Linearisation and labelling of single-stranded DNA for optical sequence analysis

Rajib Basak,† Fan Liu,† Sarfraz Qureshi,† Neelima Gupta,‡ Ce Zhang,¶ Renko de Vries,§ Jeroen A. van Kan,† S. Thameem Dheen,‡ and Johan R.C. van der Maarel†,†,‡

†Department of Physics, National University of Singapore, Singapore 117542
‡Department of Anatomy, National University of Singapore, Singapore 117594
¶Institute of Photonics and Photon-Technology, Northwest University, Xi’an, China 710069
§Laboratory of Physical Chemistry and Colloid Science, Wageningen University, The Netherlands

Received December 24, 2018; E-mail: johanmaarel@gmail.com

Abstract: Genetic profiling would benefit from linearisation of ssDNA through the exposure of the unpaired bases to gene-targeting probes. This is compromised by ssDNA’s high flexibility and tendency to form self-annealed structures. Here, we demonstrate that self-annealing can be avoided through controlled coating with a cationic-neutral diblock polypeptide copolymer. Coating does not preclude site-specific binding of fluorescence labelled oligonucleotides. Bottlebrush-coated ssDNA can be linearised by confinement inside a nanochannel or molecular combing. A stretch of 0.32 nm per nucleotide is achieved inside a channel with a cross-section of 100 nm and a two-fold excess of polypeptide with respect to DNA charge. With combing, the complexes are stretched to a similar extent. Atomic force microscopy of dried complexes on silica revealed that the contour and persistence lengths are close to those of dsDNA in the B-form. Labelling is based on hybridisation and not limited by restriction enzymes. Enzyme-free labelling offers new opportunities for the detection of specific sequences.

Stretching of DNA to a length close to its contour length is a key element in many biotechnological and biophysical applications. Examples include flow- and force-stretching to provide real-time information about protein-nucleic acid interaction at the level of single molecules.1–3 Single DNA molecules can also be stretched by confinement inside a nanochannel with a cross-sectional diameter on the order of tens to hundreds of nanometers.4–6 Advantages of nanoflow are that there is no need for chemical modification such as the attachment of tethers, the DNA molecules are in an equilibrated conformation, high throughput can be achieved by using an array of parallel channels, and integration with lab-on-chip devices. Particular applications of the nanochannel platform are mapping of large-scale genomic organisation, including restriction enzyme cutting, and denaturation mapping.7–11 The molecules of interest are usually stained and visualised with fluorescence microscopy. For genome mapping and profiling technologies, site-specific labelling by nicking or restriction endonucleases is required. Every locus needs a specific enzyme, which can be prepared with a CRISPR/Cas9 technology.12

Most of the previously reported stretching experiments are done with DNA in its double-stranded form (dsDNA). Linearisation of single-stranded ssDNA molecules on a mica surface has been demonstrated before.13 Stretching of ssDNA is compromised by its flexibility (persistence length of around 0.7 nm)14,15 and tendency to form a myriad of self-annealed structures through hybridisation of the unpaired bases. Yet, profiling of genetic information would benefit from linearisation of single DNA strands through the exposure of the unpaired bases to gene-targeting probes. Here, we propose to apply a copolymer coating to the ssDNA molecules. The advantages of this approach are two-fold. First, the concomitant increase in bending rigidity of the complex results in a stretch similar to the one of dsDNA. Second, self-annealing of the unpaired bases is largely prevented resulting in non-aggregated and unfolded single molecules. Furthermore, we will demonstrate that the coating procedure does not preclude site-specific binding of oligonucleotides. The increased bending rigidity allows linearisation of ssDNA molecules to a length similar or exceeding the contour length of the double-stranded variant inside channels with a cross-sectional diameter of around 100 nm or by molecular combing.

A suitable candidate for coating ssDNA is a cationic-neutral diblock polypeptide copolymer.16,17 The DNA binding block is relatively short and consist of 12 positively charged lysine residues (K12). The K12 block is connected to 4 repeats of an approximately 100 amino acids long polypeptide (C4).18 The amino acid composition of the C4 block is similar to that of collagen. It is also hydrophilic, net electroneutral, and behaves as a flexible polymer in aqueous solution. The entire C4K12 diblock copolymer is produced using recombinant DNA technology by large-scale expression in yeast. The copolymer is monodisperse with a total molecular weight of 38.4 kDa. As shown in previous work, complexation of dsDNA with C4K12 results in an amplified stretch inside 200-300 nm channels.19 Furthermore, it was shown that DNA metabolism is not inhibited by the copolymer coat. Here, DNA in single-stranded form is uniformly coated by C4K12 as illustrated in Figure 1. The stoichiometry of the complexes is expressed by the N/P ratio, that is the ratio of the number of positively charged amino groups
on the lysine K_{12} binding block to the negatively charged phosphate groups on the ssDNA. The linearisation of the ssDNA complexes will be gauged from nanofluidics and molecular combing experiments in tandem with fluorescence and atomic force microscopy. Finally, the potential for optical analysis will be demonstrated by targeting two specific sequences of the bacteriophage lambda genome.

