Journal of Structural Chemistry. Vol. 50, No. 5, pp. 1007-1013, 2009

CHARACTERISTICS OF ULTRASONIC CLEAVAGE OF DNA

Yu. D. Nechipurenko,[1,2] M. V. Golovkin,[1]
D. Yu. Nechipurenko,[2] I. A. Il'icheva,[1] L. A. Panchenko,[3]
R. V. Polozov,[4] and S. L. Grokhovskii[1]

UDC 577.3

The site effect (the dependence of the band intensity on the band position of gel) is the key factor that should be taken into account in quantitative analysis of the images of high-resolution electrophoretic gels containing many bands. We developed two procedures for taking this effect into account. The dependence of the relative frequencies of internucleotide bond cleavages on the type of nucleotide, obtained in the analysis of the splitting patterns of DNA fragments under the action of ultrasound or chemical factors, was evaluated by a) using the linearly sliding mean; b) describing the function of the site dependence as a third degree polynomial. The methods lead to similar results. An analysis of the splitting of different sequences, whose total length was more than 20,000 nucleotides, allowed us to determine the relative frequencies of bond cleavages in all 16 dinucleotides found in a DNA sequence. In ultrasonic treatment, the phosphodiester bond, lying between cytosine and guanine, undergoes cleavage more often than the corresponding bonds between other nucleotides in the sequence.

Keywords: ultrasonic cleavage of DNA, conformational heterogeneity of DNA, gel electrophoresis, site specificity, footprinting.

INTRODUCTION

Earlier, we suggested an experimental approach to studies of the dependence of the local conformation of the DNA molecule on its nucleotide sequence. It was shown that the frequency of cleavages in the sugar phosphate framework of DNA under the action of ultrasound depended on the nucleotide sequence in DNA [1]. The ability to undergo cleavage was found to depend on the local structural dynamic properties of the DNA molecule [2]. These properties are important gene expression regulators (see, e.g., [3-6]). Recently, an experimental method has been developed for detecting the dependence of the cleavage frequency in the sugar phosphate framework induced by certain nucleases and chemical agents on the conformation of DNA [5]. A challenge is the physical mapping of the regulator sections of DNA; i.e., finding DNA segments recognized by the genetic expression factors [6].

Analyzing data on DNA cleavage with a known sequence, we can from a database on cleavage of segments and obtain information about the nonhomogeneity of the local structural dynamic properties of the double helix of DNA.

To obtain data on cleavage of DNA, we used electrophoresis in denaturating polyacrylamide gel. The chemical splitting agent or ultrasound were used to treat the fragments of double-helix DNA with a known sequence of a few hundreds

[1]V. A. Éngelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow; nech99@mail.ru. [2]Faculty of Physics, Moscow State University. [3]Faculty of Biology, Moscow State University. [4]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences. Translated from *Zhurnal Strukturnoi Khimii*, Vol. 50, No. 5, pp. 1045-1052, September-October, 2009. Original article submitted June 21, 2008.

0022-4766/09/5005-1007 © 2009 Springer Science+Business Media, Inc.

nucleotide pairs. The fragments contained a radioactive label at the end of one of the chains. After denaturation of the fragments, the reaction mixtures were applied to the gel. The results of electrophoresis and further exposure, scanning, and digital processing of images were recorded in the form of data arrays corresponding to the individual gel tracks. The band intensity N_i, which is proportional to the number of fragments with a length of i nucleotides in the reaction mixture, is an element of this array for one track.

The gel surface generally contains defects, strongly hindering data digitization and further processing. The nonhomogeneous thickness, air bubbles, and different amounts of salts in the deposited samples lead to the bending and deformation of the tracks and shift the bands on different tracks relative to one another. The starting labeled fragment of DNA often has small admixtures of other fragments. Because of losses caused by the deposition on the walls of vials, different tracks often have different total intensity of bands.

Prolonged physical or chemical treatment of labeled fragments leads to the double impact effect, increasing the fraction of short labeled fragments in the reaction mixture. Ultrasonic splitting also leads to the unsplit ends effect: the regions that are close to the ends of the fragment are split at lower frequencies. As diffusion causes more active motion in the gel, the bands are broadened.

