Label-free, atomic force microscopy-based mapping of DNA intrinsic curvature for the nanoscale comparative analysis of bent duplexes

Renato Buzio[1,2], Luca Repetto[1,2,*], Francesca Giacopelli[3], Roberto Ravazzolo[3,4] and Ugo Valbusa[1,2]

[1]S.C. Nanobiotecnologie, National Institute for Cancer Research IST, [2]Physics Department, University of Genova, [3]Department of Pediatrics and Center of Excellence for Biomedical Research, University of Genova and [4]Laboratory of Molecular Genetics, G. Gaslini Institute, Genova, Italy

ABSTRACT

We propose a method for the characterization of the local intrinsic curvature of adsorbed DNA molecules. It relies on a novel statistical chain descriptor, namely the ensemble averaged product of curvatures for two nanosized segments, symmetrically placed on the contour of atomic force microscopy imaged chains. We demonstrate by theoretical arguments and experimental investigation of representative samples that the fine mapping of the average product along the molecular backbone generates a characteristic pattern of variation that effectively highlights all pairs of DNA tracts with large intrinsic curvature. The centrosymmetric character of the chain descriptor enables targeting strands with unknown orientation. This overcomes a remarkable limitation of the current experimental strategies that estimate curvature maps solely from the trajectories of end-labeled molecules or palindromes. As a consequence our approach paves the way for a reliable, unbiased, label-free comparative analysis of bent duplexes, aimed to detect local conformational changes of physical or biological relevance in large sample numbers. Notably, such an assay is virtually inaccessible to the automated intrinsic curvature computation algorithms proposed so far. We foresee several challenging applications, including the validation of DNA adsorption and bending models by experiments and the discrimination of specimens for genetic screening purposes.

INTRODUCTION

Atomic force microscopy (AFM) is nowadays routinely used to resolve the contour of adsorbed DNA molecules with nanoscale resolution and support in this way of important advances in fundamental and applied studies in biophysics, molecular biology, genetics, genomics and nano biomedicine. Notably, through the analysis of contours one can map the DNA intrinsic curvature and flexibility along the molecular backbone (1–6). This technique is particularly suited to address at the experimental level the impact of base-pair sequence on the local conformation of the strands (1,3–8) and also plays a pivotal role for investigations attempting to relate the inherent DNA shape and flexibility to other physical and biological properties, such as melting (9), ligand interactions (10–12), replication (13), genomic packaging and transcription regulation (14).

The wide applicability of AFM-based curvature studies demands simple and reliable experimental methods, characterized by a few processing steps for specimen preparation and minimum experimental bias on intrinsic curvature measurements. These requirements are even more important in view of the introduction of effective assays for DNA analysis fully based on AFM imaging [e.g. sizing (15), genotyping and haplotyping (16), expression profiling (17)]. Such a step may lead to envision a key role for the nanoscale curvature analysis within more complex protocols, setting up population-based genetic disease studies or solving genomic screening problems at the single-molecule level.

Current methods for the estimation of the local intrinsic curvature and flexibility start from the evidence that DNA is a long molecule whose conformation is constantly fluctuating under thermal perturbations while adsorbing

---

from a bulk solution onto a solid support. This suggests to adopt a statistical approach, based on high-resolution imaging and computer-assisted tracing of adsorbed molecules, in order to sample with accuracy the ensemble of accessible DNA conformations. These steps are followed by the estimation of the signed curvature on individual contours and by an ensemble averaging process, conducted to separate the intrinsic (static) curvature contribution from the thermally activated (dynamic) one (3). A condition for the correct conformational average is the need for a proper alignment of each contour of the ensemble, i.e. it is necessary to identify (i) which of the two contour ends corresponds to the starting point of the base-pair sequence between the two alternative choices and (ii) which of the two contour orientations with mirror curvature profiles reflect the actual helical region exposed by the adsorbed chain to the substrate. Detection of chain polarity was traditionally solved by end-labeling with bulky tags, e.g. streptavidin, streptavidin–ferritin, infrared dyes or colloidal gold (2,4,16,18). These procedures can, however, perturb the overall conformation of adsorbed molecules and represent indeed a time-consuming, labor-intensive part of the whole experiment. A different solution consists on the preparation of palindromic constructs starting from the target molecules (3). This has the advantage of bypassing the need for chain polarity discrimination due to the dyad symmetry of the base-pairs sequence. Despite such progress, specimen preparation issues very likely hamper the broad applicability of similar studies: as a matter of fact, AFM-based curvature maps are so far limited in number (1–6,8,18–21) and mostly related to the analysis of few model systems [e.g. pBR322 DNA in (2,3,18,20,21)].

Here we propose a novel method for the characterization of the local intrinsic curvature, which is inspired by the aforementioned works on computer-assisted tracing of adsorbed molecules (1–6), yet employs a new statistical chain descriptor: the ensemble averaged product of curvatures for two nanosized segments symmetrically placed on the contour of AFM-imaged chains. This peculiar choice results in a centrosymmetric statistical quantity that enables to target adsorbed strands with unknown orientation. Accordingly, conformational averages are calculated without the need for a proper alignment of the AFM-imaged molecular trajectories. In particular, we demonstrate by theoretical arguments from polymer chain statistics that the fine mapping of the average curvatures product (CP) along the molecular backbone effectively highlights all pairs of DNA tracts with large intrinsic curvature. Such conclusion is further supported by the direct investigation of representative specimens from the promoter region of the human osteopontin (OPN) coding gene and the successful comparison of experimental findings with simulations based on well-known DNA bending models. We finally contrast the results of the novel method with those obtained by the automated intrinsic curvature computation algorithms proposed so far (19). The superior response offered by our method, in terms of robustness, accuracy, flexibility and widespread applicability, justifies its potential use in novel, label-free

comparative assays of bent duplexes, aimed to detect local conformational changes of physical or biological relevance in large sample numbers. The method is intended to involve 10^2–10^3 bp long fragments that can be readily prepared for AFM imaging.

MATERIALS AND METHODS

Sample preparation

DNA samples were obtained by PCR amplification of the regulatory region of the OPN encoding gene, as described previously (22); amplicons were purified in 1% (w/w) agarose gel and electroeluted, then the solution was treated with phenol/chloroform followed by ethanol precipitation. The pellet was stabilized in Tris-EDTA buffer and stored at $-20°C$. DNA concentration, determined by absorbance at 260 nm, was in the range of 100 nM. Conventional haplotype analysis allowed us to focus primarily on a 1332 bp specimen with the nucleotide sequence reported in Figure 1. Importantly, this template does not contain extended strings of phased A-tracts or other prominent nucleotide sequences (e.g. periodic A_n/T_n groups) that could introduce anomalously large bends in the adhered DNA molecules (1–3,8) and bias our proof-of-principle investigation. For comparative purposes, a second 1335 bp specimen with point mutations at four, well-known polymorphic loci was also considered (see Supplementary Data). With the exception of the last subsection (devoted to contrast both samples), all experimental and theoretical results reported below refer to the 1332 bp sample.

The DNA adsorption was carried out according to the standard protocols reported in literature (23). A 20 μl aliquot of solution containing 4 mM Hepes (pH 7.4), 4–10 mM $MgCl_2$ and 2 nM DNA was deposited onto freshly cleaved muscovite mica (Agar Scientific); the sample drop was incubated for about 120 s and rinsed with MilliQ water. The surface was finally dried with a gentle stream of nitrogen.

Characterization of DNA local intrinsic curvature

Common practice in AFM studies on DNA structure and flexibility dictates to prepare specimens by DNA adsorption from an aqueous solution onto an atomically flat substrate. This is followed by high-resolution imaging of adsorbed species and by the use of an image-analysis software in order to reconstruct the molecular profiles and analyze the signed curvature associated to segments of given location and length (2–5,8,24). Tracing algorithms represent each molecule as a chain of xy pairs separated by a fixed contour length l. The curvature analysis along a generic trajectory proceeds through the calculation of the signed bending angles θ_i formed by the adjacent units, that are obtained from the vector product of the local tangent vectors $\vec{t}_i$ and $\vec{t}_{i+1}$ ($i = 1, 2, \ldots, N-1$ with N total number of units) (3). From the θ_i values one can define the global curvature $C_{j,m}$ for a segment

