Atomic force microscopy of DNA in solution and DNA modelling show that structural properties specify the eukaryotic replication initiation site

Monique Marilley[1],*, Pascale Milani[1], Jean Thimonier[1], José Rocca-Serra[1] and Giuseppe Baldacci[2]

[1]Régulation génique et fonctionnelle & microscopie champ proche, EA 3290, IFR 125, Faculté de Médecine, Université de la Méditerranée, 27 Bd Jean Moulin, 13385 Marseille cedex 5 and [2]Genotoxicologie et cycle cellulaire, UMR 2027, Institut Curie-CNRS, Centre universitaire, Bât. 110, 91405 Orsay, France

ABSTRACT

The replication origins (ORIs) of *Schizosaccharomyces pombe*, like those in most eukaryotes, are long chromosomal regions localized within A+T-rich domains. Although there is no consensus sequence, the interacting proteins are strongly conserved, suggesting that DNA structure is important for ORI function. We used atomic force microscopy in solution and DNA modelling to study the structural properties of the Sp*ars1* origin. We show that this segment is the least stable of the surrounding DNA (9 kb), and contains regions of intrinsically bent elements (strongly curved and inherently supercoiled DNAs). The pORC-binding site co-maps with a superhelical DNA region, where the spatial arrangement of adenine/thymine stretches may provide the binding substrate. The replication initiation site (RIP) is located within a strongly curved DNA region. On pORC unwinding, this site shifts towards the apex of the curvature, thus potentiating DNA melting there. Our model is entirely consistent with the sequence variability, large size and A+T-richness of ORIs, and also accounts for the multistep nature of the initiation process, the specificity of pORC-binding site(s), and the specific location of RIP. We show that the particular DNA features and dynamic properties identified in Sp*ars1* are present in other eukaryotic origins.

INTRODUCTION

In eukaryotes, DNA replication seems to initiate at discrete sites in the genome (1–4). Although they do not show sequence conservation, particular DNA regions—that may contain either a single site or zones of potential initiation sites—are favoured for the initiation of replication during each division cycle. The absence of sequence conservation between origins raises the question of the nature of the specificity.

The best-studied models are the replication origins from *Saccharomyces cerevisiae* where a consensus sequence of 11 nt has been identified (5,6). Except for this short sequence, the origins of *S. cerevisiae* differ in nucleotide sequence. However, it has been proposed, on the grounds of a mathematical analysis, that these regions may possess a common structural organization. Indeed, mutations that decrease or abolish the replication activity affect the higher order structural organization of DNA in regions flanking the ARS consensus sequence. Conversely, reconstitution of the structure coincides with a restoration of the replication activity (7).

In the absence of sequence similarity between the origins of *S. pombe* and higher eukaryotes, these findings justify a search for common structures. In contrast to *S. cerevisiae*, *S. pombe* origins are long DNA regions. They also appear more complex. The only specific feature found so far is that they are all within A + T-rich regions of the genome. Because they show characteristics of metazoan replicators, fission yeast origins are particularly useful as a model for origin function in higher eukaryotes (8).

Mapping of the Sp*ars1* replication origin at nucleotide resolution has shown that replication initiates at a defined position, the replication initiation site (RI) or (RIP). This site is immediately upstream from the *hus5*[+] promoter (4).

Investigations of the proteins that bind to replication origins show that a preinitiation complex (PreR) must first be formed. Building this complex involves the association of a six-subunit-protein complex (pORC) at a specific site on the DNA. It is likely that the formation of this PreR complex leads to structural changes in the DNA (9), which enable or facilitate the opening of the DNA double helix at the initiation site and, in so doing, allow the

---

*To whom correspondence should be addressed. Tel/Fax: + 33 (0)4 91 79 48 60; Email: monique.marilley@medecine.univ-mrs.fr

recruitment of the protein sets necessary for starting replication.

Consequently, the binding of the pORC complex has received considerable attention. These proteins are remarkably well conserved between species, but the nucleotide sequence of their binding site on the DNA is not conserved. There is also evidence that DNA topology is involved in the binding of the complex (10).

Various pORC-binding sites have been described by diverse approaches including DNaseI protection (11) or AFM imaging of the associated complex in air (9). The position of the pORC-binding site depends on only one subunit of the complex, spOrc4p (9,12). Numerous DNA-binding proteins and the members of the HMGI(Y) family contain AT-hook domains. This is also the case for spOrc4p, which possesses nine copies of the AT-hook amino acid consensus motif in its N-terminal part (13). There is controversy concerning whether the binding specificity depends on the AT-richness of the site or on specific motifs. In this latter case, it is proposed that either AT-hook nucleotide motifs [AAA(T/A)] (14) or optimal binding sites [AA(T/A)T] (11) are determinant, although these various motifs are frequent in the sequences analysed.

These analyses have not identified the signals on which this specificity depends. We therefore performed a structural analysis of the Sp*ars1* origin using two complementary approaches (DNA modelling and AFM imaging in liquid). The first method indicates which regions should be studied by revealing possible structural features; the second allows their analysis under near physiological conditions.

We initially focussed our attention on the thermal stability of the region, and then studied inherent properties of curvature of the Sp*ars1* DNA. We unambiguously demonstrate a very specific structural organization linked to the sequence. This structure is not unique to Sp*ars1* and is found in other eukaryotic origins, for example the origins of Sp*ars2004* and of the chorion gene locus of *Drosophila*. This study provides insight into the functional properties of these conserved structural elements.

MATERIALS AND METHODS

DNA sample preparation

The starting material was pART1 (7258 bp). The complete Sp*ars1* fragment (1204 bp) corresponding to nucleotides 4360–3157 in GenBank entry Z67961, was obtained by EcoRI digestion. The fragment was further restricted with MluI, and the remaining EcoRI end filled with *E. coli* polymerase using dNTPs and biotinylated dUTP. The resulting 1037-bp fragment was purified using a Sephaglass kit (Amersham Pharmacia Biotech). The DNA concentration was determined by absorbance measurement at 260 nm and the DNA samples stored at 4°C.

DNA–streptavidin interaction

The complex was obtained by incubating the EcoRI–MluI fragment with streptavidin (2:1 ratio with the DNA) in TE (10 mM Tris–HCl pH 7.9, 1 mM EDTA) buffer for 30 min at room temperature.

Atomic force microscopy

DNA deposition on mica and imaging in liquid using a Nanoscope IIIa microscope (Digital Instruments, Santa Barbara, CA, USA) were as previously described in Ref. (15).

For 'kinetics' studies, the scanning was limited to the 128 lines covering the object. Trace images were registered and the direction of scanning was from top to bottom.

Samples containing about 120 molecules were sufficient to obtain low fluctuation of the calculated mean of curvature, so for these experiments, samples of 150 molecules were analysed. Kinetics analysis was performed by the study of 10 series of 8–30 scans registered on the same molecule. Samples of 250 molecules were also studied in ambient air for comparison. The window size was 200 or 150 bp.

Image treatment and results. Image treatment and analysis with the 'Scanning Adventure' software were as published (16). The results were used to plot graphs of the two-dimensional (2D) trajectory and the corresponding topographic data of the molecules.

Computer analysis.
DNA trajectory. The DNA path was calculated as the result of one translation and three rotations (roll, tilt and twist) at each dinucleotide step using Bolshoy values (17). This set of parameters has been previously observed to fit well with experimental results from AFM analysis in liquid (16,18).

DNA duplex stability. The thermodynamic libraries that characterize all 10 Watson–Crick nearest-neighbour interactions in DNA provide an empirical basis for predicting the stability ($\triangle G$) of any DNA duplex region from its primary sequence. The local energy required for the strand separation of DNA segments of determined length was calculated. Each value plotted takes into account the contribution of the surrounding nucleotides. All calculations were carried out with a 1 bp step and parameters corresponding to 1 M NaCl, 25°C and pH 7 using the improved thermodynamic values from (19). The program PACS DNA, that we developed and used in the previous studies, was used for calculations (20–22).

Superhelix detection. We used our computer modelling software that detects the left-handed and the right-handed superhelical organization determined by the nucleotide sequence of a DNA molecule (7). Using this method the superhelical structure is revealed by calculating the DNA trajectory for the periodicity at which phasing of the various bent elements is obtained. When correctly phased, the intrinsically bent elements make the DNA double-helix path curve within one plane. The coordinates of the artificial trajectory of the axis of this helix are used to calculate the ENDS ratio (i.e. the ratio of the curvilinear length to the distance between the two ends). Next, using a sliding window we plot a curvature map. We either show
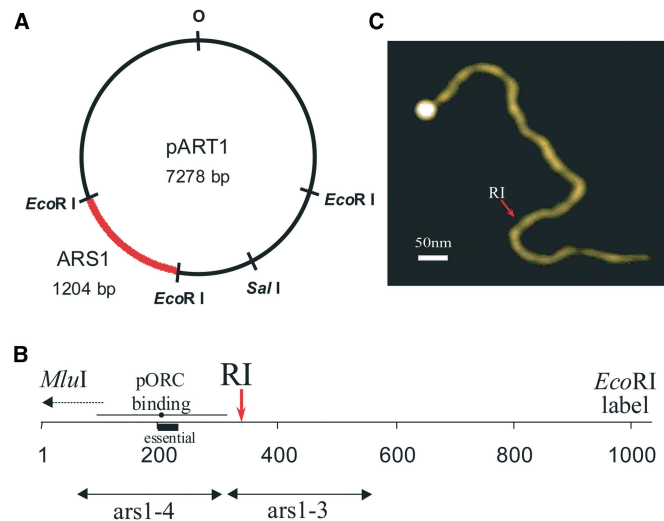
Figure 1. (**A**) Restriction sites in pART1. (**B**) The EcoRI–EcoRI 1204 bp insert was digested with MluI and labelled. Positions of segments *ars1-4* and *ars1-3* according to Ref. (11) are indicated. The region reported as essential for ARS activity by Ref. (28) is also indicated. RI is the site of initiation of replication described in Ref. (4). The pORC binding position according to Ref. (9) is indicated. The dotted arrow shows the direction of transcription. (**C**) AFM imaging in liquid of a labelled molecule and localization of the RI site.

a graphical representation limited to this simple curvature map or show a systematic scan of the different results preceding and following the phasing value (spectral analysis).

