning in vitro, we chose the 601 DNA sequence as a candidate sequence to drive nucleosome positioning over large arrays of tandem DNA, with three monomer length of respectively 167, 197 and 237 bp. First, we show that using our genome editing approach, we can easily assemble more than 15 kb-long homogeneous arrays of the 601 sequence for the three monomer tested. As expected, we show that identical arrays can position nucleosomes in vitro, at least to some extent (Supplementary Figure S5). However, the 601 sequence does not appear to force nucleosome positioning in vivo on any of the large tandem repeats. On the contrary, results of the MNase-gel and MNase-seq approaches indicate that in the yeast genome, synthetic arrays of 601 DNA do not result in a chromatin fragment with nucleosome position centered on the 601 core, nor phased accordingly to the repeat length of 197 or 237 bp. Nucleosomes are not randomly organized on these synthetic arrays (Figure 4A,B and C), but as general rule they tend to be positioned away from the 601 sequence and overlap more on the linker sequence. Given the average NRLs measured in the MNase-gel assay (171 bp vs 163 bp measured genome wide) and the peaked profiles observed in the MNase-seq analysis, our results suggest an effect of the repeat sequence (our results are not compatible with a 163 bp NRL regardless of the repeat sequence) but not a precise positioning as expected from design. There are several possible explanation for the lack of positioning observed on the 601 core in vivo. First, we tried to evaluate a possible role for transcription in overriding nucleosome positioning on the 601 core, in particular for the 197 bp and 237 bp spacing, which represent an unnatural NRL in *S. cerevisiae*. RNA sequencing of polyA enriched or total RNAs indicate a low level of transcription of RNAs containing the 601 sequence in the three experimental strains. This suggests that transcription does not play a significant role here. A definitive argument against a role for transcription would however require to study the synthesis and decay rates of these RNAs, as the measurement of RNA steady states alone does not assess well the interplay between these two kinetically coupled processes [28, 29].In addition, further studies could be carried out to identify the polymerase(s) responsible for the low level transcription of the 601 repeats and its/their occupancy level on the arrays. Another possible caveat of our attempt to force unnatural nucleosome positioning in *S. cerevisiae* is our choice of monomer length. As shown in [17], the helicoidal nucleosomal arrangement of chromatin fibers reconstituted in vitro can be either right or left handed, for respectively NRL value of 167+10*n or 172+10*n (with n=1,2 or 3). Single-linker measurements proposed quantized linkers of 10n+5 bp, which would suggest a nucleosome spacing of 162 or 172 bp rather than the 167, 197 or 237 like we tested in this work [6, 30, 31]. Based on modeling [17] and structural work [15, 32], we chose 167, 197 bp and 237 bp as repeat length to test in our assay since they were proven to allow regular right handed chromatin folding in vitro. We cannot exclude however that in yeast, chromatin assembly factors would only favor left handed stacks of nucleosomes, which would therefore be allowed to form regularly on repeats of sizes belonging to the series 162, 172, 182 bp. Therefore, we cannot rule out that constraints on chromatin folding in vivo in yeast supersedes sequence driven positioning by the 601 sequence. Our system however opens avenues to explore further such caveat.

Our results therefore emphasize the difference between in vitro and in vivo nucleosome positioning rules. The 601 sequence has been a workhorse for chromatin folding studies in vitro, and this current work serves as an additional cautionary tale in assuming the ability of the 601 sequence to impose nucleosome positioning beyond in

