Supplementary Information

Probing of nucleic acid compaction using low-frequency Raman spectroscopy

Andrey Yu. Sosorev[1,2,3*], Olga D. Parashchuk[1], Ivan V. Chicherin[4], Artem A. Trubitsyn[1], Vasiliy A. Trukhanov[1], Maria V. Baleva[4], Ulyana E. Piunova[4], Oleg G. Kharlanov[1], Piotr Kamenski[4], Dmitry Yu. Paraschuk[1]

1. *Faculty of Physics, Lomonosov Moscow State University, Leninskie Gory 1/62, Moscow 119991, Russia*
2. *Enikolopov Institute of Synthetic Polymeric Materials, Russian Academy of Sciences, Profsoyuznaya 70, Moscow 117393, Russia*
3. *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Ulitsa Miklukho-Maklaya, 16/10, Moscow 117997, Russia*
4. *Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1/12, Moscow 119234, Russia*

*e-mail: sosorev@physics.msu.ru

S1. Multiscale simulations of Raman spectra and dynamic disorder

As mentioned in the Methods section of the main text, polarizability tensors $\alpha_n^{(0)}$ of the five nucleobases n = A, G, C, T, U were calculated using DFT (Fig. S1) in their relaxed geometries, along with the derivatives $d\alpha_n^{(0)}/d\mathbf{x}_{n,i}$ with respect to the atomic coordinates. To find an isometry—a rotation matrix $R_n(t)$ and a translation vector $\mathbf{\Delta}_n(t)$—optimally mapping the equilibrium atomic positions $\mathbf{x}_{n,i}^{(0)}$ to the MD-generated instantaneous ones $\mathbf{x}_{n,i}(t)$, the following expression was used:

$$R_n = Z_n\left(Z_n^{\mathrm{T}}Z_n\right)^{-1/2}, \qquad \mathbf{\Delta}_n = \bar{\mathbf{x}}_n - \bar{\mathbf{x}}_n^{(0)},$$
$$(Z_n)_{\alpha\beta} \equiv \sum_i \left(\mathbf{x}_{n,i} - \bar{\mathbf{x}}_n\right)_\alpha (\mathbf{x}_{n,i}^{(0)} - \bar{\mathbf{x}}_n^{(0)})_\beta \,, \qquad \bar{\mathbf{x}}_n \equiv \frac{1}{N_n}\sum_i \mathbf{x}_{n,i}\,, \qquad \bar{\mathbf{x}}_n^{(0)} \equiv \frac{1}{N_n}\sum_i \mathbf{x}_{n,i}^{(0)},$$

where N_n is the number of atoms in the nucleobase (except the methyl substituent). Such a transformation minimizes the mean-squared deviation of the instantaneous nucleobase geometry from a rotated and translated reference geometry $\mathbf{x}_{n,i}^{(0)}$. With the rotation matrix thus found, the polarizabilities of the nucleobases were then calculated in the linear approximation in the residual, non-rigid transformations of the geometry, as described in the Methods section of the main text.
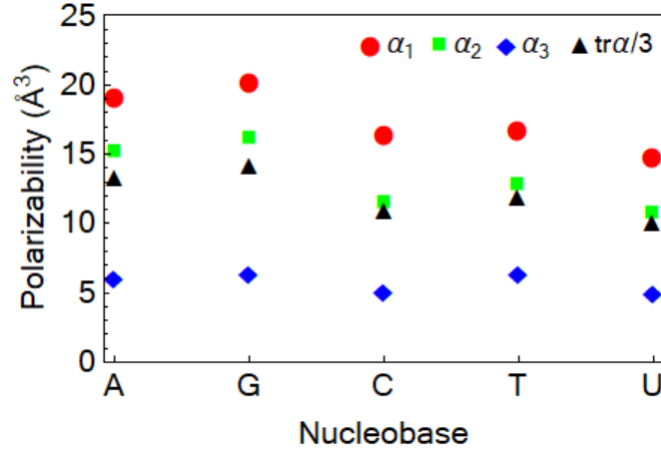
Fig. S1. Principal polarizabilities (i.e. eigenvalues of the polarizability tensor α) for various nucleobases obtained at the B3LYP/6-31G(d,p) level of DFT, in Å^3. Triangles show the mean polarizability averaged over directions in space.