We refer to the phasing pitch value as 'phasing', and we express this value as a number of base pairs per turn. These DNA maps can be used for comprehensive comparison of the supercoiled configuration of various nucleotide sequences. In simple terms, a phasing pitch value higher than 10.4 bp indicates a left-handed three-dimensional (3D) organization, a phasing of near 10.4 bp means that the curvature is roughly planar, and a value below 10.4 bp corresponds to a right-handed 3D organization.

RESULTS

Structural analysis of S*ars1* DNA

The S*ars1* fragment (1204 bp) was obtained by EcoRI digestion of pART1 (7278 bp) and further digested with MluI for DNA orientation by labelling of the EcoRI end (Figure 1). ORI regions in *S. pombe* map in A + T-rich islands (23). Thus, A + T-richness may facilitate strand separation at the origin, but it may also generate some particular base composition required for the formation of various structural features. We, therefore, examined first the thermodynamic stability of the region, and then analysed the different structures featuring S*ars1* DNA. Our studies include both a predictive and an experimental approach.

The Spars1 region has much lower thermodynamic stability than surrounding regions. The energy required for DNA melting at a replication origin is dependent on the double-helix stability. The arrangement of the nucleotides
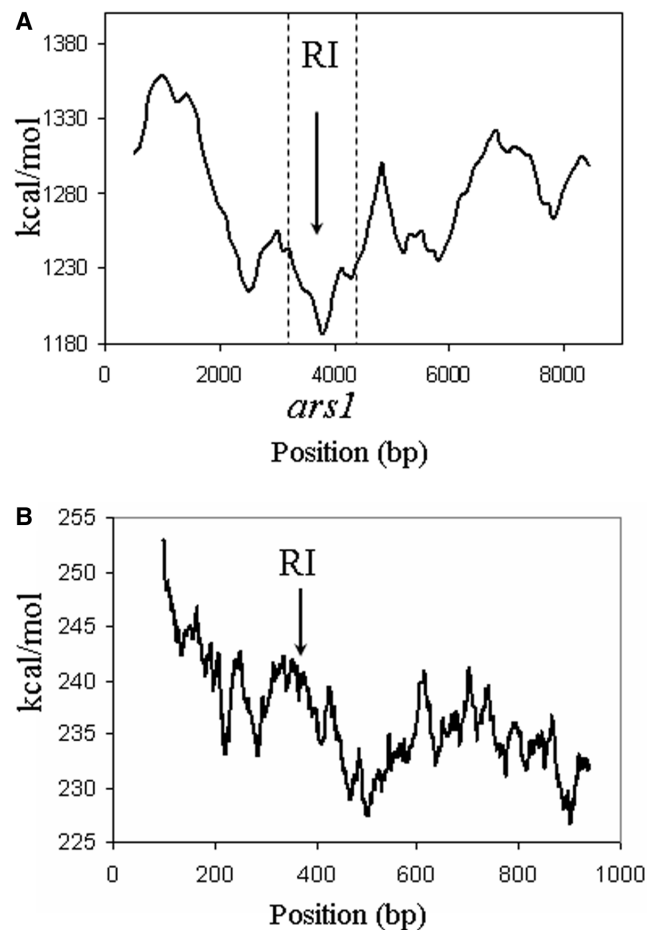
Figure 2. Thermodynamic stability within S*ars1* and surrounding nucleotides. (**A**) DeltaG variation profile of the 9000 nt surrounding the *ars1* fragment. The minimum values coincide with the *ars1* fragment. The window is 1000 bp, and each step is 10 bp. (**B**) DeltaG variation profile within the *ars1* fragment. RI is the replication initiation site. The window is 200 bp, and each step is 1 bp. Calculations were made using the thermodynamic library from (19).

is as important as their composition to stability of the DNA duplex. Thermodynamic libraries, that characterize all ten Watson–Crick nearest-neighbour interactions in DNA, provide an empirical basis for predicting the stability ($\triangle G$). Using the improved parameters from (19) and taking S*ars1* as the approximate centre, we calculated the local energy required for strand separation of 1000 bp sliding windows of DNA over a region of 9000 nt [the window size was chosen for allowing a comparison with published A + T-island data (23)]. Results are shown in Figure 2A. The A + T-islands coincide with regions of decreasing stability and, noticeably, we can see in the figure that the thermodynamic stability sharply decreases to a minimum at the precise position of S*ars1*.

Stability of the Spars1 fragment and position of the RI site. The RI site is where the initial double-strand aperture occurs *in vivo* (4). We repeated the thermodynamic analysis of S*ars1*, using a smaller window size, and found minimal thermal stability in a region near the RI site; the RI site itself did not correspond to the minimum (Figure 2B).
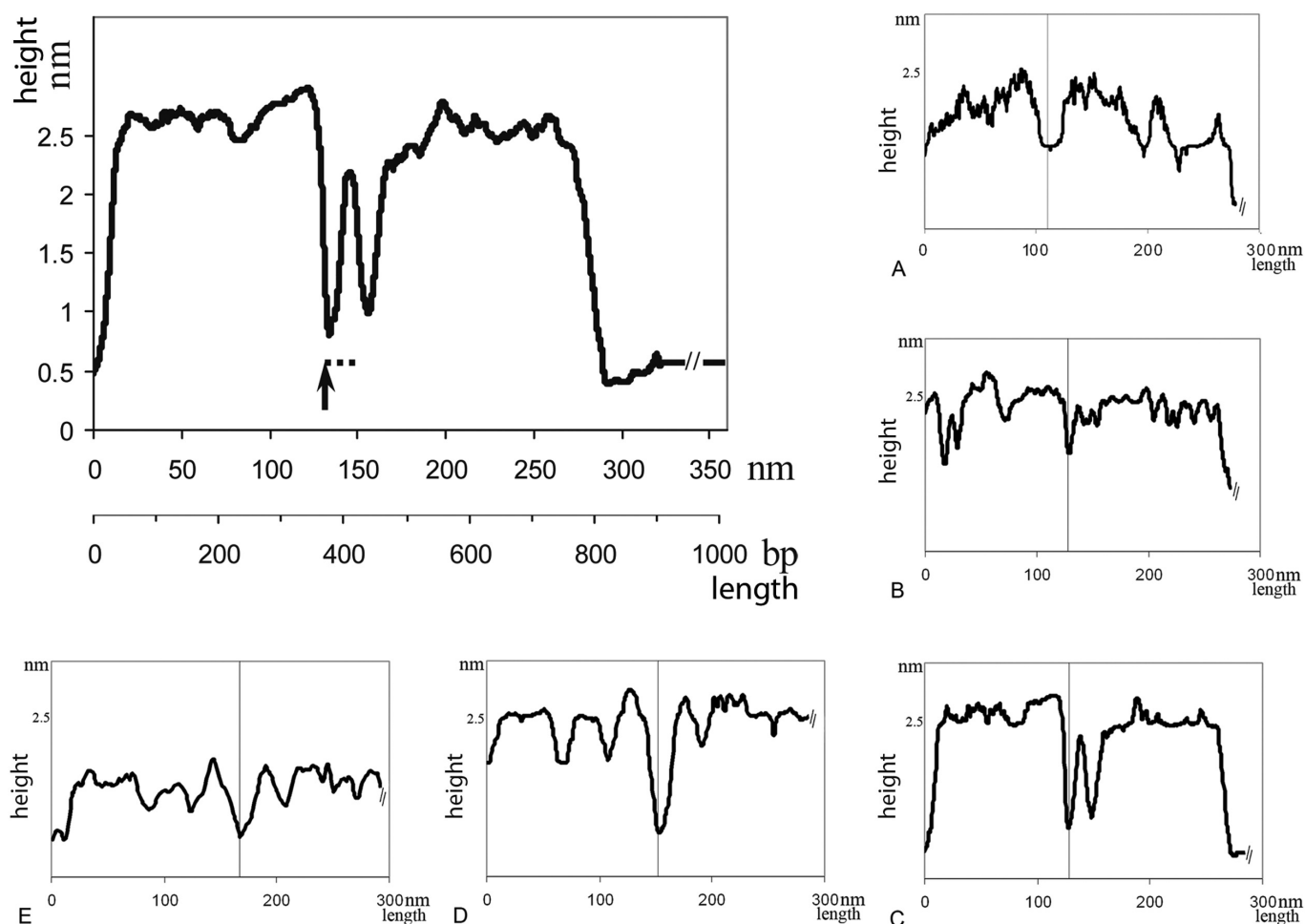
Figure 3. AFM topographical analysis of Sp*ars1* MluI–EcoRI fragment. Gradual melting by progressively increasing temperature inside the liquid cell allows the visualization of the opening of the region containing the RI site (passage from double-stranded to single-stranded DNA can be followed by the change in height of the DNA molecule). Due to possible errors in the measure caused by simultaneous denaturation at the MluI end, positions are indicated in nanometre, and base-pair correspondence is given tentatively. The initial site of the double helix opening in this region is shown by an arrow, and the range of imprecision is indicated by a dotted line. A, B, C, D, E show the correlation between position of unwinding DNA and localization of the apex of curvature (vertical line). The positions are given as measured from the MluI end (which starts to melt after the EcoRI end). Measured positions are affected by the extent of unwinding at the MluI extremity (see D and E). C corresponds to the molecule described in Supplementary Data.