Depending on channel dimensions, the nanofluidic devices are made of polydimethylsiloxane (PDMS) elastomer or poly(methyl methacrylate) (PMMA) polymer as described in the supplementary material. \(^{20-23}\) Single-stranded DNA was prepared by alkali induced denaturation of bacteriophage \(\lambda\)-DNA (48.5 kbp). The ssDNA solutions were incubated with various concentrations of C_{4}K_{12} copolymer for at least 24 h. Electrostatic binding of the cationic lysine K_{12} block on the ssDNA molecules was facilitated by a subsequent exchange of alkaline to TE-buffer (10 mM Tris/HCL, 1 mM EDTA, pH 8.0) conditions through repeated ultrafiltration. Binding of the copolymer on the single DNA strands results in the formation of a bottlebrush complex as illustrated in Figure 1. The molecules are coated with various amounts of C_{4}K_{12}, as indicated by the N/P ratio. The coating procedure prevents strand annealing but, as we will see below, does not preclude hybridisation of relatively small oligonucleotides. Prior to fluorescence imaging, the ssDNA complexes were uniformly stained through side-binding of YOYO-1 with a ratio of one dye per four bases. \(^{24-26}\) The fluorescence intensity of the thus labelled complexes is fairly weak but comparable to YOYO-1 labelled dsDNA with a ratio of one dye per 100 base-pairs (see Figure S1 in the supplementary material).

In the nanofluidics experiment, the stained complexes were brought into an array of 60 \(\mu\)m long and rectangular nanochannels by electrophoresis. After switching off the electric field, the complexes relax to their equilibrium state within 60 s. Video recording was started 2–5 min after the complexes were brought into the channels and the clips lasted for another 5 min. A montage of YOYO-1 fluorescence images of complexes with various N/P ratios and confined in channels with a cross-section of 170\(\times\)200 nm\(^2\) is shown in Figure 2A. Another montage pertaining to complexes with a coating ratio N/P = 2.0 and in channels of various cross-sectional diameter is shown in Figure 3A. All images refer to well-equilibrated conformations. For each experimental condition, that is N/P ratio and channel diameter, we have used a fresh chip and measured the stretch of 48–53 complexes by time-averaging over the duration of the video clips. Fragmented complexes (18\%) were ignored by using a cutoff being the mean stretch minus 2 times the standard deviation. In the case of naked dsDNA in nanochannels a similar amount of fragmentation was reported. \(^{27}\) The distributions in stretch are close to Gaussian and are shown in panels B–D of Figures 2 and 3. Inside 170\(\times\)200 nm\(^2\) channels, Gaussian fits give a mean stretch \(R_0 = 5.5\pm0.7, 8\pm1\) and \(10.2\pm0.9\mu\)m for N/P = 0.1, 0.5 and 2.0, respectively. With a two-fold excess of polypeptide to DNA charge (N/P = 2.0), the mean stretch takes the values \(4.6\pm0.7, 10\pm2\) and \(16\pm3\mu\)m inside 250\(\times\)250, 170\(\times\)200 and 100\(\times\)100 nm\(^2\) channels, respectively.

With decreasing channel diameter and/or increasing C_{4}K_{12} to ssDNA ratio, the stretch increases. For the wider channels and lower N/P ratios, the stretch is significantly less than the contour length of the parent \(\lambda\)-DNA molecule in its double-stranded form (16.5 \(\mu\)m). However, a maximal stretch of \(16\pm2\mu\)m is reached in 100 nm channels with N/P = 2.0. An excess of polypeptide copolymer is required to achieve a maximal stretch due to the dynamic equilibrium of binding (law of mass action). Note that the distribution in extension is fairly wide with some complexes stretched to a
length of 19 μm (Figure 3D). Furthermore, the distribution is slightly left-skewed towards smaller extensions, which is a signature of the presence of back-folded hairpin conformations.\(^{28}\) A stretch of 0.33±0.04 nm per nucleotide for ssDNA is somewhat larger than what can be achieved for dsDNA in 45 nm channels (≈ 0.29 nm/bp).\(^{10}\) The increase in stretch is mainly due to an increase in bending rigidity of the ssDNA complex. This effect will be further gauged below by atomic force microscopy. We will first demonstrate that genomic information can be obtained by hybridisation of oligomeric probes to the complexed, single-stranded DNA molecules.

