ELSEVIER

Ultramicroscopy 87 (2001) 55–66

ultramicroscopy

www.elsevier.nl/locate/ultramic

Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica

Claudio Rivetti*, Simone Codeluppi

Istituto di Scienze Biochimiche, Universitá degli Studi di Parma, Parco Area delle Scienze 23/A, I-43100 Parma, Italy

Received 13 March 2000; received in revised form 21 July 2000

Abstract

Achieving the most correct estimate of the contour length of digitized DNA molecules is a key aspect of the microscopic analysis of nucleic acids by either electron microscopy (EM) or atomic force microscopy (AFM). Six different methods, that are mathematically not too complex and suited for common, practical use, have been tested here using simulated polymers in two dimensions and real DNA molecules (564, 1054, 2049 and 4297 bp long) imaged in air by AFM. The main result is that the frequently used Freeman estimator ($L_F = n_e + \sqrt{2}n_o$) overestimates the real contour length of the polymers by about 4%. More accurate estimates are obtained with the Kulpa estimator ($L_K = 0.948n_e + 1.343n_o$) or with the corner count estimator ($L_C = 0.980n_e + 1.406n_o - 0.091n_c$). In the range of the DNA sizes and magnifications we have considered, however, the best results are obtained with an ad hoc developed routine that smoothes the DNA trace by a polynomial fitting of degree 3 over a moving window of 5 points. Under these conditions, the difference between the measured and the real contour length of the molecules is less than 0.4%. The accuracy of this procedure allowed us to reveal a discrete, size-dependent, shortening of DNA molecules deposited onto mica under low salt conditions and imaged in air by AFM. Awareness of this structural alteration, that can be attributed to a partial transition from B- to A-form DNA, may lead to a more correct interpretation of DNA molecules or protein–DNA complexes imaged by AFM.

Keywords: AFM, SFM; DNA contour length; A-form DNA; Image processing

1. Introduction

The direct assessment of the contour length of DNA fragments by either atomic force microscopy (AFM) or electron microscopy (EM) is a widely used tool for investigating the physical properties of individual or protein-bound DNA molecules. By knowing the number of base pairs and the length of a DNA fragment, it is possible to infer the helical rise per base pair [1,2] and distinguish between different DNA secondary structures, such as A, B or mixtures of them [3,4]. Another physical parameter that is often obtained from the contour length analysis of EM or AFM images is the DNA

*Corresponding author. Tel.: 39-0521-905649; fax: 39-0521-905151.

E-mail address: rivetti@unipr.it (C. Rivetti).

0304-3991/01/$ - see front matter
PII: S0304-3991(00)00064-4

persistence length, i.e. the flexibility of the polymer [5–8]. In fact, by referring to the principles of polymer chain statistics, one can use the measured contour length and the mean square end-to-end distance to calculate the persistence length of individual DNA molecules. Further extending the scope of these measurements, it has been demonstrated that the polymer worm-like chain theory can readily be applied to bent DNA molecules as well as to DNA molecules harboring regions of different flexibility [9,10]. In particular, it has been shown that the DNA bend angle, intrinsic or protein-induced, can be inferred from the position of the bend, the mean square end-to-end distance and the contour length of the molecule. More recently, systematic comparisons of the contour lengths of free and protein-associated DNA molecules have revealed the wrapping of DNA around bacterial and yeast RNA Polymerases engaged in initiation complexes [10,11]. Contour length determinations have also been used to establish whether a ligand binds to DNA in an intercalative or nonintercalative manner [12]. In this case, the observation of ligand-induced lengthening of single DNA molecules imaged by AFM was considered as a direct evidence for intercalation. Finally, the fast and reliable sample preparation and the relative ease of use of AFM imaging make it an attractive tool for the sizing of megabase DNA fragments in the construction of physical genome maps and in the mapping of protein–DNA binding sites [13–15]. It is therefore apparent, that all of the above applications require an accurate and reliable method to determine the DNA contour length.

Discrete measurements of the properties of real polymers from digital images, in which they are described as a subset of pixels in a two-dimensional grid, is a nontrivial task. This is mainly due to the fact that during the digitization process, the exact contour of the original molecule is lost and what is left is merely an approximation of it. Therefore, the contour length of the original molecule can only be estimated rather than exactly measured [16], and the reliability of such estimate will depend on both image resolution and the method employed to calculate the contour length from the string of pixels. While image resolution is an inherent property of the microscope or of the microscope-sample system [17], two operational steps of image processing can significantly affect the reliability of DNA contour length determinations from digital images. These are the correct identification of the pixels that best describe the backbone of the molecule in the image, and the choice of an image processing algorithm capable of yielding the most accurate estimate of the contour length. The first step consists in the digitization of the DNA backbone that can be done either manually, by hand tracing the molecule from the image [7], or automatically, in which a pattern recognition algorithm selects the different molecules from the image [18]. The second step involves the computation of the contour length from a discrete number pixels using a particular algorithm which is likely to be a critical aspect of DNA length determination.

Relying on manual digitization of the DNA backbone, we have systematically evaluated the measuring accuracy and reliability of six different algorithms for determining the length of the digitized contours of DNA imaged by AFM. Computer-generated images of simulated worm-like chains were used as absolute length standards for this analysis. The main result is that the best performing method in the evaluation of the DNA contour length is an *ad hoc* developed algorithm that smoothes the DNA trace by a polynomial fitting of degree 3 over a moving window of 5 points. The accuracy of this method, which yields length estimates with errors less than 0.4%, allowed us to detect a discrete shortening of the DNA molecules that could be attributed to a partial B- to A-form transition.

