ISSN 0006-2979, Biochemistry (Moscow), 2024, Vol. 89, No. 4, pp. 653-662 © Pleiades Publishing, Ltd., 2024.

Towards Development of the 4C-Based Method Detecting Interactions of Plasmid DNA with Host Genome

Alexandra P. Yan[1,2,a]*, Paul A. Salnikov[1,2], Maria M. Gridina[1,2], Polina S. Belokopytova[1,2], and Veniamin S. Fishman[1,2,b]*

[1]*Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia*

[2]*Novosibirsk State University, 630090 Novosibirsk, Russia*

[a]*e-mail: a.yan@g.nsu.ru* [b]*e-mail: minja-f@yandex.ru*

Alexandra Yan https://orcid.org/0000-0003-0305-0612

Paul Salnikov https://orcid.org/0000-0001-9470-7178

Maria Gridina https://orcid.org/0000-0002-7972-5949

Polina Belokopytova https://orcid.org/0000-0003-1390-7341

Veniamin Fishman https://orcid.org/0000-0002-5573-3100

Received October 11, 2023
Revised February 1, 2024
Accepted March 2, 2024

Abstract—Chromosome conformation capture techniques have revolutionized our understanding of chromatin architecture and dynamics at the genome-wide scale. In recent years, these methods have been applied to a diverse array of species, revealing fundamental principles of chromosomal organization. However, structural organization of the extrachromosomal entities, like viral genomes or plasmids, and their interactions with the host genome, remain relatively underexplored. In this work, we introduce an enhanced 4C-protocol tailored for probing plasmid DNA interactions. We design specific plasmid vector and optimize protocol to allow high detection rate of contacts between the plasmid and host DNA.

DOI: 10.1134/S0006297924040059

Keywords: chromatin interactions, 4C, plasmid, extrachromosomal sequences, plasmid-genome interactions, chromosome conformation capture, transfection

INTRODUCTION

Recent advances in biochemistry, molecular biology, and next-generation sequencing facilitated development of highly efficient methods for probing interphase chromatin architecture. It has become evident that organization of the vertebrate genomes is influenced by at least two distinct mechanisms. The first involves interplay between the CTCF and cohesin factors in *cis* [1], while the second pertains to organization of the genomic compartments through phase separation, mediated by interactions of chromatin proteins in *cis* and *trans* [2]. Despite a few exceptions [3, 4], these mechanisms typically operate in tandem within most vertebrate cells. This combined function makes it difficult to distinguish between the independent factors associated with each mechanism. This poses a particular challenge for studying compartment formation through phase separation, because the generic term "compartment" can be attributed to various nuclear partitions, each characterized by unique chromatin content and formed by a specific biochemical mechanism. Thus, sequence determinants, proteins involved, and additional aspects of chromatin compartmentalization phenomenon are yet to be fully elucidated [5, 6].

Another significant challenge within the field of 3D-genomics is exploration of variability in the chromatin architectures across the tree of life. While evolution of the genome architecture in vertebrates [7] and some invertebrate groups [8] has been relatively

* To whom correspondence should be addressed.

well-studied, data for other species remain limited. Furthermore, there is a lack of information regarding nuclear organization of non-genomic and exogenous sequences, including mitochondrial and plastid DNA, viral genomes, plasmids, mobile elements, and so forth. Nevertheless, existing evidence indicates that spatial packaging of these DNA molecules plays a crucial role in their function, and studying it can uncover important knowledge about the general principles of chromatin biology.

Viral DNA within the nuclei serves as a prominent example of exogenous DNA sequence. For example, it has been established that the viral DNA of herpes B virus (HBV), characterized as episomal covalently closed circular DNA, primarily contacts the A-compartment of the genome. These contacts include promoters, enhancers, and CpG islands of the promoters of highly expressed genes, including those engaged in the cellular response to infection [9]. Conversely, the transcriptionally repressed HBV is associated with the B-compartment regions of the host genome [10]. Viruses not only exist episomally, but can also integrate into the host genome, establishing new *cis*-contacts. This can result in activation of the previously silenced genes [11], or alterations of the native 3D-organization of the genome – either by integration into the existing architecture of protein binding sites or by inserting new ones. Subsequently, these changes can lead to dysregulation of gene activity and cancer development [9, 12].

Another noteworthy aspect of the spatial contacts of exogenous DNA involves their association with progressive multifocal leukoencephalopathy (PML) nuclear bodies. These bodies facilitate viral replication and transcription, a phenomenon demonstrated for various viruses including herpes simplex virus, Simian Virus 40, adenoviruses, and human papillomavirus [13, 14]. The study [15] highlighted that the chaperone HIRA depositing H3.3 promotes repression of the naked foreign DNA such as purified plasmid and viral DNA. This repression is facilitated by the DNA binding with the HIRA histone chaperone complex into the PML-bodies within the host nuclei. Concurrently, other research indicates formation of the viral replication compartments adjacent to PML, where replication and transcription processes are executed [16, 17]. Collectively, the existing body of data underscores functional importance of the three-dimensional contacts between the exogenous DNA and the genome. However, biology of extrachromosomal DNA within the nucleus, especially concerning plasmids, continues to be an area of further exploration.

Information regarding the biology of extrachromosomal DNA, particularly its interaction with proteins and assembly of the virus-associated compartments, is primarily gathered through immunostaining and microscopy. However, this approach is not without its constraints. Resolution offered by these techniques is often inadequate for smaller structures, including certain viral genomes and plasmids which are no more than a few hundred thousand base pairs in size. An alternative method for studying chromatin interactions employs chromatin conformation capture (3C) technologies [18]. These technologies offer better resolution, reaching down to hundreds or thousands of base pairs for Hi-C, and to the scale of hundreds of base pairs for circular chromatin conformation capture (4C). Studies using these methods provide a more detailed view of the spatial organization of DNA.

