Measurement 198 (2022) 111362

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Double-strand breaks quantification by statistical length analysis of DNA fragments imaged with AFM

Kamila Sofińska [a,*], Michał Cieśla [a,*], Jakub Barbasz [b], Natalia Wilkosz [a], Ewelina Lipiec [a], Marek Szymoński [a], Piotr Białas [a]

[a] Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland
[b] Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Niezapominajek 8, 30-239 Kraków, Poland

ARTICLE INFO

Keywords:
Double-strand breaks
DSB
DNA
AFM imaging
Bleomycin
DNA conformation
DNA damage
DNA lesions
Statistical analysis

ABSTRACT

DNA, a molecule carrying our genome, is constantly exposed to various endogenous and exogenous damaging factors. These factors may lead to the formation of single- or double-strand breaks. Recently, an atomic force microscopy emerged as an attractive tool for determination of the length of DNA fragments deposited at surfaces, allowing quantification of double-strand breaks (DSBs). In spite of this, the length of objects provided via AFM is strictly related to the image spatial resolution, which in consequence leads to underestimation of DSBs quantity. To correct this deficiency, we provide here a novel methodology based on statistics of DNA fragments length. This approach allows for highly accurate determination of the DSBs number. Moreover, we provide the software that enables to automatically calculate the number of DSBs according to the described methodology. Finally, the method is applied for characterization of the DNA plasmid fragmentation by bleomycin - an anticancer chemotherapeutic drug.

1. Introduction

DNA double-strand breaks (DSBs) are the most dangerous of DNA lesions because they compromise the genome integrity and not repaired may lead to cell death [1]. Moreover, in the repair process of DSBs, the numerous mutations may occur that is considered as a driving force in carcinogenesis [2–4]. On the other hand, the cytotoxic activity of several anticancer drugs is related to the ability to form DSBs, for example bleomycin, bicyclic enediynes, neocarzinostatin, or topoisomerase inhibitors like camptothecins and etoposides [5]. Precise detection and quantification of DNA double-strand breaks (DSBs) are of key importance in the field of cancer drug development, radiation biology, or genome-editing. Detailed understanding of DNA fragmentation may bring new insights into the mechanisms of action of cytotoxic drugs, opening new possibilities of cancer therapies that target specific bonds or functional groups of DNA.

The assessment of DNA fragmentation after induction of DNA double-strand breaks is mainly done by gel electrophoresis-based techniques, as well as DNA sequencing. For example, the number of DNA fragments can be determined experimentally using pulsed-field gel electrophoresis (PFGE) [6,7]. This method was applied in the past to study radiation-induced DNA double-strand breaks. However, to quantify the DNA separation and calculate DSBs/100 Mbp factor, PFGE requires DNA staining with densitometric detection, radioactive labelling, and the calibration of a fragment size against molecular weight markers as well as further calculations. Another popular approach to quantify the cellular DNA damage, regardless of its origin, is the gel electrophoresis–based approach known as the comet assay [8]. This method was first described in 1984, but since then it has been modified and now it is commercially available as a kit with fluorescent detection. To detect DNA damage at the level of a single cell, a single-cell microgel electrophoresis under alkaline conditions for X-rays exposed human lymphocytes was proposed [9]. Recently, a method of quantitative DSB sequencing (qDSB-Seq) based on induction of spike-in DSBs using a site-specific endonuclease was reported [10]. This method provides DSB frequencies per cell and a precise genomic location of DSBs. The described above methods are well-known and extensively described but most of them require an additional step for detection of DNA fragments including staining or labeling. The interaction of these additional molecules used for detection of DSBs may introduce an undesired effect like the change of DNA conformation. Such effect was described for several fluorescent dyes, namely PicoGreen, Dapi, and DRAQ5 [11]. Since DNA

* Corresponding authors.
E-mail addresses: kamila.sofinska@uj.edu.pl (K. Sofińska), michal.ciesla@uj.edu.pl (M. Cieśla).

https://doi.org/10.1016/j.measurement.2022.111362
Received 30 August 2021; Received in revised form 20 April 2022; Accepted 16 May 2022
Available online 20 May 2022

conformation is known to be strictly related to the molecule length, it means that in all research concerning the assessment of DNA fragments length, the undesired conformational rearrangements should be avoided.

Another method allowing for DNA size tracking is based on direct measurements of the length of individual DNA molecules visualized on AFM images. This approach was reported to be free of the risk to introduce undesired interaction between the DNA molecule and chemical labels. What is more, the DNA size tracking based on AFM images provides the length of short DNA fragments of a few tens to a few hundreds of nanometers with several nanometers accuracy [12]. Precise measurements of such short DNA fragments are not accessible for the other experimental techniques mentioned above. Therefore, DNA size tracking based on AFM imaging proved to be an efficient tool in studies of an influence of various molecules including fluorescent dyes [11] and chemicals used upon fixation procedure [13] on the DNA conformational properties, resulting in a change of the individual DNA molecule length. The formation of DSBs upon various physical factors including ionizing radiation such as neutrons [12], electrons, X-rays and heavy ions [14,15] have been regularly studied with AFM based DNA tracking since the 90s. Depending on the damaging factor, the distribution of DNA fragments length is different and various complex theoretical models were applied [14,16] to verify or to explore potential mechanisms of action on DNA. Wiggins *et al.* [17] showed that a direct measurement of the bending energy function based on high resolution AFM scans gives more accurate results in terms of elastic energy in the case of highly bent DNA molecules than models such as the worm-like chain (WLC) model. Thus, it is possible to estimate the length of bent DNA fragments more accurately.

Bleomycin (Blm) is an anticancer chemotherapeutic drug used in combination with other cytotoxic medications or radiotherapy to treat mainly head, neck, skin, testicular, and ovarian cancers, as well as Hodgkin lymphoma [18,19]. Blm binds to DNA and cleaves it causing both single- and double-strand breaks (SSBs and DSBs, respectively) in the presence of ferrous ions and oxygen [20,21]. The cytotoxic activity of Blm is related to its ability to create DSBs [19,22]. Blm cleaves DNA in a sequence dependent manner and recognizes the sequence of dinucleotides consisting of 5′-G- pyrimidine (5′-GT and 5′-GC) [19,20]. The activity of this metallo-glycopeptide antibiotic was shown in vitro using DNA solutions [21] and living cells as well as during clinical trials [5], however, the mechanism of action in cells still remains unresolved.

All hypotheses describing the mechanism of DNA cleavage by bleomycin stand that the process of DNA degradation is preceded by chelating metal ions (Cu and Fe) and formation of activated complex of bleomycin, e.g., bleomycin-Fe(III)–OOH in the case of Fe ions presence. This complex removes the hydrogen atom from C4′ of deoxyribose bound with pyrimidine to form C4′ radical intermediate [19,20]. The further process depends on the presence of O_2. In the presence of the oxygen molecule, the radical intermediate forms a hydroperoxy radical that leads to a Criegee-type rearrangement resulting in strand cleavage via fragmentation of the oxygenated sugars, producing 3′-phosphoglycolate and 5′-phosphate ends [19,20,23]. The alternative pathway, in the absence of oxygen, assumes the oxidation of 4′ radical intermediate that leads to a carbocation formation, which in the presence of water generates an alkali-labile lesion with 4′-oxidized abasic sites [19,23]. Those unstable DNA lesions undergo further transformation involving β-elimination to form the final DNA strand scission with a phosphate at the 5′ end and 4′-ketodeoxyribose at the 3′-end [24].

Up to date, there is no consensus concerning the origin of DSBs resulting from Blm cut. First, it was considered that DSBs are a result of the randomly accumulated SSBs [25]. The second hypothesis assumes that DSBs are generated by single Blm molecule, which regenerates before the second strand cleavage event [19,20]. In this case, the formation of double-strand breaks is considered to be strictly related to the oxygen-dependent pathway [26]. These two hypotheses do not seem to be mutually exclusive. It has also been demonstrated that Blm may cleave the DNA plasmid already immobilized on the surface [21].

