OPEN

Correspondence and
requests for materials
should be addressed to
L.R. (luca.repetto@
unige.it)

Symmetric curvature descriptors for label-free analysis of DNA

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

[1]CNR-SPIN Institute for Superconductors, Innovative Materials and Devices, Genova, Italy, [2]Department of Physics, University of Genova, Genova, Italy, [3]Department of Pediatrics and Center of Excellence for Biomedical Research, University of Genova, Genova, Italy, [4]Laboratory of Molecular Genetics, G. Gaslini Institute, Genova, Italy.

High-resolution microscopy techniques such as electron microscopy, scanning tunnelling microscopy and atomic force microscopy represent well-established, powerful tools for the structural characterization of adsorbed DNA molecules at the nanoscale. Notably, the analysis of DNA contours allows mapping intrinsic curvature and flexibility along the molecular backbone. This is particularly suited to address the impact of the base-pairs sequence on the local conformation of the strands and plays a pivotal role for investigations relating the inherent DNA shape and flexibility to other functional properties. Here, we introduce novel chain descriptors aimed to characterize the local intrinsic curvature and flexibility of adsorbed DNA molecules with unknown orientation. They consist of stochastic functions that couple the curvatures of two nanosized segments, symmetrically placed on the DNA contour. We show that the fine mapping of the ensemble-averaged functions along the molecular backbone generates characteristic patterns of variation that highlight all pairs of tracts with large intrinsic curvature or enhanced flexibility. We demonstrate the practical applicability of the method for DNA chains imaged by atomic force microscopy. Our approach paves the way for the label-free comparative analysis of duplexes, aimed to detect nanoscale conformational changes of physical or biological relevance in large sample numbers.

M icroscopical methods for the analysis of DNA contours and the mapping of its intrinsic curvature and flexibility have been developed by several groups[1–7]. These methods have been exploited for different purposes, as the experimental validation of models for DNA adsorption and bending[1,2,4–9] or the correlation of DNA shape and flexibility to melting[10], ligand interactions[11–13], replication[14], genomic packaging and transcription regulation[15,16]. The wide applicability of DNA conformational studies demands simple experimental methods, characterized by a few processing steps for specimen preparation and minimum experimental bias on curvature and flexibility measurements. These requirements are crucial for the introduction of effective assays for DNA analysis fully based on high-resolution imaging, e.g. sizing[17], genotyping and haplotyping[18], expression profiling[19,20]. Furthermore, they may lead to envision a key role for the nanoscale conformational analysis within more complex protocols, setting up population-based genetic disease studies or solving genomic screening problems at the single-molecule level[21].

Current studies on DNA structure and flexibility involve the high-resolution imaging of adsorbed species and 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[1–9,22–24]. Tracing algorithms represent each molecule as a chain of *xy* pairs separated by a fixed distance *l* along the contour (Fig. 1). The curvature analysis 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)[4]. From the θ_i values one can define the global curvature $C_{j,m}$ for a segment of *m* units, located at *j* units from one of the ends, as:

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

with $j = 1,2,\ldots,N$ and $m = 1,2,\ldots,N-j$. It is a common practice to neglect statistical correlations among neighbouring θ_i variables and to represent them as the sum of static and dynamic contributions, i.e. $\theta_i = \theta_i^0 + \theta_i^{th}$, where the thermally-induced angular fluctuations θ_i^{th} occur around the constant sequence-dependent θ_i^0 angles and are normally distributed with null mean value[1–4,22–24]. Thus the average value of the $C_{j,m}$ curvature is:
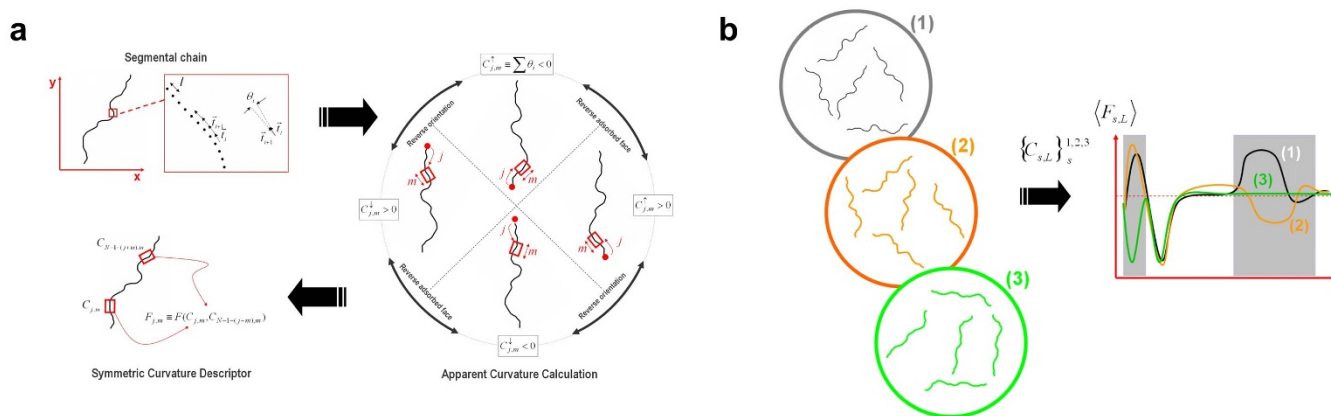
Figure 1 | Schematic illustration of the main steps of DNA conformational analysis, & presentation of a comparative essay exploiting a symmetric curvature descriptor. (a) A molecule is imaged by high-resolution microscopy then traced by an image-analysis software and represented as a chain of xy pairs 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 $\vec{t}_{i+1}$ (top left panel). One 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 (right panel). A generic symmetric curvature descriptor $F_{j,m}$ couples the signed curvatures of two m-units long segments, symmetrically placed at j units from chain ends (bottom left panel). The main advantage is that such quantity remains the same for each one of the possible orientations of the extracted contour. (b) The characteristic patterns of variation of $\langle F_{s,L} \rangle$ for some specimens – named (1), (2) and (3) – enable to highlight DNA regions with different intrinsic curvature and/or flexibility (grey boxes). This represents an original strategy to establish the comparative analysis of bent duplexes under label-free conditions.

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

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 side (l.h.s.) term might be experimentally accessed by averaging the $C_{j,m}$ realizations over a large pool of imaged molecular contours[1,3–5,25], whereas the right hand side (r.h.s.) term should be predicted computationally by well-consolidated methods (e.g. the static dinucleotide wedge models by Bolshoy *et al.*, Gorin *et al.*, Olson *et al.* or De Santis *et al.* highlighted in Ref. 26).

In line with the above arguments, the experimental curvature variance $\sigma^2_{C_{j,m}}$ can be related to the theoretical chain *flexibility*[4,25], where $\sigma^2_{C_{j,m}}$ is defined in the usual way:

$$\sigma^2_{C_{j,m}} \equiv \left\langle \left| C_{j,m} - \langle C_{j,m} \rangle \right|^2 \right\rangle \qquad (3)$$

The practical estimation of $\langle C_{j,m} \rangle$ and $\sigma^2_{C_{j,m}}$ is however a nontrivial task since it requires to orientate the sampled molecular contours in order to evaluate the curvature averages at corresponding points of the nucleotide sequence. In general, for each molecular contour extracted from a high-resolution 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 on which of the two chemically-different faces are exposed by the molecule to the substrate when collapsing on it from the bulk solution (Fig. 1a). In the case of unlabeled chains, their orientation uncertainty cannot be solved deterministically because of the lack of any distinctive topographical feature between the beginning and the end of a DNA molecule. Discrimination of chain polarity was traditionally solved by end-labelling with bulky tags[1,3,5,18]. Alternatively, palindromic dimers can be constructed starting from the target molecules[4]. An uncertainty however remains on the two orientations with mirror curvature profiles that describe DNA

adsorption on chemically-different faces. Scipioni *et al.*[25] and Sampaolese *et al.*[27] demonstrated that such orientations are not statistically equivalent if the molecules are deposited onto freshly-cleaved mica, because of a preferential adsorption of T-rich faces. This fact ultimately justifies the nonzero intrinsic curvature $\langle C_{j,m} \rangle$ obtained from the analysis of an ensemble of palindromic dimers, or even from labelled chains after proper orientation to have the same polarity.