Here we introduce a novel method to explore localization patterns of plasmid DNA, and potentially other exogenous DNA. This approach is based on the 4C-experiment, utilizing a plasmid as a target vector. Notable advantage of employing the 4C-methodology with a plasmid lies in its capability, as we suggest, to perform further screening experiments – to deliver plasmids with arbitrary insertions of DNA sequences into the cell (active and inactive genome regions, CG-rich and CG-poor regions, Polycomb group protein binding sites, etc.). We can then observe influence of these specific nucleotide sequences on the distribution of plasmids across compartments into the nucleus.

Additionally, this experimental design, involving delivery of an exogenous insert as a component of a plasmid into a cell, effectively negates the *cis*-influence originating from the adjacent genome regions with diverse epigenetic statuses. Thus, it offers a more isolated and accurate assessment of the impacts attributable directly to the inserted sequences.

MATERIALS AND METHODS

Plasmid vectors. Modified pUC19 vector was engineered from the original pUC19 by altering two bases within two *NlaIII* restriction sites using PCR of plasmid fragments with primers containing target substitution, followed by the Gibson Assembly of PCR products into a circular molecule. The p4CSCS vector was synthesized at the Cloning Facility as two sequences (991 bp and 975 bp) that were flanked by homologous regions. These sequences were joined into a circular molecule via Gibson Assembly. Maps and sequences of the plasmids are provided in the Supplementary materials (Figs. S1-S3, Online Resource 1).

Plasmids were propagated in TOP10 *E. coli* cells. We observed that propagation of the p4CSCS vector necessitated reduction in chloramphenicol concentration to, specifically, 0.02 μg of chloramphenicol per 1 ml of the medium.

Human cell culture and transfection. Experiments were conducted using HEK293T cells, which were cultured in a DMEM medium supplemented with

10% fetal bovine serum and 1% Pen Strep (all from Thermo Fisher Scientific, USA). The cells were maintained at 37°C and 5% CO_2.

One day prior to transfection, the cells were passaged into a 75 cm^2 cell culture flask to achieve approximately 70% confluency. Transfection was carried out in a serum-free DMEM medium devoid of antibiotics. For transfection mix, we combined 2.4 µl Opti-MEM (Thermo Fisher Scientific), 18 µg of plasmid DNA, and 36 µg of PEI reagent (2 µg/µl). These conditions were found to allow efficient transfection of HEK293T cells (>50% of transfected cells) in control experiments with a GFP-encoding plasmid (eGFP-n1) co-transfected with the target vector.

4C-protocol. 24 hours after transfection, cells were harvested using 0.05% trypsin-EDTA and fixed by 1% formaldehyde according to the previously published protocol [19]. Cells were counted and aliquots of 3 million cells were snap frozen. We performed key steps of the experiment: cell lysis, chromatin digestion, chromatin ligation, reversal of crosslinking, and DNA purification (2.2.1-2.2.4) according to the protocol [19] with minor modifications:

• for chromatin digestion, 50 units of *NlaIII* was used.

• biotinylation and "dangling end removal" steps were skipped.

Digestion of purified DNA. We used slightly different protocols depending on whether p4CSCS-based vectors (require digestion with *MseI*) or pUC19-based vectors (require digestion with *TaiI*) were transfected into the cells.

1) Digestion in the case of using p4CSCS-based vectors. For purified DNA digestion, the following components were mixed on ice: 5 µl 10× CutSmart Buffer, 30 units of *MseI*, 2 µg DNA, ddH_2O to 50 µl. The mixture was incubated at 37°C with shaking overnight. After *MseI* inactivation by incubation at 65°C 20 min, digested DNA was purified using KAPA Pure Beads (1×) according to the manufacturer recommendation and added to the ligation reaction (step "Ligation of purified DNA").

2) Digestion in the case of using pUC19-based vectors. In 4C-experiments with modified pUC19 we performed digestion of purified DNA by *TaiI*: a sample with 250 ng DNA was supplemented with 1 µl of 10× restriction buffer 5 units of *TaiI*, and ddH_2O to 10 µl. The mixture was incubated at 37°C with continuous shaking overnight. Digestion was terminated by heat inactivation of the restriction enzyme at 65°C for 20 min. The entire volume of the reaction mixture (10 µl) was added to the ligation reaction. Ligation of the digested DNA was performed as described in "Ligation of purified DNA".

Ligation of purified DNA. For ligation of purified DNA the following components were mixed on ice: 10 µl 10% Triton X-100, 10 µl of 10× T4 DNA ligase reaction buffer, 10 µl 25% PEg, and 1 µl of 10 mg/ml bovine serum albumin, 1 µl 100 mM ATP, 800 units of T4 ligase, 0.5-2 µg DNA, ddH_2O to 100 µl (final DNA concentration was about 5-20 ng/µl). The mixture was incubated at 16°C overnight, then DNA products were purified using KAPA Pure Beads (0.8×).

(optional) Additional digestion with DrdI. In our first rounds of experiments, which were all performed with pUC19-based vectors and *TaiI* digestion, we assume that the fraction of non-informative ligation products could be reduced by *DrdI* treatment. Although later we found that this assumption was not correct, we indicate here that after ligation, products were treated with *DrdI* under the following conditions: 100 µl of ddH_2O and 12 units of *DrdI* endonuclease were added to the ligase mixture followed by incubation at 37°C overnight; next *DrdI* was inactivated by incubation at 80°C for 20 min and DNA products were purified using KAPA Pure Beads (0.8×).