2. Materials and methods

2.1. DNA sample preparation

The 564, 1054 and 4297 bp DNA fragments were obtained by restriction digestion of plasmid pSAP (6040 bp) with HindIII endonuclease in recommended reaction conditions. The 2049 bp DNA fragment was obtained by a combined restriction digestion of plasmid pSAP with EcoRI

and SspI endonucleases. All fragments were purified in 1% (w/v) agarose gel and electroeluted by means of an Elutrap apparatus (Schleicher & Schuell, Keene NH). The DNA solution was treated with phenol/chloroform followed by ethanol precipitation. The pellet was solubilized in Tris–EDTA buffer and the DNA concentration was determined by absorbance at 260 nm. The DNA samples were stored at 4°C. Restriction endonucleases were purchased from (Amersham Pharmacia Biotech).

2.2. Atomic force microscopy

A 20 μl solution of about 1 nM DNA in deposition buffer (4 mM Hepes pH 7.4, 10 mM NaCl, 2 mM $MgCl_2$) was deposited onto freshly cleaved mica and incubated for 1–5 min. The mica disc was then rinsed with few milliliters of nanopure water and the surface was blow-dried with a flow of nitrogen. All images were collected in air with a Nanoscope III microscope (Digital Instruments, Santa Barbara, CA) equipped with a type E scanner and operating in tapping mode. Commercial diving board-type silicon cantilevers (NSCS12, SiliconMDT Moscow) were used. The AFM images were collected with a dimension of 512×512 pixels and a scan size of 2 μm. The scan rate was about three lines per second.

2.3. Simulation of polymers in two dimensions and image processing

For the purposes of this study, a DNA molecule can reasonably be considered, at the most detailed level, as chain of rigid rods each representing one base pair. The size of the rod is 0.34 nm and corresponds to the distance between two adjacent base pairs. For a chain possessing some degree of stiffness, the angle θ between two rods is a function of the flexibility of the polymer and of the curvilinear distance (ℓ) between the rods. In two dimensions, the angles between two consecutive rods are chosen from a Gaussian probability distribution centered around 0 and having a standard deviation of 4.59° as given by: $\sigma = \sqrt{\ell/P}$, where P is the DNA persistence length which has been assumed to be of 53.0 nm [7]. The number of rods composing the polymer corresponded to the number of base pairs of the different DNA fragments. In conformity with the scan size of the AFM images presented in this study, the polymers were generated within a 2×2 μm grid and taking into account excluded volume effects (i.e. the polymers were not allowed to cross themselves or other polymers already generated). For each grid, a number of polymers varying from 5 to 40, depending on their length, was generated. Binary images were obtained by assuming that each grid was made of 512×512 pixels (as the AFM images) and by setting to one any pixel through which the polymer passes. The binary images were skeletonized to eight connected lines using the *erode* function of Matlab (MathWorks, Natick, MA). The ensemble of the "on" pixels, representing the digitized polymers, were used for length determination. The n codes of the chaincode string were obtained by connecting the center of the $n+1$ grid points of each digitized polymer (Fig. 1). We will refer to such measurement as *automatic* since no influence from the imaging technique nor from the operator bias is introduced.

The binary images were then converted to AFM-like images by passing a hypothetical tip over the two-dimensional grid [19]. This was done by placing an inverted and reversed tip surface, which we have assumed to be a parabolic function

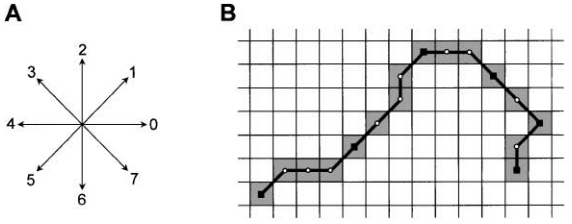

Fig. 1. (A) Scheme of the eight connected chaincode. (B) Representation of a digitized curved line. The number of boundary pixels (gray grid elements) is 16. The number of codes connecting the center of the boundary pixels (dark thick line) is 15 and from left to right the chaincode string can be written as: 100111210077756. There are 6 even codes (n_e) and the 9 odd codes (n_o). From left to right, the polygonal decomposition of this arbitrary chaincode string, using the linearity condition given in [24], results in 5 straight segments here delimited by the dark square dots. The total number of corners (n_c), obtained from the sum of the corners in each segment, is 6.