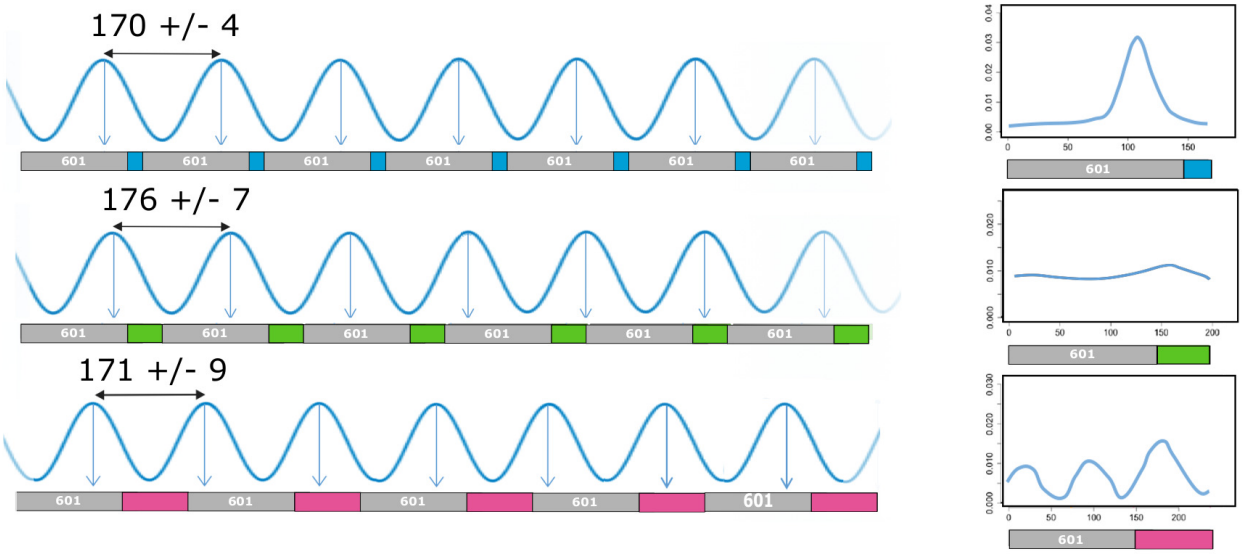

Figure 5. A model of nucleosome positioning on each of the 601 arrays in vivo. Proposition of nucleosome organization over 5-8 repeats (left panels based on both our NRL measures on gel and on the experimental distributions of nucleosomal fragment midpoints on each 601 repeats (right panels).

vitro conditions. Wu and Travers already reported that under certain in vitro conditions, the 601 sequence does not position nucleosomes well. These conditions might be in fact more relevant to physiological conditions [33].

In addition, our results are generally in line with conclusions drawn by Cole et al. when analyzing published studies of nucleosome occupancy on a single 601 sequence (or the closely related 603 sequence) in vivo, in yeast and human [21]. For example, Perales et al. showed that a single 601 element inserted in a yeast coding sequence was largely depleted of nucleosomes, and that presence of neighboring nucleosomes were influenced by transcriptional history [20]. When present episomally in the promoter of a transgene expressed in the mouse liver, strong positioning was only observed early in development when the transgene is active [18]. Therefore, both these studies hinted at the possibility that the transcriptional history or status of the region in vicinity of the inserted nucleosome positioning sequence could override the intrinsic affinity of the 601 core to a nucleosome. Our work is rather in favor of a weak affinity of the sequence in vivo alltogether.

Our results are also consistent with our recent deep learning-based mutational screen of the *S. cerevisiae* genome, which did not point at any DNA motifs that would attract and position nucleosomes in vivo [34]. Secondly, the main driver of nucleosome spacing is not the DNA sequence but the yeast chromatin remodeling factors, which yield a consistent spacing of 163 bp (Figure 3 B). On our synthetic arrays, the observed spacing is slightly larger, ranging from 166 to 183 (Figure 3 D) but is still much smaller than the spacing that would be dictated by the DNA sequence alone for the 197 and 237 bp arrays. Based on these two principles, we can explain that in the case of the 167 bp array, since the periodicity of the sequence is close to the natural periodicity of nucleosome positioning in yeast, most nucleosomes are phased with the sequence and we observe a preferential positioning of the nucleosome centers on the monomer sequence (Figure 5 top). For the 197 bp constructs on the other hand, the spacing imposed by the yeast chromatin remodeling machinery is not in phase with the DNA array and we observe a more uniform distribution of nucleosomal centers, although there is still a slight preference for positioning outside of the 601 core (Figure 5, middle). Finally, in the case of the 237 bp repeat, the impose spacing is such that 3 nucleosomes would occupy approximately 2 DNA repeats, resulting in three fuzzy peaks in the distribution of the nucleosome centers (Figure 5, bottom).