In this work, we propose a novel approach for DNA double-strand breaks analysis based on the distribution of DNA fragments length derived from AFM images. Moreover, we show that the number of DSBs can be determined in an efficient manner based only on the fraction of unbroken DNA strands. To induce DBSs in the DNA molecule, we use bleomycin at concentrations within the range of 8–300 nM. Additionally, corrections of systematic errors related to the limited AFM image resolution are provided. These corrections allow to compensate the possible underestimation of the DNA length. We also developed a simple open-source program (provided in the Supporting Information), which allows to calculate the mean/average number of DSBs automatically according to proposed methods. Since the mechanism of action of Blm on DNA is not fully explored, an appropriate model explaining the relation between the number of DSBs and DNA length distribution after damage induction has not been provided yet. Therefore, we calculated the number of DSBs according to the simple arithmetic formula proposed by Pang *et al.* [12,15] for comparison with the results of the statistical models developed in this work.

2. Materials and methods

Activated bleomycin preparation procedure. Bleomycin sulfate was obtained from TCI Europe N.V. (Tokyo Chemical Industry). First, ferric ammonium sulfate dodecahydrate was dissolved in ultrapure water. The next step of preparation procedure involved the immediate addition of a 10% molar excess of iron to bleomycin [22]. The prepared 1 mM bleomycin-iron solution was then kept till use at −20 °C.

DNA solution preparation procedure. pUC19 DNA plasmid was purchased from Thermo Fisher Scientific Inc. (NYSE: TMO). DNA plasmid was initially prepared at a concentration of 1.25 μg/mL in 20 mM HEPES buffer. Such prepared DNA solution was mixed with 10 mM $MgCl_2$ in the 1:1 ratio to achieve 0.625 μg/mL DNA in 10 mM HEPES and 5 mM $MgCl_2$ in final solution.

Cleavage of DNA by bleomycin. To study the influence of bleomycin on formation of DNA double-strand breaks, bleomycin-iron complex was added to DNA solution. We performed a series of concentration-dependent experiments to investigate the efficiency of Blm concentration in respect to a number of DNA cuts. We studied the following bleomycin concentrations: 0, 8, 17.5, 30, 40, 50, 75, 100, 200 and 300 nM. The time of bleomycin-DNA interaction was 4 min. All experiments were performed in relatively low salt concentration to ensure the presence of open circular (relaxed) form of DNA plasmid [27,28]. It enabled to avoid supercoiling of plasmid molecules, that could interfere with further length measurements based on AFM images.

DNA deposition protocol. For AFM imaging of DNA, we followed Mg^{2+} mediated deposition of DNA on a mica surface. A 10 μL droplet of DNA solution prepared as described above, containing an adequate Blm concentration was casted on a freshly cleaved mica (V1 grade mica, purchased from Ted Pella Inc., USA) and incubated for 3 min in ambient temperature. Finally, the sample was rinsed with ultrapure water and then dried using a compressed nitrogen.

AFM imaging. All AFM images were acquired using amplitude modulation (AM-AFM) mode in air dedicated for soft biological molecules. Images were collected with resolution of 512 × 512 pixels and scanning rates in a range of 0.5–1 Hz. AFM imaging was carried out using two AFM microscopes: Multimode AFM Nanoscope IIIa system (Digital Instruments, Santa Barbara, CA, USA) and the SmartSPM™ 1000 Scanning Probe Microscope (AIST-NT, Horiba, France SAS Ltd.). For Multimode imaging we used ETALON HA_NC probes (cantilever B with ca. 140 kHz resonant frequency) from NT-MDT Spectrum Instruments and RTESPA-150 probes (Bruker) for AFM imaging using AIST-NT microscope. The amplitude of cantilever oscillations was kept around 20 nm (+/- 30%), which is a standard level for visualizing biomolecules in AM-AFM mode in ambient air. We took into consideration the possible image artefacts resulting from AFM tip deformations that

could affect further length analysis. The quality/resolution of analyzed images was comparable (cross-section dimensions of DNA strands).

Image processing. The processing of AFM images relied on flattening by a 1st or 2nd order polynomial correction using Gwydion software [29] (Version 2.51).

DNA length tracking. For length tracking of DNA plasmids, based on AFM images taken before and after the interaction with bleomycin, we used an open-source FiberApp software [30]. To determine the length of the control DNA and DNA after treatment with Blm we used 3 × 3 μm AFM scans that for 512 × 512 pixels resolution gives ca. 5.9 nm per pixel.

Simulations. Numerical simulations were performed assuming that DSBs may occur under the influence of Blm at some specified sites only. We consider here only GC specific sites. Because this sequence is palindromic, it is the most susceptible to DSB formation, regardless of the mechanism of action of Blm (randomly arising SSB at both DNA strands or DSB formed by a single Blm molecule). In each of specific sites plasmid may break with equal probability. Thus, for each GC site we chose a random number from the [0,1) interval, and the DSB occurred if this number was smaller than the assumed probability. After the simulation we determined the length of the obtained DNA fragments.

3. Results and discussion

3.1. Experimental results

The distribution of fragment length in the control data, i.e., for DNA plasmid not treated with Blm is presented in Fig. 1 A. This sample contains only DNA strands of the length l_0. The distribution observed in Fig. 1 A is the result of the measurement errors related to the inability of the program and/or human operator to select individual DNA strands, e.g., in the case of strands overlapping, or multiple twisted and interlaced DNA strands. Because of this nature of the errors, modeling them pose to be challenging. A typical AFM image used to collect the fragment length data for the control DNA sample is shown in Fig. 1 B.

First, we establish the length l_0 of the full DNA plasmid deposited on a mica surface via magnesium ions. The mean fragment length of DNA in the control sample determined from the distribution presented in Fig. 1 A is $l_0 = 831$ nm. However, it should be noted that the presented histogram has significant dispersion, especially having in mind that it should contain object of the same size. It indicates that the length measurement is burdened by quite large uncertainty, which has significant impact on further analysis. pUC19 plasmid contains 2686 bp. Assuming 3.4 nm per 10.5 bp for DNA being in B-type conformation, the length of this DNA plasmid in the bulk phase is 870 nm [31]. The observed difference of our experimentally determined plasmid length is not higher than 5%. It may be related to the accuracy of determination the fragment length by the software used here and/or the conformational change because of the surface deposition or drying [13]. Therefore, we assume that the uncertainty of determining the DNA length from 3 × 3 μm AFM images is less than 5% in respect to the length of DNA plasmid in a bulk phase. DNA changes its conformation upon drying [32]. However, DNA fixation on mica via magnesium ions preserves the B-DNA conformation, in the case of the lack of free DNA ends [13]. The lack of significant change of DNA length related to the conformational transitions of the whole plasmid molecule observed here is in agreement with TERS (Tip-enhanced Raman Spectroscopy) findings on DNA conformational transitions upon surface deposition and drying [13].

pUC19 plasmid was treated with Blm of concentration in the range of 8–300 nM. The examples of AFM images for DNA treated with Blm are presented in Fig. 2. AFM images showing DNA after an application of Blm with all studied concentrations (8–300 nM) are presented in the Supporting Information.

The treatment of DNA plasmid with Blm resulted in fragmentation of DNA molecules. AFM images (Fig. 2 and S1) clearly display a fraction of cut DNA fragments, but also a fraction of unbroken plasmids can be observed, especially within the concentration range of 8–75 nM. In the following part of this work, we intend to show that the information regarding the whole distribution of DNA fragments length, as well as the information concerning only the fraction of unbroken DNA molecules can be used to determine the number of DSBs. First, we focused on unbroken plasmids. The number of unbroken molecules observed on AFM images is a parameter that can be used to estimate the number of the DSBs. We performed an analysis of whole (uncut) plasmids for samples treated with Blm at concentrations below 75 nM. For concentrations of Blm higher than 75 nM, the number of unbroken plasmids was too low to obtain their statistics and histograms. For the Blm concentrations in the range of 8–75 nM, the dependence of unbroken plasmid length on the Blm concentration was observed, as presented in Figs. 3 and 4.

The analysis of the length of unbroken plasmids allows us to determine a dependence between the mean and the median of plasmid length on Blm concentration, which is presented in Fig. 4. The decrease of the plasmid length with the increase of Blm concentration (within the concentration range of 8–75 nM) can be observed.