ATTACAATTC	GTGACTGCCT	GCCCCTCTTA	AAAAATTTCAT	AATAGTTAAC	ACACATATAG	TCCTTAAGAT		70
ACGCAGAGCA	TTTGCATCTA	ATATGTGCTA	AGCATTGCTA	GTTTAACATA	CTAATTCATT	TAAACCCCTC		140
AAAAACCCCA	TGACCTAGGT	AATAGTATTG	CATTTCATGG	ATGAGGGAAC	AAGGATAGGT	AGGCTGGGCG		210
ATTTGCCCAA	GGTTGCACAG	GTCAGCAGTG	ACACAGCGGA	ATTCAGAACC	ACGGTCTGGC	TCCTGAAGCA		280
GCCCTCTCAA	GCAGTCATCC	TTCTCTCAGT	CAGAAACTGC	TTTACTTCTG	CAACATCTAG	AATAAATTAC		350
CATTCTTCTA	TTTCATATAG	AATTTTATAT	TTTAATGTCA	CTAGTGCCAT	TTGTCTAAGT	AACAAGCTAC		420
TGCATACTCG	AAATCACAAA	GCTAAGCTTG	AGTAGTAAAG	GACAGAGGCA	AGTTCTCTGA	ACTCCTTGCA		490
GGCTTGAACA	ATAGCCTTCT	GGCTCTTCAA	TAAGTACAAT	CATACAGGCA	AGAGTGGTTG	CAGATATTAC		560
CTTTATGTTA	CTTAAACCGA	AAGAAACAAA	AATCCATTGT	ATTTAATTTT	ACATTAATGT	TTTTCCCTAC		630
TTTCTCCCTT	TTTCATGGGA	TCCCTAAGTG	CTCTTCCTGG	ATGCTGAATG	CCCATCCCGT	AAATGAAAAA		700
GCTAGTTAAT	GATATTGTAC	ATAAGTAATG	TTTTAACTGT	AGATTGTGTG	TGTGCGTTTT	TGTTTTTTTT		770
TGTTTTAACC	ACAAACCAG	AGGGGGAAGT	GTGGGAGCAG	GTGGGCTGGG	CAGTGGCAGA	AAACCTCATG		840
ACACAATCTC	TCCGCCTCCC	TGTGTTGGTG	GAGGATGTCT	GCAGCAGCAT	TTAAATTCTG	GGAGGGCTTG		910
GTTGTCAGCA	GCAGCAGGAG	GAGGCAGAGC	ACAGCATCGT	CGGGACCAGA	CTCGTCTCAG	GCCAGTTGCA		980
GCCTTCTCAG	CCAAACGCCG	ACCAAGGTAC	AGCTTCAGTT	TGCTACTGGG	TTGTGCATTC	AGCTGAATTT		1050
CATGGGGAAG	TCCAAATTCT	AAGGAAAAAT	ATTTTTAATT	GTAATGCTGT	TAAACAGACT	TAAATTTTCT		1120
AGCCTTTTTA	ATAAGCAGAT	TAGATACATT	GCAGGTCTCC	TGGAACAAAG	GTGTCTAGAT	ATTTTGAATG		1190
CCAATCAAAT	TTAAAACTTA	AAAATACTTC	CACTGGGTCC	TCAAAAGAAC	GGAAACCACC	GATGCTAATC		1260
AGAAAATAGT	AAAATTAAAT	TCACCTTTGG	AATAATTATA	CCTATATAAT	TTTCAGTGGG	GTACTGTGCA		1330
GG								

Figure 1. Sequence of the 1332 bp sample used in AFM experiments.

of m units, located at j units from one of the ends, as:

$$C_{j,m} \equiv \sum_{i=j+1}^{j+m-1} \theta_i \tag{1}$$

with $j, m = 1, 2, \ldots, N$ (Figure 2a).

It is well known that the representation of DNA molecules by segmental chains allows to compare experimental findings with predictions of polymer chain statistics, particularly with those of the worm-like chain (WLC) model in its discrete formulation (1–3,18,19,23,24). In this context, DNA is modeled by a chain of virtual bonds of length l connected by torsional-spring vertices, that are energetically uncorrelated and characterized by a harmonic local bending-energy function $E(\theta_i^{th}) = 1/2k_B T(\xi/l)(\theta_i^{th})^2$ (with k_B Boltzmann constant, T absolute temperature, $\xi \approx 50$ nm persistence length and θ_i^{th} are thermally induced angular fluctuations occurring around the constant sequence-dependent θ_i^0 angles) (1). This corresponds to represent θ_i as the sum of static and dynamic contributions, i.e. $\theta_i = \theta_i^0 + \theta_i^{th}$, where θ_i^{th} angles are normally distributed with null mean value and standard deviation $\sqrt{l/\xi}$. Thus the WLC model predicts that the average value of the $C_{j,m}$ curvature is:

$$\langle C_{j,m} \rangle = \sum_{i=j+1}^{j+m-1} \langle \theta_i \rangle = \sum_{i=j+1}^{j+m-1} \theta_i^0 \tag{2}$$

where the angle brackets $\langle \rangle$ denote an ensemble average conducted over the accessible chain conformations. Equation (2) proves that the average curvature $\langle C_{j,m} \rangle$ equals the intrinsic curvature $\sum \theta_i^0$ of the segment. Furthermore, it suggests a route for comparing the experimental values of intrinsic curvature with the theoretical ones: in fact the left-hand term might be experimentally accessed by averaging the $C_{j,m}$ realizations over a large pool of AFM-imaged molecular contours, whereas the right-hand term should be predicted computationally by well-consolidated methods [e.g. refs. (2–4,18,25)]. Unfortunately, the practical estimation of $\langle C_{j,m} \rangle$ is a non-trivial task since it requires to orientate the sampled molecular contours in order to evaluate the curvature average on corresponding points of the nucleotide sequence. In general, for each molecular contour extracted from an AFM image there are four possible spatial orientations, depending on which of the two contour ends correspond to the starting point of the base-pair sequence (the $5' - 3'$ direction) and which of the two chemically different faces are exposed by the molecule to the substrate when collapsing on it from the bulk solution (Figure 2b). For the case of unlabeled chains, their orientation uncertainty cannot be solved deterministically because there are no distinctive topographical features that allow to distinguish the beginning from the end of a DNA molecule in an AFM image. Nevertheless, a non-deterministic approach can be carried on, based on the assumption that the recorded contours share the same orientation
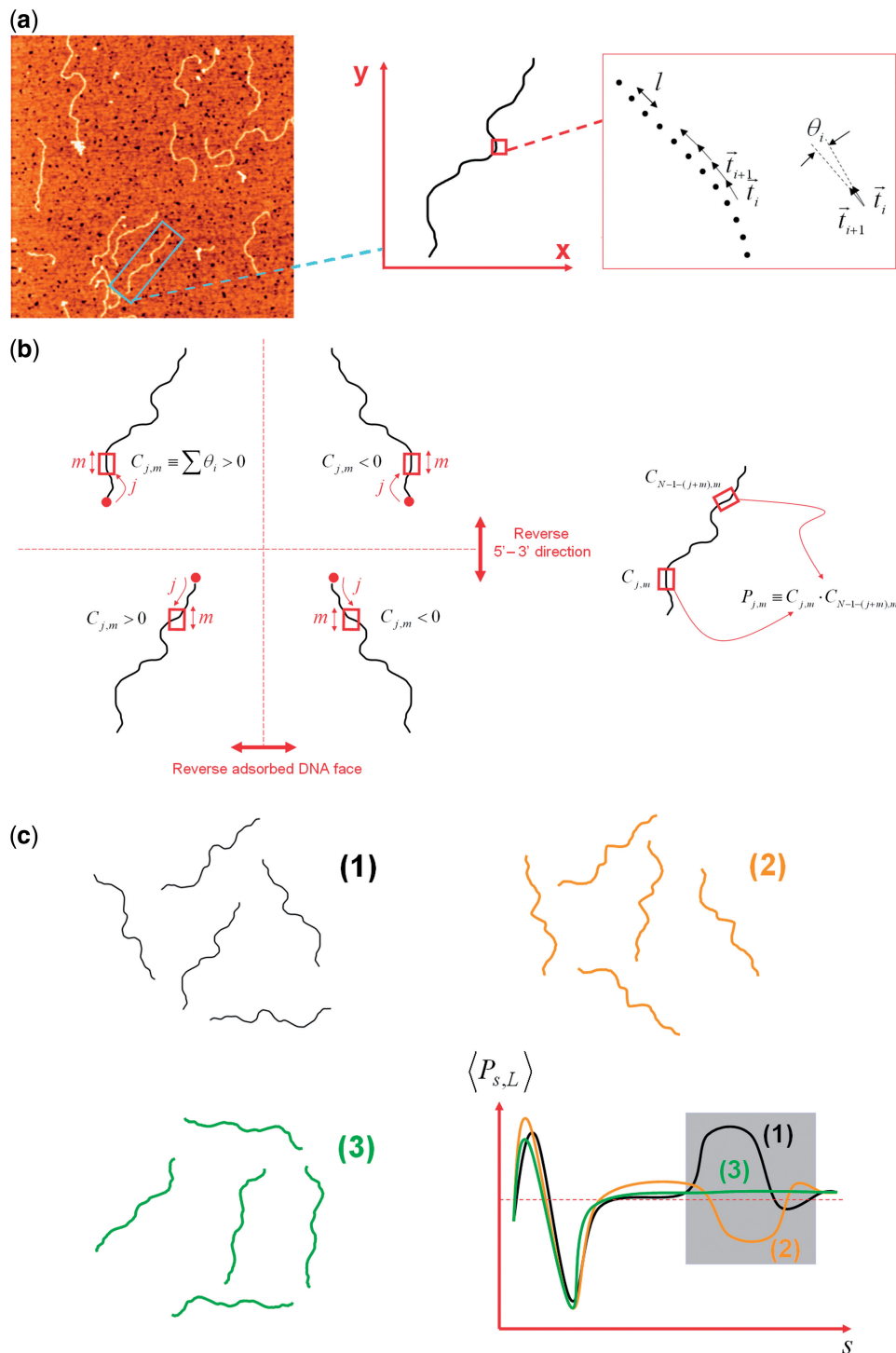
Figure 2. (**a**) Main steps of curvature analysis for a DNA chain. A molecule is imaged by AFM then traced by an image-analysis software and represented as a chain of xy pairs separated by a contour length l. The signed bending angle θ_i is obtained from the vector product of the local tangent vectors $\vec{t}_i$ and $\vec{t}_{i+1}$. (**b**) Generally speaking, we can ascribe four different spatial orientations to the extracted contour of a label-free molecule, according to the end chosen as the starting point of the nucleotide sequence (red dot) and the molecular face exposed to the substrate. As a result, the signed curvature $C_{j,m}$ changes in modulus and/or sign according to the chosen orientation. On the contrary, the CP $P_{j,m}$ is estimated by coupling the signed curvatures of two m-units long segments, symmetrically placed at j units from chain ends. Such quantity remains the same for each one of the possible orientations of the extracted contour. (**c**) The characteristic patterns of variation of $\langle P_{s,L} \rangle$ for some specimens—here named (1–3)— enable to highlight DNA regions with different intrinsic curvature (gray box). This represents an original strategy to establish the comparative analysis of bent duplexes under label-free conditions.

when we observe the minimal value of the ensemble averaged curvature variance at each point along the molecular trajectory; in this case automated computational algorithms are used to iteratively flip the orientation of the extracted molecular profiles in search for the minimum value of the overall curvature variance (19). An extensive discussion on this protocol is reported below.