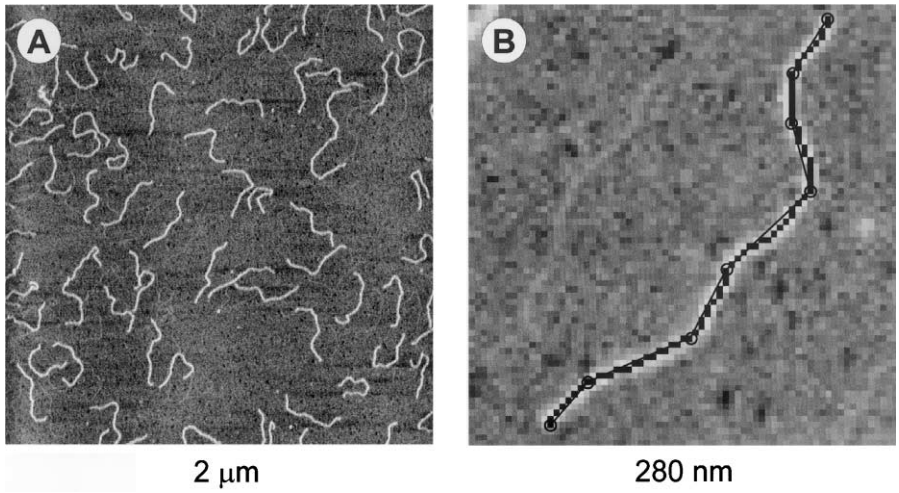

Fig. 2. (A) AFM image of the 1054 bp DNA fragment deposited onto mica under conditions that permit equilibration of the molecules onto the surface. (B) Enlargement of a DNA molecule. The thin line represents the hand-drawn trace from which the DNA backbone is found. The circles are the actual points selected by the operator. The eight connected chaincode of pixels, from which the contour length is computed, are shown in black. The gray scale corresponds to a vertical size of 2 nm from black to white.

with radius of curvature of 20 nm, on top of each grid point and forming a new image from the envelope of all such tip surfaces. To make the images similar to a real case, a random noise with an intensity between 0 and 0.1 nm was added.

These and AFM images of DNA molecules were analyzed using a *manual* procedure in which the DNA trace was digitized as follows: the two ends and several points along the DNA were selected with the mouse. A first DNA trace is obtained by interpolating the selected points with steps of one pixel. The position of each point is then adjusted to that of the pixel with the highest intensity within a five pixel wide window. Each digitized polymer string of pixels was skeletonized and the chaincode string was obtained as described above (Fig. 2B). Image processing and computations were performed with Matlab except the edge chord algorithm for which the ROOTEDGE program [20] was used.

2.4. Algorithms for length determination of digitized contours

For each DNA fragment and simulated polymer, six different algorithms have been employed to calculate the DNA contour length. The Freeman estimator (L_F), the most probable origin estimator (L_{MPO}) and the Kulpa estimator (L_K), compute the length of a chaincode string from the number of even and odd elements, denoted by n_e, n_o, respectively. This is thus an (n_e, n_o) characterization. The corner-corrected chain estimator (L_C) considers, beside even and odd elements, the number of corners (knight's moves) in the chaincode. Accordingly, this is an (n_e, n_o, n_c)-characterization. The n-point polynomial fitting (L_{PF}) takes advantage of a fitting routine to first smooth the skeleton and then determining the contour length of the resulting trace. The edge chord algorithm (L_{ECA}) draws chords along the object edge to determine its perimeter. Although the first four estimators were developed for straight strings, they can suitably be used in the length determination of arbitrary chaincode strings. In fact, by making a polygonal decomposition of the string into straight substrings, an arbitrary chaincode string can be considered as piecewise straight [16]. In practice, it is necessary to make a polygonal decomposition of the chaincode string only for the (n_e, n_o, n_c)-based estimator; in fact while the number of even and

odd codes remains the same, the number of corners changes upon polygonal decomposition.

The Freeman estimator: This estimator, introduced by Freeman in 1970 [21] and still in common use, considers the Euclidean distance of the eight connected chaincode string in which the contour length of the polymer (L_F) is given by

$$L_F = n_e + \sqrt{2}n_o = 1.000n_e + 1.414n_o, \quad (1)$$

where n_e and n_o are the number of even and odd codes of the eight connected chaincode string (Fig. 1).

The MPO estimator: This estimator [16,22], that has been shown to be very accurate in the length determination of straight segments, is given as

$$L_{MPO} = \sqrt{(n_e + n_o)^2 + n_e^2}. \quad (2)$$

The Kulpa estimator: Starting from Eq. (1), Kulpa [23] derived coefficients for the even and odd pixels that minimized the error in the measurements of long line segments. The resulting estimator is given by

$$L_K = 0.948n_e + 1.343n_o. \quad (3)$$

The corner chain estimator: Vossepoel and Smeulders [24] presented a corner-corrected chain method in which the corner count n_c counts the number of "knight's moves", consecutive odd–even or even–odd sequences in the chaincode string. The optimal coefficients for this estimator were evaluated by a probability analysis of various chain code representations of straight line segments. The L_C estimator is given by

$$L_C = 0.980n_e + 1.406n_o - 0.091n_c. \quad (4)$$

This estimator is used to determine the length of straight segments obtained by polygonal decomposition of the curved DNA trace using the linearity conditions reported in Ref. [24]. The DNA contour length is then determined by the sum of the polygon side lengths (Fig. 1).

The n-point moving polynomial fitting: This method consists in smoothing the string of pixels using polynomial fitting [7]. The coordinates of each pixel in the string are adjusted by fitting a polynomial of a certain degree over a window of n-points. The contour length is then calculated from the sum of the Euclidean distance between consecutive points in the smoothed line:

$$L_{PF} = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}. \quad (5)$$

This routine has been tested using polynomial degrees ranging from 3 to 7 and moving windows from 3 to 21 points (data not shown). The best estimates were obtained with the following degree–window combinations: 3–5, 6–11 and 7–11. Although there were no appreciable differences between these combinations of values, we decided to use the polynomial of the lowest degree and a moving window of 5 points.

The edge chord algorithm: Recently, Ewing and Kasper [25] have proposed a method to determine the perimeter and the length of digitized objects that draws chords along the object edge and computes their Euclidean distances. For objects that have a relatively constant width the length is obtained from the perimeter (P) and the area (A) using the following relation:

$$L_{ECA} = \frac{P + \sqrt{P^2 - 16A}}{4}. \quad (6)$$

The ECA, developed initially for the determination of the length of plant roots, was shown to be accurate and precise with both random and nonrandom object arrangements.