Note that, although the standard deviations of the histograms in Fig. 3 are around 100, the standard deviation of the mean values are approximately 10 times smaller, because it is divided by the square root of the observed plasmid numbers. Therefore, the observed differences in the mean plasmid length for various concentrations of Blm are significant. The fit shown in Fig. 4 allows us to estimate the median of

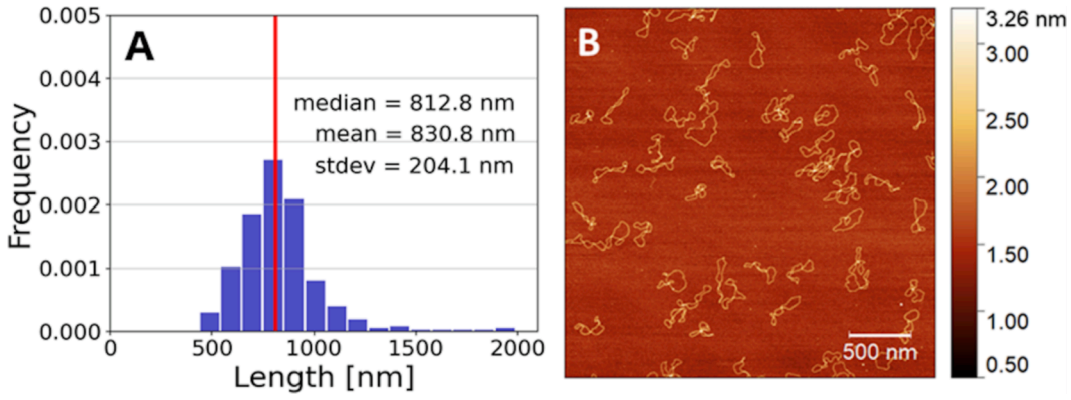

Fig. 1. (A) The distribution of fragment length in the control sample (pUC19 DNA plasmid) determined based on AFM images using FiberApp [30]. The median of this distribution is 812.8 nm, and the mean value is 830.8 nm. Red line corresponds the median of the distribution. (B) AFM image (3 × 3 μm) of pUC19 DNA plasmid showing the untreated DNA strands in the control sample. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

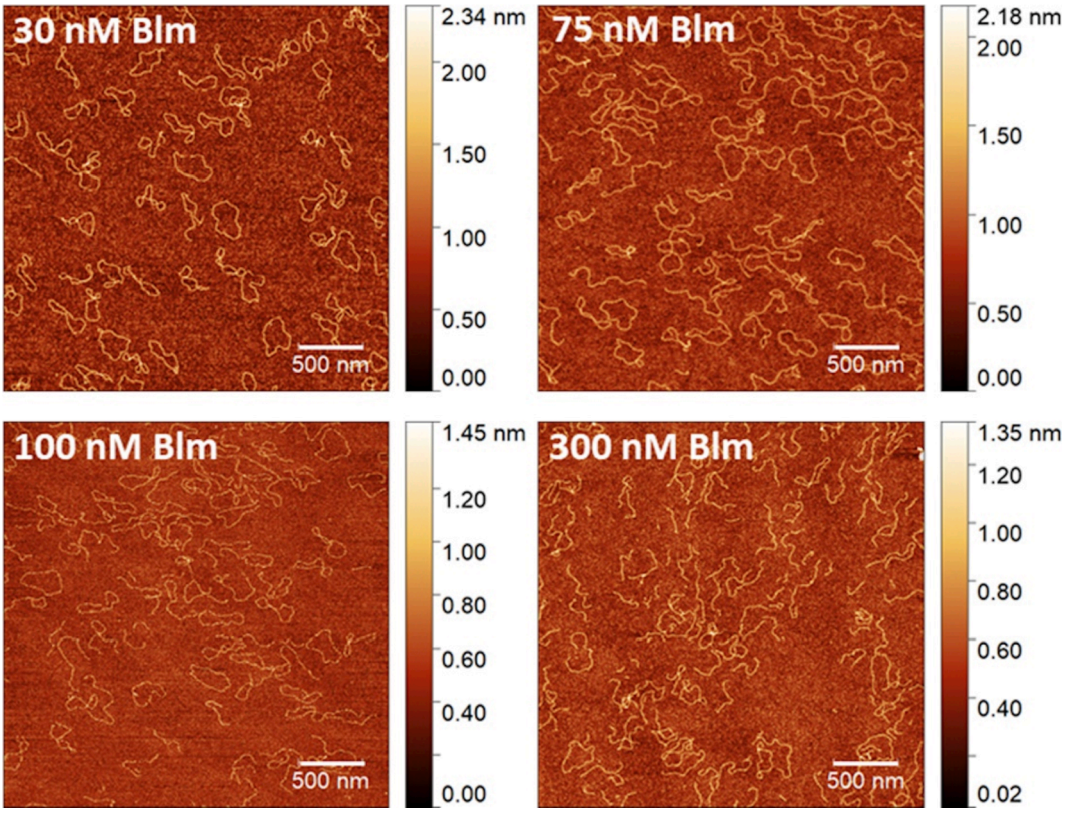

Fig. 2. AFM images (3 × 3 μm) of pUC19 DNA plasmid after the treatment with Blm of the following concentrations: 30, 75, 100, and 300 nM.

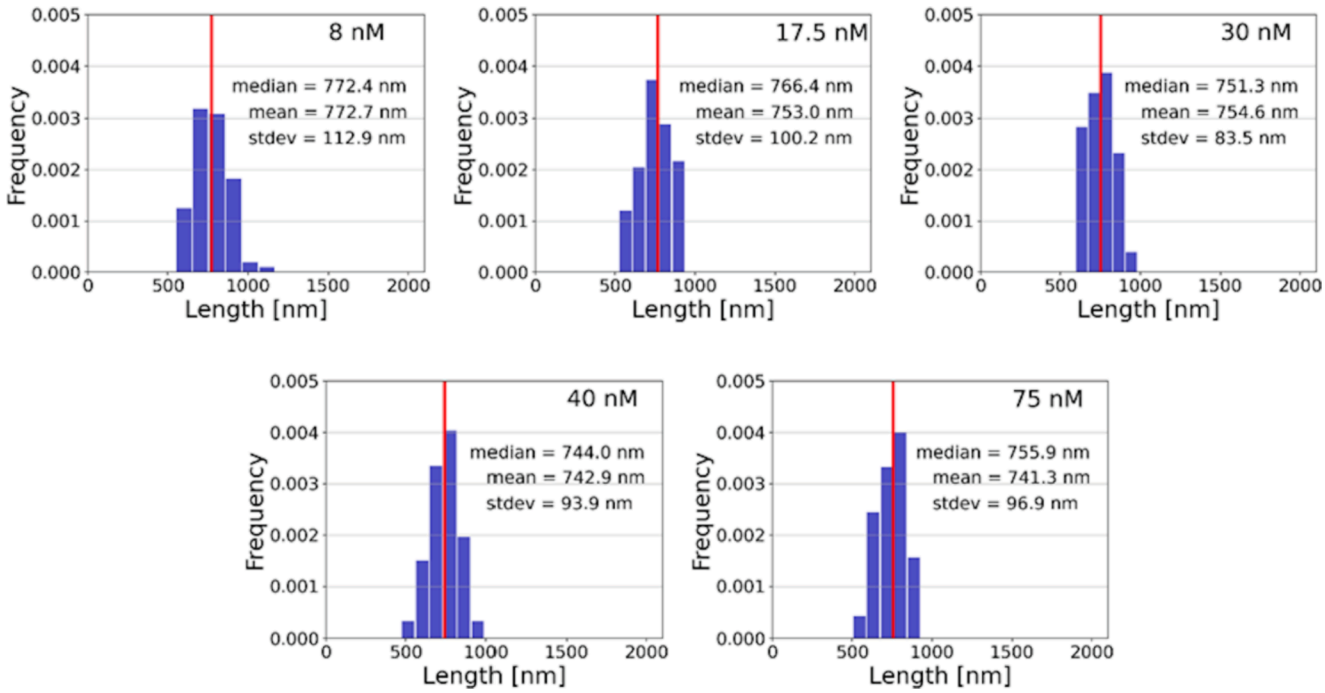

Fig. 3. Histograms of pUC19 DNA plasmid lengths after the treatment with Blm of the following concentration: 8, 17.5, 30, 40, and 75 nM.

unbroken plasmids length for Blm concentrations higher than 75 nM, which we cannot obtain experimentally.