For the reference, we also quote here how one can estimate the R ratio and the polarizability disorder from the calculated MD trajectories. Note, however, that, in application to estimation of biomolecule compaction, calculation of R (R') using MD is an overkill, since MD—if provided with an accurate enough initial geometry, force field, and necessary computational resources—is able to access the compaction states and dynamic disorder more directly (e.g., via geometric measures or RMS deviation analysis, Fig. S2). Once deviations $\delta\alpha(t)$ of the MD polarizabilities from their running means are calculated, and the Fourier transform $C(\omega)$ of their autocorrelator $\langle \text{tr}(\delta\alpha(t)\,\delta\alpha(t+\tau))\rangle$ is found, the LF Raman intensity, up to a frequency-independent prefactor, equals (see, e.g., Ref. [37] in the main text):

$$I(\omega) \propto C(\omega)\frac{\bar{n}(\omega)+1}{\langle q^2\rangle_{\text{cl}}}.$$

In this expression, the $\bar{n}(\omega)+1$ population factor involving a Planck distribution $\bar{n}(\omega) = \left(e^{\hbar\omega/k_B T} - 1\right)^{-1}$ describes a Stokes line of the Raman spectrum at temperature T ($\hbar$ and k_B are the Planck and Boltzmann constants, respectively). The denominator $\langle q^2\rangle_{\text{cl}} = k_B T/\hbar\omega$ cancels the variances of the dimensionless vibrational coordinates present in the polarizability autocorrelator, which is in our case evaluated using *classical* (as opposed to quantum) dynamics. The R ratio can be found directly from the above spectral intensity. Moreover, assuming that the HF intensity is roughly independent of the NA conformation and is proportional to the number of nucleotides, one arrives at an estimation

$$R \propto \frac{1}{N_{\text{res}}}\int_{\text{LF}}^{\square} I d\omega,$$

which can potentially be used in MD simulations of NAs. Similarly, the spectral density of the polarizability disorder can be found from the Wiener–Khinchin theorem,

$$\frac{d(\sigma_\alpha^2)}{d\omega} = C(\omega)\frac{\langle q^2\rangle}{\langle q^2\rangle_{\text{cl}}} = C(\omega) \times \frac{\hbar\omega}{2k_B T}\coth\frac{\hbar\omega}{2k_B T},$$

so that thermal fluctuations of the NA polarizability can be estimated as

$$\sigma_\alpha^2 \sim \int_{\text{LF}}^{\square}\frac{d(\sigma_\alpha^2)}{d\omega}d\omega \propto N_{\text{res}}R,$$

which additionally supports a statistically-inspired disorder metric $\sigma_\alpha/N_{\text{res}}^{1/2}$ chosen in the main text.
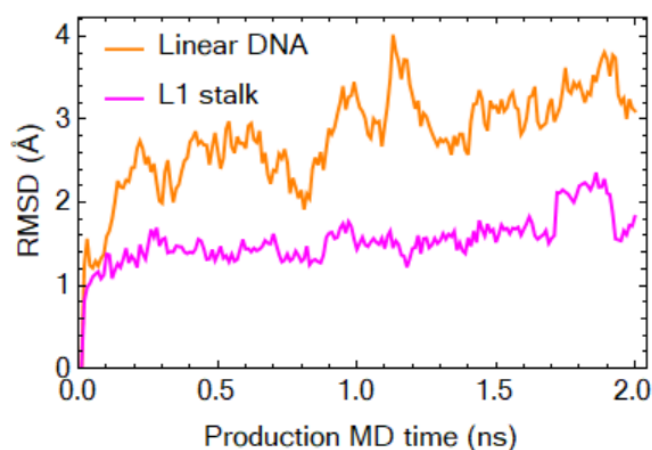
Fig. S2. RMS deviation analysis results for linear DNA and L1 stalk during 2 ns of the production MD stage. The results indicate a stronger dynamic disorder in DNA compared with L1 stalk.