3. Results

3.1. Images of real and computer-generated DNA molecules

The comparative evaluation of different contour length determination procedures, requires an absolute reference value for the real length of the molecules under examination. In the case of DNA, this could be obtained, at least in principle, by assuming a B-form structure and multiplying the number of base pairs by 0.34 nm. In most cases, however, AFM or EM images of DNA are recorded in a dehydrated environment that could

alter the twist, roll or tilt of the bases, and consequently affect the actual contour length of the molecules. Furthermore, imaging artifacts, such as the broadening effect produced by the AFM tip, can make the molecules appear longer than their real dimensions [26], while a significant shortening effect can result from overlaps caused by a harsh deposition process during the transition from solution to a two-dimensional grid [7]. Because of these limitations, we have chosen to compare the contour lengths of real DNA molecules with those of computer-generated polymers. With this procedure it is possible to distinguish between errors produced by the algorithm employed to compute the contour length, and those deriving from sample preparation or operator bias.

Four different DNA fragments of 564, 1054, 2049, and 4297 bp have been used in this study. All fragments were prepared by restriction digestion of plasmid DNA. The DNA was deposited onto freshly cleaved mica in a low salt buffer containing Mg^{2+}. In such conditions, DNA molecules are able to equilibrate on the surface, avoiding the formation of overlaps within the same or among different molecules [7]. AFM images were collected in air as described in Section 2. A typical AFM image of 1054 bp long DNA molecules is shown in Fig. 2A. The digitization of the DNA molecule was performed with the *manual* procedure in which the first and last pixel of the skeleton and a rough trace of the contour are selected by an operator (Fig. 2B).

For each of the four DNA fragments, a population of ideal worm-like chains having the same length and stiffness of the DNA molecules was generated as described in Section 2. The real contour length of the simulated polymers ($n \times 0.34$ nm) corresponds to the "ground truth" of the experiments. The digital representation of the polymers into a 512×512 pixel grid was skeletonized to eight connected lines and their length were calculated using the different procedures. By comparing the measured contour length with the "ground truth", it will be possible to evaluate the accuracy of the estimate.

The simulated polymers were also converted into AFM-like images as described in Section 2 and digitized using the manual procedure (Fig. 2B). Comparison between the automatic and the manual measurements of these polymers will help in understanding the influence of the broadening effect of the tip and the operator bias in the determination of the DNA contour length.

3.2. DNA contour length analysis

The results obtained for the four different DNA fragments are summarized in tables. Each table is divided in three panels (A, B, C) containing the contour length measurements of the simulated polymers (panels A, B) and the contour length measurements of DNA molecules (panel C). The simulated polymers in panel A have been measured with the *automatic* procedure from the binary images, whereas those in panel B and DNA molecules in panel C, have been measured with the *manual* procedure. All contour length values are the mean of the Gaussian fitting to the distribution of measurements. In all the tables, the "error" represents the difference between the measured contour length and B-form contour length divided by the B-form contour length. The "CV" is the coefficient of variation of the distributions. N is the number of molecules analyzed. As an example, the contour length distributions obtained for the 1054 bp fragment are shown in Fig. 3.

Table 1 shows the results obtained with a 564 bp DNA fragment that has a B-form contour length of 191.8 nm. About 200 simulated polymers were analyzed and their contour length measurements are reported in Table 1A.

Comparison among the different methods shows that the Freeman estimator (L_F) significantly overestimates the real contour length of the polymer. The difference between the measured value and the "ground truth" is 4.3%, which, expressed in terms of the number of 0.34 nm long segments, makes a difference of about 24. A significant improvement of the contour length estimate is obtained with the MPO and Kulpa estimators which underestimate the polymer contour length with an error of -1.6% and -1.1%, respectively. The corner count estimator and the polynomial smoothing algorithm gave a more