One readily recognizes that it would be more desirable to characterize the local DNA curvature *without any assumption* on the adsorption mechanism and preferential orientation of target chains on a given substrate. Furthermore, preparation of end-labelled duplexes and palindromic constructs represents a *time-consuming, labor-intensive* part of the whole experiment that hampers the broad applicability of similar studies. To address these limitations, we propose to perform conformational analysis of label-free duplexes using Symmetric Curvature Descriptors (SCDs). We define a SCD as a stochastic function that couples the apparent global curvatures of two segments, symmetrically placed along the DNA molecular backbone, in a way that its realizations *do not* depend - neither in modulus nor in sign - on the orientation arbitrarily assigned to the analyzed chains. In other words the SCDs are (centro)symmetric. The reader is referred to the Supplementary Note for a general discussion on their mathematical and statistical properties. Using segmental chain notation, a generic SCD $F_{j,m} \equiv F_{j,m}(C_{j,m}, C_{N-1-(j+m),m})$ couples the signed curvatures $C_{j,m}$ and $C_{N-1-(j+m),m}$ of two m-long segments, placed at j units from chain ends (Fig. 1a). In a previous work[24] we investigated the DNA intrinsic curvature using one specific descriptor, namely $P_{j,m} \equiv C_{j,m} \cdot C_{N-1-(j+m),m}$. In the present case, we use SCDs to probe *both* intrinsic curvature and differential flexibility. As several different descriptors $F_{j,m}$ might be in principle introduced to this goal, we explicitly choose a few selected descriptors defined according to criteria of simplicity and convenience. It is known that $|C_{j,m}|$, $C_{j,m}^2$ or $\cos C_{j,m}$ always depend on the intrinsic curvature and flexibility of the m-long segment[4,26], therefore we have considered basic algebraic operations involving such quantities in a symmetric arrangement. In detail, we complement $P_{j,m}$ with $COS_{j,m} \equiv \cos C_{j,m} \cdot \cos C_{N-1-(j+m),m}$ and $SS_{j,m} \equiv C_{j,m}^2 + C_{N-1-(j+m),m}^2$. It
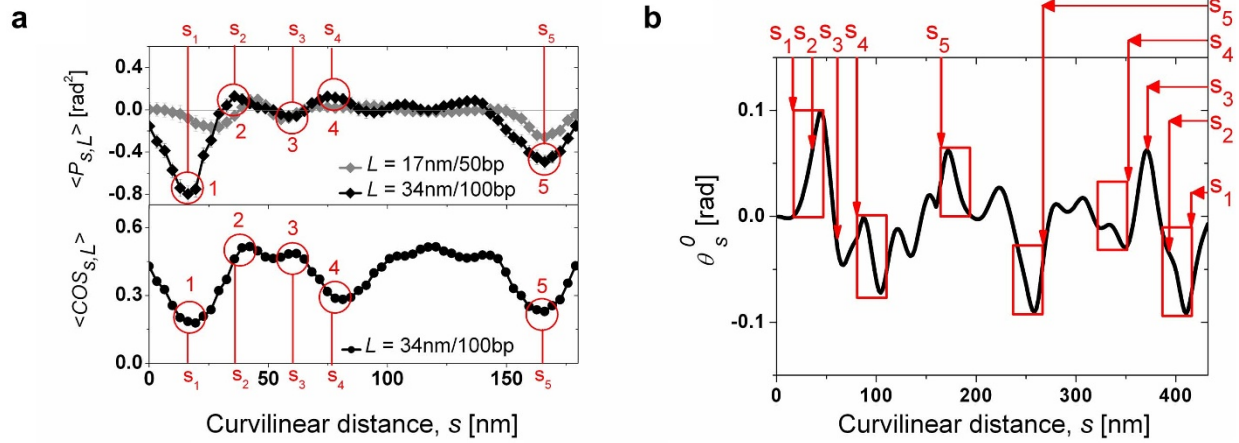
Figure 2 | Simulated conformational analysis of a bent, flexible 1332 bp duplex from the promoter region of the OPN coding gene: descriptors $\langle P_{s,L}\rangle$, $\langle COS_{s,L}\rangle$, θ_s^0. (a) Patterns of variation of $\langle P_{s,L}\rangle$ and $\langle COS_{s,L}\rangle$. Five local peaks are highlighted by circles. (b) Intrinsic curvature profile. The 34 nm-long tracts involved in the calculation of conformational averages at the sites $s_1 \ldots s_5$ are reported for completeness.

is trivial to demonstrate that these descriptors are symmetric. Accordingly, $\langle P_{j,m}\rangle$, $\langle COS_{j,m}\rangle$ and $\langle SS_{j,m}\rangle$ can be estimated by ensemble averages obtained from a large pool of molecular profiles with arbitrary relative orientation. In agreement with this picture, neither end-labelled molecules nor palindromic constructs are required.

For the case of non-overlapping fragments ($j = 1,2,\ldots,N/2$, $m = 1,\ldots,N/2 - j$):

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

$$\langle COS_{j,m}\rangle \equiv \langle \cos C_{j,m}\cdot \cos C_{N-1-(j+m),m}\rangle$$
$$= \cos\langle C_{j,m}\rangle \cdot \cos\langle C_{N-1-(j+m),m}\rangle \cdot c\left(\sigma_{C_{j,m}}^2, \sigma_{C_{N-1-(j+m),m}}^2\right) +$$
$$+ \cos^2\langle C_{j,m}\rangle \cdot \cos^2\langle C_{N-1-(j+m),m}\rangle \cdot c^*\left(\sigma_{C_{j,m}}^2, \sigma_{C_{N-1-(j+m),m}}^2\right) \quad (5)$$

$$\langle SS_{j,m}\rangle \equiv \langle C_{j,m}^2 + C_{N-1-(j+m),m}^2\rangle$$
$$= \langle C_{j,m}\rangle^2 + \langle C_{N-1-(j+m),m}\rangle^2 + \left(\sigma_{C_{j,m}}^2 + \sigma_{C_{N-1-(j+m),m}}^2\right) \quad (6)$$

where $\langle C_{j,m}\rangle$ and $\sigma_{C_{j,m}}^2$ are given respectively by equations (2) and (3). The quantities $c\left(\sigma_{C_{j,m}}^2, \sigma_{C_{N-1-(j+m),m}}^2\right)$ and $c^*\left(\sigma_{C_{j,m}}^2, \sigma_{C_{N-1-(j+m),m}}^2\right)$ are numeric constants. The last equality of each equation holds under specific conditions discussed in the Supplementary Note. Equation (4) shows that $\langle P_{j,m}\rangle$ depends *only* on the intrinsic curvatures of the two chosen segments. On the contrary, equations (5) and (6) indicate that $\langle COS_{j,m}\rangle$ and $\langle SS_{j,m}\rangle$ include *two* qualitatively different contributions, from the intrinsic curvatures of the segments ($\langle C_{j,m}\rangle$, $\langle C_{N-1-(j+m),m}\rangle$) and from their flexibility ($\sigma_{C_{j,m}}^2$, $\sigma_{C_{N-1-(j+m),m}}^2$).