Whenever the imaged chains are end-labeled with a structurally distinctive tag, the beginning and the end of the nucleotide sequence (i.e. the chain polarity) are easily inferred from AFM images; alternatively palindromic dimers can be used. However, an uncertainty remains on the two orientations with mirror curvature profiles that describe DNA adsorption on chemically different faces. Scipioni *et al.* (20) and Sampaolese *et al.* (21) demonstrated that such orientations are not statistically equivalent if palindromes are deposited onto freshly cleaved mica, because a preferential adsorption of T-rich faces occurs. This unexpected phenomenon ultimately justifies the calculation of $\langle C_{j,m} \rangle$ from an ensemble of palindromic dimers, or even labeled chains properly oriented to have the same polarity.

The alternative method described in this article characterizes the DNA intrinsic curvature through the calculation of a new statistical quantity, which couples two segments symmetrically placed along the sampled contours. In detail, we focus on the $P_{j,m}$ product between the curvatures $C_{j,m}$ and $C_{N-1-(j+m),m}$ of two tracts formed by m units and located at j units from chain ends (Figure 2b):

$$P_{j,m} \equiv C_{j,m} \cdot C_{N-1-(j+m),m} \tag{3}$$

For the case of non-overlapping fragments ($j = 1, 2, \ldots, N/2$, $m = 1, \ldots, N/2 - j$), the quantity $P_{j,m}$ is normally distributed with average value:

$$\langle P_{j,m} \rangle = \langle C_{j,m} \cdot C_{N-1-(j+m),m} \rangle = \langle C_{j,m} \rangle \cdot \langle C_{N-1-(j+m),m} \rangle \tag{4}$$

where $\langle C_{j,m} \rangle$ is given by Equation (2) and the last equality holds under the specific conditions of the present investigation (see Supplementary Equation S4 and the related discussion in the Supplementary Data). Equation (4) shows that $\langle P_{j,m} \rangle$ equals the product of the intrinsic curvatures of the two chosen segments. It is trivial to demonstrate that the realizations of the statistical chain descriptor $P_{j,m}$ do not depend on the orientation arbitrarily assigned to the DNA trajectories extracted from AFM images, in other words $P_{j,m}$ is centrosymmetric. Accordingly, $\langle P_{j,m} \rangle$ can be estimated by an ensemble average of $C_{j,m} \cdot C_{N-1-(j+m),m}$ values obtained from a large pool of molecular profiles with arbitrary relative orientation. In agreement with this picture, neither end-labeled molecules nor palindromic constructs are required.

In particular, we propose to characterize the DNA intrinsic curvature by introducing the curvilinear distance $s \equiv jl$ and by plotting s versus $\langle P_{s,L} \rangle$ at fixed L ($L \equiv ml$), which corresponds to probe the emergence of intrinsic curvature effects for pairs of segments of fixed length L, located at a given distance s from the ends. By definition, we expect to observe remarkable variations of $\langle P_{s,L} \rangle$ whenever large, intrinsic curvatures affect the trajectory of the chosen fragments, whereas $\langle P_{s,L} \rangle \approx 0$ for pairs with negligible intrinsic curvature at least at one of the two involved segments. Overall, these features contribute to generate a characteristic pattern of variation of $\langle P_{s,L} \rangle$ that can be originally exploited to set up the comparative analysis of bent duplexes (Figure 2c).

The reader is referred to the Supplementary Data to get further insight into the generality and flexibility of the present method. As a matter of fact, a whole class of new centrosymmetric curvature descriptors can be introduced to probe bent chains under label-free conditions. For the case of $P_{s,L}$, we anticipate that s versus $\langle P_{s,L} \rangle$ patterns are fully comparable—in terms of accuracy and sensitivity—with the maps of signed intrinsic curvature provided by the current experimental strategies. Moreover, simple solutions exists in order to circumvent the decrease of sensitivity that affects the sites where $\langle P_{s,L} \rangle \approx 0$, e.g. to cause a lateral shift of the centre of symmetry of the target molecules through *ad hoc* deletions of short fragments at one of the chain ends. Complementary patterns from two descriptors can be also exploited to bypass such a drawback.

We finally note that there is certainly a loss of information in the s versus $\langle P_{s,L} \rangle$ plot with respect to the s versus $\langle C_{s,L} \rangle$ data (2–4,19–21). This loss unavoidably arises from coupling pairs of DNA tracts into the definition of the statistical chain descriptor $P_{j,m}$. Nevertheless, we argue hereafter that our choice readily provides a number of important advantages, overcoming some fundamental and practical limitations of early protocols. In fact the novel method can be easily implemented on label-free molecules, therefore specimens preparation is merely reduced to standard protocols for DNA deposition onto atomically smooth substrates. Furthermore, no assumptions are done on the adsorption mechanism and preferential orientation of target chains on a given substrate (otherwise required to calculate the conformational average $\langle C_{s,L} \rangle$ (3,20,21)). The s versus $\langle P_{s,L} \rangle$ plots are also prone to an effective comparison with theoretical models [used to predict the right-hand term of Equation (4)] that impart access to the physics of DNA adsorption and sequence-dependent curvature. All these aspects are deeply explored in the following paragraphs, starting from the characterization of the intrinsic curvature of DNA molecules from the promoter region of the human OPN coding gene.

AFM imaging and analysis

Samples were imaged in air at room temperature and humidity with a Dimension 3100 AFM equipped with the closed-loop Hybrid XYZ scanner and the Nanoscope IVa control unit (Digital Instruments, Veeco). The AFM was operated in tapping mode and silicon probes (OMCL-AC160TS, Olympus) were used. The AFM images were collected with a dimension of 1024×1024 pixels and a typical scan size of $2\,\mu m$.

Our image-analysis software allowed a semi-automatic reconstruction of molecular trajectories and a straightforward analysis of the signed curvature associated to

segments of given location and length. The tracing algorithm was developed in LabView (National Instruments) following the general guidelines of ref. (24) and molecules were represented as chains of xy pairs separated by a contour length $l = 2$ nm (see Supplementary Data for further details). The positive values for the signed bending angles θ_i were arbitrarily assigned to clockwise rotations, i.e. if by progressing along the trajectory the chain turns to the right at θ_i. The signed curvatures $C_{j,m}$ were estimated from Equation (1), whereas the average product $\langle P_{j,m} \rangle$ was evaluated from the conformational average of the $C_{j,m} \cdot C_{N-1-(j+m),m}$ product over a given set of AFM-imaged molecular profiles.

We also implemented standard checks on global statistical parameters to ascertain the thermodynamic equilibration of the molecules during the deposition process onto the mica surface and to investigate the influence of intrinsic curvature on the average superstructure of the chain. According to the WLC model, the mean trajectory of an intrinsically straight chain ($\theta_i^0 = 0$ at every i location) is given by the following equation:

$$\langle R_{s,s+L}^2 \rangle = 4\xi \left[L + 2\xi \left(e^{-\frac{L}{2\xi}} - 1 \right) \right] \tag{5}$$

where $R_{s,s+L}$ is the Euclidean distance between pairs of points located at s and $s+L$ from one end of the molecule (here L increases up to the limit of the chain length) (23,24). The average $\langle \rangle$ is computed over s—up to the upper limit of the contour length of the molecule minus the contour length spacing L—and overall observed contours. We measured $\langle R_{s,s+L}^2 \rangle$ on a large ensemble of traced contours and discussed deviations of such quantity from predictions of Equation (5).

The automated fragment flipping algorithm

We compared the curvature analysis based on s versus $\langle P_{s,L} \rangle$ plots with the predictions of the first label-free method for the automated intrinsic curvature computation, originally proposed by Ficarra *et al.* (19). In this case, the angles θ_i extracted from AFM images are arranged into a curvature matrix in a way that each row represents the curvature profile of a given molecule. Since curvature profiles are loaded without any knowledge of molecules relative orientation, an automated algorithm iteratively 'flips' each row (by inverting the sign, reversing the order, combining both operations or leaving the row unchanged) in search for the optimal matrix configuration, defined as the condition in which we can observe the minimal values for columns variances; this is in fact recognized to represent the case in which all the molecules share the same orientation. Once the mean value of columns variance achieves a minimum, the columns averages are expected to provide the intrinsic DNA curvature profile, in agreement with the WLC model stating $\langle \theta_i \rangle = \langle \theta_i^0 + \theta_i^{th} \rangle = \theta_i^0$.

Initially, a custom-written code (LabView, National Instruments) implementing the automated fragment flipping (FF) algorithm was validated on a set of computer-generated and intrinsically bent chains, as explained in ref. (19) (see also the Supplementary Data).

