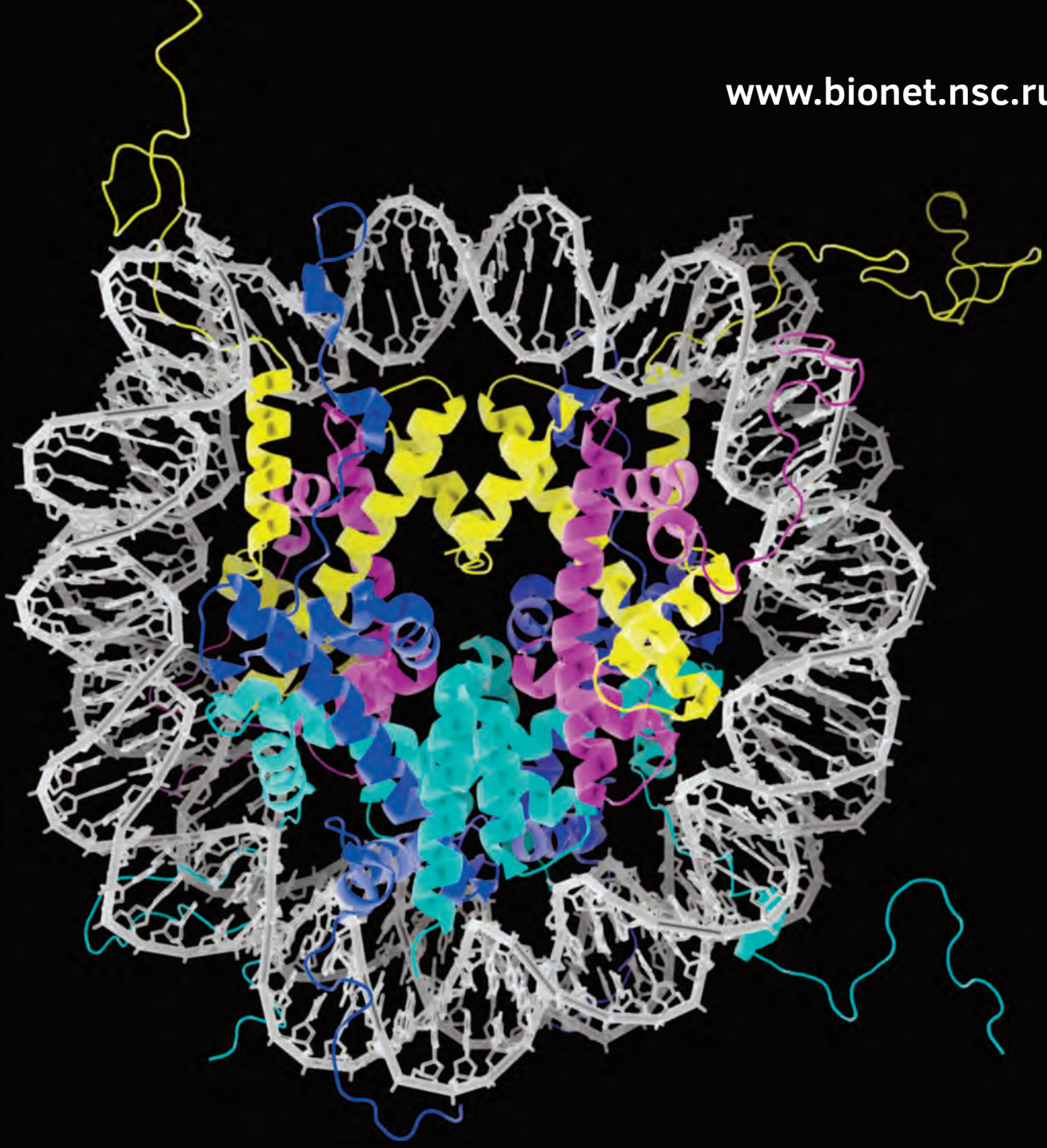
www.bionet.nsc.ru/en



Eukaryotic gene expression
a hurdle-race through chromatin

THE FEDERAL RESEARCH CENTER
Institute of Cytology and Genetics
OF SIBERIAN BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES

Eukaryotic gene expression
a hurdle-race through chromatin

In my previous article "Histone chaperones: the builders of chromatin" in the August 2015 issue of Adjacent Government (page 112) I briefly introduced a compactification problem for eukaryotic DNA, which comes down to a simple question of how to package 2 meters of human DNA into 10 μm of the cell nucleus. Here, I will expand on this challenging problem and discuss how DNA is fitted into cell nucleus, and what are the consequences of DNA packaging for gene works…

DNA packaging problem

DNA is a double helix molecule composed of A:T and G:C base pairs (bp), and a sequence of A,T,G,C nucleotides constitutes a genetic text. In humans, the genetic text is written in 23 volumes of DNA molecules and its size is over 3 billion of nucleotides. Each DNA molecule is present in two copies in diploid cells: one from the mother and one from the father. The average length of these DNA threads is ~ 4 cm and combined they total 2 m. In principle, if we place all DNA molecules in water they will coil randomly into equilibrium globule with the radius of ~400 μm. Then, with little energy this globule can be squeezed into a nuclear volume. However, in such globule the threads of DNA will be tangled and their separation in the process of cell division will be complicated. This does not look a promising solution and nature has invented a number of elegant solutions to the DNA packaging problem. The one, which is employed by most of eukaryotes, including humans, involves the wrapping of short DNA stretches of ~147 bp on protein spools, which are composed from eight histone proteins (Figure 1). In total, there are about 25 million of such spools, also known as nucleosomes, repetitively assembled on human DNA. The resulting structure resembles a collection of beads on DNA string, when visualised by electron microscopy, and thereof it is named "beads on a string". The thickness of nucleosome is ~10 nm and DNA packaged into an array of nucleosomes forms a 10 nm chromatin fiber, which provides a ground level of DNA packaging.

As soon as DNA is packaged into 10 nm of chromatin fiber, it can fold further into 30 nm helical fiber in a test tube, due to molecular interactions between neighbouring nucleosomes (Figure 1). This fiber is composed of two helically coiled stacks of nucleosomes connected by short stretches of nucleosome-free (linker) DNA, zigzagging back and forth between those stacks. The whole existence of such helix in living cells remains a contentious issue, as it is difficult to study fine chromatin structure within cells and to reproduce all cellular conditions in experiments. Nonetheless, a 30 nm chromatin helix demonstrates a pervasive propensity of biological molecules to self-assemble into various geometric shapes.

DNA packaged into nucleosome array and folded into 30 nm chromatin fiber is still too long to fit into nuclear interior. Thus additional level of compaction is required. Again, as we already noted, crumpling of all of these fibers into a bowl of spaghetti will result in extensive tangling between DNA molecules, which is undesired during cell divisions when DNA molecules need to be separated quickly. Fortunately, however, there are many alternatives to the random globule and recent studies indicate that inside nucleus chromatin fibers pack into fractal globules. Fractals are found everywhere in nature and represent curves and shapes with repeating patterns that emerge at every scale. A good representation of a fractal globule is a Koch's snowflake (Figure 1). If we zoom in on the snowflake edges we will find the same pattern we see when we glance at it from the distance. In cells, the fractal globule may result from the combination of short and long-range interactions