We carried out the DNA conformational analysis by introducing the curvilinear distance $s \equiv jl$ and by plotting s vs $\langle F_{s,L}\rangle$ at fixed L ($L \equiv ml$), that corresponds to probe the emergence of intrinsic curvature and/or flexibility 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 $\langle F_{s,L}\rangle$ whenever large intrinsic curvature and/or enhanced flexibility affect the trajectory of the chosen fragments. Overall, these features contribute to generate a *characteristic* pattern of variation of $\langle F_{s,L}\rangle$ that can be exploited to set up the comparative analysis of bent duplexes (Fig. 1b). Accordingly, we explored the characteristic patterns of variation of $\langle P_{s,L}\rangle$, $\langle COS_{s,L}\rangle$ and $\langle SS_{s,L}\rangle$, i.e. we plotted s vs $\langle P_{s,L}\rangle$, s vs $\langle COS_{s,L}\rangle$ and s vs $\langle SS_{s,L}\rangle$ at fixed L.

Results

We validated the performance of the proposed descriptors using both simulations and real Atomic Force Microscopy (AFM) data of 1332 bp strands from the promoter region of the human Osteopontin (OPN) coding gene.

Theoretical patterns of variation for human OPN coding gene. We predicted the average shape of target DNA with the static dinucleotide wedge model of De Santis *et al.*[28], and treated room-temperature bending through sequence-*independent*, Worm-Like-Chain (WLC) flexibility[22] (see Methods section). The simulated $\langle P_{s,L}\rangle$ and $\langle COS_{s,L}\rangle$ patterns are reported in Fig. 2a. For simplicity, we focus on the contour lengths $L = 17$ nm/50 bp and $L = 34$ nm/100 bp. These are larger than the DNA apparent width estimated by AFM (≈ 10 nm, see below) hence they also provide experimental patterns of variation free from AFM tip convolution artefacts. The s vs $\langle P_{s,L}\rangle$ curves are characterized by marked oscillations that persist for different L values, namely $L = 17$ nm/50 bp and $L = 34$ nm/100 bp. In particular for $L = 34$ nm one recognises three negative peaks of ≈ 0.10–0.8 rad^2 (conventionally named 1, 3 and 5) at $s_1 = 16$ nm, $s_3 = 59$ nm and $s_5 = 165$ nm, whereas smaller local maxima (named 2 and 4) occur at $s_2 = 35$ nm (≈ 0.07 rad^2) and $s_4 = 78$ nm (≈ 0.04 rad^2) respectively. Consistently with equation (4), local peaks of the s vs $\langle P_{s,L}\rangle$ plot are related to pair of segments with large intrinsic curvature. This is confirmed by a direct inspection of the static curvature profile of the 2D trajectory of the OPN-related DNA (Fig. 2b). One can notice that the signed angle θ_s^0 varies in the range -0.1 rad $< \theta_s^0 < 0.1$ rad, with few marked oscillations taking place over a length scale of ~ 100 nm. Here we properly highlighted the 34 nm-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 rad–0.10 rad. Fig. 2a reveals that the characteristic pattern of variation of $\langle COS_{s,L}\rangle$ closely resembles the s vs $\langle P_{s,L}\rangle$ one, with five local peaks at the curvilinear positions $s_1,\ldots,s_5$ for $L = 34$ nm/100 bp. This seems at first sight to contradict equation (5), which states that both the intrinsic curvature *and* flexibility affect such conformational average. However, it should be observed that our theoretical predictions are entirely based on the WLC model assuming a sequence-independent flexibility. Therefore, the theoretical s vs $\langle COS_{s,L}\rangle$ pattern reflects local variations of the intrinsic curvature and the flexibility produces a bias of the absolute values. Thus, the local intrinsic curvature ultimately drives the modulation of the $\langle COS_{s,L}\rangle$ pattern. In Fig. 3 we explore further how the WLC
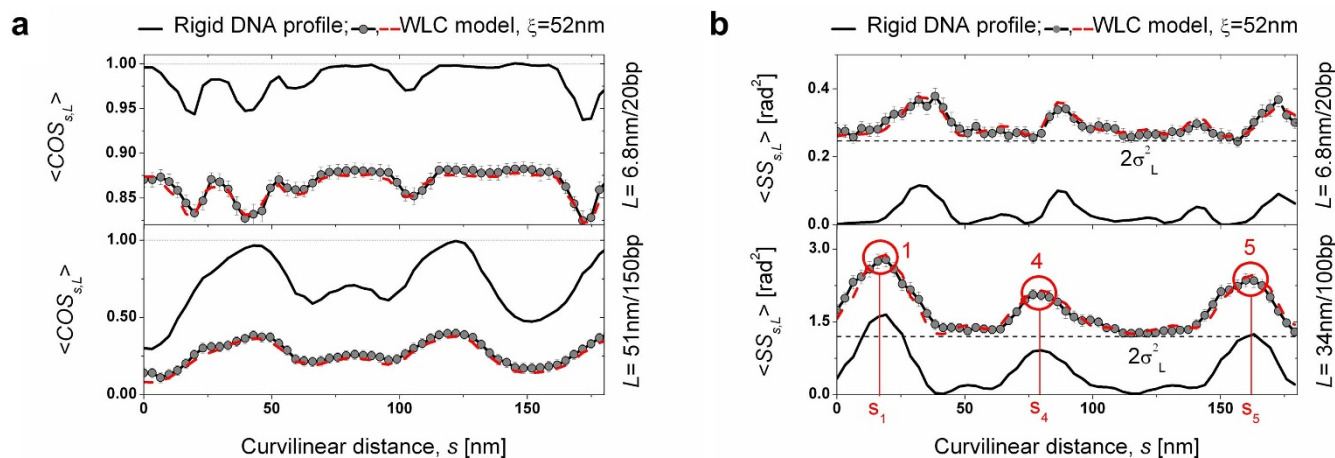
Figure 3 | Simulated conformational analysis for the 1332 bp duplex from the promoter region of the OPN coding gene: descriptors $\langle COS_{s,L} \rangle$, $\langle SS_{s,L} \rangle$.
(a) Patterns of variation of $\langle COS_{s,L} \rangle$. Results for an ideally-rigid chain are compared with WLC predictions to evaluate the role of flexibility at two different L values, i.e. 6.8 nm/20 bp (top panel) and 51 nm/150 bp (bottom panel). (b) Patterns of variation of $\langle SS_{s,L} \rangle$. Results for an ideally-rigid chain are contrasted with WLC predictions to show the flexibility contribution at two L values, i.e. 6.8 nm/20 bp (top panel) and 34 nm/100 bp (bottom panel). As a general remark, the WLC data with grey dots are obtained by direct ensemble averaging over simulated 2D chain trajectories, whereas the red, dash WLC line is obtained by approximation of the ensemble average with the Supplementary equation (S.10).

flexibility contributes to the predicted patterns of variation. In Fig. 3a we consider the s vs $\langle COS_{s,L} \rangle$ plots and contrast the reference pattern for an ideally-rigid DNA with the case of a flexible profile. We intentionally overdo comparison by introducing remarkably different contour lengths, i.e. $L = 6.8$ nm/20 bp and $L = 51$ nm/ 150 bp. For $L = 6.8$ nm/20 bp the pattern is dominated by variations of the local intrinsic curvature and the flexibility determines merely an overall, small vertical shift of the pattern of about 0.13. Indeed this is predicted by theory (see Supplementary equation (S.7) and the Supplementary Note) and corresponds to the r.h.s. term of equation (5) with $c^* \approx 0$. For $L = 51$ nm/150 bp the role of flexibility is more prominent and twofold, i.e. it determines a vertical shift of the s vs $\langle COS_{s,L} \rangle$ pattern with respect to the case of an ideally-rigid chain, and it also reduces the amplitude of the pattern modulation. As explained above, the flexibility affects the patterns uniformly along the curvilinear coordinate s in that we assumed a sequence-independent flexibility (curvature variance $\sigma_L^{*2} = L/\xi$, with $\xi = 52$ nm persistence length) according to the WLC model.