The same FF code was then used to reconstruct the intrinsic curvature profile and the CP patterns for the human OPN coding gene, starting from an ensemble of molecular profiles extracted from AFM images.

Modeling DNA intrinsic curvature and adsorption

Model chains representing the average three-dimensional (3D) shape of DNA specimens were generated by the 3DNA software (26) exploiting nearest-neighbor, static dinucleotide wedge models (27,28).

We custom developed an algorithm (LabView, National Instruments) that flattens the 3D model chain to simulate deposition. Briefly, it divides the chain into a discrete number of fragments originally lying on different planes and projects them individually. The output is a 2D chain formed by the geometric projections connected at their ends according to local continuity criteria. This procedure assumes that the 3D → 2D transformation takes place at the expense of few local twists of the molecular backbone; as a consequence it reasonably implies a minimum increase of the conformational energy of the flattened molecule with respect to the 3D counterpart.

The algorithm was implemented as follows. Geometric projection starts at one of the 3D chain ends and involves the longest fragment that can be projected onto a best fit plane while maintaining its overall fluctuations (relative to that plane) below a given threshold. Once such fragment is found, the algorithm is iterated on the remaining part of the 3D chain until the whole curve is flatted onto a unique set of preferential planes. The threshold value is chosen to match the typical range of chain—surface interaction forces, i.e. few nanometres.

The results of the above algorithm for the target DNA were found to be consistent with those obtained by a different theoretical approach, originally proposed by Scipioni *et al.* (20).

RESULTS AND DISCUSSION

Characterization of local intrinsic curvature for the human OPN coding gene

After samples preparation, a quantitative AFM analysis of molecular profiles was routinely performed in order to test the reproducibility of imaging conditions, evaluate relevant deviations of adsorbed DNA superstructure from the canonical B-form and get deeper insight on the influence of intrinsic curvatures on the local and global geometrical properties of the traced contours. Typically, measured DNA molecules displayed an average width of $\sim$10 nm and a height of $0.8 - 1.0$ nm, due respectively to AFM probe convolution effects and to the elastic deformation of the soft molecule under the repulsive forces exerted by the scanning tip (29). Molecules surface density was in the range $2 - 5\,\mu m^{-2}$. The analysis of the contour lengths for a large number of traced molecules ($\sim$400) attested a DNA contraction of 5% with respect to the B-form. This corresponds to a helix rise per base pair of 0.32 nm in excellent agreement with results of similar studies (1,8,10,23,30,31). In Figure 3a we report a representative high resolution topography of the target
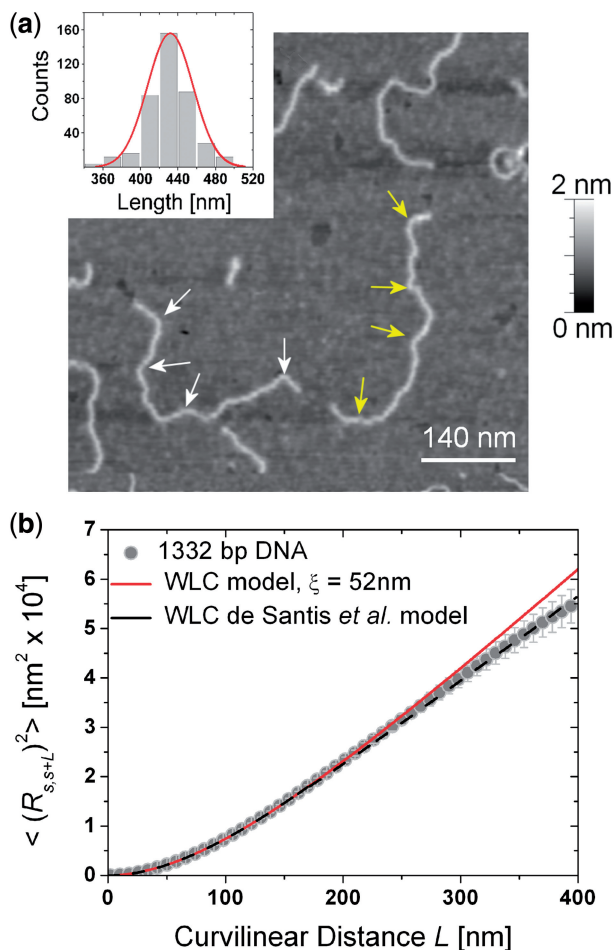
Figure 3. (**a**) Representative AFM topography of the target DNA. It shows the persistence of bends at few locations along the molecular backbone—marked by arrows—suggesting the presence of a significant intrinsic curvature at the same places. In the inset is the histogram of contour lengths. (**b**) Comparison of the experimentally measured end-to-end distance curve with the WLC model predictions for linear (red) and bent (black) chains. The chosen specimen reveals a small but systematic decrease of $\langle R^2_{s,s+L}\rangle$ at curvilinear distances above 250 nm, ascribed to an overall coiling of the chains with respect to linear DNA of comparable length. The WLC simulations on bent chains are in excellent agreement with experimental data at all curvilinear distances, confirming the key role played by intrinsic curvature.
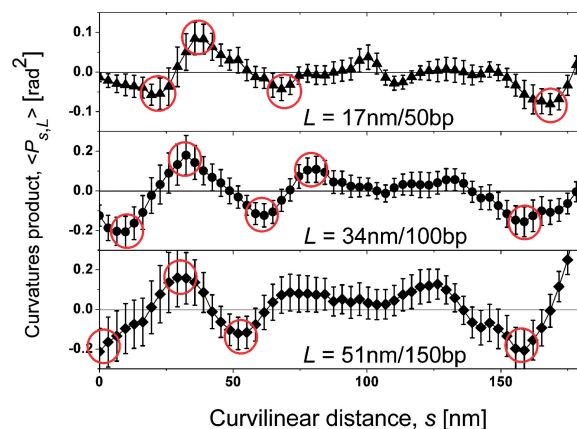

Figure 4. Characteristic patterns of variation of the average CP for the human OPN coding gene, estimated for three different L values. Circles indicate the main positive and negative peaks along each profile: they represent pairs of L-long segments with average curvature oriented respectively in the same or in the opposite direction.

DNA. As expected, it reveals the large variety of shapes assumed by DNA under the thermal stochastic perturbation of its molecular environment. By visual inspection however, one can already notice the persistence of bends at a few sites, namely in close proximity of both ends and within the central region of the chain. This fact suggests the presence of non-null intrinsic curvatures at the same places.

The first quantitative evidence for the influence of local intrinsic curvatures on the global conformation of the chains emerged when we measured the mean-squared end-to-end distance $\langle R^2_{s,s+L}\rangle$ for an ensemble of 160 molecular profiles extracted from several AFM topographies. As shown in Figure 3b, $\langle R^2_{s,s+L}\rangle$ shows a good agreement with the WLC model for $L \leq 200$ nm, as attested by the

effective interpolation of experimental data with Equation (5) in that range; in particular we estimated $\xi = 52$ nm that agrees with the DNA flexibility reported by other AFM experiments (8,23,24) and proves the thermodynamic equilibration of chains on mica for the investigated samples. Notably, a small but systematic decrease of $\langle R^2_{s,s+L}\rangle$ with respect to the WLC predictions for linear chains takes place for $L > 250$ nm. A similar behavior has been reported by Rivetti *et al.* (1) for chains with in-phase A-tracts and by Moreno-Herrero *et al.* (8) for strands with hyperperiodic sequences, and can be considered a signature of the presence of intrinsic curvatures that force DNA to assume (on the average) a more compact coil structure compared with linear DNA of the same length. We confirmed the correctness of that picture *a posteriori*, by comparing experimental data with WLC predictions for bent chains; the intrinsic curvature profile of the ensemble was evaluated through the use of the wedge model of De Santis *et al.* (28) (see also next paragraph). These arguments led us to rule out a substantial impact of excluded volume effects on the measured $\langle R^2_{s,s+L}\rangle$ (23).

Corroborated by such findings, we performed a refined characterization of the intrinsic curvature along the DNA contour by implementing the novel method; in particular, we explored three different contour lengths $L = 17$ nm (50 bp), 34 nm (100 bp) and 51 nm (150 bp) over the same ensemble of 160 profiles. The obtained patterns of variation are contrasted in Figure 4. We observe clear oscillations of s versus $\langle P_{s,L}\rangle$ curves for each L value that confirm once more the presence of intrinsic curvatures along the studied contours and can be used to locate the most significant bending sites of the molecular backbone. According to Equation (4), the main positive and negative peaks of $\langle P_{s,L}\rangle$ mark the curvilinear positions of symmetric pairs of segments with the largest intrinsic curvature. In particular, we recognize three main peaks of $0.05 - 0.1$ rad^2 for $s < 70$ nm that concern pairs of segments close to the contours ends, and a large

negative peak of $0.1 - 0.2 \, \text{rad}^2$ for $s \approx 150 - 175 \, \text{nm}$, which on the contrary regards pairs of tracts located around the middle portion of the strands. In the range $s \approx 70 - 130 \, \text{nm}$, the s versus $\langle P_{s,L} \rangle$ curves are almost completely flat and $\langle P_{s,L} \rangle \approx 0$, which means that at least one of the two symmetrically placed segments presents a negligible intrinsic curvature. Noteworthy, the curvilinear positions of the main peaks of $\langle P_{s,L} \rangle$ in Figure 4 are in good qualitative agreement with the visual inspection of DNA bends from several AFM topographies (see e.g. Figure 3a).