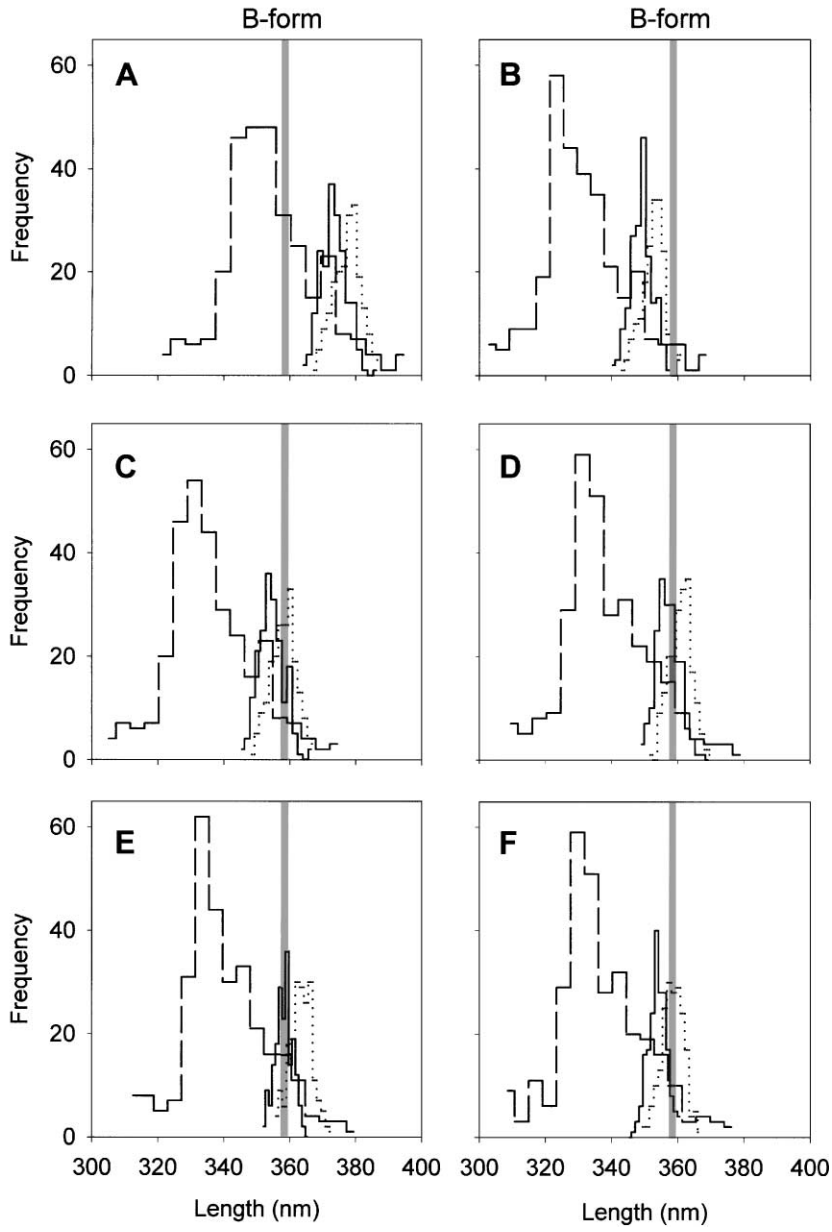

Fig. 3. Line plot of the contour length distributions obtained with the 1054 bp DNA fragment and correspondent simulated polymers. Each panel refers to one of the six methods used to compute the length. (A) Freeman estimator. (B) MPO estimator. (C) Kulpa estimator. (D) Corner count estimator. (E) Polynomial fitting algorithm. (F) ECA algorithm. In each panel are shown: simulated polymers measured with the "*automatic*" procedure (solid line), simulated polymers measured with the "*manual*" procedure (dotted line) and real DNA molecules measured with the "*manual*" procedure (dashed line). The vertical gray bar indicates the length of the full B-form DNA. The mean values obtained from the Gaussian fitting of each distribution are shown in Table 2.

Table 1
DNA 564 bp long (B-form contour length: 191.8 nm)[a]

Method	Contour length (nm)	Error (%)	CV (%)
A. *Simulated polymers: "automatic" measurements, N = 194*			
L_F	200.0	4.3	1.3
L_{MPO}	188.7	−1.6	1.4
L_K	189.6	−1.1	1.5
L_C	191.4	−0.2	1.4
L_{PF}	192.5	0.4	1.4
ECA	190.5	−0.7	1.5
B. *Simulated polymers: "manual" measurements, N = 198*			
L_F	204.7	6.8	1.9
L_{MPO}	193.5	0.9	1.8
L_K	194.2	1.3	1.9
L_C	195.8	2.1	1.7
L_{PF}	197.3	2.9	1.7
ECA	195.5	1.9	1.4
C. *DNA molecules: "manual" measurements, N = 395*			
L_F	190.0	−0.9	3.0
L_{MPO}	180.1	−6.1	2.9
L_K	180.2	−6.0	2.9
L_C	181.5	−5.3	2.8
L_{PF}	182.8	−4.7	2.9
ECA	181.6	−5.3	2.8

[a] Contour length measurements of the simulated polymers (panels A, B) and real DNA molecules 564 bp long (panel C) obtained with six different procedures. See text for details.

precise estimate of the polymer contour length, with errors of only 0.2% and 0.4%, respectively. In these cases, the difference between the measured contour length and the "ground truth" is less than 1 nm. Also the ECA algorithm is very reliable, with an error of 0.7%.

When the binary images of the simulated polymers are transformed to AFM-like images and the skeleton of the polymers is detected with the manual procedure, the results are slightly different (Table 1B). In particular, all the measured contour lengths are about 5 nm longer than those measured automatically (Table 1A). This result is attributed to the broadening effect of the AFM tip that produces an elongation of the polymers in the image. From this panel, the best evaluation of the contour length is given by the MPO estimator with an error of 0.9%. However, this value is due to a fortuitous, partial canceling of two factors: the elongation of the polymers produced by the tip convolution and the inherent contour length reduction of the MPO estimator.

Table 1C reports the measurements of DNA molecules imaged by AFM. At first glance, it appears that the Freeman estimator predicts well the real contour length of the molecules, whereas all the other methods perform poorly. However, a closer look at the data, reveals that the contour length of the DNA molecules is about 10 nm shorter than that of the simulated polymers measured automatically (Table 1A). Moreover, taking the tip effect into account, the contour length of the DNA molecules must be adjusted by subtracting the length difference between the values in Table 1A and B for each method. For example, considering the corner count method, the corrected DNA contour length would be: $181.5 - (195.8 - 191.4) = 177.1$ nm. Thus, the true length of DNA molecules deposited onto mica and imaged in air by AFM is shorter than the expected B-form contour length (191.8 nm). This implies that during the deposition process the molecules may undergo a partial B- to A-form transition. Due to a coincidence, in the case of the Freeman estimator, such contour length reduction is counterbalanced by the overestimation produced by this algorithm in computing the contour length and by the elongation of the molecules caused by the broadening effect of the tip.