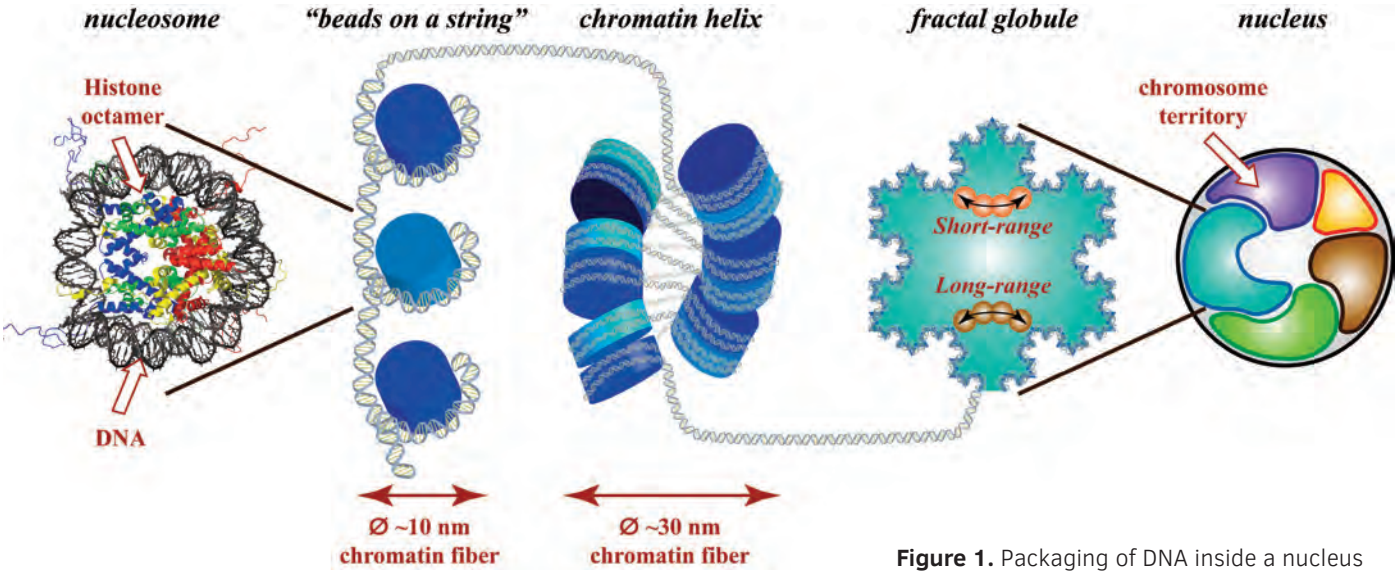


Figure 1. Packaging of DNA inside a nucleus

between pairs of loci on chromatin fiber. This is possible due to the fact that chromatin fiber attracts many non-histone nuclear proteins that tend to find each other in the space and bring together remote sites. Measurements of chromatin pairing revealed an important rule that probability of contacts between two loci depends inversely on the distance between them: $P(s) \sim s^{-1}$, where s – is the distance and P(s) is the probability of pairing. This power law is a mathematical property of fractals, and thus the packaging of DNA inside the nucleus obeys fractal geometry.

The folding of the 10 or 30 nm chromatin fiber into a fractal globule has a few advantages. Firstly, computer simulations of simultaneous folding of long DNA molecules into a fractal globule suggest that such configuration warrants the separation between DNA threads. Inside nucleus, simulated fractal chromatin globules occupy individual territories, which is consistent with microscopically observed chromosome territories in living cells (Figure 1). Secondly, within, fractal globule DNA molecules are unknotted, which simplifies their separation when cells divide. Finally, as a consequence of fractal unentanglement, fractal globule can be locally unfolded at little energy costs making DNA segments accessible for regulation on demand.

In conclusion, inside 10 μm nucleus 2 meters of DNA threads are, firstly, packaged in nucleosome arrays and, then folded into fractal globules. However, there is a trade-off between packaging and accessibility of DNA to molecules that read and copy the genetic texts.

Nucleosome positioning – setting the hurdles for molecular race

Packaging of DNA into nucleosome arrays creates a barrier for many DNA binding proteins to access their sites on DNA coiled around a histone octamer spool. Firstly, DNA covered by nucleosome bonds with histone proteins at many sites and to access these sites, the energy is needed to break the bonding between DNA and the histones. Secondly, DNA endures extreme bending upon wrapping of a short piece of DNA (~147 bp) about two times around a histone core. Eventually, this generates local distortions in DNA double helix, which are incompatible with recognition of DNA by many sequence-specific DNA binding proteins. Thus, modulation of nucleosome positioning along the DNA becomes critical for regulation of eukaryotic genome functions, such as expression of genes.