In Fig. 3b we consider the s vs $\langle SS_{s,L} \rangle$ case. The patterns for an ideally rigid chain are well structured, in particular for $L = 34$ nm/ 100 bp. one recognizes three peaks at the curvilinear coordinates s_1, s_4, s_5 that we already introduced in Fig. 2a for the s vs $\langle P_{s,L} \rangle$ and $\langle COS_{s,L} \rangle$ plots. Moreover, the WLC flexibility determines an overall, vertical shift by $2\sigma_L^{*2}$, where $\sigma_L^{*2} = L/\xi$. This agrees with the general predictions of equation (6). Thus, by analysing the descriptor $\langle SS_{s,L} \rangle$ one can easily decouple for any length L the intrinsic curvature – that tunes the modulation of $\langle SS_{s,L} \rangle$ along the curvilinear coordinate s – from the flexibility – that determines a shift of the whole pattern. We underline that in the present case $L = 6.8$ nm/201 bp is not strictly appropriate for evaluating the *experimental* patterns of variation, as these might be partially affected by AFM tip convolution effects. Therefore, data analysis is carried out below only for $L = 17$ nm(50 bp) and $L = 34$ nm(100 bp).

Experimental patterns of variation for the human OPN coding gene. In Fig. 4 we report the results from experiments on OPN-related DNA (see Methods section for details on sample preparation, AFM imaging and DNA tracing). A quantitative AFM analysis of molecular profiles was routinely performed to test the reproducibility of the imaging conditions and evaluate deviations

of adsorbed DNA superstructure from the canonical B-form. Typically, measured DNA molecules displayed an average width of ≈ 10 nm and a height of 0.8 nm–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]. Molecule surface density was in the range 2–5 μm^{-2}. The analysis of the contour lengths for a large number of traced molecules (≈ 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 the results of similar studies[2,9,11,22,30,31]. Standard checks on global statistical parameters (e.g. the mean-squared end-to-end distance) proved the thermodynamic equilibration of chains on mica and allowed us to estimate $\xi = 52$ nm, figure that agrees with the DNA flexibility reported in other AFM experiments[2,22,23]. In Fig. 4a we show a representative topography of the target 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 one can 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 suggests the presence of non-null intrinsic curvatures at the same places.

In Fig. 4b we show the experimental s vs $\langle P_{s,L} \rangle$ pattern with $L = 34$ nm/100 bp, calculated for an ensemble of 160 molecular profiles extracted from several AFM topographies. It displays oscillations that confirm the presence of intrinsic curvatures along the studied contours. We recognize four peaks for $s < 80$ nm - that concern pairs of segments close to the DNA contours ends - and an additional negative peak for $s \approx 150$ nm–175 nm, that on the contrary involves pairs of tracts located around the middle portion of the strands. In the range $s \approx 80$ nm–130 nm the s vs $\langle P_{s,L} \rangle$ curve is almost completely flat and $\langle P_{s,L} \rangle \approx 0$, which means that at least one of the two symmetrically placed segments has a negligible intrinsic curvature. Noteworthy, the curvilinear positions of the main peaks of $\langle P_{s,L} \rangle$ agree with the visual inspection of DNA bends from several AFM topographies (see for example Fig. 4a).

The comparison of the experimental s vs $\langle P_{s,L} \rangle$ pattern with the theoretical one of Fig. 2a can be effectively carried out by comparing the curvilinear positions and amplitudes of the main experimental peaks with the theoretical peaks located at $s_1, \ldots, s_5$. In detail, we recognize peak 1 also in the experimental data, biased by a small horizontal shift δs_1 of about 7 nm. Moreover, a 8 nm horizontal shift
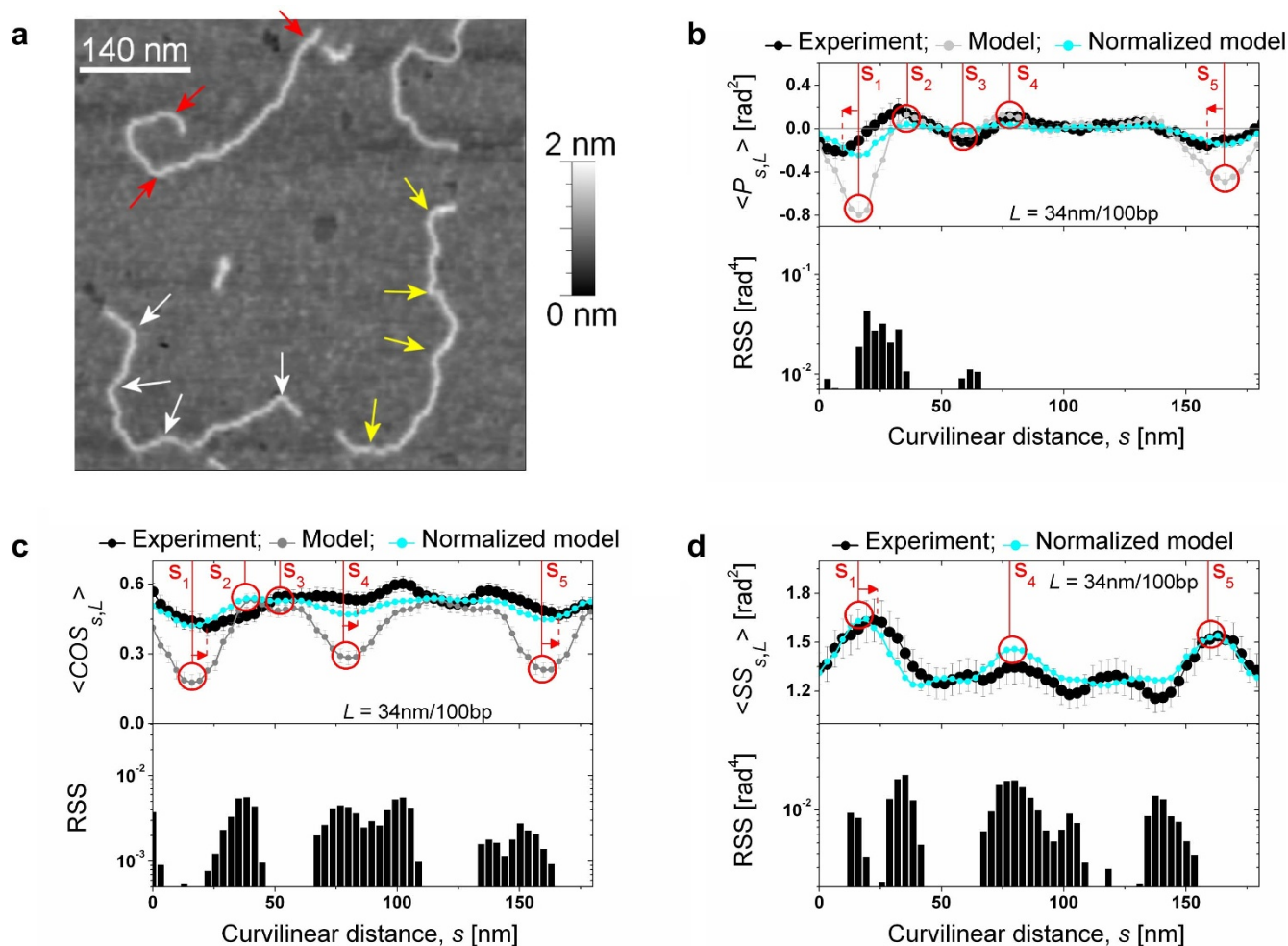
Figure 4 | Experimental conformational analysis of OPN-related DNA. (a) Representative AFM topography of the 1332 bp target DNA. It shows the persistence of bends at few locations along the molecular backbone - marked by arrows - suggesting the presence of a significant intrinsic curvature at the same places. (b) Pattern of variation of $\langle P_{s,L} \rangle$ compared with simulations (top panel). The related s vs RSS graph reveals that substantial agreement exist along the curvilinear axis, with a residual peak below 50 nm (bottom panel). (c) Pattern of variation of $\langle COS_{s,L} \rangle$ compared with simulations (top panel). The related s vs RSS graph reveals discrepancies at three distinct regions of the curvilinear axis (bottom panel). (d) Pattern of $\langle SS_{s,L} \rangle$ compared with simulations (top panel). The s vs RSS graph is very similar to that reported in (c) for the $\langle COS_{s,L} \rangle$ descriptor and suggests that the failure of sequence-independent flexibility very likely increases residual errors for $s > 50$ nm.