Melting temperature prediction has proven to give reliable results, but all available libraries rely on the thermodynamics in solution of short pieces of DNA (∼20 bp), and thus can only approximate the true melting temperature of long DNA fragments. Furthermore, very little is known concerning DNA that is not free in solution but is in interaction with supramolecular structures.

To alleviate the absence of information, we used AFM imaging in liquid buffer to study *in situ* the partial denaturation of the Sp*ars1* fragment deposited on a mica surface. This led us to develop the following approach: an experimental system involving a very thin thermocouple glued on the sample holder (24) has been used to measure the temperature–time evolution inside the liquid cell of the microscope in the absence of thermostating and revealed a gradual increase of temperature. Because of the A+T-richness of the Sp*ars1* DNA and the low-salt conditions chosen for AFM imaging, this allowed

us to perform a partial denaturation *in situ*. The very slow increase in temperature inside the liquid cell (∼3°/h) allowed us to follow the thermal denaturation of the Sp*ars1* origin of replication on the mica surface from the very beginning, until the RI region had unwound. By taking one image every 30–40 s, a complete kinetic study of the phenomenon was possible [Supplementary Data S1, and (15)].

Double-stranded DNA is thicker than single-stranded DNA, so melted and duplex regions can be distinguished on topographic maps of the Sp*ars1* images (MluI–EcoRI fragments). We show the map corresponding to partial denaturation when the RI region just opened (left, upper part of Figure 3). Because there is no reference point from which to make exact measurements (the two ends initiated melting when the central region containing the RI site was seen to unwind) the base-pair positions of the two central melted regions cannot be determined exactly.

However, by measuring from the MluI extremity, which is just starting to be denatured, we localized the RI site within the melted region (indicated by an arrow in the figure) or, in the case of an underestimation of the melted region at the MluI end, within the bridge of ∼25 bp that joins the two melted regions. In all cases, the experimental results show that the RI site is very poorly stable.

Of particular interest is the finding that the site of aperture is located at the apex of a strongly bent DNA (see Supplementary Data S1). Although acquiring data for the localization of position of the early melting regions was generally problematic because only a small proportion of scans coincide with the moment when the temperature is appropriate, we found that early melting of the internal region correlates with the apex of a curvature (Figure 3A–E).

These observations indicate that this site is liable to open. This result is not completely unexpected because Brownian Dynamic simulations have shown that sites susceptible to denaturation tend to be located at apices. Consistent with our observations, this indicates that base sequence determines the location of strand separation, both through the energetics of strand interactions and by influencing the geometry of supercoiling (25).

We next looked for possible conformational features within the Sp*ars1* region and surrounding DNA.

Presence of a strong curvature. Using Bolshoy values (17), the DNA path of the Sp*ars1* fragment was calculated as the result of one translation and three rotations (roll, tilt and twist) at each base step (Figure 4A). The RI site (365–368 bp) was mapped in the intrinsically curved region nearly 80 bp from the apex.

AFM imaging in liquid was used to confirm this strong curvature (Figure 4B): curvature was observed centred at about position 305 bp, thus, in non-denaturing conditions, the RI is located ∼60 bp from the apex of this structure (i.e. at a short distance, but not at the apex of the curvature). The presence of this strong curvature within the Sp*ars1* DNA was also confirmed by gel retardation electrophoresis (data not shown).

Presence of a supercoiled organization (theoretical and experimental approaches). Superhelical DNA is formed by tandemly repeated, intrinsically bent DNA fragments. Curved sequences shorter than 10.4 bp (helical repeat of DNA) do not allow a complete helical twist per repeated element, so the result is a left-handed superhelix; longer curved sequences result in right-handed superhelices. This property has been confirmed and clearly illustrated by cryoelectronmicroscopy of experimental molecules (26).

We used predictive and experimental approaches to investigate the supercoiling properties of the Sp*ars1* fragment.

Computer analysis

We used the approach described in Ref. (7). In this case, computer modelling of DNA trajectory can predict the superhelical conformation and its handedness. Indeed, the periodic positioning of a bend once per helical turn (or multiples of helical turns) causes the molecule
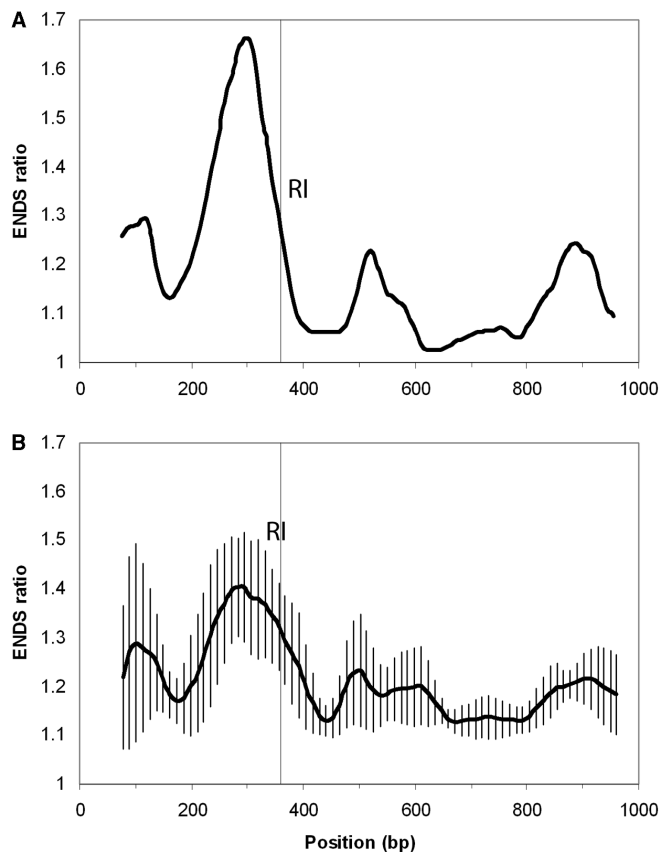
Figure 4. (**A**) Curvature map along Sp*ars1* MluI–EcoRI fragment. A strong curvature is predicted. Bolshoy dinucleotide parameters (17) were used for DNA modelling. The position of the RI site is indicated. The window is 150 bp, and each step 1 bp. (**B**) AFM imaging in liquid and Sp*ars1* trajectory analysis. ENDS ratio variation profile. Window size is 150 bp. Mean value and SD are indicated. A strong curvature is clearly visible. Position of the RI site is indicated.

to curve systematically in a plane, but bending with greater or lesser periodicity introduces a spiral into the trajectory. This type of organization was previously described in Ref. (27) by analysis of the periodicity of oligo(dA) or oligo(dT)stretches. However, because the contribution of the other nucleotides was ignored, the superhelical organization cannot always be clearly determined. In contrast, if the DNA trajectory is calculated using the 16 dinucleotide steps and the phasing is monitored by variation in the sequence period, supercoiled DNA is easily detected, since with the proper phasing the molecule is seen as plain curved DNA. This powerful approach allows comparisons of different DNA molecules (or comparison of the same molecule at different degrees of torsion). Moreover, the calculation of the 'phasing pitch' value can be used to determine the handedness of the supercoiled structure. Indeed, in order to obtain a planar shape these helices must be subject to a counteracting torsion. As a consequence, when the 'phasing pitch' that we calculate is found to be higher than 10.4 bp it indicates a left-handed superhelix, but it shows a right-handed superhelix when the 'phasing' is lower than 10.4 bp.