S2. Preparation of ribosome samples

Bacterial cells from 1 liter culture with an OD_{600} of 0.6 were collected by centrifugation, suspended in a buffer solution (50 mM Tris-HCl pH 7.4, 100 mM potassium chloride, 5 mM magnesium acetate) and lysed by sonication. The lysates were pre-clarified by centrifugation (20000 g, 15 minutes) and sedimented through a 10% sucrose cushion for 18 hours at 80000 g on the Optima XPN centrifuge, SW40Ti rotor (Beckman Coulter). The pellets were dissolved in a minimal volume (500–700 µl) of buffer solution (10 mM Tris–HCl pH 7.0; 60 mM potassium chloride, 60 mM ammonium chloride, 7 mM magnesium acetate, 7 mM β-mercaptoethanol and 0.25 mM EDTA) and separated on 15–40% sucrose gradient in similar conditions. The gradient was divided into fractions of 300 µl and A_{260} was measured. Fractions containing a peak of ribosomes were pooled and collected by ultra-centrifugation.

S3. Processing of experimental Raman spectra

Processing of the spectra, including background subtraction, was done with WIRE 3.4 software using the built-in module for cubic polynomial interpolation. The continuous broadband background was subtracted from all spectra in the range of 10–1800 cm^{-1}. All spectra, in addition to the broadband continuous background, contained a wide broadband LF signal (LFWBS); therefore, the procedures for background subtraction in the LF and HF ranges were somewhat different. Within 200–1800 cm^{-1}, the broadband background was interpolated with a set of cubic polynomials, and the resulting interpolation function was extrapolated to the range 10–200 cm^{-1}. The spectra within 10–200 cm^{-1} were analyzed both with the LFWBS and after its subtraction also by cubic polynomial interpolation, to analyze individual Raman bands in the LF range (Section 2). To stitch the LF and HF spectra together, the most intense bands were chosen in the overlap region of 100–450 cm^{-1}. Spectra recorded in the range 10–450 cm^{-1} were multiplied by such a factor that the band amplitudes were equal to the amplitudes at their maxima measured in the range of 100–1800 cm^{-1}, and all of the entire overlap region was consistent with both measurement series. To calculate the R and R' ratios, we summed the Raman intensities in the region 10–200 cm^{-1} from the spectra with ($\int_{LF} I d\omega$) and without ($\int_{LF} I' d\omega$) LFWBS, correspondingly. Also we divided every value of Raman intensities from the region 200–2000 cm^{-1} to the corresponding value of the Raman shift and summed them. Finally, we calculated the ratios $R = \int_{LF} I d\omega \ / \ \int_{HF}(I/\omega)d\omega$ and $R' = \int_{LF} I' d\omega \ / \int_{HF}(I/\omega)d\omega$, correspondingly.

S4. Raman spectra of DNA and RNA samples and onion cells

The HF Raman spectra of the DNA samples (Fig. 3a) demonstrate its characteristic bands: ~500 (PO_2 scissor), 1100 (PO_2 symmetric stretch), 728 (ring mode of A), 805 (O-P-O stretch), 1316 (ring mode of A, G), 1332, 1366, 1480 and 1580 cm^{-1} (ring mode of A, G and T) and others. Some of HF Raman are the