Plasmid molecules, after the treatment with Blm, lose their supercoiled quaternary structure observed for control sample (Fig. 1) and are observed on AFM images in a relaxed state (Figure 2 and S1). Such a phenomenon is known to be related to the induction of SSBs in DNA chain [27,33–35]. Jiang *et al.* [27] explained the relaxation of the plasmid molecule as a result of SSBs and the change of the DNA chain stiffness. DNA lesions induced by Ultraviolet Radiation (UVR) were shown to change the conformation of DNA from B- to A-DNA [36]. A-DNA conformation is 20–25% shorter than B-DNA (occurring in majority in functional cells) [32]. The dependence observed in Fig. 4 may

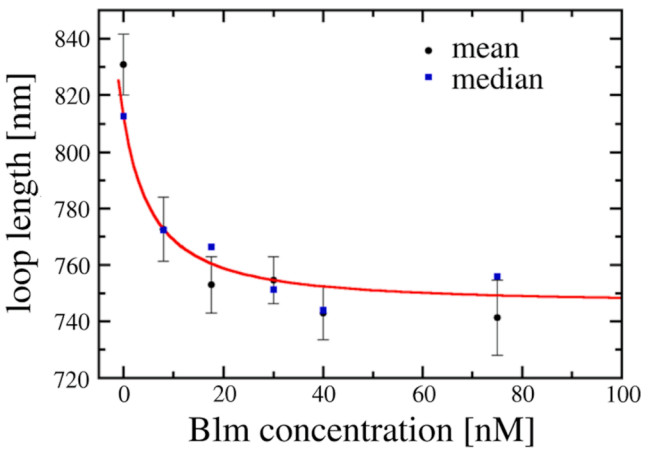

Fig. 4. Dependence of the unbroken pUC19 DNA plasmid length (loop length) on the Blm concentration. Black dots and blue squares correspond to the mean length with its standard deviation and median length, respectively. Solid red line is the fit: $median = \frac{1120}{(concentration+8.2)^{1.3}} + 746.5$. All the numerical parameters in the fit were obtained by numerical minimalization of the sum of squared deviations of the experimentally observed median values and the provided analytical function.

be related to the ability of Blm to induce single-strand breaks in DNA molecule [20,21]. These single-stranded lesions may possibly result in local conformational changes from B- to A-DNA as in the case of UVR induced DNA damage. Thus, possibly, the more Blm in the solution the higher number of SSBs that results in the shortening of the plasmid length. However, at this point, no unequivocal conclusions can be drawn, therefore this issue requires further research.

The distributions of fragments length based on AFM images for various concentrations of bleomycin are plotted in Fig. 5. Compared to the control data (Fig. 1 A) one can see a clear shift of the mean DNA fragments length towards smaller values as it could be expected.

The method used for the experimental data analysis does not distinguish between the original and the one-time cut strands, both are of the same length. The comparison of histograms in Fig. 5 to the reference histogram in Fig. 1 A shows that for small concentrations of bleomycin some fraction of shorter strands appears, but the main peak remains around 750 nm. This suggests that only tiny fraction of DNA strands was cut more than once. For concentrations of bleomycin above 40 nM, the second peak appears on the left side of the histograms. It corresponds to increasing number of relatively short fragments of DNA (below 500 nm). However, a significant number of strands still remain of their original length. With further increase of Blm concentration the peak at around 750 nm becomes smaller or even vanishes, and histograms display the only one maximum corresponding to the fraction of cut DNA fragments.

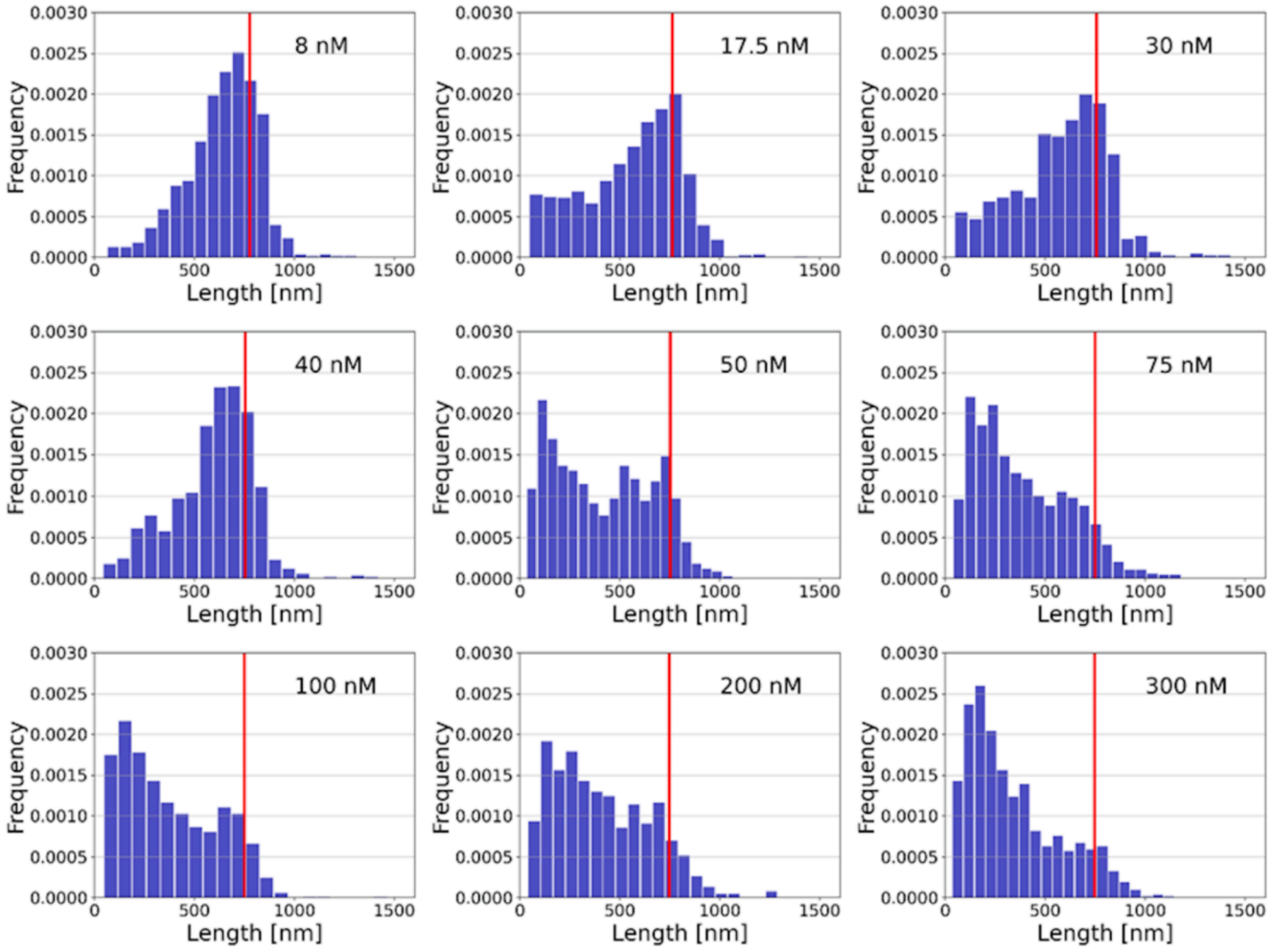

Fig. 5. The distribution of DNA fragments length for various bleomycin concentrations (8–300 nM). Red vertical line denotes the median of the circular (unbroken) plasmids length estimated from the fit obtained in Fig. 4.