The thermodynamics of the nucleosome positioning process is based on the sum of three different contributions: the energy of formation of a nucleosome on DNA, the DNA sequence affinity for the nucleosome and the long range interaction energy between nucleosomes [35, 36]. It appears from the present study that even for strong in vitro positioning sequences such as the 601, the in vivo energy of formation and long range

interactions between nucleosomes are the most important terms. This does not exclude the possibility to design new DNA sequences that would regularly position nucleosomes within a non-physiological spacing, but these sequences need to be optimized in the context of the in vivo energy terms. To conclude, our study emphasizes that the 601 sequence should not be assumed to maintain nucleosome positioning when inserted in genomic loci in vivo, and new approaches will be needed to design in vivo nucleosome positioning sequences for the engineering of designer chromatin.

Materials and Methods

Strains, plasmids, reagents and media

In vivo genomic assembly of 601 DNA repeats was achieved in the yeast strain YPH499 [37]. Strains derived from this study are listed in Supplementary Table S1. Plasmids used for the in vivo expression of spCas9 and the guide RNA targeting the *YMR262W* gene have been previously described [23] and are listed in Supplementary Table S2. Strains were grown at 30°C in yeast extract Peptone Adenine Dextrose 2% media (YPAD) or in the appropriate synthetic complete Dextrose 2% media (SCD) media minus relevant amino acids necessary to maintain plasmid borne auxotrophic markers. All media reagents were purchased from Formedium and used as recommended. Oligonucleotides used in this study were synthetized by Eurogentec. Enzymes for nucleic acids modification, including MNase, were purchased from New England Biolabs. Zymolyase 20T was purchased from Amsbio.

Assembly of 601 tandem DNA repeats in the yeast genome

Assembly of 601 DNA repeats on an episomal vector in vivo was achieved using a method adapted from the Transformation Associated Recombination method [38]. hl2 μ plasmids carrying 601 repeats assembled with this technique were used for in vitro chromatin reconstitution and single molecule fingerprinting using magnetic tweezers. The full methods relating the in vitro reconstitution and fingerprinting can be found in the Supplementary Material and Methods. Direct in vivo assembly using CRISPR/Cas9 and overlapping oligonucleotides in the Chromosome XIII of *S. cerevisiae* strain YPH499 was performed as previously described [23]. Briefly, donor DNA containing the left (YMR/601) and the right (601/YMR) genomic junction were amplified by PCR using primer couples AL-O-14/15 and AL-O-16/17 respectively. These donor DNA results, upon recombinational assembly, in the deletion of the region -129 to 232 bp of the *YMR262W* gene. All oligonucleotides used for in vivo repeat assembly are listed in Supplementary Table S3. YPH499 was first transformed with the CAS9 expressing plasmid p414-TEF1p-Cas9-CYC1t. This plasmid was a gift from George Church (Addgene plasmid ♯ 43802) [39]. The resulting strain was transformed using the LiAc technique [40] with 1 μg of gRNA expressing plasmid targeting *YMR262W* (pAL31), 100 *pmol* of each of the four or six appropriate 601-oligonucleotides (all oligonucleotides and plasmids are listed in supplementary Table S2 and S3), and 10 *pmol* of both YMR/601 and 601/YMR junction PCR. After transformation cells were plated on SCD-His-Ura to select for cells carrying both CAS9 and gRNA expressing plasmids. To test for correct in vivo assembly of 601-tandem repeated arrays, each genomic junction between the assembled repeated array and genomic DNA were analyzed by Sanger sequencing with primers AL-O-02/22 and AL-O-01/23, to amplify respectively the left YMR/601 junction or the right 601/YMR junction (see Supplementary Figure S2). To confirm assembly at the correct locus and to estimate the size of the assembled arrays, recombinant clones were also analyzed by southern-blotting. Genomic DNA from recombinant strains were digested with BamhI and DraI, which cut at each side of the insertion locus. Digested DNA was electrophoresed in 1% agarose and transferred by capillarity onto a nylon membrane (Hybond N+, GE healthcare). Membranes were hybridized in Church buffer at 68°C with two different ^{32}P-radiolabeled probes (Church and Gilbert, 1984). The 601 probe was made of the 147 *bp* long 601-core sequence. The genomic probe was a 1 *kb* DNA fragment amplified by PCR from genomic DNA using primers AL-O-24/25 (see Table S3). Probes were labeled with the prime-it II random primer labelling kit (Agilent) using α^{32P}-dCTP. For each 167, 197 and 237 design the clone carrying the longest 601 tandem repeats was named respectively ALY1, ALY2 and ALY3 and selected for further analysis (Figure2). Membranes were scanned using a FLA 9500 GE healthcare Scanner at 200 μm resolution.