δs_5 affects peak 5 whereas 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. An appreciable difference exists between the magnitude of the experimental peaks and the theoretical ones. The protocol adopted for samples preparation is certainly contributing to such discrepancies. In particular the horizontal shift 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[24,25,27]. On the other side, 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 of 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 adsorbed molecules is therefore highly probable[25,27]. Zuccheri et al.[4] and Scipioni et al.[25] showed that the theoretical predictions of the wedge model of De Santis et al.[28] are in *quantitative* agreement with experiments provided that the theoretical curvature modulus is rescaled by the empirical numerical factor ≈2.5 (see Fig. 3 in[4] and

Fig. 7 in[25] for $L = 26$ bp and 62 bp respectively). We exploited this evidence to improve the agreement between theoretical and experimental patterns of the SCDs. We rescaled $C_{s,L}$ of equations (4)–(6) by the empirical factor ≈2.0 before calculating the ensemble averages. Fig. 4b attests that in doing this we substantially improved the agreement between theory and experiment for the case of the s vs $\langle P_{s,L} \rangle$ pattern. In fact, the experimental amplitudes of peaks 1 and 5 were correctly predicted after this additional normalization step. A quantitative measure of the amount of error between the normalized wedge model and experimental data was estimated by introducing the Residual Sum of Squares (RSS). The s vs RSS plot (Fig. 4b bottom panel) shows that the main discrepancy between experimental data and the normalized wedge model is ≤0.04 and concerns a nanosized region with $s \approx 14$ nm–37 nm, whereas the remaining part of the curvilinear axis is affected by comparatively smaller errors. In view of such results, we recognize that our analysis – based on the wedge model of De Santis et al.[28] with the additional amplitude normalization step – describes the relevant features of the s vs $\langle P_{s,L} \rangle$ plot with accuracy.

The next considered SCDs probe both DNA intrinsic curvature and flexibility. The experimental s vs $\langle COS_{s,L} \rangle$ pattern is shown in Fig. 4c for $L = 34$ nm/100 bp. It displays oscillations that indicate
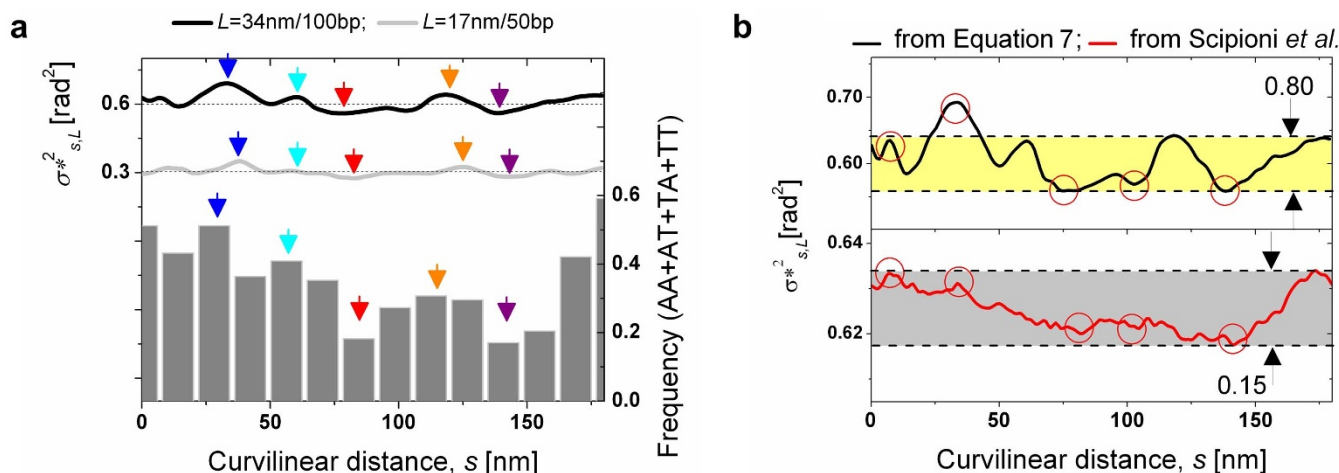
Figure 5 | Evaluation of differential flexibility of OPN-related DNA. (a) Curvature variance $\sigma^{*2}_{s,L}$ estimated from experimental data and theory using equation (7). The variance fluctuates around a constant value that depends on L and corresponds indeed to the WLC flexibility term L/ξ. The fluctuations are a robust manifestation of differential flexibility since they correlate with the profile of the cumulative local frequency of AA, TT, AT and TA dinucleotide steps. (b) Comparison of the experimental pattern of variation of $\sigma^{*2}_{s,L}$ and the theoretical one, calculated as in[26] for L = 34 nm/100 bp. Local maxima/minima at corresponding curvilinear positions are highlighted by circles. An enhancement of the experimental differential flexibility with respect to the theoretical predictions emerges and agrees with previous reports.

that intrinsic curvature and/or varying flexibility are playing a relevant role along the studied molecular contours. Comparison with the normalized model gives excellent agreement between the amplitudes of the main peaks and their relative positions ($\delta s_1 \approx \delta s_4 \approx \delta s_5 \approx 7$ nm, $\delta s_2 \approx \delta s_3 \leq 5$ nm). A more careful examination of the related s vs RSS plot reveals that residuals are localized within three distinct regions of the curvilinear axis. This contrasts with the s vs RSS curve of the $\langle P_{s,L} \rangle$ pattern and led us to ascribe different origins to the three regions. The first region ($s \approx 14$ nm–37 nm) in fact occurs also for the $\langle P_{s,L} \rangle$ pattern thus it reflects an inherent inability of the wedge model to predict the actual intrinsic curvature of the probed chains close to their ends. On the contrary the two other regions ($s \approx 75$ nm–100 nm and $s \approx 140$ nm–160 nm) are peculiar of the $\langle COS_{s,L} \rangle$ pattern and they likely reflect the local failure of the second main assumption of our theoretical analysis, i.e. the sequence-*independent* WLC flexibility. In other words, one has to assume that the nucleotide sequence of those two regions affects the persistence length ξ and causes appreciable fluctuations with respect to the mean value (52 nm). Similar effects have been reported for the well-known model system pBR322 DNA in AFM experiments on palyndromic dimers[4].