These factors point to the difficulty of quantitative analysis of cleavage data for DNA with a definite nucleotide sequence. Multiple experiments of the same type can afford the preparation of an "ideal gel," as well as standardization and inclusion of the majority of physical effects. For DNA molecules with other nucleotide sequences, however, it is necessary to choose the conditions all over again. The problem of comparing the data obtained in different series of experiments remains unresolved. Therefore, it is important to seek adequate methods for analyzing these experimental data.

ANALYSIS OF POLYACRYLAMIDE GELS

Figure 1 shows the images of polyacrylamide gels obtained by separating the reaction mixtures of radio labeled fragments of DNA after chemical splitting at purines (*a*) and ultrasonic treatment (*b*) at different treatment times. For digitizing the splitting patterns, we used the SAFA program designed by the research group from Stanford University [7]. Using this program allows one to calculate the intensities of all bands on all tracks and assign the bands to the specified nucleotide sequences after correction of the track shape on the gel.

The chemical splitting of DNA with formic acid in the presence of diphenylamine [8] leads to a loss of a purine base at the splitting site. Thus, the bands on tracks *1-10* (Fig. 1*a*) and *A+C* (Fig. 1*b*) are oligonucleotides without terminal purine.

The mobility of fragments depends nonlinearly on their molecular mass, net charge, and gel density. In the separation of fragments, the band width depends on the path of the fragment in the gel. Fragments of smaller length pass longer distances and form wider bands; thus, the band intensity is site-determined. To assess this effect, we analyzed the results of chemical splitting of fragments at purines at a reaction time when the double impact effect was negligibly small. The probability of splitting of a particular nucleotide during chemical treatment is independent of its position in the DNA fragment. It appeared, however, that after separation under the conditions indicated in Fig. 1, the band intensities at the 50th and 250th nucleotides differed strongly. This points to the site effect; during the recording of the bands containing equal amounts of the radioactive label, the intensity of light striking of the fluorescent screen depends on the band width (which, in turn, depends on the band position on the gel).

An analysis of the bands on the tracks obtained after irradiation of DNA with ultrasound for 4 min, 8 min, and 16 min showed (Fig. 1*b*) that the intensity of splitting increased with irradiation time, but the intensity ratio did not change. In this case, the double impact effect was absent, but there was a clear-cut edge effect (the ends of the molecule were split less effectively than the middle).