Chromatin digestion by Micrococcocal Nuclease (MNase)

Each strain was grown to an OD_{600} of 0.8 in 250 mL Synthetic Complete Dextrose 2 % media at 30°C with shaking at 200 rpm. Cultures were treated with a final concentration of 1.85 % formaldehyde for 30 min at 30°C. Cross-linking was stopped by addition of 105 mM Glycine (final concentration) and collected by centrifugation (6500g, 10 min). Cell pellets were washed and resuspended in 50 mL of 1 M Sorbitol, 10 mM Tris pH 7.5 supplemented with 10 mM β-mercaptoethanol and 15 mg of Zymolyase 20T. Cells were further incubated at 30°C with 1h shaking at 100 rpm. Spheroplasts were pelleted (6500g, 10 min), and resuspended in 2.4 mL solution containing 1M Sorbitol, 50 mM NaCl, 10 mM Tris pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$ and 0.75 % Igepal Ca630 freshly supplemented with 1 mM β-mercaptoethanol, 500 μM spermidine and various amounts of MNase (between 6000 and 30000 units), to allow choosing samples with the appropriate digestion pattern for further processing. The spheroplasts/MNase mixture was then incubated at 37°C for precisely 30 min, allowing breaking of the spheroplasts and digestion of chromatin. The reaction was stopped by adding 600 μL of 1% SDS, 10 mM EDTA. Reversal of crosslink and protein removal was achieved by adding 0.6 mg of Proteinase K (Invitrogen) and overnight incubation at 65°C. Samples were extracted with phenol/chloroform, and DNA was ethanol precipitated treated with DNase-free RNase.

MNase-gel analysis

The MNase-gel analysis was performed as follows. Mildly digested MNase resistant nucleosomal DNA was separated by electrophoresis in a 1.3 % agarose gel. After migration, the gel was colored with Ethidium Bromide (0.5μg/mL) and imaged in a Gbox bioImager (Syngene UK). The resulting image was analysed using a simple image analysis routine developed in-house in Matlab to calculate the size of digested nucleosomal DNA relative to Molecular Weight standards (1 kb+ DNA (Thermo Fischer Scientific)). Following gel imaging, the DNA was transferred on a nitrocellulose membrane by capillarity using the Southern blot protocol described above. Membranes were hybridized with a 147 bp α^{32P}-labelled probed and exposed against a phosphor plate for 24-48 hours. Phosphor plates were scanned using a Typhoon FLA 9500 scanner at 200 μm resolution (GE Healthcare). The image obtained was analyzed with the same image analysis routines applied to gels colored with Ethidium Bromide.