We used this approach to generate successive curvature maps (Figure 5A). These 3D representations reveal
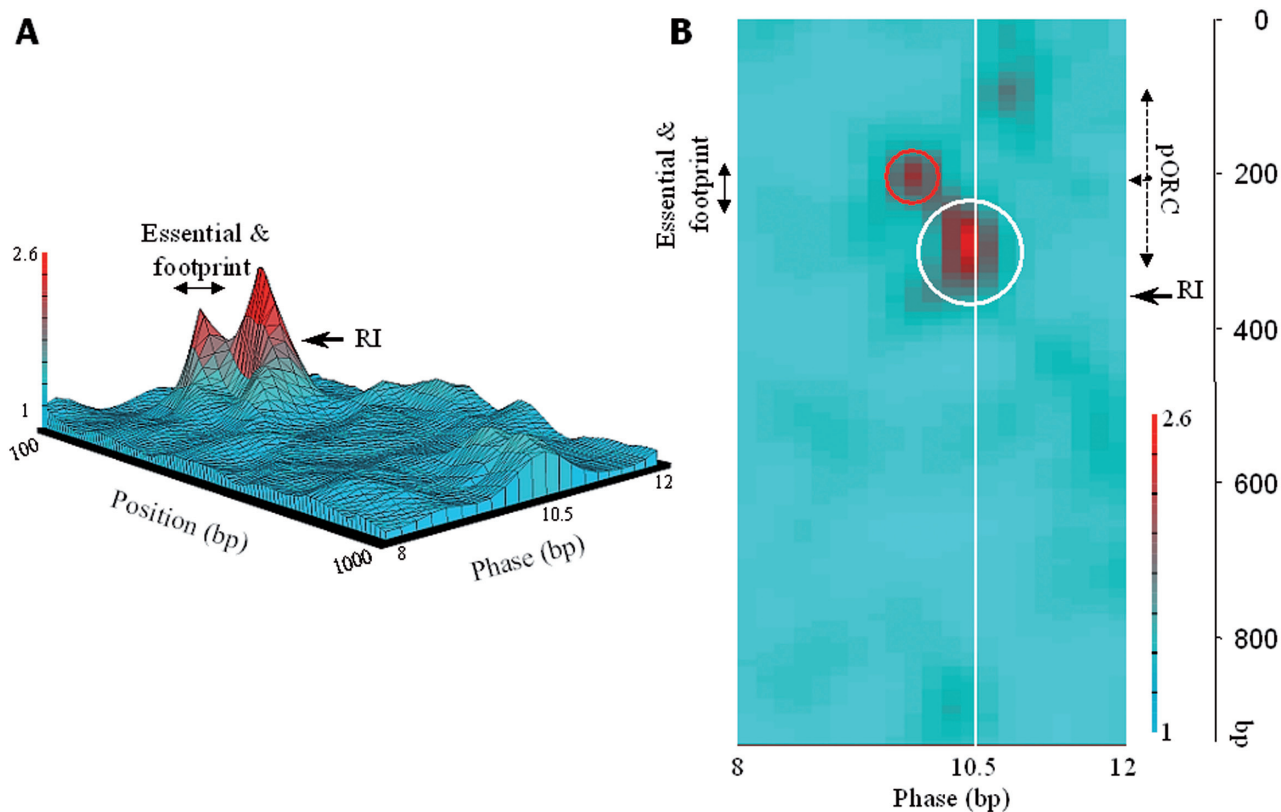
Figure 5. Diagrammatic and spectral analysis of Sp*ars1* DNA. The graphs show the presence of a strong curvature and of a region containing supercoiled DNA. The predicted curvature (white circle) is associated with the RI site. The curvature is right-handed. The right-handed supercoiled DNA region co-maps with the essential region and a pORC-binding footprint from Ref. (11). pORC-binding position '838' described in Ref. (9) is centred on the supercoiled structure (red circle). Window size is 200 bp, and each step 10 bp; analysed phases are between 8 and 12 with each step being 0.2.

inherently supercoiled DNA in addition to the previously described curvature. Spectral analysis (Figure 5B) clearly showed that this DNA structure (red circle) is right-handed and is close to the previously reported curvature (marked by a white line). This superhelical DNA is seen to co-localize with the region essential for *ars* activity described in (28) and with pORC footprint (9).

The analysis provides supplementary information concerning the unique strong curvature, which we find here to be right-handed and not strictly planar (Figure 5B). All these features may be relevant to function because torsional stress may affect the structures. Indeed, converting a curved region into a superhelix or the reverse would modify the geometry of the region, possibly substantially.

### Experimental analysis—curvature and kinetics of conformational changes

We used AFM imaging in liquid to validate our theoretical predictions: this technique revealed both the strong curvature and the supercoiled right-handed organization (Figure 6). The positions of this latter structure was measured in nanometre and converted into base pair: the superhelical structure was found to encompass ~200 bp (approximately from position 80 to 290 bp of the MluI end of the DNA fragment).

When AFM experiments were performed with shorter times between two successive images by changing the tip velocity and with the image limited to only the 128 lines covering the molecule, we succeeded in recording one image of the same single molecule every 30–40 s. Because of the reduced molecular plasticity when a molecule is interacting strongly with the support (despite working in liquid and using conditions of low adhesion to the support), these imaging conditions allowed the kinetics of conformational changes to be studied.

In order to perform the analysis, we took advantage of the spatial dimensionality of the molecule that needs to adjust continuously its trajectory to maintain a 2D structure under thermal agitation. Changes in the DNA conformation involve modifications of roll, tilt or twist angle values. In particular, superhelicoidal DNA can easily adopt a planar shape by changing twist values (21). Thus, by adjusting the different angular parameters, the molecule may find multiple equivalent solutions for staying confined on the mica surface. We succeeded in visualizing some of these conformational changes: for example, Figure 6 and Figure S2 (Supplementary Data 2) show the formation of a positive loop.

We analysed several series of 8–30 successive images of the same molecule. As a result of thermodynamic equilibrium, the measured length of the molecule oscillated continuously. The amplitude of the main curvature varied and, as described for temperature-driven DNA unwinding (15), the position of the apex also varied. The positively supercoiled region changed its superhelix
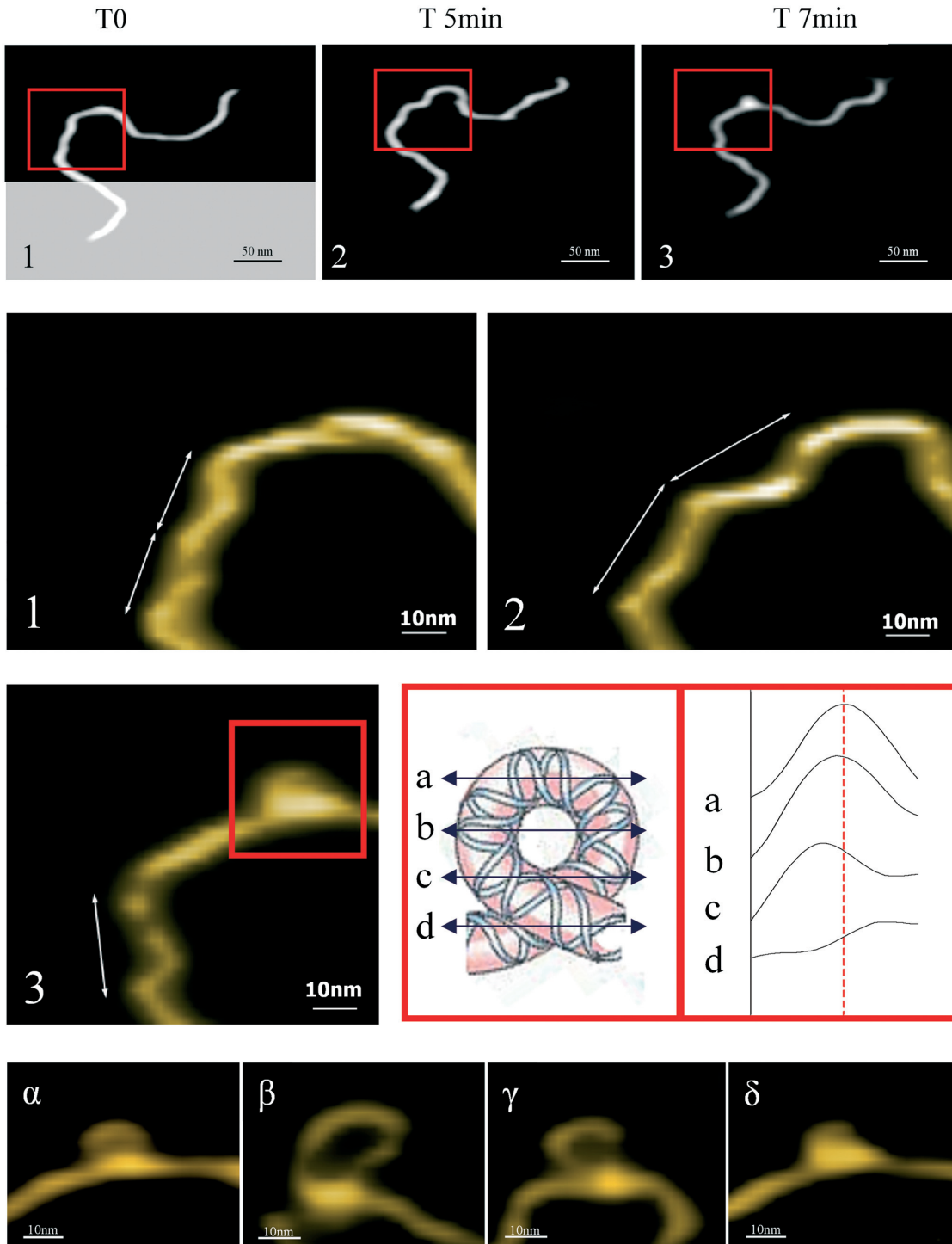
Figure 6. Formation of a right-handed loop. Images 1, 2 and 3 are AFM images of the MluI–EcoRI segment taken 0, 5 and 7 min, respectively, after the beginning of the observation. The red rectangles correspond to the enlarged regions beneath. White arrows correspond to the superhelix pitch. This pitch increases or decreases with time. Image number 3 shows the newly formed supercoiled loop. Insert: height variation on successive slices (right) corresponding to measurements on the loop as described in the scheme (left). These unambiguously show the positive organization of this loop. The loop size is ~130 bp. Last line, α, β, γ, δ are four independent examples of right-handed loops at the same position on the molecule. One of them, loop (β), formed before the first scan, is larger.