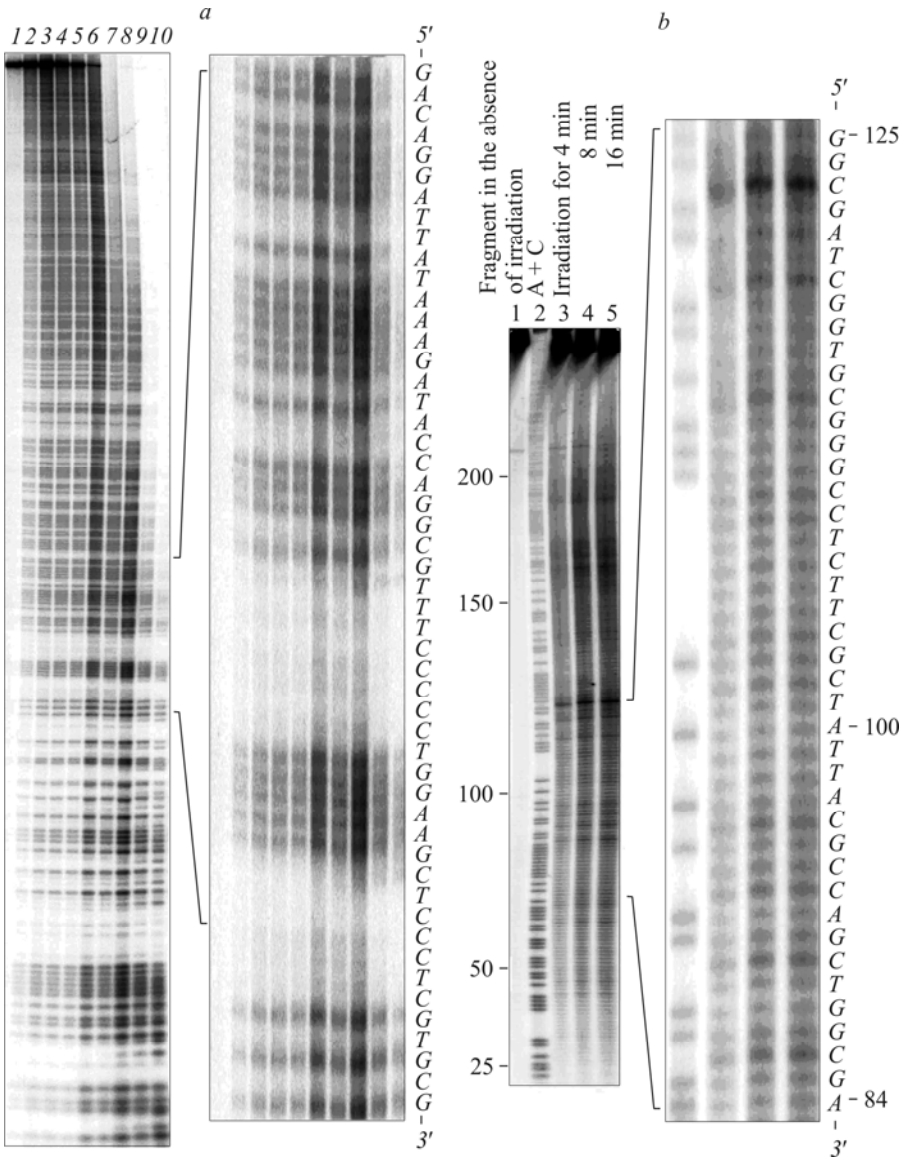

Fig. 1. Profiles of DNA fragment splitting in 6% denaturating polyacrylamide gel: (*a*) results of chemical splitting at purines: track *1* — fragment without treatment; tracks *2-10* — treatment with formic acid in the presence of diphenylamine for 15 s and 30 s; 1 min, 2 min, 4 min, 8 min, 16 min, 32 min, and 64 min, respectively; (*b*) results of ultrasonic treatment at a frequency of 22 kHz: track *1* — fragment without treatment; *2* — chemical splitting at purines; *3*, *4*, *5* — irradiation of a fragment with ultrasound for 4 min, 8 min, and 16 min, respectively. The right-hand part of figures (*a* and *b*) shows a magnified section of the gel and the corresponding sequence.

SLIDING WINDOW ANALYSIS

The band intensity N_i o a track is proportional to the concentration of the fragment that ends with the *i*th nucleotide. To take into account the component of the intensity that depends on the band position on the track, we introduce r_i, the relative intensity of the band corresponding to the fragment of *i* nucleotides among the fragments whose length is from $i - k$ to $i + k$ nucleotides,

$$r_i = N_i \frac{2k+1}{N_{i-k} + N_{i-k+1} + \ldots + N_{i+k}}, \quad (1)$$

where N_i is the intensity of the band of a fragment of i nucleotides; $2k + 1$ is the size of the window with the band corresponding to this fragment at the center ($i > k$). This procedure allows for the difference in the band intensities on different gels and site effects. Varying the size of the window showed that for ultrasonic splitting, 31 nucleotides was the optimum value ($k = 15$); when it decreases, the scatter of data increases, while its further increase does not affect the values of r_i (but the amount of data decreases after this analysis).

ANALYSIS BY THE POLYNOMIAL METHOD

The dependence of N_i on the length of the i fragment can be represented as polynomial. In a simple case, this is polynomial of degree 3 (as indicated by the asymmetry of DNA splitting with ultrasound along its length [2]),

$$N_i = R_i(ai^3 + bi^3 + ci + d), \tag{2}$$

where a, b, c, and d are the coefficients of the polynomial; R_i is the relative intensity of the ith band (for the polynomial method).

The a, b, c, and d coefficients of the polynomial are determined by the least squares procedure, which minimizes the sum of the squares of the deviations of $\{N_i\}$ from the corresponding values of the polynomial ($ai^3 + bi^3 + ci + d$). After the coefficients were determined, we can calculate R_i from (3),

$$R_i = \frac{N_i}{ai^3 + bi^3 + ci + d}. \tag{3}$$

Below the method for calculating the R_i by approximating the site dependence of N_i is called the polynomial method, and the procedure for calculating the r_i by using the sliding window is called the sliding window method.