The experimental s vs $\langle SS_{s,L} \rangle$ pattern is compared with its theoretical counterpart for $L = 34$ nm/100 bp in Fig. 4d. The curvilinear positions and magnitude of the peaks 1,4 and 5 are again useful to drive the comparison. We find an excellent correspondence of the curvilinear positions of the experimental and theoretical peaks at s_1, s_4 and s_5 ($\delta s_1 \approx 7$ nm, $\delta s_4 \approx \delta s_5 \leq 5$ nm) and we observe full overlap of the patterns in the s ranges 0–18 nm and 145 nm–180 nm. However the two graphs have different trends and amplitudes on the remaining part of the curvilinear axis. This is evident when considering that the experimental maximum at $\sim$122 nm and minima at $\sim$100 nm and 140 nm do not have a clear correspondence in the theoretical curve. Importantly, the s vs RSS curve closely resembles the corresponding graph for the $\langle COS_{s,L} \rangle$ descriptor, with the residuals localized within the same three regions of the curvilinear axis. Again, the two regions for $s > 50$ nm very likely indicate that the assumption of the constant WLC flexibility breaks at those places.

To investigate the sequence-*dependent* flexibility of the OPN-related DNA we focused on the s vs $\langle SS_{s,L} \rangle$ pattern and we used the following equation:

$$\sigma^{*2}_{j,m} = \frac{\langle SS_{j,m} \rangle - \langle C_{j,m} \rangle^2 + \langle C_{N-1-(j+m),m} \rangle^2}{2}$$
$$= \frac{\langle SS_{j,m} \rangle - \langle SS_{j,m} \rangle_{Rigid}}{2} \tag{7}$$

The quantity $\sigma^{*2}_{j,m} \equiv 1/2 \cdot \left(\sigma^2_{C_{j,m}} + \sigma^2_{C_{N-1-(j+m),m}} \right)$ is the average curvature variance for a pair of tracts placed symmetrically along the studied contours. Equation (7) shows that $\sigma^{*2}_{s,L}$ can be obtained by subtracting from the experimental average $\langle SS_{s,L} \rangle$ the corresponding theoretical quantity calculated for an ideally-rigid molecular profile $\langle SS_{s,L} \rangle_{Rigid}$. The latter corresponds indeed to the profiles of Fig. 3b (solid lines) with the additional amplitude renormalization of intrinsic curvature by the factor 2.0 (see above). The results of a similar analysis are reported in Fig. 5a for $L = 34$ nm/100 bp and $L = 17$ nm/50 bp. We note that in both cases $\sigma^{*2}_{s,L}$ fluctuates around a constant value that depends on L and corresponds indeed to the WLC flexibility term L/ξ, i.e. $\sim$0.6 for $L = 34$ nm/100 bp and $\sim$0.33 for $L = 17$ nm/50 bp respectively. The fluctuations are in the range of 5%–20% of the WLC term and indicate that along the molecular backbone variations of the local flexibility are taking place. To probe that $\sigma^{*2}_{s,L}$ fluctuations are a robust manifestation of the chain flexibility - rather than the result of trivial statistical discrepancies between the theoretical and experimental patterns $\langle SS_{s,L} \rangle_{Rigid}$ and $\langle SS_{s,L} \rangle$ - we reported in the same graph the profile of the cumulative local frequency of AA, TT, AT and TA dinucleotide steps. The comparison shows that the modulation of $\sigma^{*2}_{s,L}$ follows very closely the variations in the cumulative content of dinucleotide steps, i.e. the local chain flexibility is modulated by the local content of the AA, AT, TA and TT dinucleotide steps. This finding agrees with the results originally reported for palindromic constructs in that it demonstrates the AT-rich regions are more flexible that GC-rich sequences[4]. Scipioni et al.[25] have explained such a correlation from the point of view of the thermodynamic stability of the DNA chain, i.e. they have connected the sequence-dependent flexibility of a DNA tract to the dinucleotide melting temperatures. Fig. 5b shows the comparison of the experimental pattern of variation of $\sigma^{*2}_{s,L}$ and the theoretical one, calculated as in[25] for L = 34 nm/100 bp. We recognize several similarities in the trends of the two graphs, with

local maxima/minima at corresponding curvilinear positions. Notably, the theory predicts an overall fluctuation of $\sim$0.15rad^2 whereas the experimental differential flexibility estimated by our analysis fluctuates by $\sim$0.80rad^2. An enhancement by a factor $\sim$3 of the experimental differential flexibility with respect to the theoretical predictions has been reported by others[4,25] therefore we are in line with previous reports.

Discussion

The reported results demonstrate that the oscillations of s vs $\langle P_{s,L} \rangle$, s vs $\langle COS_{s,L} \rangle$ and s vs $\langle SS_{s,L} \rangle$ curves can be used to locate the most significant bending sites of the DNA backbone. Moreover, the vertical shifts of the s vs $\langle COS_{s,L} \rangle$ - for short L values - and of s vs $\langle SS_{s,L} \rangle$ pattern - for any L value - can be exploited to investigate flexibility effects. We underline that informative patterns are obtained for DNA templates that do not contain extended strings of phased A-tracts, or other prominent nucleotide sequences responsible for large stereospecific bends. Apart from the system discussed above, we already demonstrated that random DNA ($\sim$25% content of A,T,G and C respectively) is still characterized by an appreciable (nonzero) intrinsic curvature over 20 bp–100 bp long fragments, and provides relevant signals in terms of the SCDs patterns[24]. The same conclusions hold for the 937 bp EcoRV-PstI fragment of pBR322 DNA[1,3,4,25,27], a well-known template discussed in the Supplementary Note and Supplementary Fig. S2. This confirms that the applicability of proposed method goes definitely beyond a few specific cases, and regards 10^2-10^3 bp long fragments that can be readily prepared and imaged at the nanoscale by various microscopy techniques. Fig. 4 and 5 give evidence of the rich information achieved by complementing the $\langle P_{s,L} \rangle$ and $\langle SS_{s,L} \rangle$ patterns. It is clear that the characterization of the differential flexibility in terms of s vs $\sigma^{*2}_{s,L}$ patterns (equation (7)) ultimately relies on the use of models to predict the sequence-dependent intrinsic curvature and to simulate DNA adsorption on the solid substrate. As shown above, one can preliminarily estimate the degree of accuracy of a given model by comparing the experimental $\langle P_{s,L} \rangle$ pattern with the theoretical one. In the present work and in[24] we have demonstrated that static dinucleotide wedge models are in good agreement with the experimental results. An improved analysis will certainly benefit of future developments to the models dealing with DNA sequence-dependent curvature and surface adsorption.

We finally note that there is a loss of information in the study of SCDs with respect to the case of s vs $\langle C_{s,L} \rangle$ (or s vs $\sigma^2_{C_{j,m}}$) data highlighted in previous works[3–5,25,27,32]. This loss arises from coupling pairs of DNA tracts into the definition of $P_{j,m}$, $COS_{j,m}$ and $SS_{j,m}$. Nevertheless this choice provides a number of advantages, overcoming some fundamental and practical limitations of early protocols. First, the novel method can be implemented on label-free molecules, therefore specimens preparation is merely reduced to standard protocols for DNA deposition onto atomically smooth substrates. Second, the patterns of variation of SCDs are prone to an effective comparison with theoretical models (used to predict the r.h.s. of equations (4),(5),(6)), that impart access to the physics of DNA sequence-dependent curvature and flexibility.