We show in the next paragraphs that the patterns of variation of Figure 4 depend on the nucleotide sequence and the adsorption mechanisms. For simplicity, we focus on the contour lengths $L = 17 \, \text{nm}$ and $34 \, \text{nm}$ that provide patterns of variation free from AFM tip convolution artifacts; in fact these are expected to affect the experimental CP values whenever L becomes comparable with (or smaller than) the DNA apparent width ($\sim 10 \, \text{nm}$, see Supplementary Data). The substantial lack of novel features in the CP pattern for $L = 51 \, \text{nm}$ further justifies our interest for the shorter contour lengths.

Patterns of variation of the average product of curvatures: experiment versus theory

A theoretical model suitable for the interpolation of the experimental results of Figure 4 should in principle account for the sequence-dependent static curvature of DNA and chain dynamics during adsorption and the subsequent surface relaxation, and should as well provide indications on the most important parameters governing the reorganization of superstructure under realistic experimental conditions. This is certainly a complex task since the long-range van der Waals forces and the short-range double-layer ones that control the adsorption process, apart from inducing an adjustments of DNA segments positions in order to adopt the equilibrium distance from the surface, can also tune the appearance of out-of-equilibrium, long-lived alterations of chain architecture, including kinks, over(under)twists, local B to A transitions and even melting (6,31). As a result adsorption can dramatically affect the standard chain geometry and statistics, as already demonstrated by a number of works based on Monte-Carlo (MC) and molecular dynamics simulations (24,32–34).

Whereas the implementation of a comprehensive model for the adsorption of an intrinsically curved DNA is out of the scopes of the present article, we note that the need for a straightforward comparison of electron microscopy and AFM data with theoretical models often lead to the practice to treat the average 3D shape of DNA by means of nearest-neighbor, static dinucleotide wedge models and reduce adsorption to a simple geometric projection of the 3D trajectory onto one or more preferential planes (3,4,18). This solution is of course prone to errors, only partially mitigated by taking into account bent and approximately planar DNA molecules (4). Nevertheless, it can be considered a first-order approximation to the analysis of the intrinsic curvature profile of any 3D chain geometry. For such reasons the same solution was
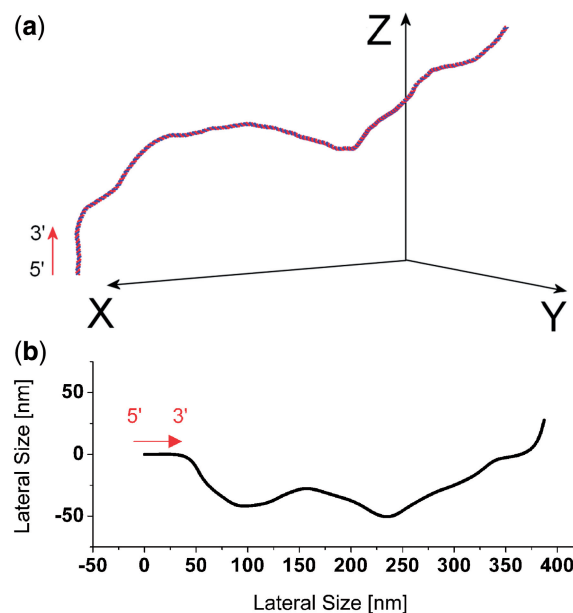

Figure 5. (a) Representation of the average shape of the human OPN coding gene according to the static dinucleotide wedge model of De Santis *et al*. (b) 2D trajectory of the DNA obtained by flattening the 3D model of (a) and twisting a portion of the backbone region (see also Supplementary Data).

adopted in the present case. In doing this we recognize that the obtained theoretical framework provides an oversimplified picture of the real DNA structure and dynamics, as recently confirmed by the results of extensive molecular dynamics simulations [e.g. ref. (35) and references therein]. On the other side, experimental results (1–6,8,18–21) demonstrated that modeling adsorption in terms of projections onto best-fit planes results in a final DNA configuration that satisfyingly approximates the actual, average conformation of equilibrated adsorbed chains.

The 3D intrinsic structure of the target DNA was investigated, respectively, by means of the wedge models of De Santis *et al*. (28) and Bolshoy *et al*. (27) that already demonstrated a good agreement with the DNA intrinsic curvature data accessed by AFM imaging (3,4). In such case, the local static curvature is computed by summation of the differential deviation angles of the helix axis at individual dinucleotide steps and the average shape and intrinsic curvature profile of DNA in bulk solution are readily obtained from mere knowledge of the whole nucleotide sequence.

Visual inspection of the superstructure predicted by the model of De Santis *et al*. for the human OPN coding gene reveals the presence of local bends that extend over several helix turns and clearly impart a 3D shape to the studied strand (Figure 5a).

The result of the 3D $\rightarrow$ 2D transformation mimicking deposition is shown in Figure 5b (see also Supplementary Data for details). There is an astonishing resemblance of the 2D chain with several AFM-imaged molecules, as already attested by comparing Figures 5b and 3a.

The 2D trajectory of Figure 5b was used to simulate the room temperature bending of DNA, describing chain
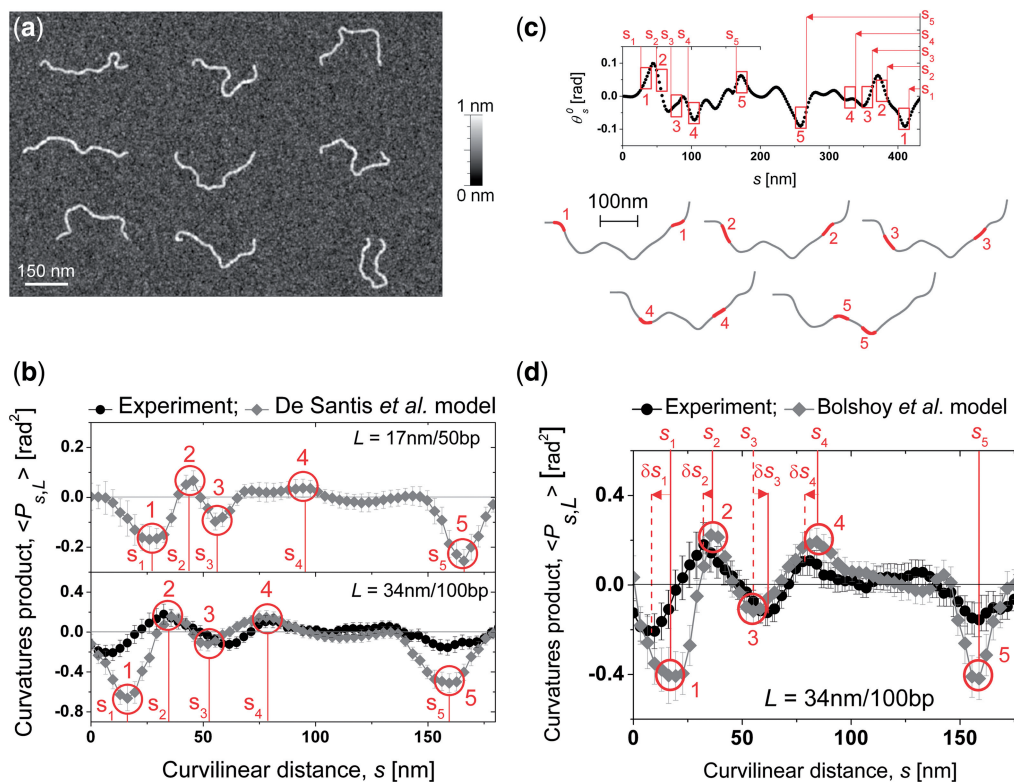
Figure 6. (a) Representative conformations of six chains generated by MC methods from the intrinsic 2D trajectory predicted by De Santis *et al* model for the target DNA. A randomly flat substrate has been intentionally added to generate topographies resembling as close as possible those obtained by AFM (b) Theoretical pattern of variation of the CP for two different values of sliding windows length L; encircled are the main peaks of the plots. Experimental results are reported for comparison with $L = 34$ nm. (c) Intrinsic curvature of the 2D trajectory predicted by De Santis *et al* model, with marked positions of the pairs of segments of length $L = 17$ nm (50 bp) related to the peaks highlighted in (b). Positions are shown for clarity also on the 2D chain. (d) Comparison of experimental results with the theoretical pattern of variation of the CP predicted with the model of Bolshoy *et al.*

lateral motion onto the mica surface. To this purpose, it was sampled at the spacing $l_{WLC} = 0.32$ nm (corresponding to the experimentally found helix rise per base-pair) and thermal effects (on bending) were implemented by adding to the angles among neighbor segments a fluctuation chosen by a MC method from normally distributed numbers with mean zero and variance of l_{WLC}/ξ ($\xi = 52$ nm). The new trajectories were superimposed on a randomly flat substrate (roughness 0.1 nm) and dilated by a parabolic tip (36) in order to generate topographies resembling as close as possible those obtained by AFM (Figure 6a). These were finally analyzed with the tracing algorithm in order to assure a bias—due to random and systematic angular distortions—comparable to that affecting experimental data. In Figure 6b, we report the obtained results for the two sliding windows of size $L = 17$ nm and $L = 34$ nm, respectively.