pitch (Figure 6, details in frames 1 and 2). Although the ionic conditions were chosen to allow the molecule to reorientate, it interacted substantially with the support, imposing a 'viscous' environment on the DNA molecule (29). This, as discussed earlier, slows down the conformational fluctuations, moreover, despite the DNA being linear, the length of the molecule delays the diffusion of torsional stress (29,30).
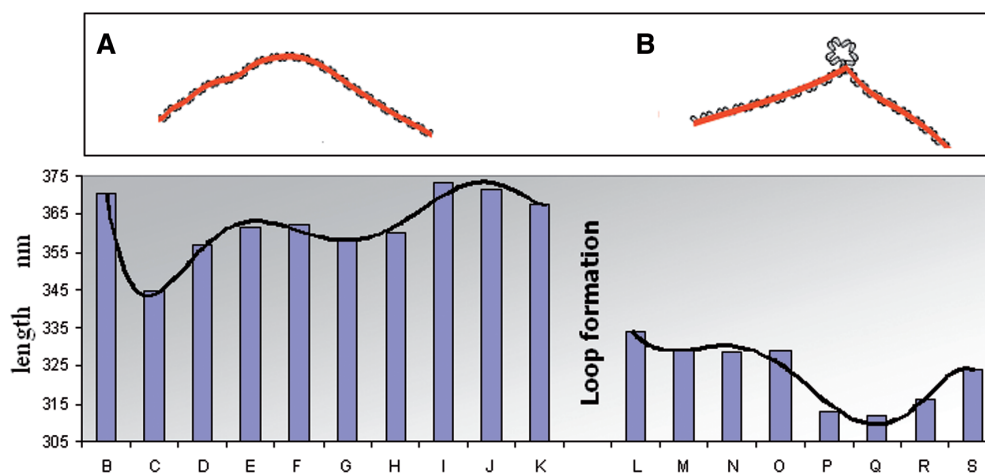
Figure 7. Variations of dimension values for the Sp*ars1* MluI–EcoRI fragment as measured by the 'Scanning adventure' software with time (red trace). Successive frames are labelled from **A** to **S**. Because the first image captured is incomplete, the total length of **A** is not reported in graph A. Graph A is the variation of length before the formation of the right loop. Graph B is the variation of length after the formation of the loop (the software does not trace the DNA trajectory inside the loop).

The AFM *in situ* analysis shows how even small changes in supercoiling can cause substantial structural modifications as, for instance, in the example shown in Supplementary Data S2 and analysed in Figures 6 and 7. In the series of recorded images, the permanent twisting and untwisting of the superhelix, necessary for it to explore different conformations while maintaining its two-dimensionality, make it possible to follow how the supercoiled organization generates a distant positive loop. The difference in length before looping (Figure 7A), and after looping (Figure 7B), indicates that the 'blobby' image on the molecule trajectory is really a loop. The distance between the start and the end of the loop indicate that the size was ~130 bp. Topographical slices of the loop (insert, Figure 6) unambiguously show that it was right-handed. Four additional examples are shown (taken at the same location); in all cases the right-handedness of the newly formed loops is clear (bottom of Figure 6).

Untwisting appeared to increase the amplitude of the curvature and to shift the position of the apex of the curvature such that the RI site was progressively drawn into the apex. This would be expected to have thermodynamic consequences (25). Indeed, during temperature-driven melting, DNA opening has been observed at the apex (Supplementary Data S1).

Searching for curved and supercoiled structures within the surrounding sequence

We extended our analysis to the 9000 nt surrounding Sp*ars1* using the conditions described in Figure 5. The substantial curvature described earlier is the only major curvature in the region (Figure 8); the same applies to the positive superhelical region.

A similar structural organization is found in another well-studied *S. pombe* origin

The RI in *S. pombe ars2004* has been mapped to an 800-bp segment of DNA (3). Deletion and substitution studies
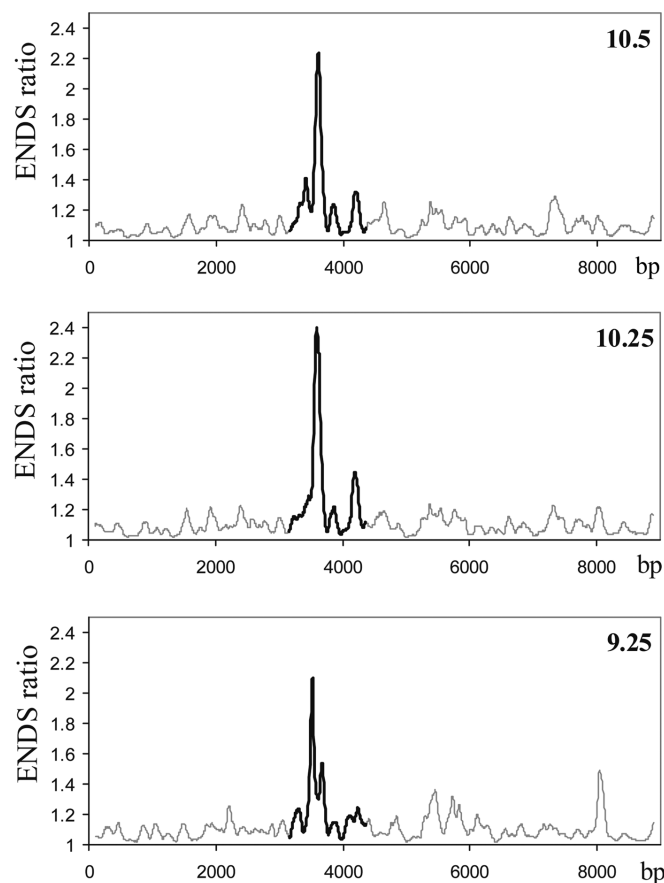
Figure 8. Curved and supercoiled structures are unique. The Sp*ars1* sequence is shown by black solid lines, and grey solid lines indicate sequences outside Sp*ars1*. Phasing is between 8 and 12 with steps of 0.25. The sliding window is 200 bp and each step 10 bp. For simplification, only the informative curves corresponding to the phases 10.5, 10.25 and 9.25 are shown. Exploration of 9000 nt around Sp*ars1* shows that the main curvature and the positively supercoiled region are unique.
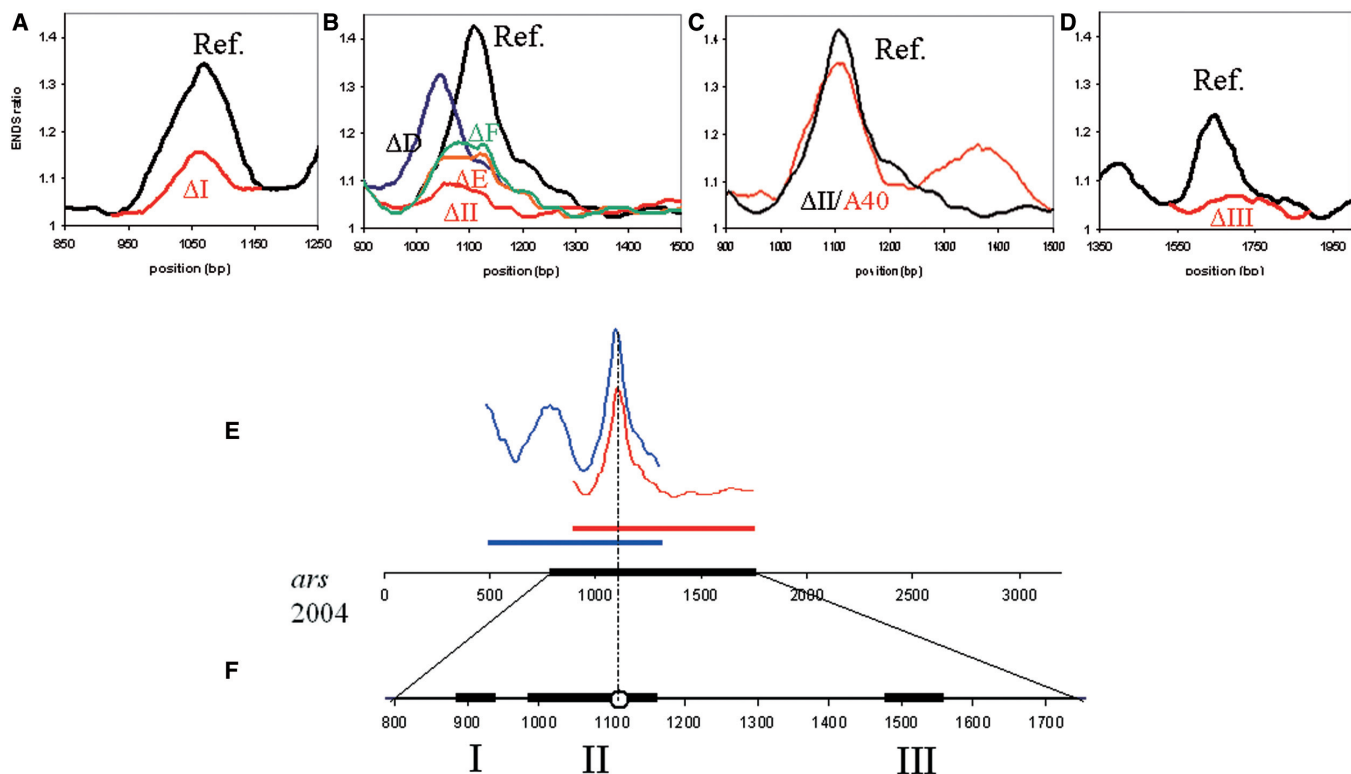
Figure 9. Analysis of the regions required for autonomous replication in *ars* 2004. (**A** and **D**) Modelling the region predicts right-handed superstructures co-mapping with regions I and III (3,25) essential for *ars* activity. Deletion of these regions also abolishes the structure (ΔI and ΔIII): see solid red line. Maximum of curvature is detected for phasing pitches of 8.9 and 6.1, respectively (window 200, step 2). (**B**) Modelling of region II shows a strong curvature; deletion of the region eliminates the structure (ΔII); see solid red line. Short deletions within the region II (ΔD, ΔE, ΔF), which were shown by Okuno to abolish or greatly reduce *ars* activity, alter the structure. (**C**) Replacing region II with a short A-stretch (A40) restores the activity and also clearly induces a strong curvature with identical positioning of the apex. (**E**) The 800-bp segment (blue line) that contains the RIP, identified by comparing 2D gel electrophoresis results (3), and the 835 bp fragment (red line), analysed in (23), that both contain a replication origin, are modelled. The two fragments overlap and the overlapping region corresponds precisely to the region of strong curvature that cannot be deleted without abolishing the *ars* function. (**F**) Mapping of the three regions (regions I, II and III) required for autonomous replication and enlargement (31). Thick black lines are essential regions. The open circle is the predicted opening site.