Similar considerations can also be made for the other three DNA fragments analyzed in this study (Tables 2–4). In all cases, very accurate estimates are obtained with the corner count estimator and with the polynomial smoothing algorithm with errors always below 0.5%. On the other hand the Freeman estimator is the worst in all cases. Both the MPO and the ECA methods work reasonably well for the shorter DNA fragments but they loose accuracy as the DNA size increases.

It is interesting to notice that for the 4297 bp DNA fragment the contour length of the simulated polymers measured with the manual procedure (Table 4B) is shorter than that measured automatically (Table 4A), a reversed situation compared to the smaller DNA fragments. This is a consequence of the fact that to use our measuring procedure, the operators have worked at lower magnification in order to keep the whole molecule

Table 2
DNA 1054 bp long (B-form contour length: 358.4 nm)[a]

Method	Contour length (nm)	Error (%)	CV (%)
A. *Simulated polymers: "automatic" measurements, N = 192*			
L_F	373.2	4.1	1.1
L_{MPO}	349.1	−2.6	0.9
L_K	354.1	−1.2	1.1
L_C	356.8	−0.4	1.0
L_{PF}	358.4	0.0	0.8
ECA	354.0	−1.2	0.8
B. *Simulated polymers: "manual" measurements, N = 198*			
L_F	377.8	5.4	1.1
L_{MPO}	353.3	−1.4	0.9
L_K	358.4	0.0	1.1
L_C	361.2	0.8	1.0
L_{PF}	363.7	1.5	0.9
ECA	358.7	0.1	0.9
C. *DNA molecules: "manual" measurements, N = 305*			
L_F	351.8	−1.8	3.2
L_{MPO}	329.1	−8.2	3.1
L_K	333.5	−6.9	3.1
L_C	336.0	−6.2	3.2
L_{PF}	338.1	−5.6	3.1
ECA	334.5	−6.7	3.1

[a] Contour length measurements of the simulated polymers (panels A, B) and real DNA molecules 1054 bp long (panel C) obtained with six different procedures. See text for details.

Table 3
DNA 2049 bp long (B-form contour length: 696.7 nm)[a]

Method	Contour length (nm)	Error (%)	CV (%)
A. *Simulated polymers: "automatic" measurements, N = 181*			
L_F	725.8	4.2	0.8
L_{MPO}	674.9	−3.1	0.8
L_K	688.6	−1.2	0.8
L_C	694.0	−0.4	0.7
L_{PF}	696.1	−0.1	0.5
ECA	686.2	−1.5	0.6
B. *Simulated polymers: "manual" measurements, N = 195*			
L_F	725.4	4.1	0.9
L_{MPO}	674.4	−3.2	0.9
L_K	688.4	−1.2	0.9
L_C	693.8	−0.4	0.8
L_{PF}	698.1	0.2	0.7
ECA	688.1	−1.2	0.7
C. *DNA molecules: "manual" measurements, N = 248*			
L_F	701.2	0.7	2.1
L_{MPO}	652.0	−6.4	2.1
L_K	665.3	−4.5	2.1
L_C	670.5	−3.8	2.2
L_{PF}	674.2	−3.2	2.2
ECA	665.7	−4.4	2.3

[a] Contour length measurements of the simulated polymers (panels A, B) and real DNA molecules 2049 bp long (panel C) obtained with six different procedures. See text for details.

visible on the screen and therefore, there was the tendency to select the end points of the molecule slightly before the very edge.

3.3. Estimate of the B- to A-form transition

From the analysis of DNA molecules deposited onto mica and imaged in air by AFM it has been found that the molecules have a reduced contour length compared to that of B-form DNA. This result suggests a partial transition from B- to A-form DNA. As determined by X-ray crystallography, we can assume a helical rise of 0.34 and 0.26 nm/bp for B- and A-form DNA, respectively [27]. Therefore, from the contour length measurements it is possible to determine the fraction of DNA that has switched to A-form using the relation

$$fA_{DNA} = \frac{0.34N - L_{PF}}{0.34N - 0.26N}, \quad (7)$$

where N is the number of base pairs.

In Table 5, the contour length measurements of the four DNA fragments obtained with the polynomial smoothing method are summarized together with the value of the contour length adjusted by the tip effect and the operator bias. Also reported is the average fraction of base pairs adopting an A-form conformation. The 564 and 1054 bp fragments show that about 30% of the bases adopt an A-form, whereas, about 15% of the bases are in A-form for the 2049 and 4297 bp fragments.

4. Discussion and conclusion

The algorithm used to determine the DNA contour length from digitized contours is often not considered an important issue in the analysis of DNA images obtained by either EM or AFM. In many cases the attention is focused on the method employed to retrieve the DNA contour from the

Table 4
DNA 4297 bp long (B-form contour length: 1461.0 nm)[a]

Method	Contour length (nm)	Error (%)	CV (%)
A. *Simulated polymers: "automatic" measurements, N = 175*			
L_F	1522.8	4.2	0.6
L_{MPO}	1411.1	−3.4	0.6
L_K	1444.7	−1.1	0.6
L_C	1456.5	−0.3	0.5
L_{PF}	1459.8	−0.1	0.3
ECA	1437.1	−1.6	0.4
B. *Simulated polymers: "manual" measurements, N = 186*			
L_F	1510.1	3.4	0.6
L_{MPO}	1399.5	−4.2	0.6
L_K	1432.9	−1.9	0.5
L_C	1445.0	−1.1	0.5
L_{PF}	1453.2	−0.5	0.5
ECA	1431.8	−2.0	0.5
C. *DNA molecules: "manual" measurements, N = 187*			
L_F	1454.7	−0.4	2.3
L_{MPO}	1348.1	−7.7	2.2
L_K	1380.3	−5.5	2.3
L_C	1390.8	−4.8	2.3
L_{PF}	1397.4	−4.3	2.1
ECA	1376.3	−5.8	2.2

[a] Contour length measurements of the simulated polymers (panels A, B) and real DNA molecules 4297 bp long (panel C) obtained with six different procedures. See text for details.