The aim of large scale genome mapping is not to determine the genetic code at the level of single bases, but to provide a map of the genome at the larger scale. This map can be used for genome characterisation as well as to sort out large scale variations such as repeats, insertions, inversions, and translocations. For the demonstration of the accessibility of the unpaired bases, we have added an equal mixture of two Alexa Fluor 546 probes with different oligonucleotide sequences to the denatured DNA solution prior to bottlebrush-coating (see Figure 1). These probes were selected to target the 8.0 and 35.8 kb sites of the same single strand of the λ-DNA molecule (there are, hence, no probe sites present on the complementary strand). Note that, in principle, any site can be selected by specification of the nucleotide sequence. In order to obtain a maximal stretch, we have used a two-fold excess of polypeptide to DNA charge (N/P = 2.0).

The doubly labelled (YOYO-1 and Alexa Fluor 546) complexes were brought into the array of 100×100 nm\(^2\) channels by electrophoresis. The complexes were equilibrated and imaged with fluorescence microscopy using the appropriate optical filters and excitation wavelength. As can be seen in panel A of Figure 4, individual Alexa Fluor labels are discernible at the targeted sites. For comparison, in panel C of Figure 4, we have displayed the corresponding fluorescence images of combed complexes.\(^{29}\) The distributions in probe distance in populations of 50 molecules pertaining to the nanoarraying and combing experiments are shown in panels B and D of Figure 4, respectively. In the nanoarraying and combing experiments, the complexes are stretched to a similar extent of 0.32±0.04 nm per nucleotide. The stretch per nucleotide as inferred from the separation of the Alexa Fluor probes is in perfect agreement with the one obtained from the measurement of the overall stretch of the complex through measurement of the YOYO-1 fluorescence. This confirms that fragmentation has not biased our length measurements. With a typical optical resolution of around 200 nm, the resolution in site position is about 1 kb. This resolution is similar to what can be achieved for dsDNA in 45 nm channels.\(^{8,10}\) The resolution might be increased by using super-resolution imaging strategies and/or combinations of probes of different colour.\(^{29}\)

In order to assess the uniformity of stretching along the channel, we have also measured the extensions of the sections between the Alexa Fluor labels and the end-points of the ssDNA molecules. The distributions in extension of the 0–8 kb and 35.8–48.5 kb segments are shown in Figure 5. For the 0–8 kb and 35.8–48.5 kb segments, the stretch per
nucleotide takes the value 0.33±0.03 nm and 0.34±0.03 nm, respectively. These values of the stretch are similar to the one pertaining to the middle section of the molecule between the Alexa Fluor labels (0.32±0.04 nm per nucleotide). The calculated variation in stretch among the end and middle sections of the linearised single-stranded λ-DNA (N/P = 2.0) molecules in 100×100 nm² channels is 9%.

In both the nanoarraying and combing experiments, occasionally Alexa Fluor labels are missing. Out of a pool of 240 ssDNA molecules, 80 and 20 of the molecules show one and two hybridised labels, respectively. This corresponds with a labelling efficiency of 67% and 17% for one and two labels, respectively (only 50% of the single DNA strands are potentially labelled since there are no probe sites present on the complementary strand). The relatively low double label efficiency may be related to differences in binding affinity between the 20 and 30 nucleotide probes targeting the 8.0 and 35.8 kb site, respectively. Other reasons for missing labels are photo induced transitions to dark states including photobleaching (no anti-bleaching agent was used).

The increase in stretch induced by the copolymer coating can be rationalised in terms of an increase in persistence and/or contour length of the ssDNA complexes. These characteristic lengths were obtained from analysis of atomic force microscopy images of the complexes on silica. In order to image the complexes in their entirety with a field of view of 3×3 μm², we have used 1 kb dsDNA as the source material for the preparation of bottlebrush-coated ssDNA. No additions to the buffer are necessary to promote adhesion, because the neutral C₁ blocks weakly adsorb to silica. Due to the weak adhesion as shown by the easy removal by flushing with water, the complexes are not kinetically trapped in a 3D conformation. The molecules were allowed to equilibrate on the surface in a 2D conformation for 5 min followed by the removal of excess copolymer through flushing with ultra-pure water. Subsequently, all specimens were N₂-dried. A montage of images of the complexes with N/P = 2.0 is shown in Figure 6A (an example of a raw image is shown in Figure S2 in the supplementary material). Individual complexes are visible, without signs of inter-complex aggregation nor intramolecular hybridisation.