3.2. Data analysis

Typically, the analysis of the above histograms is focused on determining the average number of cuts n_c per single DNA strand. The most common approach uses the relation derived by Pang *et al.* [12], which is based on the total number of DNA fragments N and their total length L. Note that n_c breaks of single DNA plasmid produce n_c DNA fragments. Because the original number of plasmid molecules was L/l_0, the mean number of DSBs can be estimated by:

$$n_c = \frac{N}{L/l_0} \tag{1}$$

Note that in the nominator of the above relation, N counts only DNA fragments after at least one cut, excludes the unbroken molecules. Thus, when counting all fragments in the image, all the circular (unbroken) molecules must be omitted.

The above relation is very simple and accurate assuming a precise measurement of DNA fragments length and their quantity. However, in most of experiments based on AFM imaging, an observation and taking into consideration objects or fragments smaller than twice the pixel size is prevented or largely prohibited by the Nyquist limit [37]. Objects of a size being close to the image resolution cannot be unambiguously classified as the objects of interest. Therefore, in the above formula we underestimate N, and thus, obtained n_c values will be understated. Therefore, we develop alternative methods of estimating n_c that may help to overcome this issue.

3.3. Statistical model for determination of n_c

The process of DNA cutting can be simulated numerically to visualize the fragmentation process of DNA plasmid under the influence of Blm in the case of infinite image resolution. Assuming a specific sequence of nucleobases (GC), which is recognized by Blm in the DNA strand and a given probability of cutting at each specific site, one can generate histograms of DNA fragments length. Some examples, which differ in an average number of DSB (cuts) per one DNA plasmid n_c (theoretically assumed) are shown in Fig. 6.

In contrast to the experiment, the determination of DNA fragments length in numerical simulations is very accurate. Therefore, here, the peak corresponding to untreated DNA plasmid is very narrow and tall.

Whereas, according to the Fig. 1 A, we can assume that measurement of the length of DNA fragment gives a random number around the real value according to Gaussian distribution.

It appears that the fraction of unbroken DNA strands can be derived analytically. Assuming that double-strand break may occur at each of n_{bs} site carrying a specific sequence recognized by bleomycin with the same probability p_c, the probability that the DNA strand will not brake at a given site is $(1-p_c)$. The probability that the whole strand remains circular (without DSBs) is $(1-p_c)^{n_{bs}}$. Similarly, the probability that the strand will break only at one site and will have full original length is $n_{bs}(1-p_c)^{(n_{bs}-1)}p_c$. When $n_{bs} \gg 1$, and simultaneously $p_c \ll 1$, the model can be described by one parameter only - the average number of cuts per DNA n_c:

$$n_c = p_c \cdot n_{bs}. \tag{2}$$

Then $(1-p_c)^{n_{bs}} \approx e^{-n_c}$, and thus, the fraction of DNA strands that have been cut at most once is:

$$q' = (1-p_c)^{n_{bs}} + n_{bs}\,p_c\,(1-p_c)^{(n_{bs}-1)} \approx (1+n_c)e^{-n_c}. \tag{3}$$

If at the beginning of the experiment, there are N strands, after Blm treatment, only $q'N$ of them (at average) will remain unchanged. Because at average Nn_c DSBs occur, and each DSB produces a new DNA fragment with except of the first break of an initial plasmid, therefore, the total mean number of fragments is equal:

$$N(n_c + e^{-n_c}), \tag{4}$$

where the second term corresponds to circular DNA molecules (without DSBs) that did not break. Thus, the mean fraction of strands of original length is:

$$q(n_c) = \frac{q'}{n_c + e^{-n_c}} = \frac{1+n_c}{1+n_c\,e^{n_c}}, \tag{5}$$

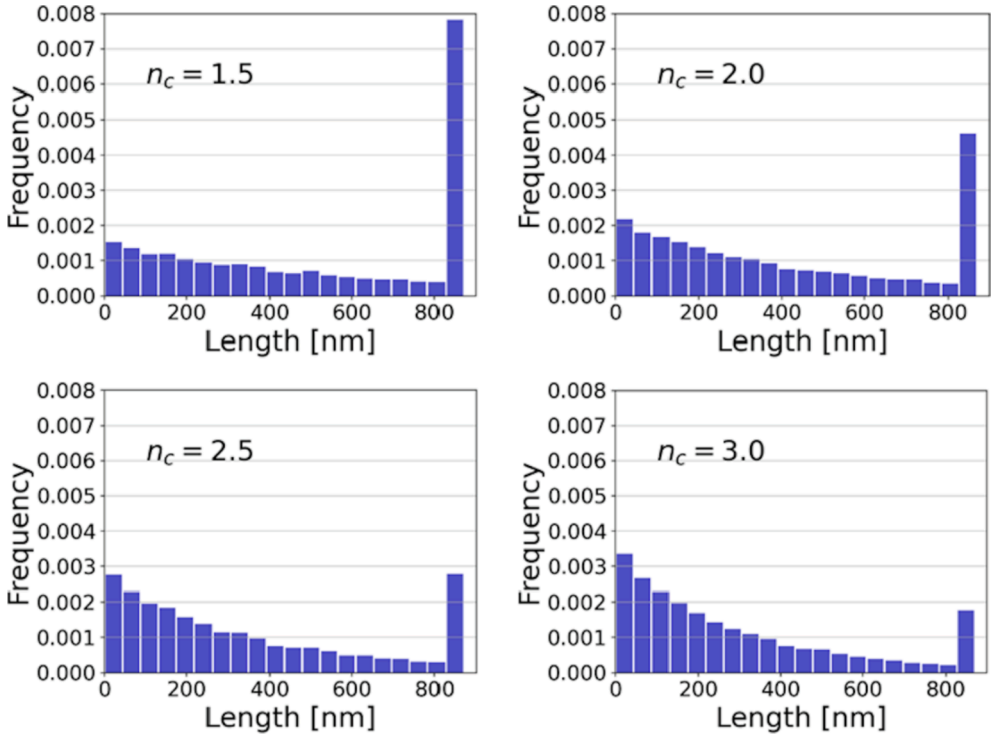

Fig. 6. The distribution of DNA fragments length derived in numerical simulations for the following number of DSBs, nc: 1.5, 2, 2.5, and 3.

and the fraction of unbroken DNA molecules (circular) is:

$$q_{loop}(n_c) = \frac{1}{1 + n_c e^{n_c}}. \quad (6)$$

These equations follow directly to another relation that can be used to fast n_c estimation:

$$n_c = \frac{q(n_c)}{q_{loop}(n_c)} - 1 \quad (7)$$

It is worth noting that it is possible to analytically calculate the whole distribution of DNA fragments using the above model. It is given by:

$$p(l) = (1-q)\cdot\frac{\lambda e^{-\lambda l}}{1 - e^{-\lambda l_0}} + q\cdot\delta(l - l_0), \quad (8)$$

where l_0 is the length of DNA plasmid (control). The parameter λ can be numerically calculated from the following equation:

$$\frac{1}{\lambda} = l_0\frac{1 - (n_c + 1)e^{-n_c}}{n_c(1 - e^{-n_c})} + \frac{l_0}{e^{\lambda l_0} - 1} \quad (9)$$

To derive the distribution of DNA fragments length, additional assumption that the DSB sites are randomly distributed along the DNA strand is required, which is not true in the case of the plasmid studied here. However, the agreement between the above distribution and numerically obtained histograms shows that here, the DBS sites are irregularly distributed. As long as this assumption is satisfied, the model will provide correct results regardless of specific order of bases in DNA chain, and specific interaction between DNA and a factor that causes DSB.

3.4. The estimation of the average number of DSB

The results predicted by the model or numerical simulations can be compared to experimental histograms using the fraction of strands q, which were not cut more than once. Thus, we can determine the relation between Blm concentration and the mean number of DSBs, n_c.

According to the statistical model, the mean number of DSB (n_c) can be determined directly from Eq. (5), providing the fraction of the longest fragments q. However, estimation of this fraction from experimental histogram is not obvious because as it is seen in Fig. 1 A, even for the sample containing only circular DNA the distribution is spread from 500 nm up to 2000 nm. Therefore, to estimate fraction of circular (without DSBs) DNA fragments q from the experimental data, we propose the following procedure. First, we calculate the median of the unbroken (circular) plasmid length. Then for each distribution we count the number of fragments greater than this median. This is illustrated in Fig. 5, where the red vertical line marks the median – different for each concentration of Blm. We assume that all fragments in the peak are distributed equally on both sides of the median, and all fragments above the median are from the peak only. Therefore, we estimate the peak size as twice of the number of fragments above the median. The data obtained from analysis of obtained strands for different concentration of Blm are summarized in Table 1. These results enable us to solve numerically Eqs. (5) and (6) to obtain n_c.