This step was omitted when using p4CSCS-based vectors.

Preparation of 4C-libraries for NGS. PCR with primers containing Illumina adapter sequences was performed for 15-20 cycles. Sequences of the primer are provided in the Supplementary materials (Online Resource 1). PCR products were purified using KAPA Pure Beads (0.8×). Distribution of DNA-fragments in the obtained 4C-libraries was evaluated by visualization of the gel after electrophoresis in 1% agarose.

DNA concentration was measured with a Qubit fluorimeter. For experiments with modified pUC19 plasmid we prepared 4C-libraries for NGS by nested PCR: we performed the first PCR round with i_F1, i_R1 primers (see Online Resource 1), then PCR products were purified using KAPA Pure Beads (0.8×) and one third of the amount of PCR products served as a DNA matrix for the second PCR round with P5 and P7 primers for Illumina sequencing, see Online Resource 1).

Experiments with control chromatin and control plasmids. HEK293T cells were transfected by a modified pUC19 vector using PEI and fixed 24 h post-transfection, as described above. Fixed *Mus musculus* fibroblasts in approximately equal amounts were introduced to the fixed, transfected HEK293T cells and co-lysis was conducted as described in the 4C-protocol above. 200 ng of the control plasmid (modified pUC19 with a cytomegalovirus (CMV) promoter) was added to the sample before *NlaIII* digestion. Next, we performed 4C-experiment as described above for all these components together.

4C-libraries with Klenow treatment. To distinguish products formed due to incomplete digestion by *NlaIII*, and digested but self-ligated products, we included Klenow treatment step into the 4C-protocol described above. For this purpose, after *NlaIII* inactivation,

centrifugation, and supernatant removal, the following components were mixed on ice: 20 µl ×10 NEBuffer 3.1, 1.5× µl of each 10 mM dNTPs (dATP, dTTP, dCTP, dGTP), 50 units of Klenow, ddH_2O to 200 µl.

The mixture was added to the precipitate, resuspended, and incubated at 16°C for 4 h. Then the experiment was continued from the stage of "Blunt end ligation" of the Hi-C 2.0 protocol [19] up to and including the DNA purification stage and further according to the above described 4C-protocol with plasmid.

4C-data processing and analysis. Raw reads were processed using cutadapt and then aligned to a reference that included the human hg38 genome and plasmid vector sequences (either modified pUC19 or p4CSCS). For experiments involving control chromatin, the reference also incorporated the mouse mm10 genome, modified pUC19, and modified pUC19 with the CMV-promoter sequence. Alignment was performed using the *BWA mem*. Alignment statistics were gathered with *samtools idxstats*. To visualize the alignment results, we employed IGV 2.16.1 software.

In our protocol, two ligation events should occur: first ligation at the *NlaIII* site at the fixed chromatin condition and second ligation at the *MseI* or *TaiI* site during the circularization step. Products of the first ligation reaction represent spatial colocalization of DNA fragments in the nucleus, whereas the second ligation can either occur as intramolecular ligation producing circular DNA or intermolecular ligation between random DNA ends. Thus, products of the second ligation do not add information about spatial contacts of the plasmid, but may reflect random collisions resulting from diffusion processes. Therefore, we discarded sequences ligated to the *MseI* or *TaiI* site, and only used sequences ligated to the *NlaIII* site. According to the plasmid vector design, these sequences always start from the p7-adapter (Figs. S1-S3, Online Resource 1).

RESULTS

Design of 4C-experiment with plasmid. We aimed to develop an approach based on using a 4C-method to identify contacts between the plasmid DNA and the host genome. Design of the experiment is shown in Fig. 1a. Initially, cells are transfected with a plasmid vector. We anticipate that within the nuclei, plasmids are secured by host cellular proteins, and biophysical characteristics of these proteins dictate localization of the plasmid (I). Subsequently, spatial contacts within chromatin, including interactions between the plasmid and DNA, are fixed with formaldehyde (II). According to the 4C-protocol (detailed in methods), sequential digestion and proximity ligation reactions are conducted on the fixed chromatin (III-IV). After this, the cross-links are reversed; the DNA is digested with a second restriction enzyme and ligated in solution (V). This results in formation of circular molecules, containing a fragment of the plasmid DNA and a segment of the genome with which the plasmid was in contact (VI). By amplification and sequencing of these circular templates using NGS technology, we expect to detect interactions between the plasmid DNA and the genome.

According to the described design, the vector used in the 4C-experiment should include two distinct restriction sites: the first is hydrolyzed during the initial DNA digestion in the fixed chromatin (E1, Fig. 1a), and the second during the subsequent digestion of the plasmid DNA following cross-links reversal (E2, Fig. 1a). A crucial aspect to consider is placement of the PCR primer annealing region, which should be located between these two restriction sites, near to their positions. Furthermore, there must be neither E1 nor E2 restriction sites between the primer annealing regions. Moreover, both E1 and E2 should be frequent cutter sites, i.e., 4-base cutters, to allow high resolution of the 4C-analysis (determined by E1 site frequency), and efficient amplification of 4C-fragments (determined by E2 site frequency). Finally, the E1-specific enzyme has to be able to digest formaldehyde-fixed chromatin.

To validate the proposed design, we performed 4C-experiments with the pUC19 plasmid vector. We used *NlaIII* and *TaiI* restriction sites as E1 and E2 target sites of the experiment, respectively. In addition to the two designated *NlaIII* restriction sites, the pUC19 vector contains additional *NlaIII* and *TaiI* sites (nine and three, respectively), including those present in the region for PCR primer annealing (Fig. 1b). We expected that additional *NlaIII* and *TaiI* sites outside the target region (where binding sites of the primers used for library amplification are located) would not interfere with the 4C-experiment, as opposed to the sites located between the primers. Therefore, we engineered a modified vector based on pUC19, wherein the *NlaIII* (-CATG-) restriction sites located between the primers were removed (Fig. 1b).