Table 5
Estimate of the B- to A- form transition[a]

DNA (bp)	L_{PF} (nm)	Tip lengthening and user bias (nm)	Corrected L_{PF} (nm)	fA
564	182.8	4.8	178.0	0.30
1054	338.1	5.3	332.8	0.30
2049	674.2	2.0	672.2	0.15
4297	1397.4	−6.6	1404.0	0.17

[a] The tip lengthening and the user bias quantity was obtained from the difference between measurements in panels (B) and those in panels (A). The fraction of DNA that has switch to A-form was calculated from Eq. (7).

image and the length is usually determined as the sum of the Euclidean distances between consecutive points. When dealing with a string of eight connected pixels this reduces to the use of the Freeman estimator. In this paper it is emphasized that, measuring contours from digitized images is not a geometrical problem but rather an estimation problem and that the accuracy and precision of the contour length determination will depend on the choice of the measuring algorithm.

Although the (n_e, n_o)- and (n_e, n_o, n_c)-based length estimators used here were optimized for straight strings, their applicability to arbitrary strings is remarkable. In this study it was found that, out of the six different methods tested, and for all the DNA sizes analyzed, the corner count estimator (L_C) and the polynomial smoothing routine (L_{PF}) are the most accurate in estimating the real DNA contour length. The error is very small, with values always below 0.5%. Good estimates are also obtained with the Kulpa (L_K) estimator which has the advantage of being simple for implementation. Likewise, the MPO and the ECA methods produce acceptable results but their accuracy decreases for the longer DNA fragments. In all cases, the Freeman estimator (L_F) is biased with overestimates of up to 4.3%.

In this study it is also shown that the broadening effect of the AFM tip produces a lengthening of the DNA molecules. Clearly, this effect is constant for the different DNA sizes but for short fragments its relative influence on the DNA contour length is higher. It must be emphasized that the observed tip lengthening effect derives from a comparison of simulated polymers before and after convolution with a parabolic function. This is purely a geometrical transformation where both the tip and the DNA are considered as infinitely rigid bodies. Consequently, the possibility that the DNA ends have a different structure or a different resistance to compression with respect to the rest of the DNA molecule, cannot be ruled out.

Long DNA fragments measured with the manual procedure are also susceptible to the bias from the operator that, because of the lower magnification used during the hand-tracing procedure, tends to shorten the molecules. In principle, this problem could be avoided with a digitization procedure in which the DNA ends are detected with an intensity threshold analysis of the end pixels [18].

The contour length analysis of DNA reveals a shrinking of the molecules that could be accounted by the transition of molecular domains to A-form,

upon the deposition onto the mica surface. The possibility of a B- to C-form transition of the DNA is not considered since it has only been observed at low relative humidity and in the presence of high Li^+ salt [28]. Bacterial DNA at low (75%) relative humidity in NaCl displays features of the A-form [29]. It is interesting to observe that the B- to A-form transition is more conspicuous for the shorter DNA fragments (564 and 1054 bp). This result is in agreement with a previously proposed model in which mica can be thought as an array of binding sites capable of interacting with the DNA [1,7]. Long DNA fragments bind to mica at many sites, restricting the transition to A-form. On the contrary, short DNA fragments have fewer constraints with the surface, allowing a higher number of domains to switch to the A-form. Interestingly, previous data show that the contour length of 100 and 200 bp DNA fragments, imaged in air by AFM, was consistent with that of A-form DNA [1].

However, it is not clear whether the structural transition is triggered by the drying step of the deposition procedure, that causes a partial dehydration of the DNA, or whether it is induced prior to the drying step by the entering of the DNA into the mica/solution interface, where a high concentration of Mg^{2+} [30] could produce dehydration of the DNA and stabilize the A-form. In a recent study, comparison of the contour length of DNA molecules ($\sim$1000 bp long) imaged in air and under buffer has shown that DNA imaged in air shrink by about 4% with respect to the same DNA imaged under buffer [2]. This indicates that the A to B transition of the DNA is induced, at least in part, by the drying of the mica surface during the sample preparation.

B to A transition of DNA confined to mica in the presence of Mg^{2+} and exposed to ethanol, have been recently reported [3,4]. In particular it was shown that mica facilitates the ethanol-induced transition from B-form to A-form DNA. In solution a concentration of 70–85% of ethanol is required for the B to A shift whereas on mica, the same shift is reached at 30% ethanol. In these studies, DNA (1968 bp) in the absence of ethanol was assumed to be in a B-form conformation as judged from its contour length. However, Fang et al. computed the DNA length using the Freeman estimator which, as shown here, introduces an overestimation bias of about 4.2%. Furthermore, for some DNA fragments the contour length decreases with increase in the ethanol concentration but it seems to plateau at a length slightly larger than all A-form DNA. Again, this could be due to the low accuracy of the measuring method, which produces a shift of all DNA contour length measures towards larger values. In view of the results presented in this paper, it seems that mica not only facilitates the B to A transition in the presence of ethanol but also is capable of inducing such transitions in the absence of ethanol.