The theoretical s versus $\langle P_{s,L} \rangle$ curves are characterized by marked oscillations that persist at similar curvilinear distances for both L values. In particular, for $L = 17$ nm three negative peaks of $\approx 0.10 - 0.25$ rad^2 (conventionally named 1, 3 and 5) appear at $s_1 = 25$ nm, $s_3 = 55$ nm and $s_5 = 168$ nm, whereas smaller local maxima (named 2 and 4) occur at $s_2 = 43$ nm(≈ 0.07 rad^2) and $s_4 = 95$ nm(≈ 0.04 rad^2), respectively. A similar trend is

found for $L = 34$ nm. Consistent with Equation (4), local peaks of s versus $\langle P_{s,L} \rangle$ plots are related to pairs of segments with large intrinsic curvature. This is confirmed by a direct inspection of the static curvature profile of the projected chain, reported in Figure 6c: here we properly highlighted the L-long tracts involved in the calculation of $\langle P_{s,L} \rangle$ at the sites $s_1, \ldots, s_5$, demonstrating that each one of them holds appreciable curvatures θ_s^0, whose magnitude varies in the range $0.02 - 0.10$ rad. Notably, the model can trace each tract back to its base-pairs content, therefore the local peaks point out those parts of the primary sequence that impart local and persistent nanoscale curvatures to the adsorbed chains.

We explored the dependence of model predictions from the chosen set of dinucleotide parameters by performing a new data analysis based on the model of Bolshoy *et al.* (27). The latter originates from a large body of curvatures data from circularization and gel electrophoresis mobility experiments, whereas the De Santis *et al.* model is primarily based on theoretical calculations of the minimum energy structure of the DNA strand successively refined to improve the correlation with experimental results. The theoretical pattern for $L = 34$ nm is reported in Figure 6d: here we observe the same peaks of Figure 6b at similar curvilinear positions but with an appreciable variation of

their magnitude. This confirms the largely accepted opinion that one dinucleotide model is as good as another in determining the structure and mechanical properties of DNA in bulk solution (25) and supports the conclusion that peaks curvilinear positions are consistently predicted by our analysis with negligible dependence on the chosen model. On the contrary, the estimated amplitude of the $\langle P_{s,L} \rangle$ peaks is sensibly affected by the specific dinucleotide parameters set and by the flattening process mimicking DNA adsorption. Such fact suggests to exploit the comparison of experimental and theoretical CP patterns on DNA model systems in order to systematically contrast the response of several, well-known DNA bending models proposed so far (25).

The five peaks at $s_1, \ldots, s_5$ can be exploited to drive the comparison between theory and experiment. Assuming $L = 34$ nm for simplicity, we recognize peak 1 also in the experimental data, biased by a small horizontal shift δs_1 of about 8 nm. The shifts for the remaining peaks are negligible compared to the positional errors (<5 nm) affecting the molecular trajectories extracted from the tip-convoluted AFM images (Figure 6d). The protocol adopted for samples preparation is certainly contributing to the observed discrepancy at s_1. In particular, the horizontal shifts δs_1 might be ascribed to a structural reorganization of adsorbed DNA at one or both ends, involving local variations of the helix rise, nanosized deletions or out-of-equilibrium alterations that are not properly resolved by AFM imaging and that can be induced by sample drying (31). Moreover, the reduced magnitude of the peaks in the experimental pattern with respect to the theoretical counterpart (mostly at s_1 and s_5) may be attributed to the rinsing the samples with pure water after DNA adsorption on mica: this step in fact reduces the ionic strength of the solution and consequently enhances the electrostatic repulsion of charged phosphate groups. A net decrease of the absolute curvature of the already adsorbed molecules is therefore highly probable (20,21).

The overall satisfying agreement shown in Figure 6d is achieved also when contrasting data with predictions based on De Santis *et al.* model, but with slightly different δs_i values (bottom panel of Figure 6b). In view of such results, we recognize that our analysis effectively describes the relevant features of the patterns of variation introduced by the new method with a simple and sound theoretical framework. We are able to predict the curvilinear position and amplitude of the main local peaks in a s versus $\langle P_{s,L} \rangle$ plot and if necessary find out those parts of the primary sequence that impart a persistent bending to the target DNA. Due to the consistent response offered by several dinucleotide and trinucleotide bending models (25), the peaks curvilinear positions show robustness against variations of the angular parameters, thus they can be used to compare model predictions with experimental data as well as to gain a deeper insight into the physical processes characterizing DNA adsorption. A quantitative measure of the amount of error between the individual wedge models and experimental data is also settled by introducing the residual sum of squares (RSS). We find $\mathrm{RSS}_{\mathrm{De\,Santis}} \approx 2.3\,\mathrm{rad}^2$ and $\mathrm{RSS}_{\mathrm{Bolshoy}} \approx 1.0\,\mathrm{rad}^2$

due to the better interpolation of the amplitude of peaks 1 and 5 offered by the Bolshoy *et al.* model. Nevertheless, the two models are comparable in the restricted range $30\,\mathrm{nm} < s < 130\,\mathrm{nm}$ (peaks 2, 3 and 4) where $\mathrm{RSS}_{\mathrm{De\,Santis}} \approx \mathrm{RSS}_{\mathrm{Bolshoy}} \approx 0.15\,\mathrm{rad}^2$.

Comparison of the novel method with the automated FF algorithm

The intrinsic curvature profile of the human OPN coding gene was reconstructed by applying the FF algorithm to an ensemble of 100 molecular profiles extracted from AFM images. Figure 7a contrasts the reconstructed θ_s^0 profile with De Santis *et al.* model predictions. The two profiles show very comparable features, in particular two regions of large curvature at $s \approx 50\,\mathrm{nm}$ and $s \approx 170\,\mathrm{nm}$, respectively, (peaks 2 and 5) and a well-defined sequence of smaller local peaks at similar curvilinear positions (peaks 8–12).

The reconstructed θ_s^0 values were used to calculate $\langle P_{s,L} \rangle$ by means of Equations (2) and (4), and the result was finally compared to that obtained with our method (that gives a virtually exact $\langle P_{s,L} \rangle$ value in that it works directly on the ensemble average of $P_{s,L}$ realizations over
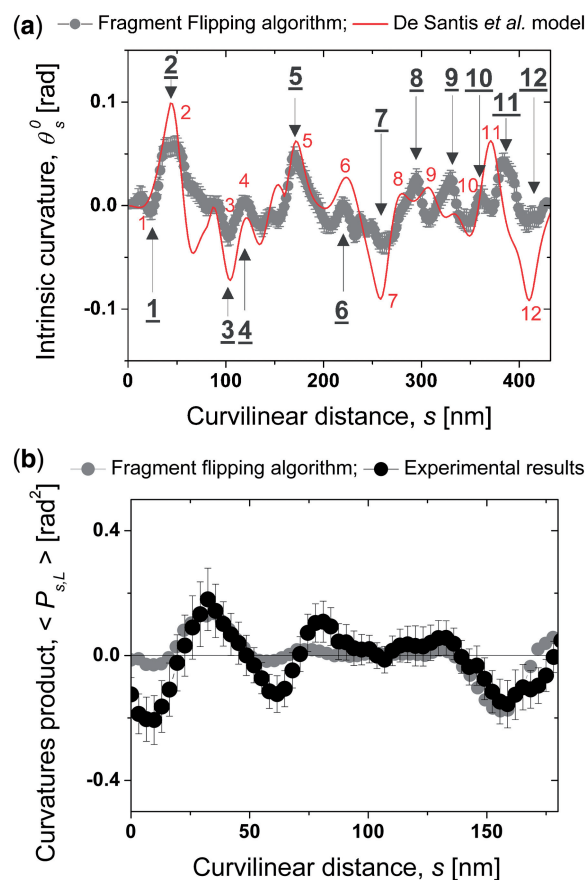
Figure 7. (a) The intrinsic curvature profile estimated by the FF algorithm is compared with the corresponding curve predicted by the theoretical approach described in previous subsection. (b) The CP pattern computed from the intrinsic curvature profile in (a) is compared with the CP profile directly estimated on experimental DNA trajectories by our protocol.

experimental trajectories). As shown in Figure 7b, there is a reasonable agreement of the two data sets for $20\,\text{nm} < s < 50\,\text{nm}$ and $s > 100\,\text{nm}$, whereas appreciable discrepancies occur in the two regions $0 < s < 20\,\text{nm}$ and $50\,\text{nm} < s < 100\,\text{nm}$. This fact demonstrates that the FF algorithm fails to recover the whole intrinsic curvature information displayed by the experimental CP pattern, with the exception of the position and magnitude of the two main peaks located, respectively, at $s \approx 30\,\text{nm}$ and $s \approx 160\,\text{nm}$.

To this purpose, we report few notes of caution on the use of the FF algorithm. Its response sensibly depends on the relative position and orientation of the rows (i.e. the experimental curvature profiles) arranged into the starting curvature matrix, and in fact we observed relevant deviations in the θ_s^0 magnitude according to the chosen initial conditions. This ambiguity is due to the well-known possibility for a hill-climbing optimization routine to provide solutions representing local minima of the objective function (here the mean value of columns variance) instead of the global one (19). For such reason, we reported in Figure 7a the profile with the smallest mean value of columns variance from a set of 10 curvature profiles, obtained by iteratively shifting the starting point of the automated FF algorithm within the ordered ensemble of molecular profiles. It is thus not surprising to see that there is—generally speaking—a systematic uncertainty in the accuracy of θ_s^0 profiles and CP patterns predicted by means of the FF algorithm. Moreover, the pit-fall of local minima stops is likely to become highly probable for very large sets of molecular trajectories. A second crucial drawback of the FF algorithm consists in the fact that it does not give indication on the alignment of the reconstructed curvature profile with respect to chain polarity (i.e. the $5' - 3'$ direction), which complicates any comparison of experimental results with theoretical models and definitely hampers the implementation of assays contrasting the profiles of a large number of samples. As mentioned earlier, this serious limitation is encompassed by the introduction of the statistical chain descriptor $P_{s,L}$ that leads to an orientation-independent description of local intrinsic curvature through the CP pattern.