COMPARISON OF THE TWO METHODS

For chemical splitting, digitized tracks *2-10* in Fig. 1*a* should led to similar relative intensities for each particular band after division by the mean value of the band intensity and further treatment by one of the suggested procedures. One can expect that the tracks that differ in the amount of the label or the exposure time will lead to similar relative intensities for all bands after the use of one of the methods (these values cannot coincide completely because the double impact effect can be compensated by these methods only partially).

The difference of the relative intensity on one track from that of the corresponding band on another track detected after using one method for both tracks can be indicative of track contamination or the ability of this method to allow for site effects.

Figure 2 shows data for two tracks corresponding to chemical splitting (Fig. 1*a*, tracks *3* and *8*) before treatment (*a* and *b*) abd after the use of the sliding window method. This method allows the double impact effect to be compensated (at least, partially). Figure 3 presents the data obtained by scanning track *4* (Fig. 1*b*) corresponding to ultrasonic splitting (*a* without treatment; *b* and *c* after inclusion of site effects by the sliding window and polynomial procedures, respectively). Both procedures led to approximately equal results.

We compared these methods in the following way. We considered the relative intensities and calculated the mean intensity for each band for nine tracks corresponding to chemical splitting (Fig. 1*a*) and the mean square deviations. Summing all deviations, we obtained a characteristic of the method that defines its ability to summarize all data on one standardized track with the mean band intensities. A similar procedure was applied to ultrasonic splitting data (Fig. 1*b*). The results of comparison are presented in Table 1. The mean square deviations were obtained by comparing the tracks by the sliding window and polynomial methods,

$$D = \sum_{i=1}^{Q} \sqrt{\frac{\sum_{j=1}^{m} (x_{ij} - \langle x_i \rangle)^2}{m(m-1)}},$$

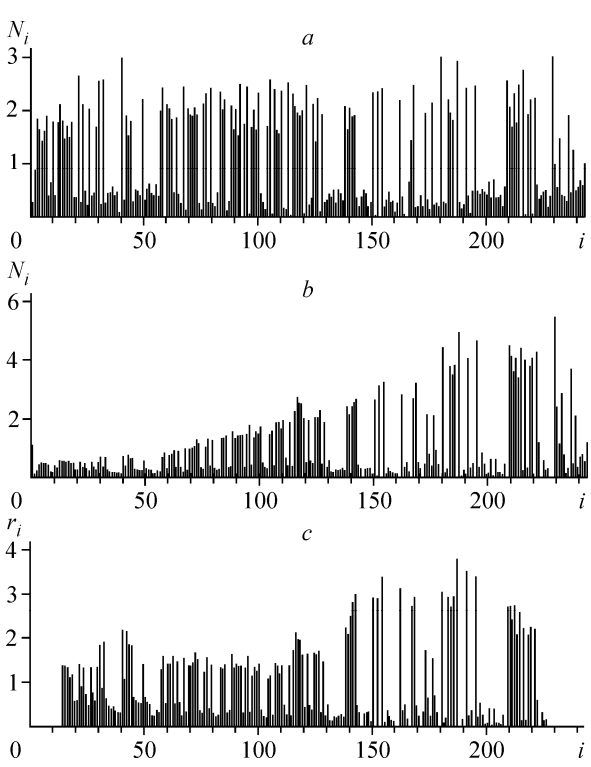

Fig. 2. Band intensities of gel electrophoresis for chemical splitting of DNA (Fig. 1*a*): *a* and *b* correspond to treatment times 30 s and 16 min; *c* — relative intensities after data processing in Fig. 2*b* by the sliding mean method.