What determines the placement of nucleosomes on DNA is an actively investigated topic[1]. In general, all we know is that a number of forces shape nucleosome arrays to achieve desired configuration with respect to nucleosome positions. Firstly, like many DNA-binding proteins, histones have some preferences towards certain DNA sequences. DNA

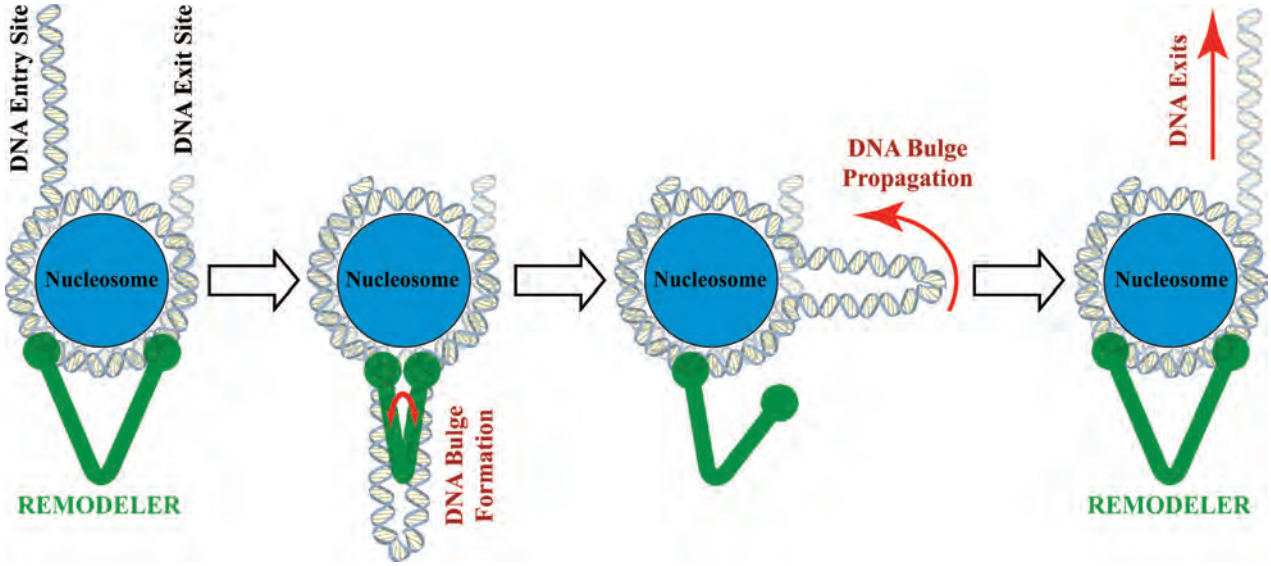


Figure 2. Repositioning of a nucleosome by ATP-dependent chromatin remodelling enzyme

must bend sharply to embrace a histone spool, thus DNA molecules susceptible to curve are favoured for nucleosome assembly. A degree to which DNA string can flex depends on the nucleotide sequence and therefore nucleosome positioning depends on DNA sequence. In fact, bioinformatics analysis makes it possible to predict a distribution of nucleosome particles along the DNA with good accuracy just from its sequence.

Secondly, however, nucleosome arrays in cells are subjected to forces that move nucleosomes from the preferred positions[2]. One of such forces comes from the action of ATP-dependent chromatin remodelling enzymes (remodelers). Fuelled by the energy from ATP hydrolysis, remodelers grab DNA wrapped on a histone spool and pull it causing DNA from the entry site to slide inward. This extra DNA bulges from the surface of nucleosome and propagates around exiting on the other side. As a result the nucleosome shifts to a new position or falls off the DNA (Figure 2). The action of remodelers exposes DNA sequences wrapped on a histone spool to DNA-binding proteins. This, in turn, facilitates all cellular processes requiring access to DNA, such as gene expression, DNA replication and so on.

Finally, nucleosome arrays assembled on transcriptionally active genes are synchronised and phased relative to the transcription start site – TSS (Figure 3). Such phasing cannot be explained according to the sequence rules and it is not observed on silenced genes. However, such alignment of nucleosome core particles can be described by borrowing analogies and methodologies from physics. As an example, all planets in the solar system are kept in orbits by the gravity well of Sun, or more generally – by its potential well. If we remove Sun, all planets will disarray. On active genes, RNA polymerase, a large macromolecular machinery which is responsible for reading the genetic text and synthesis of RNA, lands on DNA at TSS. This causes the +1 nucleosome, the first nucleosome from TSS, to clamp creating a potential well, against which all neighbouring (+2, +3, etc.) nucleosomes align (Figure 3). In fact, by using a toolbox of statistical physics, we can model nucleosome phasing on active genes reproducing experimentally observed nucleosome arrays configuration on active genes[3]. Nucleosome phasing occurs also on many other sites on DNA occupied by DNA-binding protein complexes. Thus, phasing of nucleosomes off wells and barriers along with the remodelers activity counteracts sequence-dependent nucleosome positioning. Combined these forces shape nucleosome landscape in living cells.