3.5. Statistical and systematic errors

Of course, such a procedure is subjected to both statistical and systematic errors. At first, let us discuss the statistical errors, which have two sources. One corresponds to determination of median of circular DNA (without DSBs) length, and the second comes from determining the number of fragments in the right peak based on experimental histogram obtained after application of Blm. Both of them are estimated by bootstrapping [38]. Bootstrapping uses a set of values that build a given histogram, to create another version of this histogram. Another version uses the same number of values as the original one, and each of these values are randomly selected from the original set. Thus, in the new set some original values can repeat, while others may not be present at all. This new set is then used to determine the same parameter as the original one. Creating independently a number of such new sets we get a statistics of measured parameters, which are used to determine standard deviation of the measured characteristics. Here we apply this procedure to all acquired data (for the control sample and samples of DNA treated with bleomycin). Then we perform the procedure of estimating n_c. We repeat this 100 times. The standard deviation of n_c obtained in this way is presented in Fig. 7.

Table 1
Statistics of DNA fragments. The number of circular DNA molecules (without DSBs) was determined manually and the fraction of long strands (that was cut at most once) was determined using above-described procedure.

Blm [nM]	number of fragments	minimal length [nm]	number of circular (unbroken) plasmids	fraction of circular (unbroken) plasmids q_{loop}	long strands fraction q
8	896	64.8	658	0.734	0.576
17.5	1066	47.2	523	0.491	0.482
30	555	47.5	288	0.441	0.463
40	832	41.2	439	0.528	0.416
50	657	35.3	177	0.269	0.204
75	852	41.3	94	0.110	0.155
100	1032	47.3	145	0.140	0.153
200	587	41.1	114	0.185	0.194
300	943	35.5	73	0.077	0.155

3.6. Estimation of systematic errors

The systematic errors are more challenging to estimate than statistical errors. We have distinguished two sources of bias: the missing short fragments and leakage of longer fragments into the right peak.

The first source of systematic errors is related to the missing short fragments. When comparing the simulated data presented in Fig. 5 and the experimental data in Fig. 3 one can notice that the simulated distribution has a maximum at zero length while the experimental data has no fragments shorter then $l_{min} = 50$ nm. This is related to limitation of the measurement procedure. For data analysis, we acquired 3 × 3 μm AFM images (512 × 512 pixels). Thus, one pixel corresponds to 5.9 nm. However, short fragments may fold and form the structures occupying only a few pixels. It makes them indistinguishable from the other small artifacts on an image. Therefore, some of such fragments are considered neither by the analyzing program nor by the human operator. In

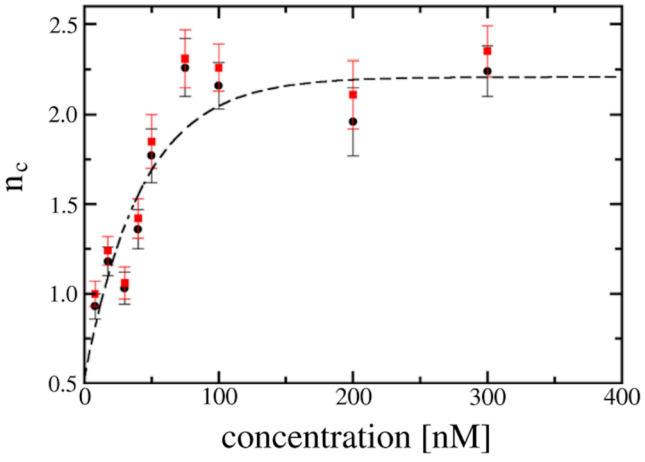

Fig. 7. The average number of DSB per one DNA plasmid strand estimated from the size of the peak (black dots), corrected for missing short fragments (red squares). Errors were estimated using bootstrap technique. Dashed line is to guide the eye.

consequence, the ratio q is overestimated. In order to compensate for this systematic error using the model, one may proceed as follows. The q parameter may be estimated as described above and used to obtain the parameter n_c. Then the number of fragments smaller than some l_{min} is calculated from the model and it can be used to correct the parameter q. One needs to iterate until convergence. The results of the described procedure for various values of the parameter l_{min} and n_c are collected in Table 2.

The percentage value of this correction in the studied range can be roughly estimated as $0.12 \cdot l_{min}$, where l_{min} is given in [nm]. Interestingly, it hardly depends on n_c.

The second source of systematic errors results from the fact that longer fragments, because of measurement errors, can be wrongly attributed to the peak. Again, one can estimate this effect using our model. We proceed in the similar way as described before. We started without any corrections and estimated n_c. Using our model, one can estimate how many fragments fall in the interval $(l_{med} - \Delta; l_{med})$ and subtract them from them from the peak. Then the new q value can be calculated. This process is repeated until convergence. The influence of this error on the results obtained here cannot be easily estimated, because the experimental procedure does not give any hints concerning fraction of such longer fragments and value of Δ. However, like the previous correction this one will increase the measured value of n_c.

3.7. The number of DSBs - results

Using the method described above, we estimated the mean number of DSBs for the studied DNA plasmid for different concentrations of Blm. As an input we used experimentally measured histograms of DNA lengths presented in Figs. 1 and 3. The result is shown in Fig. 7.

The number of DSBs increases with Blm concentration in the range of 8–50 nM. Interestingly, the amount of DSBs saturates for Blm concentration slightly above 50 nM. The experimentally obtained distributions of DNA fragments length do not differ much for higher concentrations. Since the mechanism of action of bleomycin is not known, interpretation of the observed saturation is ambiguous. One possible reason may be related to local conformation changes induced at site of DNA lesion. This molecular transition alters the distance between DNA base pairs modifying the size of minor and major grooves [32]. Molecules such as bleomycin interact with DNA via the minor groove of the B-DNA helix [39,40]. Therefore, change of its size may prohibit further interaction of bleomycin with DNA.

The number of DSBs derived from Pang (Eq. (1)), Eq. (5), and Eq. (6) is summarized in Table 3. The results calculated according to Pang are underestimated as expected because of the applied here AFM spatial resolution, which was too low to resolve short DNA fragments (theoretically below ca. 12 nm according to Nyquist theorem). In practice, as shown in Table 1, the shortest observed length of DNA fragments was at least three times larger.

It can be noticed that results obtained using Eqs. (1) and (6) are similar, especially when the missing small fragments are not considered. Both these estimations use number of uncut DNA plasmids, that are counted manually, which is inefficient and error prone. Additionally, Pang's method – Eq. (1) is very sensitive to number of short fragments, that cannot be observed because of applied here AFM resolution. Using above-described procedure, we estimate that the Eq. (1) underrates the mean number of DSB by approximately 30%, which is 2–4 times larger than for the method based on Eq. (5). Underestimation of DSBs number calculated from Eq. (6) is likely related to relatively low statistics (low number of unbroken molecules) due to the rapid decrease of the number of unbroken plasmids with the increasing Blm concentration.

Table 3
The average number of DSBs per plasmid according to three different calculation procedures. Values in brackets contain the systematic error corresponding to the lack of short DNA fragments in the observed samples. The statistical error estimated using bootstrapping is below 0.2 for all provided values.