Preparation of DNA libraries for MNase-seq analysis

For the preparation of fragment libraries, 500 ng of gel purified mononucleosomal DNA were repaired using the PreCR Repair Mix Kit (New England Biolabs) with 100 μM dNTPs in a 50 μL reaction, as recommended by the manufacturer. The reaction was incubated for 30 minutes at 37°C, purified using the QIAquick PCR Purification Kit (Qiagen) and eluted in 80 μL H_2O. Nucleosomal DNA was 5' phosphorylated by adding 333 μM dNTPs, 50 units of T4 polynucleotide kinase (NEB), 15 units of T4 DNA polymerase (NEB) and 5 units of Klenow DNA polymerase (NEB) to 80 μL of repaired mononucleosomal DNA to a final reaction volume of 120 μL. The reaction was incubated 30 minutes at room temperature, purified using the QIAquick PCR Purification Kit and eluted in 30 μL H2O. To add a 3'-dA to the nucleosomal DNA the 30 μL eluate was completed to 50 μL with 0.2 mM dATP and 15 units Klenow DNA polymerase (3'-5' exo-). The reaction was incubated 30 minutes at 37°C and inactivated 10 minutes at 65°C. Reaction was purified using the QIAquick PCR Purification Kit, and eluted in 20 μL. Adapters were then ligated to the nucleosomal DNA by completing the reaction to 30 μL with 10 mM adapters and 1200 units of T4 DNA ligase. The reaction was incubated at 16°C overnight, followed by inactivation for 20 minutes at 65°C. Ligated DNA was purified by gel electrophoresis. The adapter-ligated nucleosomal DNA was amplified by PCR prior to sequencing, using Illumina PE1.0 and PE2.0 primers. We used only 8 to 15 cycles to minimize possible bias due to PCR amplification. PCR were as follows: 3 μL of adapter-ligated nucleosomal DNA were used in a 50 μL reaction volume containing 5 μL of each Illumina paired-end PCR primers at 2 μM, 1 μL of 10mM dNTPmix and 2 μL of Phusion polymerase (NEB). PCR products were purified using the QIAquick PCR purification column and eluted with 30 μL Qiagen Elution Buffer. DNA concentration was determined using a Qubit Fluorometer. Libraries were multiplexed and then sequenced on a HiSeq 2000 device.

MNase-seq analysis

For each 601-containing strains four individual libraries were sequenced, yielding 5 to 28 Million reads for each library. For the three 601 designs we constructed a reference genome carrying two full repeats plus 65 bp of a third one. Indeed, the construction of the chimeric genome is required to take into account all possible alignments of the reads. Reads for which one mate overlaps the linker and the beginning of the following 601-core will align on a second repeat in our reference genome. Moreover, extreme cases where the 2 mates of a read overlap two linkers will align on the beginning of a third repeat (65 bp were chosen as the length of our read -1 bp) (see Supplementary Figure S1). After removing of barcodes, paired-end reads of 65 nt each end were mapped against the appropriate 601 (167, 197 or 237) reference genome using Bowtie2 (version 2.3.5.1) [41,42] allowing at most 2 alignments per read (-k 2) and fragments length ranging from 100 to 200 bp for 167 and 197 strains, and from 120 to 250 bp for 237 strains (-I and -X parameters). By choosing a reference genome with two repeats and 65 bp of an additional 601, a large proportion of the reads of interest should be able to map twice. For this reason we allowed 2 hits per read using Bowtie2. Concordant reads were specifically selected using samtools (version 1.10). Reads presenting multiple hits on the repeat region were filtered to select the most $5'$ alignment using a python script. Coverage per base-pair was determined in both the whole-genome (WG) and the 601-repeats in each individual library, to calculate the enrichment factor in the repeats (see Supplementary Figure S6). For all the fragments overlapping the 601 area their midpoint positions were localized on the repeats using a python script. Finally, these midpoint coordinates were merged on a unique repeat to represent the averaged nucleosomal dyad density at each position of a single repeat using a custom R script.