A case study: sensitivity of CP patterns to point mutations in the OPN encoding gene

In view of the excellent response offered by our method in terms of robustness, accuracy and flexibility, we foresee several challenging applications for the CP patterns. To this purpose, we first note that the typical target specimens should consist of 10^2–10^3 bp long chains: these are readily deposited by standard protocols on atomically smooth substrates and can be routinely imaged by AFM (1–13,15–21). High-resolution AFM (8,24) is mandatory in order to explore the intrinsic curvature of shorter chains ($<10^2$ bp) and achieve reproducible estimates of the angular parameters for fragments as small as 15 bp ($L = \sim5\,\text{nm}$). Secondly, we emphasize that the applicability of the new method goes definitely beyond the case of the OPN encoding gene. In fact we demonstrated above

that any DNA strand with a non-zero s versus $\langle C_{s,L} \rangle$ profile is described as well by a CP profile with comparable accuracy and sensitivity. This follows directly from the definition of the CP descriptor [Equations (3), (4), Supplementary Equations (S10), (S11)]. Moreover, we carried out additional simulations on two model systems, namely 500 bp random sequences and the 937 bp EcoRV-P*st*I fragment of pBR322 DNA (see Supplementary Data for details). The obtained results attest that DNA templates that do not contain prominent nucleotide sequences responsible for large bends are nevertheless characterized by informative CP patterns. All together such arguments make us confident of the applicability of our method to a broad class of intrinsically bent duplexes.

One interesting possibility of application might regard the systematic use of CP maps to deeply explore the predictions of DNA adsorption and bending models. An insight into this topic was provided in the sections above and significant improvements are expected to come from state of the art modeling (as Brownian dynamics and molecular dynamics simulations) going beyond the nearest-neighbor approximation in conformational analysis or describing the non-equilibrium processes of DNA adsorption and relaxation on the atomically flat substrate (24,35,37,38). For example, a tight comparison of experimental and theoretical CP patterns might allow to identify the presence of restricted regions where out-of-equilibrium alterations of the chain architecture systematically take place during adsorption. This information might be eventually related to the local base pairs sequence and/or exploited to tune DNA adsorption according to the needs of novel comparative essays. Another challenge might involve the use of CP patterns to routinely detect small conformational changes in large sample numbers. The capability to relate DNA structural variations to physical or biological causes (e.g. mutations at one or more base-pairs) might eventually contribute to develop new assays and even genetic screening protocols for highly bent duplexes. Interestingly, some studies might explore the ultimate sensitivity of CP patterns to point mutations and mismatched base-pairs and largely contribute to the discovery of physical methodologies for molecular haplotyping (16,39). Within this context we offer a concrete example on the CP patterns sensitivity to single nucleotide polymorphisms (SNPs) in the OPN encoding gene. In detail, we contrast two homozygous specimens having different SNPs at four, well-known polymorphic sites. To date, there is a well documented functional effect of such SNPs on the OPN gene transcriptional activity (22), and they play a useful role as genetic markers to characterize patients with oligoarticular juvenile idiopathic arthritis (40). In Figure 8a, we show the 3D model chains predicted for the two specimens. It appears that the insertion (or deletion) of an individual G base at the sequence site 762 (marked by the vertical arrows) dramatically affects the whole DNA bending close to the centre of the chain, in fact inducing a variation of the relative orientation of the $5'$ half with respect to the $3'$ end. This is confirmed by the corresponding CP patterns, evaluated through theory and experiment as described above. In particular, Figure 8b attests the
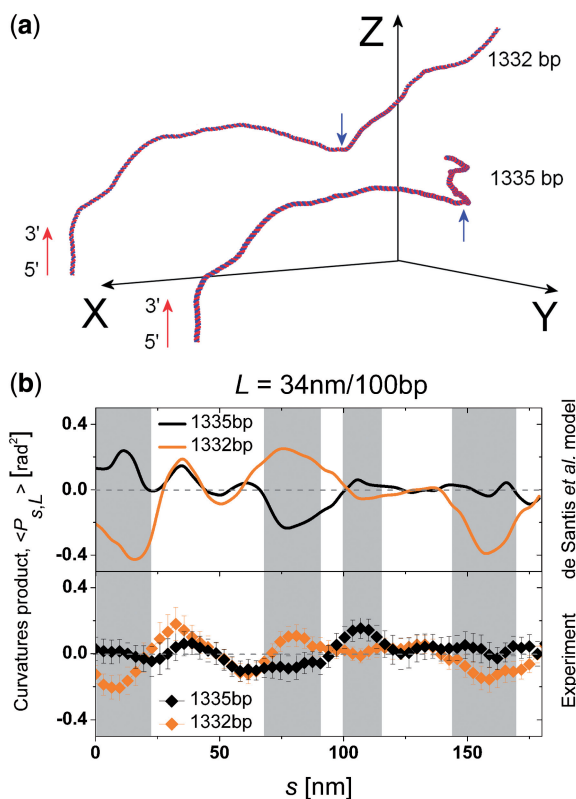
Figure 8. (a) Representation of the average 3D shape of two homozygous samples according to the wedge model by De Santis *et al.* The SNP at the sequence site 762 (marked by arrows) impacts the overall relative orientation of the 5′ half with respect to the 3′ end, whereas the other three SNPs do not substantially affect the DNA shape. (b) Top: theoretical patterns of variation of the CP for the two chains in (a) with $L = 34$ nm. Bottom: experimental patterns of variation for the CP. Gray regions highlight statistically relevant differences between the two specimens, in excellent agreement with theoretical predictions.

emergence of statistically relevant differences for the experimental CP values at four main regions of the curvilinear distance s (highlighted in gray). The experimental pattern of the 1335 bp specimen also shows less marked amplitude variations with respect to the 1332 bp counterpart. A detailed analysis of the overall fluctuation of the CP signal for different L values (50–120 bp) confirms that this feature systematically occurs in both theory and experiment (see Supplementary Data); it thus represents a robust sequence-dependent property of the samples that is successfully captured by the CP method. We underline that the most relevant message of Figure 8b is to document the practical feasibility of the label-free comparative analysis envisaged in Figure 2c. Such an essay represents the crucial advantage offered by the CP method with respect to other conformational methods and is reported in the present investigation—to the best of our knowledge—for the first time.

We strongly believe that the possibility to routinely achieve similar results, through the use of symmetric curvature descriptors operating under label-free conditions, should boost the applicability of AFM conformational analysis in novel, genetic screening tests.

We finally note that further attractive developments might come from the evaluation of CP patterns to address the structural properties of DNA fragments complexed with intercalating dyes and binding drugs (12,41) or even proteins. In fact the CP patterns might be quite useful to complement current AFM studies on the formation of protein–DNA complexes [e.g. ref. (42)], where the position distribution of protein binding along unlabeled DNA fragments is calculated relative to the closest DNA terminus. Indeed this choice statically couples binding events occurring on symmetrically placed tracts, in analogy with the curvatures coupling contained in the $P_{s, L}$ definition. As a result, a visual correlation of s versus $\langle P_{s, L} \rangle$ and s versus protein-binding-frequency plots would easily point out the existence of helix sites where local intrinsic curvature drives the so called 'indirect' DNA recognition or competes with other binding mechanisms (25,43). This is certainly of dramatic interest for fundamental investigations addressing the ability of proteins to locate specific sites or structures among a vast excess of non-specific, intrinsically bent DNA, as in the relevant case of mismatch repair proteins interrogating DNA to find out biosynthetic errors and promote strand-specific repair (11).

CONCLUSIONS

In this article we proposed a novel method to characterize the local intrinsic curvature of adsorbed DNA molecules by AFM. It relies on the fine mapping of a statistical chain descriptor that highlights all pairs of intrinsically bent segments symmetrically placed along the helix chain. This peculiar choice provides a number of advantages overcoming some fundamental and practical limitations of early protocols. It is in fact well known that such protocols generate intrinsic curvature maps starting from the contours of end-labeled molecules or palindromes and those conformational averages are carried on under the assumption that a preferential DNA adsorption takes place. More importantly, none of the current methods is expected to readily manage comparative assays involving a large number of samples. On the contrary, we demonstrated, both theoretically and experimentally, that the novel method can be implemented on label-free molecules with unknown orientation, in fact reducing specimen preparation to standard procedures for DNA deposition. Accordingly, neither end-labeled molecules nor palindromic constructs are strictly required and no *a priori* assumptions or additional evidences on the DNA adsorption mechanisms are necessary. Experimental uncertainty affecting the new curvature patterns is comparable to that already discussed in early works and derives from AFM tip convolution effects and the specific algorithm used for DNA tracing from AFM topographies.

We therefore conclude that the novel method paves the way for a reliable, unbiased, label-free comparative analysis of bent duplexes, aimed to detect local conformational changes of physical or biological relevance in large sample numbers. To this purpose, we suggested few relevant examples that should boost the applicability

of AFM-based curvature studies, e.g. validating DNA adsorption and bending models by experiments, uncovering DNA interactions with proteins, intercalating dyes and drugs, setting up population-based genetic disease studies or solving genomic screening problems at the single-molecule level.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table S1, Supplementary Figures S1-S9, Supplementary Materials and Methods and Supplementary Equations S1-S11.