The hurdle-race for RNA polymerase through nucleosome barriers

As soon as nucleosomes are placed and phased on DNA, a hurdle-race begins. Racing through a gene RNA polymerase (RNAP) has to overcome

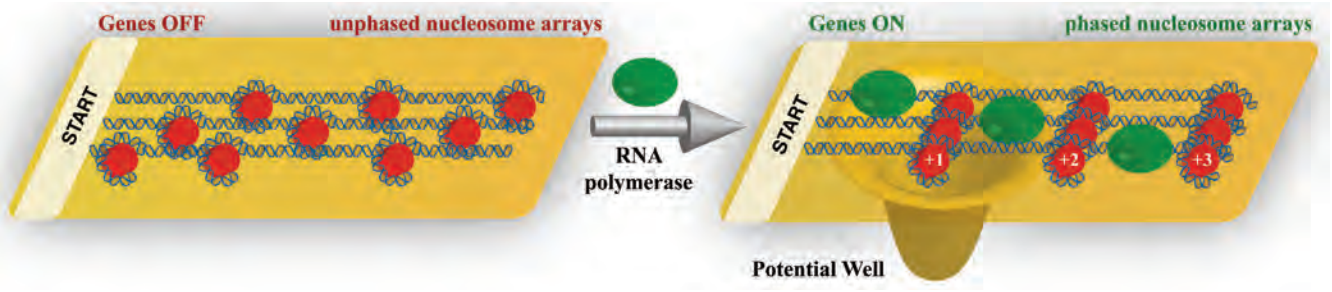


Figure 3. On active genes nucleosomes are phased off a potential well established at +1 nucleosome relative to genes' starts

nucleosome barriers. Traversing through chromatin alone, RNAP is blocked by nucleosomes and needs assistance from histone chaperones, which partially disassemble histone spool to lower the nucleosome barrier. Once the barrier is lowered, RNAP takes it by spooling through nucleosome particles and continues to the next hurdle. Then, the process is repeated and the hurdle-race goes on until RNAP reaches the finish (gene end).

It is reasonable to suggest that repositioning the hurdles will interfere with a speed at which RNAP travels through a gene and, thereof, with gene expression rates. At start sites (TSS) of active genes, nucleosomes are phased against the potential well; however, spacing between them simply depends on a ratio of histone spools to DNA length. The number (concentration) of histone spools and the spacing between nucleosomes varies between species and cell types. Thus regulation of histone protein levels may play an important role in adaptation of gene expression levels to cellular demands. Besides this, remodelers exert a significant impact on nucleosome positioning and spacing on active genes. Consequently, removal of remodelers from cells affects the expression of thousands of genes and often leads to cell malfunction or even death.

Repositioning of nucleosome hurdles in disease

From what we've just discussed, it must become apparent that the positioning of nucleosomes along the DNA plays a pivotal role in cells ability to read and copy the genetic texts and thus to function normally. The alterations in nucleosome positioning may cause a number of diseases such as cancer. Indeed, mutations in the genes encoding for remodelers are often found in the patients diagnosed with cancer. For example, a strict association has been established between mutations in *SNF5* gene encoding for SWI/SNF remodeler subunit and the development of malignant rhabdoid tumor, one of the most aggressive malignancies in pediatric oncology. In fact, this is one of the rare examples when tumor development occurs in children and is caused by mutation in a single gene. Thus, investigation of the principles governing nucleosome positioning is not only an important task for basic science, but it may also help to understand how human diseases develop and to combat them.