RNA libraries preparation

Total RNA was extracted by starting from 25 mL culture of each strain grown to an OD_{600} of 0.5 in SCD media at 30°C with shaking at 200 rpm. Cells pellets were resuspended in 500 μL of Nucleazol solution (Macherey Nagel, 740404.200) followed by 20 minutes agitation in MN Bead Tubes Type C (Macherey Nagel, 740813.50). Total RNA was then purified using the NucleoSpin® RNA Set for NucleoZOL (Macherey Nagel, 740406). RNA concentration and quality was determined using respectively Qubit™ RNA HS assay kit and Qubit™ RNA IQ Assay kit with a Qubit® fluorometer. The first RNA sequencing was performed using purified poly-A mRNA fragmented, reverse-transcribed into 300 bp cDNA before single-end Illumina adaptor ligation. Libraries were sequenced on a HiSeq 4000 (1x50bp) SR device by the Eurofins sequencing platform (INVIEW Transcriptome Explore). Another round of sequencing was performed using total RNAs without poly-A enrichment. After agmentation, RNAs were reverse-transcribed into 300 bp cDNA before single-end Illumina adaptor ligation. cDNA Libraries were then sequenced on a NovaSeq 6000 S4 PE150 XP (2x150bp) by the Eurofins sequencing platform (NGSelect RNA on Illumina).

RNA-seq analysis

We sequenced RNA from the three 601-strain and the wild-type stain during exponential growth phase in YPAD media. Reads were pseudoaligned to the reference protein-coding cDNA collection of the *S. cerevisiae* strain S288C via Kallisto [43] (version 0.46.2) in the single-end mode. The *YMR262W* cDNA sequence of the fasta files used to create index was replaced by the sequence of four 601 repeats of the appropriate design when aligning the three 601 strains. For each of the four biological replicates cpm (counts-per-million) values for every gene were calculated with the edgeR (version 3.30.3) package of Bioconductor [44]. cpm values were normalized using edgeR to correct for sample-specific variation typically introduced by differences in library size. To verify that 601 repeats do not disturb the global transcriptome we calculated the $log2(cpm)$ value of every gene (except the ones for which all samples displayed less than 1 *cpm*) to determine expression fold change between the wild-type strain and the three 601 strains (see Supplementary Table S4 (polyA-RNA) and Supplementary Table S5 (total RNAs)). To estimate the 601 transcription rate we deduced in each 601-strain from the cpm value in the repeats the rank of expression of the 601 region over all genes analyzed in each library (Table 1).

Acknowledgments

The authors thank the members of the Genome Structure and Instability unit for discussions and Dr Romain Koszul for providing assistance with sequencing. The authors also thank the four anonymous reviewers for their insightful comments on the original manuscript. This work was supported by a grant from the french national research agency (MuSeq, ANR-15-CE12-0015), Sorbonne Université (CONVERGENCE - CVG1110) and by core funding from CNRS, INSERM and MNHN to the Structure and Genome instability unit. French Museum of National History to the Genome Structure and Instability unit. AL was supported by a doctoral fellowship from the National Museum of Natural History. Work in the group of VC is part of "Institut Pierre-Gilles de Gennes" ("Investissements d'Avenir" program ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-31) and the Qlife Institute of Convergence (PSL Univesité).

Appendix A. Supplementary Data

Supplementary Methods and Figures are available online at XXX.

Conflict of interest statement.

The authors declare no conflict of interest.