Blm [nM]	Eq. (1)	Eq. (5)	Eq. (6)
8	0.31 (0.37)	0.93 (1.00)	0.27 (0.29)
17.5	0.68 (0.84)	1.18 (1.24)	0.58 (0.59)
30	0.73 (0.89)	1.03 (1.06)	0.66 (0.70)
40	0.60 (0.71)	1.36 (1.42)	0.53 (0.52)
50	1.31 (1.62)	1.77 (1.85)	1.00 (1.01)
75	1.73 (2.22)	2.26 (2.31)	1.61 (1.65)
100	1.71 (2.23)	2.16 (2.26)	1.44 (1.49)
200	1.48 (1.87)	1.96 (2.11)	1.26 (1.30)
300	2.01 (2.50)	2.24 (2.35)	1.86 (1.84)

4. Conclusions

We have provided a new method for estimating the number of double-strand breaks (DSBs), induced by various damaging factors, basing on the histogram of DNA fragments length. Additionally, we observe the influence of SSBs on the DNA conformation. The arising number of SSBs inducted by bleomycin in the individual DNA molecule results in consistent shortening of DNA chain length with the increase of Blm concentration. This effect is probably related to local conformational transitions induced by SSBs. Comparing to previously used method (arithmetic formula proposed by Pang, Eq. (1)) our method is more robust, especially for relatively high bp DNA molecules like pUC19 as it relies on the fraction of the longest fragments observed in the sample. Our statistical model takes into consideration all DNA fragments and provides more relevant values of DSBs than the formula provided by Pang. The results confirmed that Pang formula is affected significantly by AFM resolution, and thus, the mean number of DSBs is underestimated. Nevertheless, we introduced the numerical procedure allowing to estimate the correct value considering missing short fragments of DNA strands. The provided here method is more resistant to systematic errors and provides results comparable to Pang's equation. Moreover, we attached in the Supporting Information the program that allows to automatically calculate the number of DSBs according to all

Table 2
Systematic error corrections of n_c for different values of the microscope resolution and the mean number of cuts per strand estimated from Eq. (5). The values were obtained numerically with the precision of 0.01.

l_{min} [nm]	n_c									
	0.5	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**
10	0.50	1.01	1.51	2.02	2.53	3.03	3.54	4.05	4.55	5.06
20	0.51	1.02	1.53	2.04	2.56	3.07	3.58	4.10	4.61	5.12
30	0.51	1.03	1.55	2.07	2.59	3.11	3.62	4.15	4.66	5.18
40	0.52	1.04	1.57	2.09	2.62	3.14	3.67	4.20	4.72	5.24
50	0.52	1.05	1.58	2.11	2.65	3.18	3.71	4.25	4.78	5.28
60	0.53	1.06	1.60	2.14	2.68	3.21	3.76	4.30	4.84	5.35
70	0.53	1.07	1.61	2.17	2.71	3.25	3.80	4.35	4.90	5.42
80	0.53	1.08	1.63	2.19	2.74	3.28	3.84	4.40	4.96	5.49
90	0.54	1.09	1.65	2.21	2.77	3.32	3.89	4.45	5.01	5.56
100	0.54	1.1	1.66	2.23	2.80	3.35	3.94	4.50	5.06	5.64

presented and discussed methods.

Our method allows to estimate the accurate number of DSB in the case of limited image resolution which is determined by a simultaneous observation of the whole DNA length distribution, from few nanometers to few hundreds of nanometers. To calculate the number of DSBs correctly for numerous potential cut targets, it is essential to get a reasonable statistic. It is also crucial to choose the appropriate ratio between the molecule size / image resolution / the number of images for reasonable statistics / experiment time. Therefore, to observe ca. 830 nm DNA plasmid and its fragments after the induction of DSB, we applied AFM imaging with the resolution of 5.9 nm per pixel that allows to observe several plasmid molecules for one 3 × 3 μm image, but shorter DNA fragments may be not visible or not possible to be interpreted as a short DNA fragment. Additionally, the finite resolution of AFM imaging as presented here, which affects the observation of shorter fragments should be systematically considered in the analysis of the obtained results.

The main advantage of the presented method is the fact that it is based on the statistics of the obtained length of DNA fragments from AFM images. It does not require highly precise measurement of the length of DNA fragments (high-resolution AFM). In order to obtain reliable results, the resolution of AFM images needs to be comparable and reasonable. Relaying on the statistics, we showed that it is possible to take into consideration even those DNA fragments that are too short to be visualized with applied resolution of AFM images. Thus, the precision of obtained number of DSBs is increased in comparison to results obtained only based on numerically obtained histogram from experiments. Therefore, the presented method could be applied in biologically relevant systems of relatively long DNA strands, where an application of high-resolution AFM imaging is impossible. Moreover, the presented method could be applied in research on the DSBs formation upon exposure to various factors, such as drugs, enzymes, ionizing radiation, free radicals or other. However, an application of macromolecules as the factor inducing DSBs may interfere with the DNA length measurement, since macromolecules and DNA fragments displaying comparable dimensions could be easily mistaken on AFM topography images.

Funding

This work was supported by the National Science Centre, Poland under the “OPUS 16” project [Reg. No. UMO-2018/31/B/ST4/02292].

CRediT authorship contribution statement

Kamila Sofińska: Conceptualization, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Michał Cieśla:** Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Jakub Barbasz:** Conceptualization, Validation, Data curation, Writing – review & editing. **Natalia Wilkosz:** Investigation, Data curation. **Ewelina Lipiec:** Writing – original draft, Writing – review & editing, Project administration. **Marek Szymoński:** Writing – review & editing, Funding acquisition. **Piotr Białas:** Methodology, Software, Formal analysis, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Science Centre, Poland under the “OPUS 16” project [Reg. No. UMO-2018/31/B/ST4/02292].

Appendix A. Supplementary material

Supporting Information containing Figure S1. AFM images of DNA treated with 8, 17.5, 40, 50, and 200 nM of Blm. (PDF). Program for calculation of the DSBs number. (.zip). Supplementary data to this article can be found online at https://doi.org/10.1016/j.measurement.2022.111362.