We foresee several challenging applications for the use of SCDs. One interesting possibility might regard the systematic use of $\langle P_{s,L} \rangle$ and $\langle SS_{s,L} \rangle$ maps to explore in detail the predictions of DNA adsorption and bending models. An insight into this topic is provided in the present work and significant improvements are expected to come from state of the art modelling (as Brownian dynamics and molecular dynamics simulations) going beyond the nearest-neighbour approximation in conformational analysis or describing the non-equilibrium processes of DNA adsorption and relaxation on the atomically flat substrate[23,33–38]. For example, a tight comparison of experimental and theoretical patterns might allow us to identify the presence of nanosized 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 $\langle P_{s,L} \rangle$ and $\langle SS_{s,L} \rangle$ patterns to detect automatically small conformational changes in large sample numbers. The capability of relating 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 such patterns to point mutations and mismatched base-pairs and largely contribute to the discovery of physical methodologies for molecular haplotyping[18,39]. Within this context, we already offered a concrete example on the $\langle P_{s,L} \rangle$ patterns sensitivity to single nucleotide polymorphisms in the OPN encoding gene[24]. Further attractive developments might come from the evaluation of $\langle P_{s,L} \rangle$ and $\langle SS_{s,L} \rangle$ patterns to address the structural properties of DNA fragments complexed with intercalating dyes and binding drugs[13,40] or even proteins. In fact the patterns of variation might be useful to complement current microscopy studies on the formation of protein-DNA complexes (e.g.[41]), where the position distribution of protein binding along unlabelled 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 SCDs definition. As a result, a visual correlation of s vs $\langle P_{s,L} \rangle$ (or $\langle SS_{s,L} \rangle$) and s vs protein-binding-frequency plots would easily point out the existence of helix sites where local intrinsic curvature (or flexibility) drives the so called 'indirect' DNA recognition or competes with other binding mechanisms[25,42]. 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[12]. One might also apply SCDs to support investigations on the interference of DNA-intercalating agents on DNA intrinsic curvature[43,44], on the enhancement of DNA flexibility by sequence non-specific DNA-binding proteins[45], and complement DNA investigations recently addressed by high-speed, real-time or hybrid AFM techniques[46–48].

In conclusion, we presented a novel method to characterize the intrinsic curvature and flexibility of adsorbed DNA molecules, starting from the topographies obtained by high-resolution microscopy. The method relies on mapping along the molecular backbone a selection of symmetric curvature descriptors, i.e. stochastic functions that couple pairs of segments symmetrically placed along the helix chain in a way that their realizations do not depend on chains polarity. We demonstrated the practical applicability of this approach for the relevant case of AFM-imaged DNA, through theoretical and experimental arguments. Our strategy provides a number of advantages overcoming some fundamental and practical limitations of early protocols, e.g. there is no need to prepare end-labelled molecules or palindromes and no assumptions are done on the preferential DNA adsorption mechanisms. More importantly, our approach can readily manage comparative assays involving a large number of samples. We suggested examples where curvature studies based on SCDs might complement several existing investigations.

Methods

Modeling DNA intrinsic curvature, adsorption & room-temperature bending. Model chains representing the average three-dimensional (3D) shape of DNA specimens were generated by the 3DNA software[49] exploiting nearest-neighbor, static dinucleotide wedge models[28] (see Supplementary Methods and Supplementary Fig. S3,S4).

We custom developed an algorithm (LabView, National Instruments) that flattens the 3D model chain to simulate deposition[24]. Briefly, it divides the chain into a discrete number of fragments originally lying on different planes and projects them individually. The output is a two-dimensional (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.[25].

The 2D chain formed by geometric projection was used to simulate the room temperature bending of DNA, describing chain lateral motion onto the mica surface. According to the WLC model, DNA can be modelled by a chain of virtual bonds of length l_{WLC} connected by torsional-spring vertices, that are energetically uncorrelated and characterized by a harmonic local bending-energy function $E\left(\theta_i^{th}\right) = 1/2 k_B T (\xi/l) \left(\theta_i^{th}\right)^2$ (with k_B Boltzmann constant, T absolute temperature, ξ persistence length and θ_i^{th} are thermally-induced angular fluctuations occurring around the constant sequence-dependent θ_i^0 angles)[2]. We sampled the 2D chain at the spacing $l_{WLC} = 0.32$ nm (corresponding to the experimentally found helix rise per base-pair, see Results section) and thermal effects (on bending) were implemented by adding to the angles among neighbour segments a fluctuation chosen by a Monte-Carlo (MC) method from normally distributed numbers with mean zero and variance of l_{WLC}/ξ. The new trajectories were superimposed on a randomly flat substrate (roughness 0.1 nm, grid spacing 1.95 nm) and dilated by a parabolic tip[50] in order to generate topographies resembling as close as possible those obtained by AFM. These were finally analysed with the tracing algorithm (see below) to assure a bias - due to random and systematic angular distortions - comparable to that affecting experimental data.

We used the procedure above to predict the theoretical patterns of variations reported in Fig. 2,3 and Supplementary Fig. S2.

Sample preparation. The 1332 bp DNA fragments were obtained by PCR amplification of the regulatory region of the OPN encoding gene[51]; 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. Importantly, this template does not contain extended strings of phased A-tracts or other prominent sequences (e.g. periodic A_n/T_n groups) that could introduce anomalously large bends in the adhered DNA molecules and bias our proof-of-concept investigation (see Supplementary Methods and Supplementary Fig. S3 for the base-pairs sequence)[3–5,9]. The DNA adsorption was carried out onto freshly cleaved muscovite mica according to the standard protocols reported in literature[22].

AFM imaging & 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 μm.

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 and molecules were represented as chains of xy pairs separated by a contour length $l = 2$ nm. Briefly, the AFM images were processed by a first-order, line by line flattening followed by filtering with a 3×3 median pixel filter[30,50]. Chain tracing was initiated at a set of user-defined trial points, that were located along the molecular backbone by hand and linearly interpolated to obtain a constant spacing. The coordinates of each point were then automatically adjusted as explained in[23]. After this transformation the trajectory consisted in a series of points with non-integer coordinates on the digitized grid. The final processing step consisted in the trajectory interpolation with a cubic polynomial curve, with square error below a user-defined threshold (typically 0.1 nm). 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 quantities $\langle P_{j,m} \rangle$, $\langle COS_{j,m} \rangle$, $\langle SS_{j,m} \rangle$ were evaluated respectively from the conformational average of $C_{j,m} \cdot C_{N-1-(j+m),m}$, $\cos C_{j,m} \cdot \cos C_{N-1-(j+m),m}$ and $C_{j,m}^2 + C_{N-1-(j+m),m}^2$ over a given set of AFM imaged molecular profiles.

We preliminarily tested the tracing algorithm by analysis of intrinsically *bent* 2D chains that were computer-generated using WLC statistics (see Supplementary Methods and Supplementary Fig. S5).

1. Muzard, G., Théveny, B. & Révet, B. Electron microscopy mapping of pBR322 DNA curvature. Comparison with theoretical models. *EMBO J.* **9**, 1289–1298 (1990).

2. Rivetti, C., Walker, C. & Bustamante, C. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* **280**, 41–59 (1998).

3. Cognet, J. A. H., Pakleza, C., Cherny, D., Delain, E. & Le Cam, E. Static curvature and flexibility measurements with microscopy. A simple renormalization method, its assessment by experiment and simulation. *J. Mol. Biol.* **285**, 997–1009 (1999).