Data availability

Sequencing data are available here: https://www.ncbi.nlm.nih.gov/sra/PRJNA700072

References

1. R. D. Kornberg and Y. Lorch, "Primary Role of the Nucleosome," *Molecular Cell*, vol. 79, pp. 371–375, aug 2020.
2. A. L. Olins and D. E. Olins, "Spheroid chromatin units (v bodies)," *Science*, vol. 183, no. 4122, pp. 330–332, 1974.
3. R. D. Kornberg, "Chromatin structure: A repeating unit of histones and DNA," *Science*, vol. 184, no. 4139, pp. 868–871, 1974.
4. R. D. Kornberg, "Structure of chromatin.," *Annual review of biochemistry*, vol. 46, pp. 931–954, 1977.
5. E. N. Trifonov and J. L. Sussman, "The pitch of chromatin DNA is reflected in its nucleotide sequence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 7, pp. 3816–3820, 1980.
6. J. Widom, "A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 1095–9, feb 1992.
7. P. T. Lowary and J. Widom, "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.," *Journal of molecular biology*, vol. 276, pp. 19–42, feb 1998.
8. P. J. J. Robinson and D. Rhodes, "Structure of the '30 nm' chromatin fibre: a key role for the linker histone.," *Current opinion in structural biology*, vol. 16, pp. 336–43, jun 2006.
9. S. J. Correll, M. H. Schubert, and S. a. Grigoryev, "Short nucleosome repeats impose rotational modulations on chromatin fibre folding.," *The EMBO journal*, vol. 31, pp. 2416–26, may 2012.

10. A. Bancaud, N. Conde E Silva, M. Barbi, G. Wagner, J. F. Allemand, J. Mozziconacci, C. Lavelle, V. Croquette, J. M. Victor, A. Prunell, and J. L. Viovy, "Structural plasticity of single chromatin fibers revealed by torsional manipulation," *Nature Structural and Molecular Biology*, vol. 13, pp. 444–450, may 2006.

11. A. Bancaud, G. Wagner, E. S. N. Conde, C. Lavelle, H. Wong, J. Mozziconacci, M. Barbi, A. Sivolob, E. Le Cam, L. Mouawad, J. L. Viovy, J. M. Victor, and A. Prunell, "Nucleosome chiral transition under positive torsional stress in single chromatin fibers," *Mol Cell*, vol. 27, no. 1, pp. 135–147, 2007.

12. H. Meng, K. Andresen, and J. Van Noort, "Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers," *Nucleic Acids Research*, vol. 43, pp. 3578–3590, mar 2015.

13. D. Vasudevan, E. Y. Chua, and C. A. Davey, "Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence," *Journal of Molecular Biology*, vol. 403, pp. 1–10, oct 2010.

14. A. Kaczmarczyk, A. Allahverdi, T. B. Brouwer, L. Nordenskiöld, N. H. Dekker, and J. Van Noort, "Single-molecule force spectroscopy on histone H4 tail-cross-linked chromatin reveals fiber folding," *Journal of Biological Chemistry*, vol. 292, pp. 17506–17513, oct 2017.

15. A. Routh, S. Sandin, and D. Rhodes, "Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 8872–8877, jul 2008.

16. M. Barbi, J. Mozziconacci, and J. M. Victor, "How the chromatin fiber deals with topological constraints," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, p. 031910, mar 2005.

17. H. Wong, J.-M. M. Victor, and J. Mozziconacci, "An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length.," *PLoS ONE*, vol. 2, p. 8, jan 2007.

18. L. E. Gracey, Z.-Y. Chen, J. M. Maniar, A. Valouev, A. Sidow, M. A. Kay, and A. Z. Fire, "An in vitro-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo.," *Epigenetics & chromatin*, vol. 3, p. 13, jan 2010.

19. A. Subtil-Rodríguez and J. C. Reyes, "BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo," *EMBO Reports*, vol. 11, pp. 751–757, oct 2010.

20. R. Perales, L. Zhang, and D. Bentley, "Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history.," *Molecular and cellular biology*, vol. 31, pp. 3485–96, aug 2011.

21. H. a. Cole, V. Nagarajavel, and D. J. Clark, "Perfect and imperfect nucleosome positioning in yeast.," *Biochimica et biophysica acta*, vol. 1819, pp. 639–43, jul 2012.

22. S. A. Chakraborty, A. A. Kazi, T. M. Khan, and S. A. Grigoryev, "Nucleosome-Positioning Sequence Repeats Impact Chromatin Silencing in Yeast Minichromosomes.," *Genetics*, vol. 198, pp. 1015–1029, sep 2014.