3

modes of the sugar-phosphate backbone, e.g., ~790 cm^{-1} (the O–P–O vibration in the sugar–phosphate backbone, although probably obscured by the contraction and expansion of the cytosine ring at a close frequency) and 1100 cm^{-1} (PO$_2$ symmetric stretch), — while some others originate from alteration of π-conjugation in the nucleobases, e.g., the ring modes of A and G at ~1332, 1480 and 1580 cm^{-1}, as well as the C2=O vibration in thymine at ~1670 cm^{-1}. While the frequencies of the abovementioned bands are virtually the same for the two DNA species (native and digested), the intensities slightly differ. Specifically, we observe that relative intensity of sugar-phosphate modes as compared to nucleobase modes is higher for native DNA than for digested DNA. This can be explained by the fact that the latter species is more tightly packed, and sugar-phosphate backbones interact stronger resulting in decrease of the vibrational amplitude. The same is observed for unfolded ribosome and native ribosome (Fig. 3b), where sugar-phosphate modes show higher intensity for the former species. This can also be explained by tighter packing of the native ribosome.

The HF regions of the spectra of two nuclei from onion cells (Fig. 4a) exhibit several characteristic DNA peaks: ~500 (PO$_2$ scissor), 1100 (PO$_2$ symmetric stretch), 728 (ring mode of A), 805 (O-P-O stretch), 1316 (ring mode of A, G), 1366 cm^{-1} (ring mode of A, G and T). Some of them differ from those of the pristine DNA (e.g., the one at 790 cm^{-1} nearly disappear, but that at 805 cm^{-1} increases) because of interaction with the proteins. The HF parts of the two spectra are quite similar, even though these spectra belong to the nuclei of different cells from different samples.
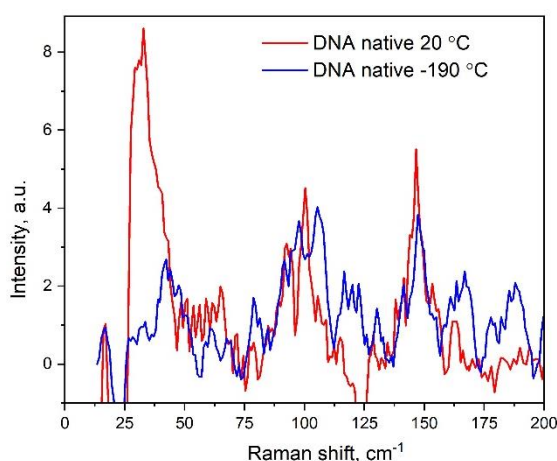

Fig.S3. LF Raman spectra with LFWBS subtracted for native DNA measured at different temperatures pointed in the graphs.