References

[1] S.P. Jackson, J. Bartek, The DNA-damage response in human biology and disease, Nature 461 (2009) 1071–1078, https://doi.org/10.1038/nature08467.
[2] C. Richardson, M. Jasin, Frequent chromosomal translocations induced by DNA double-strand breaks, Nature 405 (2000) 697–700, https://doi.org/10.1038/35015097.
[3] A.S. Bader, M. Bushell, DNA:RNA hybrids form at DNA double-strand breaks in transcriptionally active loci, Cell Death Dis. 11 (2020) 1–7, https://doi.org/10.1038/s41419-020-2464-6.
[4] S. Agarwal, A.A. Tafel, R. Kanaar, DNA double-strand break repair and chromosome translocations, DNA Repair (Amst). 5 (2006) 1075–1081, https://doi.org/10.1016/j.dnarep.2006.05.029.
[5] W.J. Cannan, D.S. Pederson, Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin, J. Cell. Physiol. 231 (2016) 3–14, https://doi.org/10.1002/jcp.25048.
[6] J.M.R. De Almodóvar, G.G. Steel, S.J. Whitaker, T.J. Mcmillan, A comparison of methods for calculating DNA double-Strand break induction frequency in mammalian cells by pulsed-field gel electrophoresis, Int. J. Radiat. Biol. 65 (1994) 641–649, https://doi.org/10.1080/09553009414550751.
[7] B. Cedervall, R. Wong, N. Albright, J. Dynlacht, P. Lambin, W.C. Dewey, Methods for the quantification of DNA double-strand breaks determined from the distribution of DNA fragment sizes measured by pulsed-field gel electrophoresis, Radiat. Res. 143 (1995) 8–16, https://doi.org/10.2307/3578920.
[8] P.L. Olive, J.P. Banáth, The comet assay: A method to measure DNA damage in individual cells, Nat. Protoc. 1 (2006) 23–29, https://doi.org/10.1038/nprot.2006.5.
[9] N.P. Singh, M.T. McCoy, R.R. Tice, E.L. Schneider, A simple technique for quantitation of low levels of DNA damage in individual cells, Exp. Cell Res. 175 (1988) 184–191, https://doi.org/10.1016/0014-4827(88)90265-0.
[10] Y. Zhu, A. Biernacka, B. Pardo, N. Dojer, R. Forey, M. Skrzypczak, B. Fongang, J. Nde, R. Yousefi, P. Pasero, K. Ginalski, M. Rowicka, qDSB-Seq is a general method for genome-wide quantification of DNA double-strand breaks using sequencing, Nat. Commun. 10 (2019) 1–11, https://doi.org/10.1038/s41467-019-10332-8.
[11] A. Japaridze, A. Benke, S. Renevey, C. Benadiba, G. Dietler, Influence of DNA binding dyes on bare DNA structure studied with atomic force microscopy, Macromolecules. 48 (2015) 1860–1865, https://doi.org/10.1021/ma502537g.
[12] D. Pang, B.L. Berman, S. Chasovskikh, J.E. Rodgers, A. Dritschilo, Investigation of neutron-induced damage in DNA by atomic force microscopy: Experimental evidence of clustered DNA lesions, Radiat. Res. 150 (1998) 612–618, https://doi.org/10.2307/3579883.
[13] A. Japaridze, D. Vobornik, E. Lipiec, A. Cerreta, J. Szczerbinski, R. Zenobi, G. Dietler, Toward an Effective Control of DNA’s Submolecular Conformation on a Surface, Macromolecules. 49 (2016) 643–652, https://doi.org/10.1021/acs.macromol.5b01827.
[14] E. Gudowska-Nowak, K. Psonka-Antończyk, K. Weron, T. Elsässer, G. Taucher-Scholz, Distribution of DNA fragment sizes after irradiation with ions, Eur. Phys. J. E. 30 (2009) 317–324, https://doi.org/10.1140/epje/i2009-10522-7.
[15] K. Psonka-Antonczyk, T. Elssser, E. Gudowska-Nowak, G. Taucher-Scholz, Distribution of double-strand breaks induced by ionizing radiation at the level of single DNA molecules examined by atomic force microscopy, Radiat. Res. 172 (2009) 288–295, https://doi.org/10.1667/RR1772.1.
[16] D. Pang, A.R. Thierry, A. Dritschilo, DNA studies using atomic force microscopy: Capabilities for measurement of short DNA fragments, Front. Mol. Biosci. 2 (2015), https://doi.org/10.3389/fmolb.2015.00001.
[17] P. Wiggins, T. Van Der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, P.C. Nelson, High flexibility of dna on short length scales probed by atomic force microscopy, Nat. Nanotechnol. 1 (2006) 137–141, https://doi.org/10.1038/nnano.2006.63.
[18] L.V. Liu, C.B. Bell, S.D. Wong, S.A. Wilson, Y. Kwak, M.S. Chow, J. Zhao, K. O. Hodgson, B. Hedman, E.I. Solomon, Definition of the intermediates and mechanism of the anticancer drug bleomycin using nuclear resonance vibrational spectroscopy and related methods, Proc. Natl. Acad. Sci. USA 107 (52) (2010) 22419–22424.
[19] V. Murray, J.K. Chen, L.H. Chung, The interaction of the metallo-glycopeptide antitumour drug bleomycin with DNA, Int. J. Mol. Sci. 19 (2018), https://doi.org/10.3390/ijms19051372.
[20] J. Chen, J.A. Stubbe, Bleomycins: Towards better therapeutics, Nat. Rev. Cancer. 5 (2005) 102–112, https://doi.org/10.1038/nrc1547.
[21] O. Piétrement, D. Pastré, F. Landousy, M.O. David, S. Fusil, L. Hamon, A. Zozime, E. Le Cam, Studying the effect of a charged surface on the interaction of bleomycin with DNA using an atomic force microscope, Eur. Biophys. J. 34 (2005) 200–207, https://doi.org/10.1007/s00249-004-0443-y.

[22] W. Li, W.E. Antholine, D.H. Petering, Kinetics of reaction of DNA-bound Fe(III) bleomycin with ascorbate: Interplay of specific and non-specific binding, J. Inorg. Biochem. 90 (2002) 8–17, https://doi.org/10.1016/S0162-0134(02)00368-9.
[23] S.M. Hecht, Bleomycin: New perspectives on the mechanism of action, J. Nat. Prod. 63 (2000) 158–168, https://doi.org/10.1021/np990549f.
[24] M. Pitié, G. Pratviel, Activation of DNA carbon - Hydrogen bonds by metal complexes, Chem. Rev. 110 (2010) 1018–1059, https://doi.org/10.1021/cr900247m.
[25] B. Roy, S.M. Hecht, Hairpin DNA sequences bound strongly by bleomycin exhibit enhanced double-strand cleavage, J. Am. Chem. Soc. 136 (2014) 4382–4393, https://doi.org/10.1021/ja500414a.
[26] J. Chen, M.K. Ghorai, G. Kenney, J.A. Stubbe, Mechanistic studies on bleomycin-mediated DNA damage: Multiple binding modes can result in double-stranded DNA cleavage, Nucleic Acids Res. 36 (2008) 3781–3790, https://doi.org/10.1093/nar/gkn302.
[27] Y. Jiang, C. Ke, P.A. Mieczkowski, P.E. Marszalek, Detecting ultraviolet damage in single DNA molecules by atomic force microscopy, Biophys. J. 93 (2007) 1758–1767, https://doi.org/10.1529/biophysj.107.108209.
[28] Y.L. Lyubchenko, L.S. Shlyakhtenko, Visualization of supercoiled DNA with atomic force microscopy in situ, Proc. Natl. Acad. Sci. USA 94 (1997) 496–501, https://doi.org/10.1073/pnas.94.2.496.
[29] D. Nečas, P. Klapetek, Gwyddion: an open-source software for SPM data analysis, Open Phys. 10 (2012) 181–188, https://doi.org/10.2478/s11534-011-0096-2.
[30] I. Usov, R. Mezzenga, FiberApp: An open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects, Macromolecules. 48 (2015) 1269–1280, https://doi.org/10.1021/ma502264c.
[31] A. Wolffe, Chromatin: Structure and Function, third ed., Academic Press, 1999.
[32] B.R. Wood, The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues, Chem. Soc. Rev. 45 (2016) 1980–1998, https://doi.org/10.1039/c5cs00511f.
[33] E. Lipiec, R. Sekine, J. Bielecki, W.M. Kwiatek, B.R. Wood, Molecular characterization of DNA double strand breaks with tip-enhanced Raman scattering, Angew. Chemie - Int. Ed. 53 (2014) 169–172, https://doi.org/10.1002/anie.201307271.
[34] C. Ke, Y. Jiang, P.A. Mieczkowski, G.G. Muramoto, J.P. Chute, P.E. Marszalek, Nanoscale detection of ionizing radiation damage to DNA by atomic force microscopy, Small. 4 (2008) 288–294, https://doi.org/10.1002/smll.200700527.
[35] Y. Jiang, M. Rabbi, P.A. Mieczkowski, P.E. Marszalek, Separating DNA with different topologies by atomic force microscopy in comparison with gel electrophoresis, J. Phys. Chem. B. 114 (2010) 12162–12165, https://doi.org/10.1021/jp105603k.
[36] E. Lipiec, K.R. Bambery, P. Heraud, W.M. Kwiatek, D. McNaughton, M.J. Tobin, C. Vogel, B.R. Wood, Monitoring UVR induced damage in single cells and isolated nuclei using SR-FTIR microspectroscopy and 3D confocal Raman imaging, Analyst. 139 (2014) 4200–4209, https://doi.org/10.1039/c4an00838c.
[37] H. Nyquist, Certain Topics in Telegraph Transmission Theory, Trans. Am. Inst. Electr. Eng. 47 (1928) 617–644, https://doi.org/10.1109/T-AIEE.1928.5055024.
[38] B. Efron, R.J. Tibshirani (Eds.), An Introduction to the Bootstrap, Springer US, Boston, MA, 1993.
[39] J. Kuwahara, Y. Sugiura, Sequence-specific recognition and cleavage of DNA by metallobleomycin: minor groove binding and possible interaction mode, Proc. Natl. Acad. Sci. USA 85 (1988) 2459–2463, https://doi.org/10.1073/pnas.85.8.2459.
[40] K.D. Goodwin, M.A. Lewis, E.C. Long, M.M. Georgiadis, Crystal structure of DNA-bound Co(III)·bleomycin B2: Insights on intercalation and minor groove binding, Proc. Natl. Acad. Sci. USA 105 (2008) 5052–5056, https://doi.org/10.1073/pnas.0708143105.