have identified regions, called region I, region II and region III (Figure 9F), that are essential for replication function: the three regions appear to act in concert (31). We modelled the *ars2004* element and found positive supercoils at positions corresponding to regions I and III, and a strong curvature at a position corresponding to region II. The model of an *ars2004* from which these regions are deleted (ΔI, ΔII or ΔIII) shows the abolition of the structures (see red traces and compare with the reference curves in Figure 9A, B and D). This agrees well with Okuno's *et al.* results showing that these deletions abolish the *ars* activity (31).

Short deletions within region II (ΔD, ΔE or ΔF) have various effects on the *ars* activity (31). Models of these changes also predicted various effects on the structure: the effects on *ars* activity seem to correspond to the changes in size (ΔE or ΔF) or position of the structure (ΔD).

Replacing region II with an A stretch (A40) was found to restore, at least partially, *ars* activity (31). We modelled this construction (Figure 9C, ΔII/A40) and the model predicts that insertion of this sequence promotes a roughly equivalent curvature at the same position as in the non-mutated region (see reference graph Figure 9B).

Thus, the predictions of our models for *ars2004* agree with, and even explain, experimental findings, and are thus consistent with the involvement of DNA structures in replication function.

We next looked on previous 2D gel electrophorectic studies of the *ars2004* origin (3,31) and (23) and found that they did not refer to exactly the same segment of DNA. However, in both cases, it was shown that replication initiates within the analysed fragment. We found that the fragments studied overlap by ~400 bp (Figure 11, red and blue lines), which may explain the similarity of their findings; it also suggested that the replication initiation site maps within these 400 bp. Modelling of the two candidate regions shows that the overlapping 400 bp region co-localizes exactly with the strongly curved DNA feature.

These structural components are very similar to those in *ars1*. Thus, in *ars2004* as in *ars1*, increasing bending by torsion of the structure may facilitate a localized melting by simultaneously lowering the unstacking energy and accumulating energy within the sugar phosphate backbone. This energy may be subsequently released to open the DNA double helix; a precise site in the DNA, the apex of the curvature, may be specified, defining the RIP position.
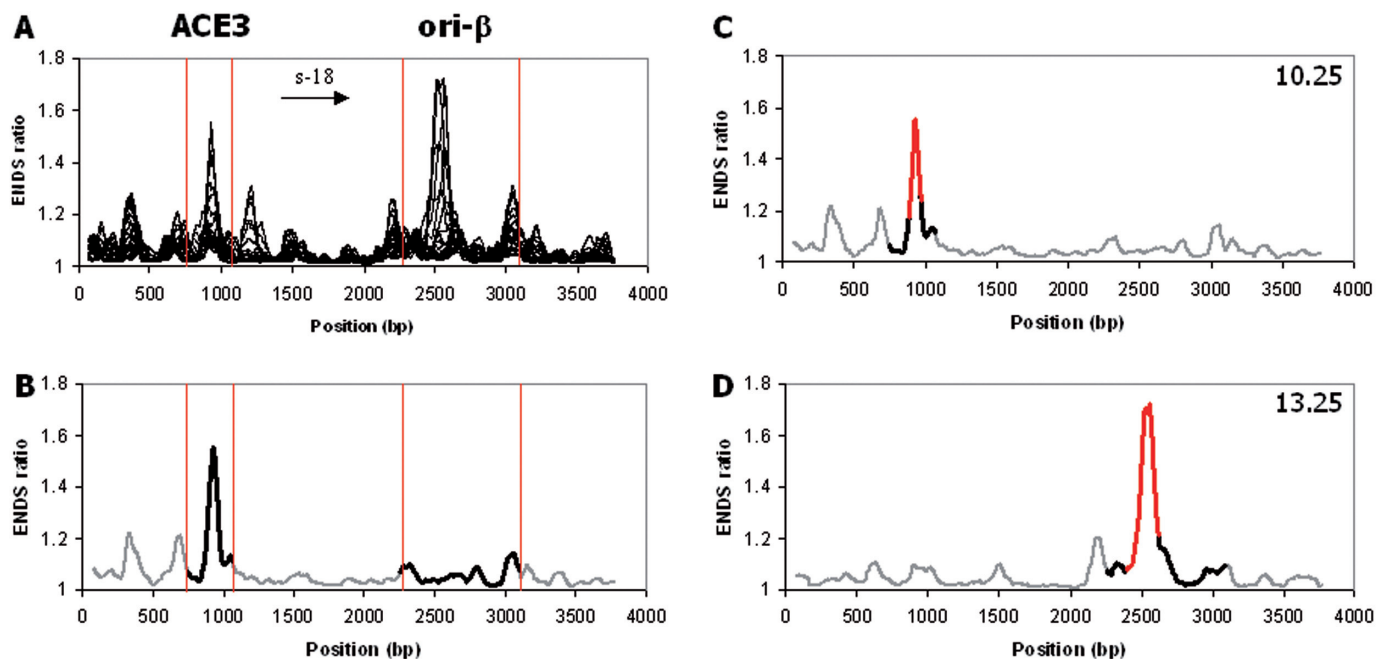
Figure 10. Presence of curved and supercoiled organization in a *Drosophila* replication origin: analysis of the DNA region containing Ace 3 and ori-beta in the *Drosophila* genome. (**A**) Unusual structures co-map with ACE3 and ori-beta. (**B**) Curvature map of reference: a strong curved region is associated with ACE3. (**C**) An evolutionarily conserved 142 bp core sequence (red trace) maps with right-handed DNA. (**D**) The beta-region described in Ref. (35), as a required component for amplification function (red trace) maps with a superhelical DNA (negative superhelicity). The direction of S-18 chorion gene transcription is indicated. Phasing is between 8 and 14 with steps of 0.25. The window is 150 bp, and each step is 2 bp.

It is not possible to extend this type of study to the 321 *S. pombe* origins recently published by Ref. (32), or to the 401 origins published by Ref. (33) because of the low resolution of the mapping. In the first case, the origins tug1 and rum1, although 5000 bp apart, cannot be discriminated. In the second case, the average interprobe distance is 1.3 kb: this may explain why the results are not consistent with the known positions of the RIP in *ars2004* or *ars1*.

Similar organization is found within a *Drosophila* origin

In the chorion gene locus of *Drosophila*, the amplification control element 3 (ACE3) functions as a replicator and the 3800 bp SalI fragment containing it has been studied in great detail. This segment contains two striking sequence elements, the alpha region upstream from the *S18* chorion gene and the beta region downstream from it. Recently, the 320 bp ACE3 sequence element (34) was shown to contain an evolutionarily conserved 142 bp core region, and deletion mapping of the 840 bp ori-beta identified a required region of 226 bp that was called the beta region (35). We modelled the 3800-bp segment and found two regions of unusual structural features including curved and inherently supercoiled DNAs and also found that they coincide with the 320 and 840 bp segments (Figure 10A and B). Moreover, the essential and evolutionarily conserved core region corresponds exactly to the predicted structure (Figure 10C) whereas the 'beta-region' co-maps with the supercoiled DNA (Figure 10D). Note that an analysis of the localization of the DmORC protein complex showed the influence of DNA torsion stress on its binding (10): *in vivo*, the 320 bp ACE3 fragment is sufficient to

localize DmORC in follicle cells. This is also the case for ori-beta (36).

Thus, this functional organization is very similar to the Sp*ars1* model in that structure and function are linked. Preliminary analyses with other eukaryotic models provide further support for this model (Marilley manuscript in preparation).

DISCUSSION

The initiation of eukaryotic DNA replication begins when the origin recognition complex (ORC) binds to DNA, recruiting the necessary helicases, polymerases and cofactors. Although the biochemical mechanism and factors appear to be conserved, the primary DNA sequence at which these events take place is not. In *S. pombe*, however, ORIs coincide with A + T-rich islands that differentiate them from the rest of the genome (23) but further specific DNA features had not, until now, been described.