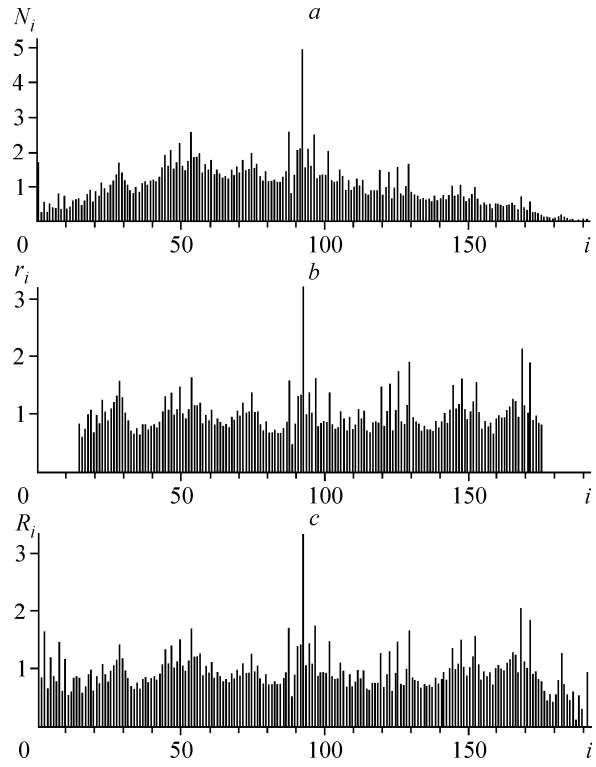

Fig. 3. Band intensities of gel electrophoresis for ultrasonic splitting of DNA (Fig. 1*b*): *a* — band intensities of track *4* without preliminary treatment; *b*, *c* — relative intensities after data processing by the sliding window and polynomial methods, respectively.

TABLE 1. Comparison of the Procedures

Type of treatment	Number of bands	*D*	
		Sliding window method	Polynomial method
Chemical splitting	212	21.14	23.80
Ultrasonic treatment	162	6.02	5.99

where m is the number of tracks; Q is the number of bands being considered on a track; x_{ij} is the relative intensity of the ith band of the jth track (for the ith band, x_{ij} corresponds to the relative intensity r_i for the sliding window method and the relative intensity R_i for the polynomial method); $\langle x_i \rangle$ is the mean intensity of the ith band over all m tracks.

When applied to chemical splitting data for nine tracks, the sliding window method led to a slightly smaller mean square deviation than the polynomial method (the difference being ~10%), while using both procedures for ultrasonic splitting data gave approximately the same deviations. Inclusion of the specificity of DNA splitting with ultrasound (see [2]) allows us to decrease, using self-consistency, the mean square deviation by 20% for the polynomial method.

Note that the discrepancy in the mean square deviations (Table 1) defines the ability of the method to compensate the site effects and standardize data, but not the error of the method. Both procedures gave approximately the same results.

The disadvantage of the sliding window method is the narrowing of the initial data base on intensities (the data corresponding to the half-width of the window are lost at each end of the fragment; Fig. 2*c* and 3*b*).

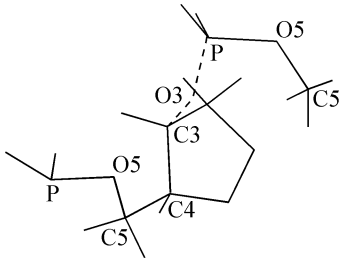

Fig. 4. Fragment of the sugar phosphate chain of DNA. The dashed line shows the bonds at which DNA splitting can take place.

TABLE 2. Relative Frequencies of Bond Cleavage at the 5′ End of Different Nucleotides

Type of nucleotide	Number of nucleotides	Relative frequencies of cleavage	
		Sliding window method	Polynomial method
A	38	0.88±0.20	0.86±0.22
C	39	1.34±0.43	1.28±0.45
G	55	0.95±0.18	0.92±0.18
T	30	0.89±0.15	0.86±0.16

Based on the analysis of data on ultrasonic splitting of one fragment of DNA using the above methods (track *4*, Fig. 1*b*), we also obtained the relative frequencies of cleavage of the phosphodiester bonds lying at the *5′* end from *A*, *T*, *G*, and *C* nucleotides (Table 2).

Note that these data correlate with the data of our evaluation [2] of the relative frequencies of bond cleavage between different dinucleotides during ultrasonic treatment (the database on splitting of 2500 dinucleotides was processed in this study by the sliding window procedure).