ACKNOLEDGEMENTS

FUNDING

REFERENCES

1. Rivetti,C., Walker,C. and Bustamante,C. (1998) Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.*, **280**, 41–59.
2. Cognet,J.A.H., Pakleza,C., Cherny,D., Delain,E. and Le Cam,E. (1999) Static curvature and flexibility measurements with microscopy. A simple renormalization method, its assessment by experiment and simulation. *J. Mol. Biol.*, **285**, 997–1009.
3. Zuccheri,G., Scipioni,A., Cavaliere,V., Gargiulo,G., De Santis,P. and Samorì,B. (2001) Mapping the intrinsic curvature and flexibility along the DNA chain. *Proc. Natl Acad. Sci. USA*, **98**, 3074–3079.
4. Marilley,M., Sanchez-Sevilla,A. and Rocca-Serra,J. (2005) Fine mapping of inherent flexibility variation along DNA molecules. Validation by atomic force microscopy (AFM) in buffer. *Mol. Gen. Genomics*, **274**, 658–670.
5. Moukhtar,J., Fontaine,E., Faivre-Moskalenko,C. and Arneodo,A. (2007) Probing the persistence in DNA curvature properties by atomic force microscopy. *Phys. Rev. Lett.*, **98**, 178101–4.
6. Faas,F.G.A., Rieger,B., van Vliet,L.J. and Cherny,D.I. (2009) DNA deformations near charged surfaces: electron and atomic force microscopy views. *Biophys. J.*, **97**, 1148–1157.
7. Moukhtar,J., Faivre-Moskalenko,C., Milani,P., Audit,B., Valliant,C., Fontaine,E., Mongelard,F., Lavorel,G., St-Jean,P., Bouvet,P. *et al.* (2010) Effect of genomic long-range correlations on DNA persistence length: from theory to single molecule experiments. *J. Phys. Chem. B*, **114**, 5125–5143.
8. Moreno-Herrero,F., Seidel,R., Johnson,S.M., Fire,A. and Dekker,N.H. (2006) Structural analysis of hyperperiodic DNA from *Caernorhabditis elegans*. *Nucleic Acids Res.*, **34**, 3057–3066.
9. Marilley,M., Milani,P. and Rocca-Serra,J. (2007) Gradual melting of replication origin (*Schizosaccharomyces pombe* ars1): *in situ* atomic force microscopy (AFM) analysis. *Biochimie*, **89**, 534–541.
10. Dame,R.T., Van Mameren,J., Luijsterburg,M.S., Mysiak,M.E., Janicijevic,A., Pazdzior,G., van der Vliet,P.C., Wyman,C. and Wuite,G.J.L. (2005) Analysis of scanning force microscopy images for protein-induced DNA bending using simulations. *Nucleic Acids Res.*, **33**, e68.
11. Gorman,J., Chowdhury,A., Surtees,J.A., Shimada,J., Reichman,D.R., Alani,E. and Greene,E.C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell*, **28**, 359–370.
12. Adamcik,J., Valle,F., Witz,G., Rechendorff,K. and Dietler,G. (2008) The promotion of secondary structures in single-stranded DNA by drugs that bind to duplex DNA: an atomic force microscopy study. *Nanotechnology*, **19**, 384016–384023.
13. Marilley,M., Milani,P., Thimonier,J., Rocca-Serra,J. and Baldacci,G. (2007) Atomic force microscopy of DNA in solution and DNA modelling show that structural properties specify the eukaryotic replication initiation site. *Nucleic Acids Res.*, **35**, 6832–6845.
14. Garcia,H.G., Grayson,P., Han,L., Inamdar,M., Kondev,J., Nelson,P.C., Phillips,R., Widom,J. and Wiggins,P.A. (2006) Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, **85**, 115–128.
15. Fang,Y., Spisz,T.S., Wiltshire,T., D'Costa,N.P., Bankman,I.N., Reeves,R.H. and Hoh,J.H. (1998) Solid-state DNA sizing by atomic force microscopy. *Anal. Chem.*, **70**, 2123–2129.
16. Wooley,A.T., Guillemette,C., Li Cheung,C., Housman,D.E. and Lieber,C.M. (2000) Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat. Biotechnol.*, **18**, 760–763.
17. Reed,J., Mishra,B., Pittenger,B., Magonov,S., Troke,J., Teitell,M.A. and Gimzewski,J.K. (2007) Single molecule transcription profiling with AFM. *Nanotechnology*, **18**, 044032–044046.
18. Muzard,G., Théveny,B. and Révet,B. (1990) Electron microscopy mapping of pBR322 DNA curvature. Comparison with theoretical models. *EMBO J.*, **9**, 1289–1298.
19. Ficarra,E., Masotti,D., Macii,E., Benini,L., Zuccheri,G. and Samorì,B. (2005) Automated intrinsic DNA curvature computation from AFM images. *IEEE Trans. Biomed. Eng.*, **52**, 2074–2085.
20. Scipioni,A., Anselmi,C., Zuccheri,G., Samori,B. and De Santis,P. (2002) Sequence-dependent DNA curvature and flexibility from scanning force microscopy images. *Biophys. J.*, **83**, 2408–2418.
21. Sampaolese,B., Bergia,A., Scipioni,A., Zuccheri,G., Savino,M., Samori,B. and De Santis,P. (2002) Recognition of the DNA sequence by an inorganic crystal surface. *Proc. Natl Acad. Sci. USA*, **99**, 13566–13570.
22. Giacopelli,F., Marciano,R., Pistorio,A., Catarsi,P., Canini,S., Karsenty,G. and Ravazzolo,R. (2004) Polymorphisms in the osteopontin promoter affect its transcriptional activity. *Physiol. Genomics*, **20**, 87–96.
23. Rivetti,C., Guthold,M. and Bustamante,C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, **264**, 919–932.
24. Wiggins,P.A., Van der Heijden,T., Moreno-Herrero,F., Spakowitz,A., Phillips,R., Widom,J., Dekker,C. and Nelson,P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.*, **1**, 137–141.
25. Crothers,D.M. (1998) DNA curvature and deformation in protein–DNA complexes: a step in the right direction. *Proc. Natl Acad. Sci. USA*, **95**, 15163–15165.
26. Lu,X. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
27. Bolshoy,A., McNamara,P., Harrington,R.E. and Trifonov,E.N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl Acad. Sci. USA*, **88**, 2312–2316.
28. De Santis,P., Palleschi,A., Savino,M. and Scipioni,A. (1988) A theoretical model of DNA curvature. *Biophys. Chem.*, **32**, 305–317.
29. Ebeling,D., Holscher,H., Fuchs,H., Anczykowski,B. and Schwarz,U.D. (2006) Imaging of biomaterials in liquids: a comparison between conventional and Q-controlled amplitude modulation ('tapping mode') atomic force microscopy. *Nanotechnology*, **17**, S221–S226.

30. Sanchez-Sevilla,A., Thimonier,J., Marilley,M., Rocca-Serra,J. and Barbet,J. (2002) Accuracy of AFM measurements of the contour length of DNA fragments adsorbed on mica in air and in aqueous buffer. *Ultramicroscopy*, **92**, 151–158.

31. Sushko,M.L., Shluger,A.L. and Rivetti,C. (2006) Simple model for DNA adsorption onto a mica surface in 1:1 and 2:1 electrolyte solutions. *Langmuir*, **22**, 7678–7688.

32. Cerda,J.J. and Sintes,T. (2005) Stiff polymer adsorption: onset to pattern recognition. *Biophys. Chem.*, **115**, 277–283.

33. Semenov,A.N. (2002) Adsorption of a semiflexible wormlike chain. *Eur. Phys. J. E Soft Matter*, **9**, 353–363.

34. Stepanow,S. (2001) Adsorption of a semiflexible polymer onto interfaces and surfaces. *J. Chem. Phys.*, **115**, 1565–1568.

35. Lavery,R., Zakrzewska,K., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T. III, Dixit,S., Jayaram,B., Lankas,F., Laughton,C. *et al*. (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.

36. Horcas,I., Fernández,R., Gómez-Rodríguez,J.M., Colchero,J., Gómez-Herrero,J. and Baro,A.M. (2007) WSXM: a software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.*, **78**, 013705–013713.

37. Curuksu,J., Zacharias,M., Lavery,R. and Zakrzewska,K. (2009) Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.*, **37**, 3766–3773.

38. Lankas,F., Spackova,N., Moakher,M., Enkhbayar,P. and Sponer,J. (2010) A measure of bending in nuclei acids structures applied to A-tract DNA. *Nucleic Acids Res.*, **38**, 3414–3422.

39. Kwok,P.Y. and Chen,X. (2003) Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.*, **5**, 43–60.

40. Marciano,R., Giacopelli,F., Dovizia,M.T., Gattorno,M., Felici,E., Pistorio,A., Martini,A., Ravazzolo,R. and Picco,P. (2006) A polymorphic variant inside the osteopontin gene shows association with disease course in oligoarticular juvenile idiopathic arthritis. *Ann. Rheum. Dis.*, **65**, 662–665.

41. Günther,K., Mertig,M. and Seidel,R. (2010) Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Res.*, **38**, 6526–6532.

42. Yang,Y., Sass,E.S., Du,C., Hsieh,P. and Erie,D.A. (2005) Determination of protein-DNA binding constants and specificities from statistical analysis of single molecules: MutS-DNA interactions. *Nucleic Acids Res.*, **33**, 4322–4334.

43. Dickerson,R.E. and Chiu,T.K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers*, **44**, 361–403.