DNA melting at replication origin: a thermodynamic point of view

It is now well established that not only the base composition, but also the nearest neighbours in a sequence have to be taken into account for the determination of thermal stability. Methods for calculating this stability have been continuously improved and their reliability has been proven. Using one such set of thermodynamic parameters, we show that the Sp*ars1* sequence is nested within a region of low stability and that the RI site is close to the point of lowest theoretical stability. This suggests
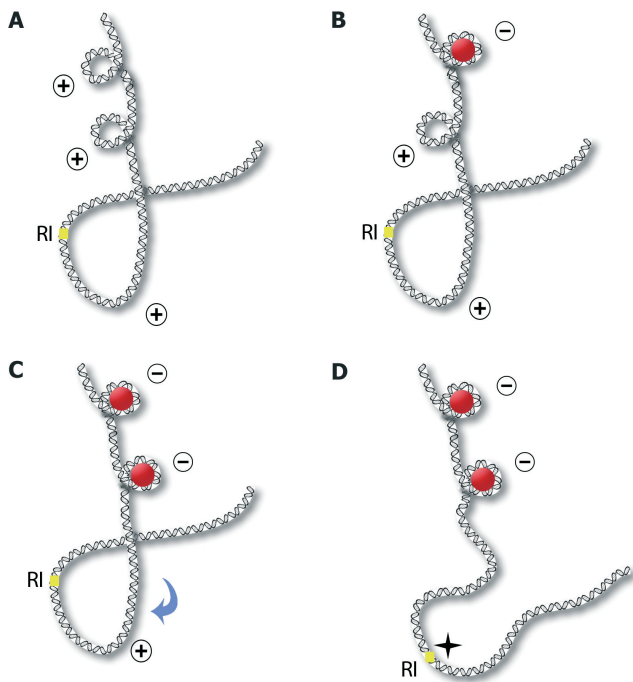
Figure 11. Replication preinitiation model. (**A**) Poorly stable DNA with inherent structures made of curved and supercoiled features. (**B**) A positively supercoiled region (+) is recognized and bound by a component of pORC, Orc4p (red circle) which organizes the DNA into a negative loop (−). (**C**) A second site (+) is recognized and bound by a second Orc4p, and a second negative loop is generated (−). (**D**) These binding events cooperate in unwinding DNA and displace the apex of the curvature to the RI site. Although the region is not melted, strand opening at this site may be greatly facilitated (star).

that the DNA at the RI site might not need much additional energy to open. These characteristics are unique in the surrounding 9000-base region analysed. We found a similar strong melting potential in all eukaryotic origins we have analysed so far (data not shown here). Initiation of DNA replication at CpG islands in mammalian chromosomes (37) is also consistent with our results. Because the G + C-rich islands are embedded in much more G + C-rich regions, these origins have lower stability than the surrounding sequences.

Inherent superhelicity and curvature

A nucleotide sequence having low stability is advantageous because reduced energy is required to open the RI site; however, it may increase the risk of uncontrolled melting. Thus, it is important to resolve the apparent paradox: how to maintain a replication origin closed until firing? The control may be exerted by proteins interacting with the region, but the DNA structure may also help to prevent unscheduled DNA strand separation: this could be the role of the energy barrier around the RI site. Similarly, the local DNA structures that we found in S*pars1* and which are completely absent from the 9000 surrounding nucleotides may also be involved. Indeed, because of its inherent geometry, supercoiled DNA can help maintain the two DNA strands in close contact.

We found a strong curvature with right-handedness and an intrinsically positively supercoiled sequence. These two structures are characterized by high A + T content. Earlier work by Okuno *et al.* (31), showed that essential regions of *ars*2004 could be replaced with poly(dA/dT) tracts indicating that this kind of sequence plays an important role in origin activity. Moreover, it has been shown that regular spacing of A/T tracts, but with a periodicity out of phase of the helix pitch, leads to inherent supercoiling (38). Thus, the oligo(dA)/oligo(dT) tract phasing might be important and not the A + T content.

Because the initial strand separation for initiation of replication is expected to occur when the initiation complex is assembled, this suggests that surface effects should not be ignored. Therefore, experimental conditions in which DNA is not free in solution but is lying on a surface were useful to provide additional information. Our experimental study shows that inherent curvature or supercoiling may be determinant for the geometry of the *S. pombe ars1* region. Moreover, these structures may change substantially according to induced superhelical stresses. For instance, a supercoiled region may become more or less supercoiled but also may give rise to strongly curved DNA (21). Alternatively, curved DNA may be transformed into positively or negatively supercoiled DNA according to the kind of superhelical stress applied. It follows that control of DNA melting at the origin may involve, in addition to the energetics of strand interactions, the geometry of the site.

A binding site for pORC

That inherently curved DNA might play a role in controlling local strand stability is an interesting possibility but does not exclude this DNA from having other important functions. Here we show that the *S. pombe ars1* possesses all the features required of a specific binding site for the pORC complex.

pORC binds to specific sites in ORIs. The site selection, at least *in vitro*, is determined solely by the Orc4p subunit (12) and the binding site is essential for viability *in vivo* (39). The N-terminal domain of this protein contains AT-hook motifs essential for the ATP-independent-*ars1*-binding activity (11). They might be involved in S*pars1* origin selection (39). A/T tracts are ubiquitous, and it was therefore suggested by (14) that specificity might result from all nine AT-hook motifs 'acting in concert'.

Work with HMG-I proteins (40,41) showed that substrate structure influences the binding of proteins containing AT-hook domains. The A/T tracts need to be properly phased for high-affinity binding and the high-affinity binding is lost when the spacing is greater than 10 bp. Strikingly, in Ref. (26) the presence of five homopolymeric dA.dT bases pairs causes significant bending of the DNA axis. Moreover, their repetition every 11 bp makes the DNA adopt a right-handed trajectory. When the spacing is greater than 10 bp, the length of the repeat is increased. Thus, the molecule remains right-handed, and only the diameter of the spiral is changed. The loss of high-affinity binding when the spacing is greater than 10 bp shows the importance of controlling the precise shape

(handedness and radius of curvature) of the binding site. Thus, specificity of the binding site may rely on the precise shape of the site and not on a consensus sequence. This kind of structure is only found at particular sites in Sp*ars1* and may therefore specify pORC positioning.

Interestingly, a torsional stress under negative super-coiling can be an important factor for pORC binding (10). This suggests that A/T tracts need to have the appropriate phasing, and/or that enlarging narrow minor grooves is important for binding and/or that torsional stress may facilitate the wrapping of DNA around pORC.

pORC binding and DNA unwinding at the replication origin

Although reports describing the number of pORC-binding sites and their precise locations are not in complete accord (9,11,12), one position at least is clearly established; the so-called '838' binding site. The position and functional relevance of this site has been extensively documented by: determination of spORC-binding sites by DNase I footprinting (11); localization of sequences that are protected by Orc4p (12); AFM imaging of spORC complexes bound to DNA (11); and co-localization with an essential region as demonstrated by deletion analysis (28).

Thus, it is clear that binding occurs within an intrinsically, positively supercoiled region. One consequence of binding is the formation of a DNA–protein complex in which the DNA wraps negatively around the protein core. Binding also induces topological changes in a closed plasmid (9). However, this DNA is resistant to both P1 nuclease and oxidation of T and/or C residues by $KMnO_4$ demonstrating that single-stranded DNA was not produced. Thus, wrapping around one spORC may change the DNA structure without strand separation, and we have seen here that this may be a consequence of this specific DNA structure. This phenomenon may also be explained on the basis of previous results (42): a study of mammalian HMG-I(Y) proteins, reports a concentration effect. Unexpectedly, when binding the protein to closed circular DNA at high molar ratios of protein to DNA, HMG-I(Y) induces a negative supercoiling, but when the ratios are low, positive supercoiling is induced. This effect may be due to the underlying positively supercoiled structure which is the substrate of the HMG-I(Y) proteins. For a molar ratio of protein to DNA of 3 to 1 [as used in Ref. (9)] a negative supercoiling is predicted and is observed. However, due to the super helical turn reserve (buffer effect), no strand opening is observed; however, this does not mean that the DNA has not, at least partially, lost its supercoils.

It is possible that other sites co-localize with intrinsically supercoiled left-handed DNA, as seems to be the case for pORC '1148'. This would explain why pORC '838' and pORC '1148' were found to be mutually exclusive (9); however, the second site was close to the vector sequence, so exogenous effects cannot be completely excluded.

Multiple ORC-binding events are required for origin firing

Multiple ORC binding within the *ori2004* locus is essential for origin firing (43). Cooperation between ORC complexes to attain the critical concentration is also consistent with gradual changes in ORI topology that may favour successive binding events. Binding of one pORC complex onto Sp*ars1* unwinds the DNA, but its effect is not sufficient to lead to local strand separation. Therefore, multiple binding may have an additive effect, and/or, may be a prerequisite for the association of the MCM complex.

Replication–transcription interplay

Numerous eukaryotic replication origins are located near transcription promoters, upstream from genes. This is the case for Sp*ars1* where the RI site is 261 bp upstream from the start site of the $hus5^+$ promoter (4); thus, there is positively coiled DNA lying between the two.

Because of this co-positioning, transcriptional activity may have various effects. For instance, the negative wave following RNA polymerase progression could encounter the positive superhelix and thus progressively alter the phasing of the bent elements (mainly A/T-tracts). This could increase the superhelical pitch until a planar curvature is achieved, or even, reverse the handedness of the curvature. Through structural effects, the region may acquire the function of an enhancer element. Indeed, the release of negative superturns buffered by the structure is likely to facilitate polymerase progression during transcription. On the other hand, abolishing, or simply reducing the inherent positive structure may in turn facilitate strand separation at the replication origin. However, at a preinitiation step, the organization of the structure into a negative loop, due to the binding of pORC, may slow down, or even switch off transcription, rather than enhancing it. Thus, both replication and transcription may be affected by the structural organization.