RESULTS OF DNA SPLITTING ANALYSIS

We have shown that both procedures used for analyzing the splitting patterns gave approximately the same results. The sliding mean method was used with a window of 31 nucleotides for analyzing the database on the splitting of 20,588 dinucleotides. The experiments were run under the standard conditions: 0.1 M NaOAc, pH 6.0 (the procedure for obtaining and processing data was described in [1, 2]). Table 3 lists the relative frequencies of cleavage of phosphodiester bonds in 16 possible dinucleotides occurring in DNA. The data for 138 tracks on different gels were summarized (several fragment sequences overlapped).

It follows from the table that the frequency of ultrasonic splitting of the 5′-CpG-3′ dinucleotide was ~1.5 times higher than for all other dinucleotides (with significant statistical reliability). In general, we can draw the conclusion that the phosphodiester bond that follows the sugar of cytosine is more liable to cleavage during ultrasonic treatment. We can attribute this to the local structural dynamic properties of the double helix of DNA. Figure 4 shows the section of the sugar phosphate framework of DNA; the dashed line shows the bonds, at one of which the dinucleotide is split.

We have considered high-resolution methods of quantitative treatment of electrophoretic gels, which allowed us to eliminate the site-determined distortions in band intensities on gel. We used the sliding mean and polynomial methods for describing the site dependence. It was noted that the double impact effect could be compensated as a result of the use of the given procedures. The relative frequencies of cleavage of different internucleotide bonds were evaluated. Cleavages of the

TABLE 3. Relative Frequencies of Bond Cleavage of Different Dinucleotides

Type of nucleotide	Number of dinucleotides	Mean relative frequency of cleavage	Minimum and maximum relative frequencies of cleavage	Standard deviation
AA	1636	0.919	0.570; 1.289	0.129
AC	1076	0.913	0.574; 1.242	0.128
AG	1028	0.900	0.575; 1.266	0.124
AT	1374	0.904	0.590; 1.257	0.119
CA	1265	1.160	0.686; 1.829	0.209
CC	1141	1.007	0.646; 1.416	0.144
CG	1230	1.444	0.823; 2.627	0.334
CT	1077	1.130	0.714; 1.718	0.198
GA	1153	0.970	0.620; 1.359	0.133
GC	1317	0.954	0.554; 1.368	0.146
GG	1168	0.922	0.584; 1.312	0.145
GT	1101	0.952	0.623; 1.327	0.126
TA	1065	0.973	0.670; 1.297	0.120
TC	1173	0.912	0.532; 1.262	0.131
TG	1305	0.979	0.645; 1.361	0.126
TT	1672	0.932	0.616; 1.275	0.127

sugar phosphate framework of the DNA molecule most often occur at the phosphodiester bond connecting cytosine and guanine (in the direction of 5′-CpG-3′).

The experimental methods and materials are described in [1, 2].

This work was financially supported by the Russian Foundation for Basic Research, grant Nos. 07-04-01031, 07-03-00492, and 08-04-01739 and the Program on Molecular and Cell Biology of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. S. L. Grokhovskii, *Molek. Biol*., **40**, 317-325 (2006).
2. S. L. Grokhovskii, I. A. Il'icheva, D. Yu. Nechipurenko, et al., *Biofizika*, **53**, 417-425 (2008).
3. Yu. D. Nechipurenko, R. V. Polozov, D. Yu. Nechipuenko, et al., "Mathematical models of gene expression regulation: mechanical perturbations in the structure of DNA," Mathematics. Computer. Education, in: *A Collection of Treatises*, Vol. 2, G. Yu. Riznichenko (ed.), Regular and Chaotic Dynamics Center, Izhevsk, (2006), pp. 392-402.
4. Yu. D. Nechipurenko, D. Yu. Nechipurenko, R. V. Polozov, et al., DNA-Based Nanostructures: Changes of Mechanical Properties of DNA Upon Ligand Binding, Nanomaterials for Application in Medicine and Biology (NATO Workshop in Bonn), Springer (2008), pp. 59-67.
5. J. A. Greenbaum, S. C. J. Parker, and T. D. Tullius, *Genome Res*., **17**, 940-946 (2007).
6. J. A. Greenbaum, Bo Pang, and T. D. Tullius, *ibid*., 947-953.
7. R. Das, A. Laederach, S. M. Pearlman, et al., *RNA*, **11**, 344-354 (2005).
8. W. P. Tate and G. B. Petersen, *Biochem. J*., **147**, 439-445 (1975).