In conclusion, when measuring DNA molecules from digital images, the two methods of choice should be: the corner count estimator or the polynomial smoothing with degree 3 over a moving window of 5 points. A valid and simple method is also the Kulpa estimator, whereas the commonly used Freeman estimator should be avoided when accuracy is required. Clearly, the applicability of the different methods depends on both the size of the DNA and the resolution of the image (i.e. the size of one pixel). In this study we have considered DNA fragments in the range 500–4000 bp and images of 512×512 pixels spanning an area of $2 \times 2\,\mu m^2$. These conditions or comparable ones, should be encountered in most of the EM or AFM experiments dealing with images of DNA deposited onto a two-dimensional grid.

Acknowledgements

We are grateful to Simone Ottonello, Claudio Oleari and Martin Guthold for discussion and critical reading of the manuscript, and to the Centro Interfacoltá Misure (CIM) of the University of Parma for the Atomic Force Microscope facility. Technical advice from one of the reviewers is also gratefully acknowledged. This work was supported by the Ministry of University and of Scientific and Technological Research (Cofin '99) and by the National Research Council of Italy.

References

[1] H.G. Hansma, I. Revenko, K. Kim, D.E. Laney, Nucleic Acids Res. 24 (1996) 713.
[2] A. Schulz, N. Mucke, J. Langowski, K. Rippe, J. Mol. Biol. 283 (1998) 821.
[3] J.E. Coury, J.R. Anderson, L. McFail-Isom, L.D. Williams, L.A. Bottomley, J. Am. Chem. Soc. 119 (1997) 3792.
[4] Y. Fang, T.S. Spisz, J.H. Hoh, Nucleic Acids Res. 27 (1999) 1943.
[5] C. Frontali, E. Dore, A. Ferrauto, E. Gratton, A. Bettini, M.R. Pozzan, E. Valdevit, Biopolymers 18 (1979) 1353.
[6] J. Bednar, P. Furrer, V. Katritch, A.Z. Stasiak, J. Dubochet, A. Stasiak, J. Mol. Biol. 254 (1995) 579.
[7] C. Rivetti, M. Guthold, C. Bustamante, J. Mol. Biol. 264 (1996) 919.
[8] J.A. Cognet, C. Pakleza, D. Cherny, E. Delain, E. Le Cam, J. Mol. Biol. 285 (1999) 997.
[9] C. Rivetti, C. Walker, C. Bustamante, J. Mol. Biol. 280 (1998) 41.
[10] C. Rivetti, M. Guthold, C. Bustamante, Embo J. 18 (1999) 4464.
[11] T.-K. Kim, T. Lagrange, Y.-H. Wang, J.D. Griffith, D. Reinberg, R.H. Ebright, Proc. Natl. Acad. Sci. USA 94 (1997) 12268.
[12] J.E. Coury, L. McFail-Isom, L.D. Williams, L.A. Bottomley, Proc. Natl. Acad. Sci. USA 93 (1996) 12283.
[13] D.P. Allison, P.S. Kerper, M.J. Doktycz, T. Thundat, P. Modrich, F.W. Larimer, D.K. Johnson, P.R. Hoyt, M.L. Mucenski, R.J. Warmack, Genomics 41 (1997) 379.
[14] D.V. Klinov, I.V. Lagutina, V.V. Prokhorov, T. Neretina, P.P. Khil, Y.B. Lebedev, D.I. Cherny, V.V. Demin, E.D. Sverdlov, Nucleic Acids Res. 26 (1998) 4603.
[15] L.I. Pietrasanta, D. Thrower, W. Hsieh, S. Rao, O. Stemmann, J. Lechner, J. Carbon, H. Hansma, Proc. Natl. Acad. Sci. USA 96 (1999) 3757.
[16] L. Dorst, A.W.M. Smeulders, Comput. Graphics Image Process. 40 (1987) 311.
[17] C. Bustamante, C. Rivetti, Ann. Rev. Biophys. Biomol. Struc. 25 (1996) 395.
[18] T.S. Spisz, Y. Fang, R.H. Reeves, C.K. Seymour, I.N. Bankman, J.H. Hoh, Med. Biol. Eng. Comput. 36 (1998) 667.
[19] D.J. Keller, F.S. Franke, Surf. Sci. 294 (1993) 409.
[20] T.C. Kaspar, R.P. Ewing, Agron. J. 89 (1997) 932.
[21] H. Freeman, in: B.S. Lipkin, A. Rosenfeld (Eds.), Picture Processing and Psychopictorics, Academic Press, New York, 1970, pp. 241–266.
[22] W.L. Pan, R.P. Bolton, Agron. J. 83 (1991) 1047.
[23] Z. Kulpa, Comput. Graphics Image Process. 6 (1977) 434.
[24] A.M. Vossepoel, A.W.M. Smeulders, Comput. Graphics Image Process. 20 (1982) 347.
[25] R.P. Ewing, T.C. Kaspar, J. Comput. Assist. Microsc. 7 (1995) 91.
[26] C. Bustamante, D. Keller, G. Yang, Curr. Opin. Struct. Biol. 3 (1993) 363.
[27] W. Saenger, Principles of Nucleic Acid Structure, Springer, New York, 1984.
[28] D.M. Loprete, K.A. Hartman, J. Biomol. Struct. Dyn. 7 (1989) 347.
[29] H. Kang, W.C. Johnson, Biochemistry 33 (1994) 8330.
[30] R.M. Pashley, J.N. Israelachvili, J. Colloid Interface Sci. 97 (1984) 446.