A model

The structural organization of the Sp*ars*1 origin is also found in other eukaryotic origins (including the amplification control chorion gene element three in *Drosophila*). Therefore, we propose a model in which this structural organization explains: (i) DNA opening at a specific site and protection against uncontrolled melting, (ii) pORC binding to a specific site despite the absence of a consensus sequence, (iii) the function of multiple ORC-binding events and (iv) interrelation between transcription and replication and the advantage of coupling the two processes.

In this preinitiation model (Figure 11), the replication origin region consists of DNA of low stability. It contains substantially curved DNA, and close to it, an inherent supercoiled organization preventing uncontrolled DNA opening. The supercoiled region described here is right-handed, and is recognized by pORC (a left-handed supercoiled region might also function if it could bind pORC). Next, a component of pORC, Orc4p, organizes the DNA into a negative loop. The process is facilitated by negative torsional stress. This event may be followed by the binding of a second Orc4p to a second site (recognition and binding might be facilitated by the first one), and consequently a second negative loop could be generated. These binding events cooperate to unwind the

DNA, resulting in the increase in amplitude of the curvature and the apex of the curvature being repositioned to the RI site, thus fragilizing the site. Although the region is not melted at this stage (9), strand opening at the apex of the curvature is potentiated. DNA opening will then depend on additional effects of the other initiation complexes whose recruitment becomes possible because of the modified local structure.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Caddle,M.K. and Calos,M.P. (1994) Specific initiation at an origin of replication from Schizosaccharomyces pombe. *Mol. Cell. Biol.*, **14**, 6348–6357.
2. Dubey,D.D., Zhu,J., Carlson,D.L., Sharma,K. and Huberman,J.A. (1994) Three ARS elements contribute to the ura4 replication origin region in the fission yeast Schizosaccharomyces pombe. *EMBO J.*, **13**, 3638–3647.
3. Okuno,Y., Okazaki,T. and Masukata,H. (1997) Identification of a predominant replication origin in fission yeast. *Nucleic Acids Res.*, **25**, 530–536.
4. Gomez,M. and Antequera,F. (1999) Organization of DNA replication origins in the fission yeast genome. *EMBO J.*, **18**, 5683–5690.
5. Broach,J.R., Li,Y.-Y., Feldman,J., Jayaram,J., Abraham,K., Nasmyth,A. and Hicks,J.B. (1983) Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1165–1173.
6. Newlon,C.S. and Theis,J.F. (1993) The structure and function of yeast ARS elements. *Curr. Opin. Gen. Dev.*, **3**, 753–758.
7. Marilley,M. (2000) Structure-function relationships in replication origins of the yeast Saccharomyces cerevisiae: higher-order structural organization of DNA in regions flanking the ARS consensus sequence. *Mol. Gen. Genet.*, **263**, 854–866.
8. Forsburg,S.L. (2005) The yeasts Saccharomyces and Schizosaccharomyces pombe: models for cell biology research. *Grav. Sp. Biol.*, **18**, 3–10.
9. Gaczynska,M., Osmulski,P.A., Jiang,Y., Lee,J.K., Bermudez,V. and Hurwitz,J. (2004) Atomic force microscopic analysis of the binding of the Schizosaccharomyces pombe origin recognition complex and the spOrc4 protein with origin DNA. *Proc. Natl Acad. Sci. USA*, **101**, 17952–17957.
10. Remus,D., Beall,E.L. and Botchan,M.R. (2004) DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J.*, **23**, 897–904.
11. Lee,J.-K., Moon,K.-Y., Jiang,Y. and Hurwitz,J. (2001) The Schizosaccharomyces pombe origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc. Natl Acad. Sci. USA*, **98**, 13589–13594.
12. Kong,D. and DePamphilis,M.L. (2001) Site-specific binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit. *Mol. Cell. Biol.*, **21**, 8095–8103.
13. Chuang,R.-Y. and Kelly,T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc. Natl Acad. Sci. USA*, **96**, 2656–2661.
14. Kong,D., Coleman,T.R. and DePamphilis,M.L. (2003) Xenopus origin recognition complex (ORC) initiates DNA replication preferentially at sequences targeted by Schizosaccharomyces pombe ORC. *EMBO J.*, **22**, 3441–3450.
15. Marilley,M., Milani,P. and Rocca-Serra,J. (2007) Gradual melting of a replication origin (Schizosaccharomyces pombe ars1): in situ atomic force microscopy (AFM) analysis. *Biochimie*, **89**(Special issue), 534–541.
16. 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.
17. 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.
18. 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. Genet. Genomics*, **274**, 658–670.
19. SantaLucia,J.Jr, Allawi,H.T. and Seneviratne,P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.
20. Marilley,M. and Pasero,P. (1996) Common DNA structural features exhibited by eukaryotic ribosomal gene promoters. *Nucleic Acids Res.*, **24**, 2204–2211.
21. Roux-Rouquié,M. and Marilley,M. (2000) Modeling of DNA local parameters predicts encrypted architectural motifs in Xenopus laevis ribosomal gene promoter. *Nucleic Acids Res.*, **28**, 3433–3441.
22. Marilley,M., Radebaugh,C.A., Geiss,G.K., Laybourn,P.J. and Paule,M.R. (2002) DNA structural variation affects complex formation and promoter melting in ribosomal RNA transcription. *Mol. Genet. Genomics*, **267**, 781–791.
23. Segurado,M., deLuis,A. and Antequera,F. (2003) Genome-wide distribution of DNA replication origins at A + T-rich islands in Schizosaccharomyces pombe. *EMBO Reports*, **4**, 1048–1053.
24. Astier,J.P., Bokern,D., Lapena,L. and Veesler,S. (2001) alpha-amylase crystal growth investigated by in situ atomic force microscopy. *J. Crystal Growth*, **226**, 294–302.
25. Mielke,S.P., Gronbech-Jensen,N., Krishnan,V.V., Fink,W.H. and Benham,C.J. (2005) Brownian dynamics simulations of sequence-dependent duplex denaturation in dynamically superhelical DNA. *J. Chem. Phys.*, **123**, 124911.
26. Dubochet,J., Bednar,J., Furrer,P., Stasiak,A.Z., Stasiak,A. and Bolshoy,A. (1994) Determination of the DNA helical repeat by cryo-electron microscopy. *Nat. Struct. Biol.*, **1**, 361–363.
27. Marini,J.C., Levene,S.D., Crothers,D.M. and Englund,P.T. (1982) Bent helical structure in kinetoplast DNA. *Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.
28. Clyne,R.K. and Kelly,T.J. (1995) Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *EMBO J.*, **14**, 6348–6357.
29. Nelson,P. (1999) Transport of torsional stress in DNA. *Proc. Natl Acad. Sci. USA*, **96**, 14342–14347.
30. Kouzine,F., Liu,J., Sanford,S., Chung,H.J. and Levens,D. (2004) The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol.*, **11**, 1092–1100.
31. Okuno,Y., Satoh,H., Sekiguchi,M. and Masukata,H. (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Mol. Cell. Biol.*, **19**, 6699–6709.
32. Feng,W., Collingwood,D., Boeck,M.E., Fox,L.A., Alvino,G.M., Fangman,W.L., Raghuraman,M.K. and Brewer,B. (2006) Genomic mapping of single strand DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat. Cell. Biol.*, **8**, 148–155.
33. Heichinger,C., Penkett,C.J., Bähler,J. and Nurse,P. (2006) Genome-wide characterization of fission yeast DNA replication origins. *EMBO J.*, **25**, 5171–5179.

34. Orr-Weaver,T.L., Johnston,C.G. and Spradling,A.C. (1989) The role of ACE3 in Drosophila chorion gene amplification. *EMBO J.*, **8**, 4153–4162.

35. Zhang,H. and Tower,J. (2004) Sequence requirements for function of the Drosophila chorion gene locus ACE3 replicator and ori-beta origin elements. *Development*, **131**, 2089–2099.

36. Austin,R.J., Orr-Weaver,T.L. and Bell,S.P. (1999) Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev.*, **13**, 2639–2649.

37. Delgado,S., Gomez,M., Bird,A. and Antequera,F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.*, **17**, 2426–2435.

38. Eckdahl,T.T. and Anderson,J.N. (1990) Conserved DNA structures in origins of replication. *Nucleic Acids Res.*, **18**, 1609–1612.

39. Chuang,R-Y., Chrétien,L., Dai,J. and Kelly,T.J. (2002) Purification and characterization of the Schizosaccharomyces pombe origin recognition complex. *J. Biol. Chem.*, **277**, 16920–16927.

40. Maher,J.F. and Nathans,D. (1996) Multivalent DNA-binding properties of the HMG-I proteins. *Proc. Natl Acad. Sci. USA*, **93**, 6716–6720.

41. Reeves,R. and Wolffe,P. (1996) Substrate structure influences binding of the non-histone protein HMG-I(Y) to free and nucleosomal DNA. *Biochemistry*, **35**, 5063–5074.

42. Nissen,M.S. and Reeves,R. (1995) Changes in superhelicity are introduced into closed circular DNA by binding of high mobility group protein I/Y. *J. Biol. Chem.*, **270**, 4355–4360.

43. Takahashi,T., Ohara,E., Nishitani,H. and Masukata,H. (2003) Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *EMBO J.*, **22**, 964–974.