In order to estimate the thickness of the brush, the cross-sectional profiles taken at 10 different and randomly chosen positions of the complexes were analysed. A gaussian fit yields a cross-sectional radius of gyration of 5 nm. This value should be taken as an indication, because the measured profiles are broadened by the width of the tip and the brushes are dried and spread on the silica surface. A comparison with the atomic force microscopy images of bottlebrush-coated dsDNA in otherwise the same experimental conditions (N/P = 2.0) reveals that the double-stranded complex is much thicker with a cross-section of around 30 nm.¹⁹

We traced the centrelines of a population of 27 individual complexes. The distribution in contour length is shown in Figure 6B. The average contour length is 0.33 ± 0.04 μm, which corresponds with a rise per nucleotide of 0.33±0.04 nm. The latter value is in perfect agreement with the values obtained from the total stretch of the complexes (YOYO-1) and the separation between the oligonucleotide probes (Alexa Fluor 546) in the nanoarraying and combing experiments. Furthermore, it confirms that the stretch is somewhat inhomogeneous with a spread of about 10%.

From the centrelines, we also obtained the tangent vector correlation function \( \langle \cos \theta_{s,s+L} \rangle \), where \( \theta \) is the angle between tangent vectors at points \( s \) and \( s + L \), by averaging \( s \) along the contour. For weakly adsorbed complexes equilibrated in 2D conformation, the tangent vector correlation follows \( \langle \cos \theta_{s,s+L} \rangle = \exp (-L/(2P)) \) with the persistence length \( P \). As can be seen in Figure 6C, reasonable agreement is obtained for \( P \geq 60 \pm 2 \) nm. The persistence length of the bottlebrush-coated ssDNA molecules is close to the \( P \) value of around 50 nm for dsDNA.²² For ssDNA in saline solutions \( P \approx 0.65 \) nm.¹⁵ The primary reason for the bottlebrush coating induced increase in stretch inside the nanochannels is, hence, the huge increase in bending rigidity of the intrinsically rather flexible ssDNA molecule.

A secondary reason is the increased cross-sectional diameter of the complex. For bottlebrush-coated dsDNA in the same experimental conditions and obtained with the same methodology, \( P \) takes a value of 240±10 nm.¹³ The changes in cross-sectional diameter and persistence length following our denaturation protocol confirm that the complexes are indeed in the single-stranded form.

We have shown that ssDNA can be linearised through coating with a diblock polypeptide copolymer. The copolymer coat prevents self-annealing of the unpaired bases and concomitant effects such as intramolecular folding and intramolecular aggregation but does not preclude site-specific binding of fluorescence labelled oligonucleotide probes. A stretch of 0.32 nm per nucleotide can be achieved with a two-fold excess of copolymer with respect to DNA charge either inside a channel with a cross-sectional diameter of 100 nm or by molecular combing. To the best of our knowledge this is the first report of linearisation of ssDNA in nanofluidic channels. Atomic force microscopy of dried ssDNA complexes on silica revealed that the contour and persistence lengths are similar to those of dsDNA in the B-form. An advantage of linearising ssDNA is that, in principle, any site can be targeted by a specific oligonucleotide probe. There is, hence, no need for labelling with the help of endonucleases. The easy customisation of the sequence and density of probes bound on DNA may offer a substantial improvement in the detection of genomic variation at the larger scale.
Important features are the possibilities to image multiple targets in close proximity and/or genomic regions without appropriate restriction site motifs. The proposed technology can potentially be used in a wider range of applications by employing more probe colour combinations. Besides relatively simple, enzyme free labelling, our technology offers the additional possibility to image a single-stranded genome including RNAs.

Sample preparation, chip fabrication, fluorescence imaging (image), molecular combing, atomic force microscopy (image)

Acknowledgement This research was supported by grant MOE2014-T2-1-001 of Singapore’s Ministry of Education.

References

(22) van Kan, J. A.; Shao, P. G.; Wang, Y. H.; Malar, P. Proton beam writing a platform technology for high quality three-dimensional metal mold fabrication for nanofluidic applications. Microsyst. Technol. 2011, 17, 1519–1523.
(23) van Kan, J. A.; Zhang, C.; Malar, P. P.; van der Maarel, J. R. C. High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies. Biomicrofluidics 2012, 6, 066502.