Estimating background noise level in 4C-protocol with plasmid. The 4C-experiment design with plasmid assumes that restriction and ligation stages occur under conditions of fixed chromatin, therefore the resulting products reflect spatial co-localization of the plasmid DNA and the host genome inside the nucleus. However, it is also possible that part of the contacts is formed in solution as a result of diffusion processes after chromatin fixation, causing random collisions and ligations of the plasmid and the genomic DNA fragments. These contacts do not reflect biological preferences of the plasmid distribution in the nucleus.

To estimate the scale of such ligation events, we performed a control experiment, shown in Fig. 2a. Human HEK293T cells, transfected with the modified

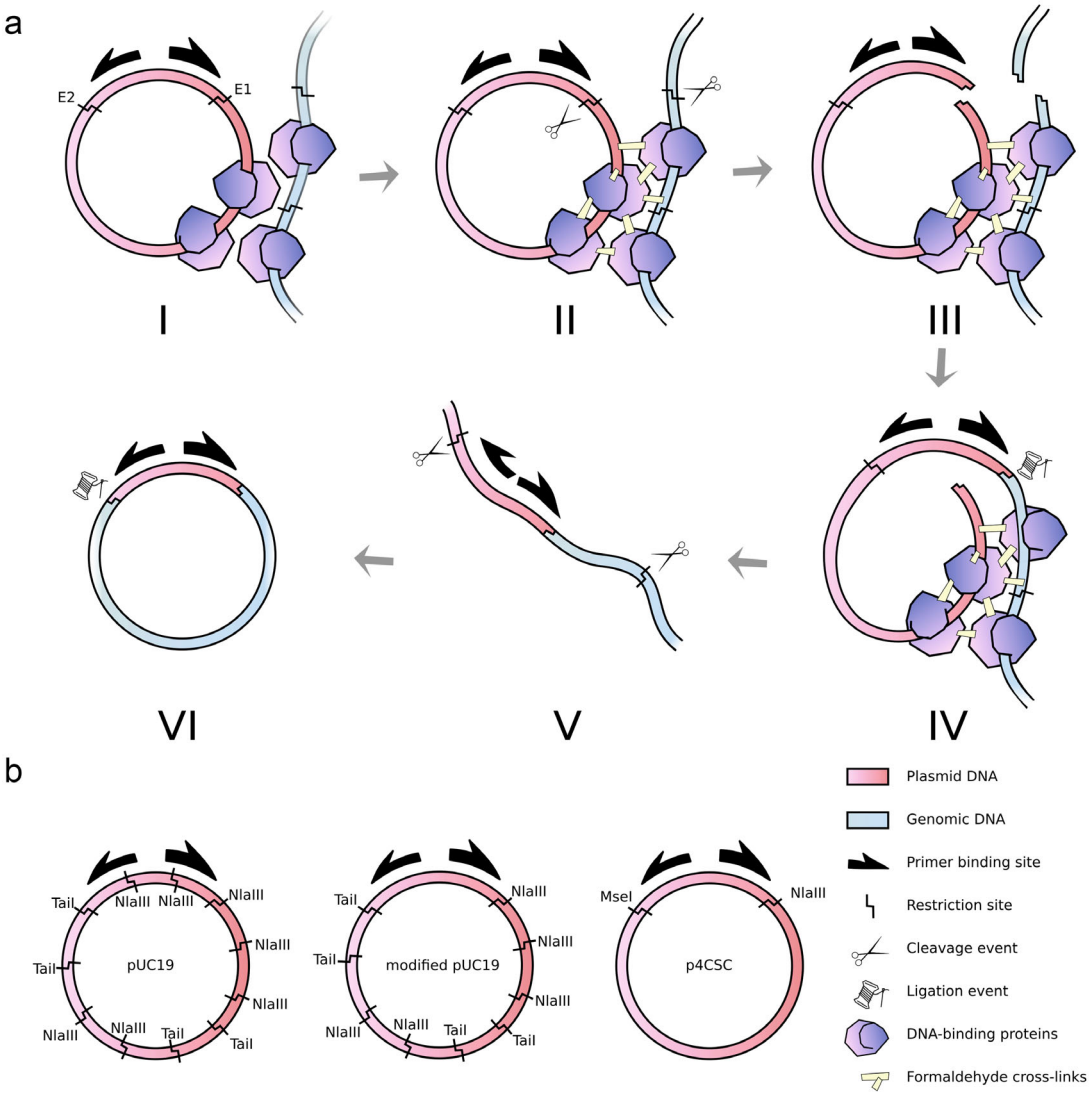

Fig. 1. Design of 4C-experiment with plasmid vector. a) [I-VI] overall experiment design. b) Restriction sites scheme of the plasmid vectors generated in this work. Plasmid maps and sequences are provided in the Supplementary materials (Figs. S1-S3, Supplementary text 2, see Online Resource 1).

pUC19 vector, underwent chromatin fixation 24 h post-transfection. Next, the control chromatin (fixed chromatin of a non-human species) and the control plasmid (plasmid with the sequence different from pUC19 vector) were added to these fixed HEK293T samples. Finally, we prepared 4C-libraries as described in the experiment design section above (Fig. 1a). Given that the control chromatin of non-human species was added after fixation, any ligations of the pUC19 vector (which was transfected pre-fixation) with non-human chromatin represent random collision events. Similarly, ligations of the control plasmid (added post-fixation) with the HEK293T chromatin and with chromatin of a non-human species are due to random collisions as well.