Fig. S3 shows the LF Raman spectra for native DNA measured at 20 and −190 °C. It is clearly seen from the graphs that LF Raman intensity decreases with cooling, in accordance with the weakening of the dynamic disorder.
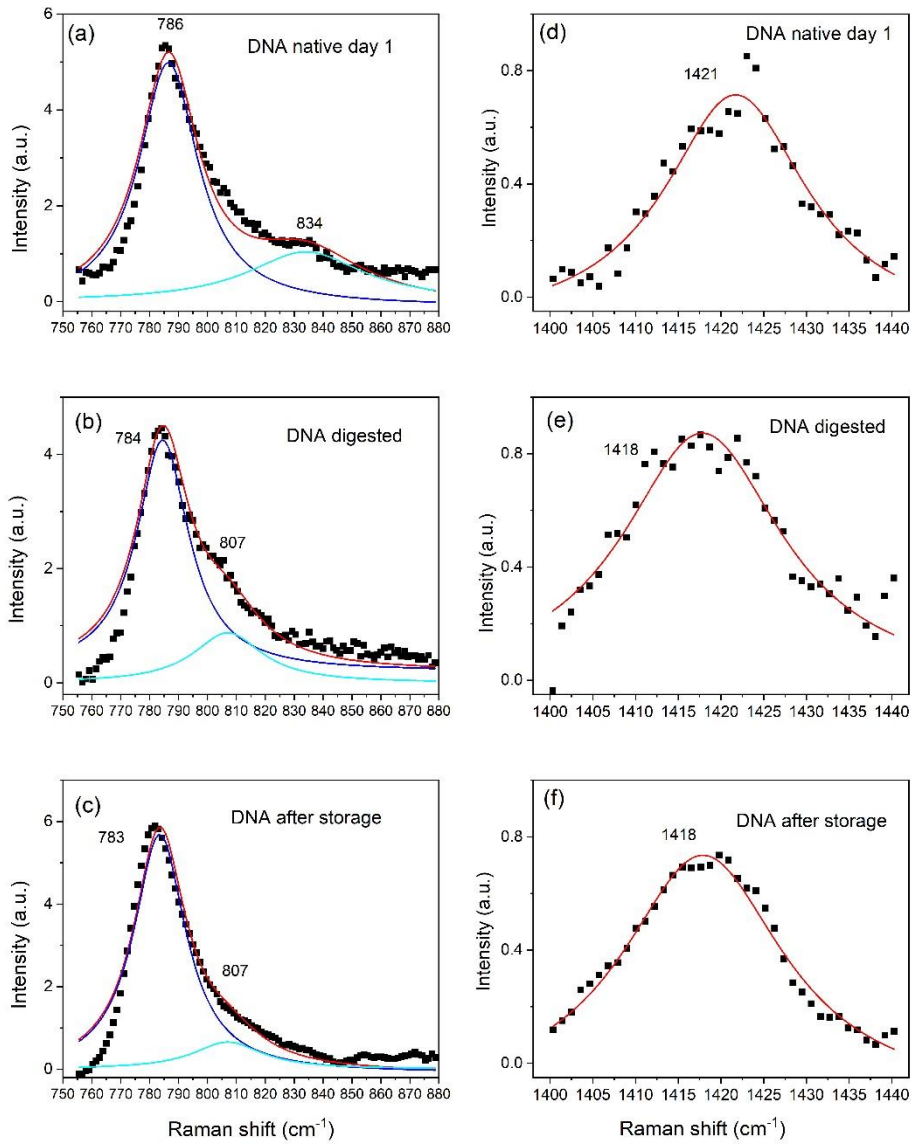
Fig S4. The Raman bands at ~784 cm^{-1}/~1421 cm^{-1} for native DNA on day 1 (a/d, correspondingly), digested DNA on day 1 (b/e, correspondingly) and native DNA after storage (c/f, correspondingly). The bands are approximated by the sum of two (a,b,c) and by the one (d, e, f) Lorentz peak functions. At the graphs (a, b, c) blue and cyan curves are the two single Lorenz peak functions with maximum position pointed in the graphs, whereas the red ones are the resulted superpositions of the Lorentz peak functions.
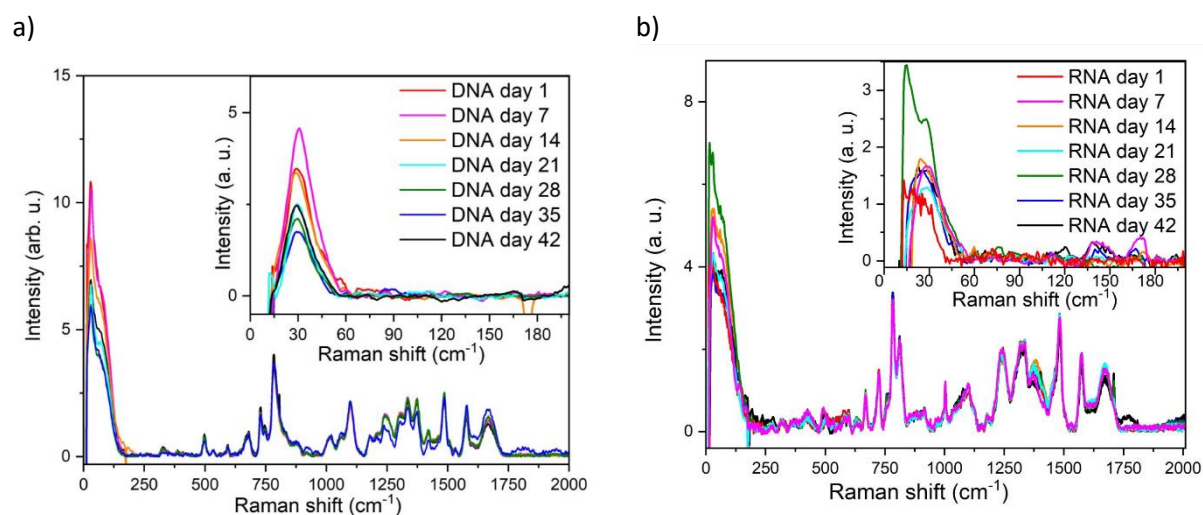
a)