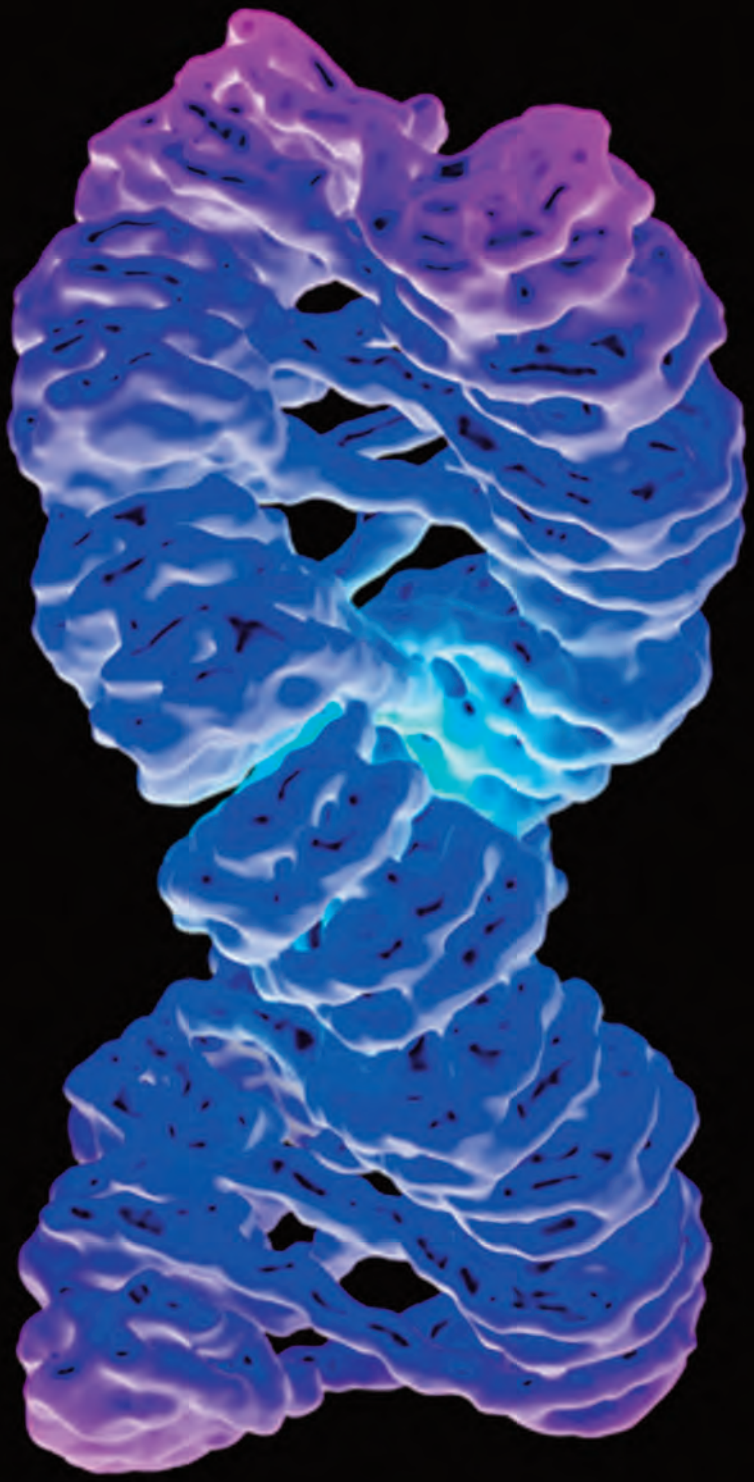
1 Y.M. Moshkin, Chromatin – a global buffer for eukaryotic gene control, AIMS Biophysics, (2015) 2(4): 531-554.

2 Y.M. Moshkin et al., Remodelers organise cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner, Molecular and Cellular Biology, (2012) 32: 675-688.

3 R.V. Chereji et al., Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster, Nucleic Acids Research (2015) doi: 10.1093/nar/gkv978.

Funding: This article has been sponsored by Intellect Partnership Fund for Biomedical Research

Dr Yuri Moshkin
Department of Translational Genetics
The Federal Research Center
Institute of Cytology and Genetics
Prospekt Lavrentyeva 10
630090 Novosibirsk
Russian Federation
yury.moshkin@gmail.com



www.bionet.nsc.ru/en

Department of Translational Genetics
The Federal Research Center
Institute of Cytology and Genetics
Prospekt Lavrentyeva 10
630090 Novosibirsk
Russian Federation