Therefore, the contacts of plasmid no. 1 (modified pUC19 transfected into the nucleus) reflect both spatial localization of the plasmid in living cells (by ligation with human chromatin) and possible random diffusion processes (by ligation with chromatin of human and non-human species). The contacts of plasmid no. 2 added to the cells after fixation with chromatin are not conditioned by biological regularities and are only determined by random diffusion processes (by ligation with human chromatin and with chromatin of a non-human species).

We compared how often the plasmid no. 1, transfected into human cells, and plasmid no. 2, added to the cells after fixation, contact with the human and non-human chromatin. Our results, presented in Fig. 2b, demonstrate that 80% of the plasmid no. 1 contacts are with the human chromatin, whereas the plasmid no. 2 interacts with the human chromatin only in 64% of cases (percentages are average for three replicas;

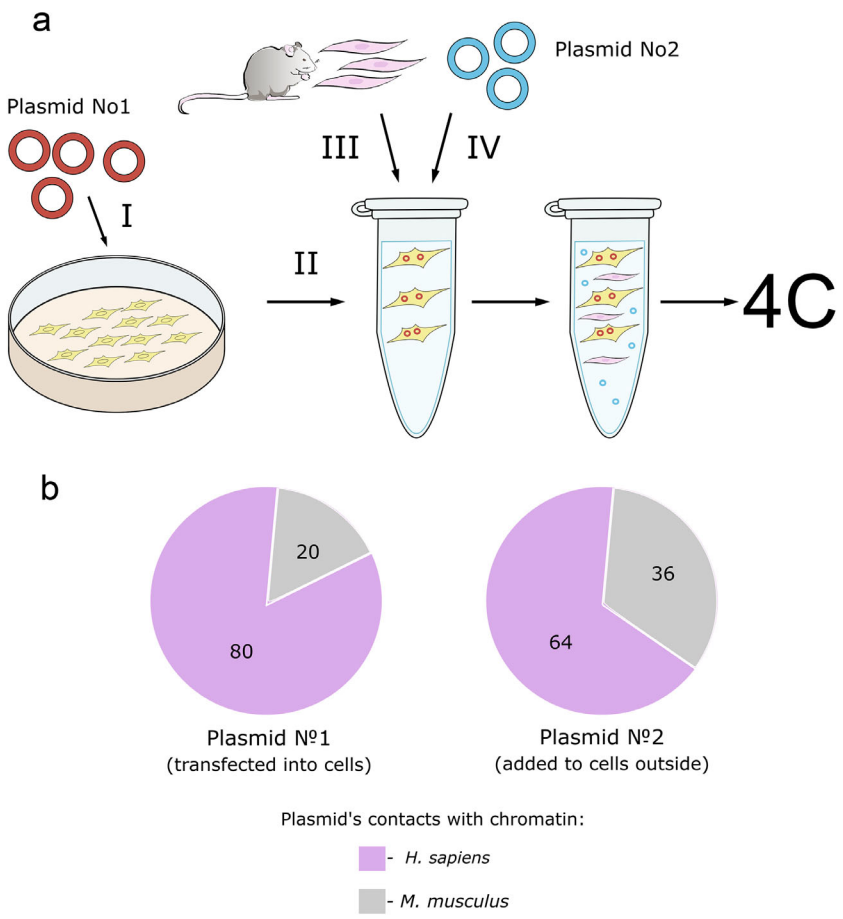

Fig. 2. Estimation of "random" contacts fraction. a) Design of the experiment: plasmid no. 1 was transfected into HEK293T cells (I), then spatial contacts of DNA and proteins in the nucleus were fixed with formaldehyde (II). Fixed *Mus musculus* fibroblasts (III) and plasmid No. 2 (IV) were added to the obtained sample, and the 4C-experiment was performed. b) Plasmid no. 1 and plasmid no. 2 differ in the ratio of the number of contacts with *H. sapiens* and *M. musculus* chromatin The diagrams show percentage of the plasmid contacts with the corresponding chromatin, averaged over three independent experiments. Results for the individual replicates are presented in Fig. S4 in the Online Resource 1. Total number of contacts is taken as 100%.

for all three replicas difference of contact frequencies was significant according to Fisher's exact test, *p*-value is negligible; see Supplementary notes for details, Online Resource 1). These observations suggest that part of the contacts between the plasmid transfected into cells and the chromatin detected in the experiment is due to the spatial proximity in the nucleus, i.e., cannot be explained by random collisions of the plasmid DNA with DNA in solution. We also quantify the fraction of random collisions (see Supplementary notes, Online Resource 1).

4C-vector optimization. Analyzing the data obtained in the 4C-experiment with the modified pUC19 we faced the problem that the majority (about 90%) of the reads did not contain the human genome sequence. Instead, the reads aligned to the specific regions of the plasmid. This issue was not only observed when using control chromatin and control plasmid, but also in the experiments involving HEK293T cells transfected with the modified pUC19 vector and subjected to the 4C-protocol without controls. Detailed analysis of the reads distribution and structure revealed that this phenomenon could be traced back to the ligation of various plasmid fragments during the in-chromatin ligation stage. It occurs because the plasmid DNA is not only hydrolyzed at the intended *NlaIII* and *TaiI* restriction sites, but also at other *NlaIII* and *TaiI* sites present in the plasmid. Under conditions promoting proximity ligation, fragments originating from a single plasmid molecule are more likely to undergo ligation in *cis*, a phenomenon we have named "self-ligation." Consistent with the significant prevalence of *cis*-interactions over the *trans*-contacts, which is typical for the 3C-data [20], reads containing only plasmid sequences make up to about 90% of the total data.