b)

Fig. S5. Time of evolution of the Raman spectra for the native DNA (a) and the native ribosome (b) with storage time. Insets show the Raman spectra in LF region with LFWBS subtracted. The slight differences in the HF region result from inhomogeneity of the samples and difficulties in focusing on the same point of the sample on different measurement days.
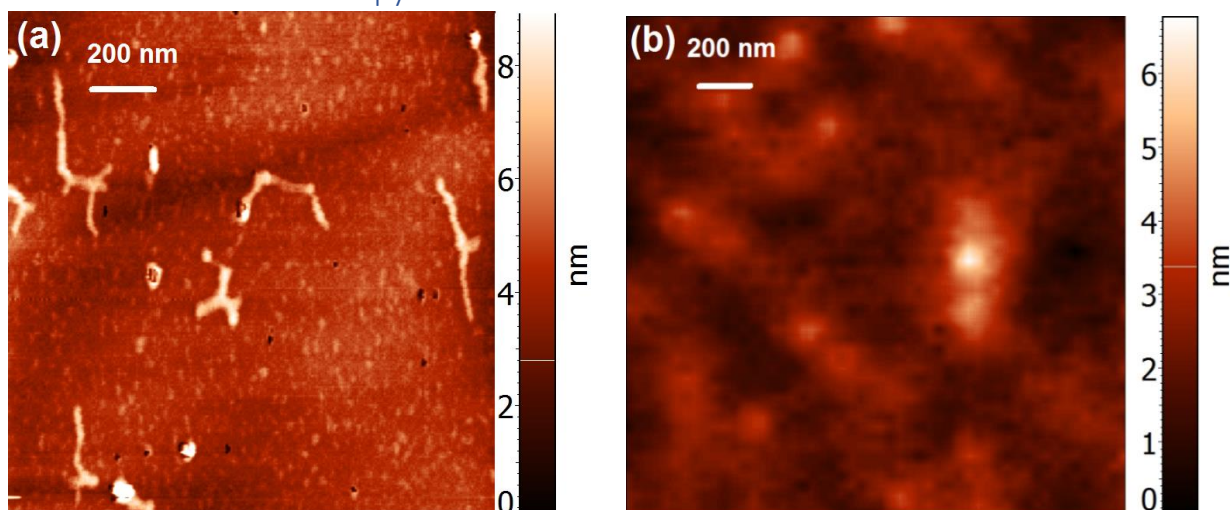
S5. Atomic-force microscopy data

Fig. S6. Atomic-force microscope maps of mica substrates with native (a) and digested (b) DNA deposited from 4 ng/µl solutions. Atomic-force microscope (NTEGRA Spectra, NT-MDT) was used in the semicontact mode with uncoated probes (XSC11, MikroMasch; typical resonance frequency 150 kHz). For the native DNA (a), separate thread-like structures about 2 nm high are observed, whereas, for the digested DNA (b), globular structures are observed.