23. A. Lancrey, A. Joubert, and J.-B. Boulé, "Locus specific engineering of tandem DNA repeats in the genome of Saccharomyces cerevisiae using CRISPR/Cas9 and overlapping oligonucleotides," *Scientific Reports*, vol. 8, p. 7127, may 2018.

24. A. L. Hughes, Y. Jin, O. J. Rando, and K. Struhl, "A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning : A Unifying Model for Establishing the Genome-wide Pattern," *Molecular Cell*, vol. 48, pp. 5–15, oct 2012.

25. T. N. Mavrich, I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, J. Qi, S. C. Schuster, I. Albert, and B. F. Pugh, "A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome," *Genome research*, vol. 18, pp. 1073–1083, jul 2008.

26. A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire, and A. Sidow, "Determinants of nucleosome organization in primary human cells.," *Nature*, vol. 474, pp. 516–20, jun 2011.

27. J. G. Hedley, V. B. Teif, and A. A. Kornyshev, "Nucleosome induced homology recognition in chromatin," *bioRxiv*, p. 2021.04.29.441844, apr 2021.

28. D. Wichtowska, T. W. Turowski, and M. Boguta, "An interplay between transcription, processing, and degradation determines tRNA levels in yeast," *Wiley interdisciplinary reviews. RNA*, vol. 4, pp. 709–722, nov 2013.

29. B. Neymotin, R. Athanasiadou, and D. Gresham, "Determination of in vivo RNA kinetics using RATE-seq," *RNA (New York, N.Y.)*, vol. 20, pp. 1645–1652, oct 2014.

30. J.-P. Wang, Y. Fondufe-Mittendorf, L. Xi, G.-F. Tsai, E. Segal, and J. Widom, "Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae.," *PLoS computational biology*, vol. 4, p. e1000175, jan 2008.

31. K. Brogaard, L. Xi, J. P. Wang, and J. Widom, "A map of nucleosome positions in yeast at base-pair resolution," *Nature*, vol. 486, pp. 496–501, jun 2012.

32. P. J. J. Robinson, L. Fairall, V. a. T. Huynh, and D. Rhodes, "EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 6506–11, apr 2006.

33. C. Wu and A. Travers, "Relative affinities of DNA sequences for the histone octamer depend strongly upon both the temperature and octamer concentration," *Biochemistry*, vol. 44, pp. 14329–14334, nov 2005.

34. E. Routhier, E. Pierre, G. Khodabandelou, and J. Mozziconacci, "Genome-wide prediction of DNA mutation effect on nucleosome positions for yeast synthetic genomics," *Genome Research*, p. gr.264416.120, dec 2020.

35. J. Riposo and J. Mozziconacci, "Nucleosome positioning and nucleosome stacking: two faces of the same coin," *Molecular bioSystems*, vol. 8, pp. 1172–1178, mar 2012.

36. D. A. Beshnova, A. G. Cherstvy, Y. Vainshtein, and V. B. Teif, "Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions," *PLoS computational biology*, vol. 10, no. 7, 2014.

37. R. S. Sikorski and P. Hieter, "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae.," *Genetics*, vol. 122, pp. 19–27, may 1989.

38. N. Kouprina, "Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes," *Nucleic Acids Research*, vol. 31, pp. 922–934, feb 2003.

39. J. E. DiCarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, and G. M. Church, "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.," *Nucleic acids research*, vol. 41, pp. 4336–43, apr 2013.

40. J. Hill, K. A. Donald, D. E. Griffiths, and G. Donald, "DMSO-enhanced whole cell yeast transformation.," *Nucleic acids research*, vol. 19, p. 5791, oct 1991.

41. B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nature Methods*, vol. 9, pp. 357–359, apr 2012.

42. B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biology*, vol. 10, mar 2009.

43. N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nature Biotechnology*, vol. 34, pp. 525–527, may 2016.

44. M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, pp. 139–140, nov 2009.