High frequency of ligations occurring between the plasmid DNA fragments, as described above, can be attributed to the predominance of *cis* (within the plasmid) over *trans* (plasmid-genome) spatial contacts and therefore is unavoidable in the 3C-experiment.

However, this problem could be overcome if the plasmid vector lacks multiple digestion sites. In such cases, *cis*-ligation would restore the original plasmid molecule, which would then be excluded during the PCR and NGS steps due to its length (>1000 bp). Therefore, to solve the problem of plasmid self-ligation, we aimed to redesign the plasmid sequence to contain only single *NlaIII* and *MseI* restriction sites (in this vector design, we use *MseI* instead of *TaiI* at the purified DNA digestion step). This task is challenging because both enzymes are frequent cutters (which is essential for the resolution and detectability of contacts, see above). Consequently, multiple cut sites are present in all functional elements of the available plasmid vectors.

To overcome this limitation, we substituted the ampicillin resistance gene with the chloramphenicol resistance cassette, devoid of restriction sites in the promoter region. However, the replication origin and body of the chloramphenicol resistance gene still contained *NlaIII* and *MseI* restriction sites. We introduced single-nucleotide substitutions within the gene body, ensuring the amino acid sequence remained unaltered. Choice of alternative nucleotides was based on the codon frequencies in *E. coli* [21]. In the replication origin, the substitutions were made arbitrarily, given the absence of the straightforward criterion to predict their functional implications. As a result, we designed a sequence of a new plasmid adapted for 4C-experiment, containing only two target restriction sites, p4CSCS: (plasmid for 4C with Single Cutter Sites), depicted in Fig. 1b (plasmid map and sequence see in Fig. S3 and in supplementary text 2, Online Resource 1). It contains primer binding regions and essential elements for reproduction in *E. coli*, such as replication origin and chloramphenicol resistance gene. Total length of the vector is 1826 bp.

We used this improved plasmid vector in the 4C-experiment with HEK293T cells. The obtained data contained a substantial proportion of contacts between the plasmid and the genome, approximately 81%. Thus, we confirm that the observed in previous experiments high frequency of reads mapped to the plasmid was due to the presence of multiple restriction sites in the vector. Obtained results clearly show the advantage of the redesigned plasmid sequence for 4C-experiments.

Although contacts between the plasmid and the genome are well represented, reads including exclusively plasmid sequences persist, accounting for about 20% of the data. According to visualization in IGV, the majority of these reads are mapped upstream of the *NlaIII* restriction site. This observation raises the question of the origin of these reads: whether they are formed due to incomplete digestion by *NlaIII*, or because the plasmid molecules were hydrolyzed but then ligated back to form a whole plasmid molecule again?

To confirm which hypothesis is accurate, we prepared 4C-libraries where chromatin was treated with the Klenow fragment of *E. coli* DNA polymerase I after digestion by the *NlaIII* enzyme. Hydrolysis with the *NlaIII* enzyme produces 3′-ends that are removed by the 3′-5′ exonuclease activity of the Klenow fragment. Subsequent ligation of the plasmid ends generates a specific sequence signature distinct from the undigested plasmid sequence. Therefore, sequencing the ligation products should distinguish between the molecules from the undigested plasmid (containing the sequence 5′...TCTGAC-CATG-AGGAGA...3′ with the original *NlaIII* 5′...-CATG-...3′ restriction site subsequence) and the molecules that were hydrolyzed by *NlaIII* but then ligated back (containing the Klenow ends resection signature 5′...TCTGAC-AGGAGA...3′, without the *NlaIII* restriction site).

Sequencing data from six independent experiments showed that the number of non-hydrolyzed products (with a mean fraction of 4.82% of all reads analyzed) was almost 30 times greater than the number of molecules that were hydrolyzed and ligated back together (mean fraction of 0.17% of all reads analyzed). This indicates that the reads containing an intact plasmid sequence are mainly due to inefficient hydrolysis by the *NlaIII* enzyme.

DISCUSSION

In this study, we present a method that enables detection of the contacts between the plasmid DNA introduced into the cell and the host genome. Our preliminary experiments show promising results: the 4C-protocol allows efficient enrichment of the contacts between the plasmid and host genome DNA. Although interactions representing contacts between the genomic loci are about a million times more frequent than between the genome and plasmid (according to the ratio of human genome length to the approximate length of both modified pUC19 and p4CSCS plasmid), a large portion of the contacts detected in the 4C-experiment contains plasmid sequences. Moreover, optimal design of the plasmid vector allows to enrich the plasmid-genome interactions over the plasmid-plasmid interactions, though the latter occur in *cis* and therefore is expected to have higher frequency. Looking ahead, precision and informational yield of this method could be augmented through refinement of the chromatin digestion process. This assumption is supported by our observation that the fraction of non-informative products originates from the undigested plasmid.

In addition, it should be noted that the noise in 3C-data is expected to be uniformly distributed and, therefore, should not hinder detection of the increased frequency of contacts between the plasmid and specific

genomic loci. However, high noise level may require deeper sequencing to detect specific interactions.

We envision several applications of the developed method.

• First, the proposed method can be used to investigate localization of the plasmid molecules themselves. Despite the widespread usage of plasmids as vectors for exogenous gene expression in genetic engineering and research, distribution of plasmids in the nucleus remains largely unknown [22-24].