4. Zuccheri, G. *et al.* Mapping the intrinsic curvature and flexibility along the DNA chain. *Proc. Natl. Acad. Sci. USA* **98**, 3074–3079 (2001).

5. Marilley, M., Sanchez-Sevilla, A. & Rocca-Serra, J. Fine mapping of inherent flexibility variation along DNA molecules. Validation by atomic force microscopy (AFM) in buffer. *Mol. Gen. Genomics* **274**, 658–670 (2005).

6. Moukhtar, J., Fontaine, E., Faivre-Moskalenko, C. & Arneodo, A. Probing the persistence in DNA curvature properties by atomic force microscopy. *Phys. Rev. Lett.* **98**, 178101–1/4 (2007).

7. Faas, F. G. A., Rieger, B., van Vliet, L. J. & Cherny, D. I. DNA Deformations near Charged Surfaces: Electron and Atomic Force Microscopy Views. *Biophys. J.* **97**, 1148–1157 (2009).

8. Moukhtar, J. *et al.* Effect of Genomic Long-Range Correlations on DNA Persistence Length: From Theory to Single Molecule Experiments. *J. Phys. Chem. B* **114**, 5125–5143 (2010).

9. Moreno-Herrero, F., Seidel, R., Johnson, S. M., Fire, A. & Dekker, N. H. Structural analysis of hyperperiodic DNA from Caernorhabditis elegans. *Nucleic Acids Res.* **34**, 3057–3066 (2006).

10. Marilley, M., Milani, P. & Rocca-Serra, J. Gradual melting of replication origin (Schizosaccharomyces pombe ars1): in situ atomic force microscopy (AFM) analysis. *Biochimie* **89**, 534–541 (2007).

11. Dame, R. T. *et al.* Analysis of scanning force microscopy images for protein-induced DNA bending using simulations. *Nucleic Acids Res.* **33**, e68/1–7 (2005).

12. Gorman, J. *et al.* Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. *Mol. Cell* **28**, 359–370 (2007).

13. Adamcik, J., Valle, F., Witz, G., Rechendorff, K. & Dietler, G. 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 (2008).

14. Marilley, M., Milani, P., Thimonier, J., Rocca-Serra, J. & Baldacci, G. 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 (2007).

15. Garcia, H. G. *et al.* Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers* **85**, 115–128 (2006).

16. De Santis, P. & Scipioni, A. Sequence-dependent collective properties of DNAs and their role in biological systems. *Physics of Life Reviews* **10**, 41–67 (2013).

17. Fang, Y. *et al.* Solid-state DNA sizing by atomic force microscopy. *Anal. Chem.* **70**, 2123–2129 (1998).

18. Wooley, A. T., Guillemette, C., Li Cheung, C., Housman, D. E. & Lieber, C. M. Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat. Biotech.* **18**, 760–763 (2000).

19. Reed, J. *et al.* Single molecule transcription profiling with AFM. *Nanotechnology* **18**, 044032–044046 (2007).

20. Billingsley, D. J., Bonass, W. A., Crampton, N., Kirkham, J. & Thomson, N. H. Single molecule studies of DNA transcription using atomic force microscopy. *Phys. Biol.* **9**, 021001 (2012).

21. Angeli, E. *et al.* Nanotechnology applications in medicine. *Tumori* **94**, 206–215 (2008).

22. Rivetti, C., Guthold, M. & Bustamante, C. Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration versus Kinetic Trapping Studied by Statistical Polymer Chain Analysis. *J. Mol. Biol.* **264**, 919–932 (1996).

23. Wiggins, P. A. *et al.* High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotech.* **1**, 137–141 (2006).

24. Buzio, R., Repetto, L., Giacopelli, F., Ravazzolo, R. & Valbusa, U. Label-free, atomic force microscopy-based mapping of DNA intrinsic curvature for the nanoscale comparative analysis of bent duplexes. *Nucleic Acids Res.* **40**, e84 (2012).

25. Scipioni, A., Anselmi, C., Zuccheri, G., Samori, B. & De Santis, P. Sequence-dependent DNA curvature and flexibility from scanning force microscopy images. *Biophys. J.* **83**, 2408–2418 (2002).

26. Crothers, D. M. DNA curvature and deformation in protein–DNA complexes: a step in the right direction. *Proc. Natl. Acad. Sci. USA* **95**, 15163–15165 (1998).

27. Sampaolese, B. *et al.* Recognition of the DNA sequence by an inorganic crystal surface. *Proc. Natl. Acad. Sci. USA* **99**, 13566–13570 (2002).

28. De Santis, P., Palleschi, A., Savino, M. & Scipioni, A. A theoretical model of DNA curvature. *Biophys. Chem.* **32**, 305–317 (1988).

29. Ebeling, D., Holscher, H., Fuchs, H., Anczykowski, B. & Schwarz, U. Imaging of biomaterials in liquids: a comparison between conventional and Q-controlled amplitude modulation ('tapping mode') atomic force microscopy. *Nanotechnology* **17**, S221–S226 (2006).

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

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

32. Ficarra, E. *et al.* Automated intrinsic DNA curvature computation from AFM images. *IEEE Trans. Biomed. Eng.* **52**, 2074–2085 (2005).

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

34. Semenov, A. N. Adsorption of a semiflexible wormlike chain. *Euro. Phys. J. E* **9**, 353–363 (2002).

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

36. Lavery, R. *et al.* 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 (2010).

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

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

39. Kwok, P. Y. & Chen, X. Detection of Single Nucleotide Polymorphisms. *Curr. Issues Mol. Biol.* **5**, 43–60 (2003).

40. Günther, K., Mertig, M. & Seidel, R. Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Res.* **38**, doi:10.1093/nar/gkq434 (2010).

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

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

43. Winzer, A. T., Kraft, C., Bhushan, S., Stepanenko, V. & Tessmer, I. (2012) Correcting for AFM tip induced topography convolutions in protein–DNA samples. *Ultramicroscopy* **121**, 8–15 (2012).

44. Tan, H. K. *et al.* Interference of intrinsic curvature of DNA by DNA-intercalating agents. *Org. Biomol. Chem.* **10**, 2227–2230 (2012).

45. Zhang, J., McCauley, M. J., Maher III, L. J., Williams, M. C. & Israeloff, N. E. Mechanism of DNA flexibility enhancement by HMGB proteins. *Nucleic Acids Res.* **37**, 1107–1114 (2009).

46. Kobayashi, M., Sumitomo, K. & Torimitsu, K. Real-time imaging of DNA–streptavidin complex formation in solution using a high-speed atomic force microscope. *Ultramicroscopy* **107**, 184–190 (2007).

47. Sanchez, H., Kanaar, R. & Wyman, C. Molecular recognition of DNA–protein complexes: A straightforward method combining scanning force and fluorescence microscopy. *Ultramicroscopy* **110**, 844–851 (2010).

48. Fronczek, D. N. *et al.* High accuracy FIONA–AFM hybrid imaging. *Ultramicroscopy* **111**, 350–355 (2011).

49. Lu, X. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).

50. Horcas, I. *et al.* WSXM: a software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.* **78**, 013705–013713 (2007).

51. Giacopelli, F. *et al.* Polymorphisms in the osteopontin promoter affect its transcriptional activity. *Physiol. Genomics* **20**, 87–96 (2004).

Acknowledgments

Author contributions

U.V. and R.R. supervised the project. R.B. and L.R. developed the method based on symmetric curvature descriptors and ran simulations. F.G. prepared the DNA samples. R.B. carried out AFM experiments. R.B. and L.R. performed data analysis. All authors wrote the manuscript.

Additional information