• Second, the proposed experiment design allows us to test how DNA sequences with different epigenetic properties (active and inactive genome regions such as GC-rich, GC-poor regions, Polycomb, and HP1 binding sites, etc.) embedded in the plasmid will affect plasmid localization. This may be applicable to the study of compartments formed as a result of phase separation in the nuclei. Although it is clear that the active and inactive chromatin segregate in the nucleus, it is still an open question how many phases (or subcompartments) are formed in total [25, 26]. Phase separation of the nucleus is being actively studied from the protein perspective. Compartmentalization can also be considered from the viewpoint of DNA motifs to which proteins involved in phase separation are bound. Several DNA motifs were shown to be essential for the loci compartmentalization. For example, DNA sequences that define compartmentalization profile for the Polycomb-repressed loci in Drosophila embryos were identified using the Hi-C and ChIP-seq methods and confirmed using genome editing [27]. Besides that, genetic determinants of compartmentalization remain unknown. We suggest that the 4C-method with plasmid could be applied to identify genetic determinants of compartmentalization by testing effect of different DNA sequences on the pattern of plasmid contacts with the genome. In perspective, the method could be used as a basis for large-scale screening studies and will allow us to compile a profile of the contribution of each locus to chromatin compartmentalization throughout the genome.

• DNA methylation alters three-dimensional organization of the genome and chromatin accessibility by recruiting proteins such as HP1, which form heterochromatin blocks as a result of phase separation [25, 28, 29]. Methylation of CTCF binding sites disrupts its association with DNA, leading to changes in the genome architecture [30]. Overall, CpG methylation affects the ability of TFs to bind to DNA in both negative and positive ways *in vitro* [31]. However, this has not been confirmed *in vivo* [28]. We propose integration of the 4C-method with the *in vitro* 5mC-methylated plasmid DNA as a strategic approach for *in vivo* localization of the methylated DNA loci and clarification whether they form distinct subcompartments by clustering together.

• The 4C-method with plasmid can be used to study DNA repair. A plasmid with modified nucleotides, such as 8-oxoguanine inclusions or other modifications, is likely to be localized in the specific DNA repair compartments assembly of which is initiated by the activated PARP1 and FUS [32]. The 4C-experiment can be used to capture the loci attracted to the DNA repair compartment, as well as dynamics of its formation.

To summarize, we are optimistic that the refined 4C-experiment with plasmid, as described in this study, would be instrumental in the domains of 3D-genomics, transcription regulation, and epigenetics.

Supplementary information. The online version contains supplementary material available at https://doi.org/10.1134/S0006297924040059.

Acknowledgments. We acknowledge infrastructure and resources provided by the collective usage center of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 121031800061-7 (Mechanisms of genetic control of development, physiological processes, and behavior in animals) for running next generation sequencing experiments. Computational data analysis was performed using the high-throughput computing nodes of the Novosibirsk State University (supported by the Ministry of Science and Higher Education of the Russian Federation, grant no. FSUS-2024-0018).

Contributions. V.F., P.S., and A.Y. conceived the study. A.Y. performed experiments with the help from P.S. and M.G. P.B. developed a computational data analysis pipeline. A.Y. performed data analysis with the help from P.B., P.S., and V.F. supervised the study and analyzed the data. A.Y., P.S., and V.F. wrote the manuscript. All authors edited the manuscript and approved its final version.

Funding. This work was supported by the Russian Science Foundation (project no. 22-14-00242).

Ethics declarations. This work does not contain any studies involving human and animal subjects. The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Kabirova, E., Nurislamov, A., Shadskiy, A., Smirnov, A., Popov, A., et al. (2023) Function and evolution of the loop extrusion machinery in animals, *Int. J. Mol. Sci.*, **24**, 5017, doi: 10.3390/ijms24055017.
2. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation, *Proc. Natl. Acad. Sci. USA*, **115**, E6697-E6706, doi: 10.1073/pnas.1717730115.

3. Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., et al. (2019) 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin, *Nucleic Acids Res.*, **47**, 648-665, doi: 10.1093/nar/gky1103.
4. Ryzhkova, A., Taskina, A., Khabarova, A., Fishman, V., and Battulin, N. (2021) Erythrocytes 3D genome organization in vertebrates, *Sci. Rep.*, **11**, 4414, doi: 10.1038/s41598-021-83903-9.
5. Razin, S. V., and Gavrilov, A. A. (2020) The role of liquid-liquid phase separation in the compartmentalization of cell nucleus and spatial genome organization, *Biochemistry (Moscow)*, **85**, 643-650, doi: 10.1134/S0006297920060012.
6. Kantidze, O. L., and Razin, S. V. (2020) Weak interactions in higher-order chromatin organization, *Nucleic Acids Res.*, **48**, 4614-4626, doi: 10.1093/nar/gkaa261.
7. Nuriddinov, M., and Fishman, V. (2019) C-InterSecture-a computational tool for interspecies comparison of genome architecture, *Bioinformatics (Oxford, England)*, **35**, 4912-4921, doi: 10.1093/bioinformatics/btz415.
8. Lukyanchikova, V., Nuriddinov, M., Belokopytova, P., Taskina, A., Liang, J., et al. (2022) Anopheles mosquitoes reveal new principles of 3D genome organization in insects, *Nat. Commun.*, **13**, 1960, doi: 10.1038/s41467-022-29599-5.
9. Dias, J. D., Sarica, N., Cournac, A., Koszul, R., and Neuveut, C. (2022) Crosstalk between hepatitis B virus and the 3D genome structure, *Viruses*, **14**, 445, doi: 10.3390/v14020445.
10. Tang, D., Zhao, H., Wu, Y., Peng, B., Gao, Z., et al. (2021) Transcriptionally inactive hepatitis B virus episome DNA preferentially resides in the vicinity of chromosome 19 in 3D host genome upon infection, *Cell Rep.*, **35**, 109288, doi: 10.1016/j.celrep.2021.109288.
11. Sokol, M., Wabl, M., Ruiz, I. R., and Pedersen, F. S. (2014) Novel principles of gamma-retroviral insertional transcription activation in murine leukemia virus-induced end-stage tumors, *Retrovirology*, **11**, 36, doi: 10.1186/1742-4690-11-36.
12. Razin, S. V., Gavrilov, A. A., and Iarovaia, O. V. (2020) Modification of nuclear compartments and the 3D genome in the course of a viral infection, *Acta Naturae*, **12**, 34-46, doi: 10.32607/actanaturae.11041.
13. Everett, R. D. (2013) The spatial organization of DNA virus genomes in the nucleus, *PLoS Pathog.*, **9**, e1003386, doi: 10.1371/journal.ppat.1003386.
14. Corpet, A., Kleijwegt, C., Roubille, S., Juillard, F., Jacquet, K., et al. (2020) PML nuclear bodies and chromatin dynamics: catch me if you can!, *Nucleic Acids Res.*, **48**, 11890-11912, doi: 10.1093/nar/gkaa828.
15. Rai, T. S., Glass, M., Cole, J. J., Rather, M. I., Marsden, M., et al. (2017) Histone chaperone HIRA deposits histone H3.3 onto foreign viral DNA and contributes to anti-viral intrinsic immunity, *Nucleic Acids Res.*, **45**, 11673-11683, doi: 10.1093/nar/gkx771.
16. Schmid, M., Speiseder, T., Dobner, T., and Gonzalez, R. A. (2014) DNA virus replication compartments, *J. Virol.*, **88**, 1404-1420, doi: 10.1128/JVI.02046-13.
17. Charman, M., and Weitzman, M. D. (2020) Replication compartments of DNA viruses in the nucleus: location, location, location, *Viruses*, **12**, 151, doi: 10.3390/v12020151.
18. Kempfer, R., and Pombo, A. (2020) Methods for mapping 3D chromosome architecture, *Nat. Rev. Genet.*, **21**, 207-226, doi: 10.1038/s41576-019-0195-2.
19. Belaghzal, H., Dekker, J., and Gibcus, J. H. (2017) Hi-C 2.0: an optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation, *Methods (San Diego, Calif.)*, **123**, 56-65, doi: 10.1016/j.ymeth.2017.04.004.
20. Gridina, M., Mozheiko, E., Valeev, E., Nazarenko, L. P., Lopatkina, M. E., et al. (2021) A cookbook for DNase Hi-C, *Epigenet. Chromatin*, **14**, 15, doi: 10.1186/s13072-021-00389-5.
21. Gvritishvili, A. G., Leung, K. W., and Tombran-Tink, J. (2010) Codon preference optimization increases heterologous PEDF expression, *PLoS One*, **5**, e15056, doi: 10.1371/journal.pone.0015056.
22. Prajapati, H. K., Kumar, D., Yang, X.-M., Ma, C.-H., Mittal, P., et al. (2020) Hitchhiking on condensed chromatin promotes plasmid persistence in yeast without perturbing chromosome function, *bioRxiv*, doi: 10.1101/2020.06.08.139568.
23. Gracey Maniar, L. E., Maniar, J. M., Chen, Z.-Y., Lu, J., Fire, A. Z., et al. (2013) Minicircle DNA vectors achieve sustained expression reflected by active chromatin and transcriptional level, *Mol. Ther.*, **21**, 131-138, doi: 10.1038/mt.2012.244.
24. Dean, D. A. (1997) Import of plasmid DNA into the nucleus is sequence specific, *Exp. Cell Res.*, **230**, 293-302, doi: 10.1006/excr.1996.3427.
25. Mladenova, V., Mladenov, E., and Russev, G. (2009) Organization of plasmid DNA into nucleosome-like structures after transfection in eukaryotic cells, *Biotechnol. Biotechnolog. Equip.*, **23**, 1044-1047, doi: 10.1080/13102818.2009.10817609.
26. Hildebrand, E. M., and Dekker, J. (2020) Mechanisms and functions of chromosome compartmentalization, *Trends Biochem. Sci.*, **45**, 385-396, doi: 10.1016/j.tibs.2020.01.002.
27. Erdel, F., and Rippe, K. (2018) Formation of chromatin subcompartments by phase separation, *Biophys. J.*, **114**, 2262-2270, doi: 10.1016/j.bpj.2018.03.011.
28. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M., and Cavalli, G. (2018) Polycomb-dependent chromatin looping contributes to gene silencing during Drosophila development, *Mol. Cell*, **71**, 73-88. e5, doi: 10.1016/j.molcel.2018.05.032.

29. Mattei, A. L., Bailly, N., and Meissner, A. (2022) DNA methylation: a historical perspective, *Trends Genet.*, **38**, 676-707, doi: 10.1016/j.tig.2022.03.010.
30. Rountree, M. R., and Selker, E. U. (2010) DNA methylation and the formation of heterochromatin in *Neurospora crassa*, *Heredity*, **105**, 38-44, doi: 10.1038/hdy.2010.44.
31. Phillips, J. E., and Corces, V. G. (2009) CTCF: master weaver of the genome, *Cell*, **137**, 1194-1211, doi: 10.1016/j.cell.2009.06.001.
32. Singatulina, A. S., Hamon, L., Sukhanova, M. V., Desforges, B., Joshi, V., Bouhss, A., Lavrik, O. I., and Pastré, D. (2019) PARP-1 activation directs FUS to DNA damage sites to form PARG-reversible compartments enriched in damaged DNA, *Cell Rep.*, **27**, 1809-1821, doi: 10.1016/j.celrep.2019.04.031.