

CHALMERS



Investigations of a Confined DNA-Protein Complex

Erasmus Mundus Complex Systems Science M1 Project

Toby St Clere Smithe (920208-C296)

Department of Physics & Engineering Physics
Non-linear Dynamics & Statistical Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
July 10, 2015

Abstract

Coating double-stranded DNA (dsDNA) with the protein RecA increases the persistence length of the polymer so that the full conformations are accessible to light microscopy. Here, we analyse light microscopy recordings of nanochannel confinement experiments on the dsDNA/RecA complex to determine the dependence of the extension on contour length and channel width. We compare these results with simulations of a wormlike chain model for confined DNA, and find less backfolding in the experiments than predicted. We hypothesize that this is an artefact of the conformational kinetics and the experimental duration, and therefore investigate the nucleation time of hairpin bends.

Acknowledgements

I am immensely grateful to my supervisor, Bernhard Mehlig, for providing such interesting problems on which to work, and for the support and guidance that has meant that this report covers only one part of the work that has been done in the duration of this project. It is a great testament to said support that I, with no prior formal training in physics, have managed all of this.

I would also like to thank Erik Werner, for countless clarifications and pieces of advice on many aspects of this work, and Vitalii Iarko, for working beside me.

Finally, this project would not have been the same without the work of Kevin Dorfman, who provided the simulations against which the experimental results are compared, and who had the insight to derive hairpin nucleation times from Odijk's theory.

Toby St Clere Smithe, Finistère, France, July 2015

Contents

1	Introduction	1
2	Methodology	4
2.1	Experimental setting and microscopy	4
2.2	Image analysis	5
2.3	Simulations	5
3	Results and discussion	6
3.1	Hairpin nucleation time	11
3.2	Conclusions	14
A	Kymographs	15
A.1	130603-RecA-T4-wide-2	15
A.2	130603-RecA-T4-wide-3	16
A.3	130603-RecA-T4-wide-4	16
A.4	130603-RecA-T4-wide-5	17
A.5	130603-RecA-T4-wide-7	18
A.6	130607-RecA-T4-narrow-10	19
A.7	130607-RecA-T4-narrow-7	20
A.8	130607-RecA-T4-narrow-8	21
A.9	130607-RecA-T4-narrow-9	22
A.10	130607-RecA-T4-wide-1	23
A.11	130607-RecA-T4-wide-2	24
A.12	130607-RecA-T4-wide-3	25
A.13	130607-RecA-T4-wide-4	26
A.14	130607-RecA-T4-wide-5	27
A.15	130607-RecA-T4-wide-6	28
A.16	130905-RecA-T4-narrow-1	29
A.17	130905-RecA-T4-narrow-2	30
A.18	130905-RecA-T4-narrow-3	31
A.19	130905-RecA-T4-narrow-4	31
A.20	130905-RecA-T4-narrow-5	32
A.21	130905-RecA-T4-narrow-6	32
A.22	130905-RecA-T4-narrow-7	33

A.23	130905-RecA-T4-wide-1	34
A.24	130905-RecA-T4-wide-2	35
A.25	130905-RecA-T4-wide-3	36
A.26	130905-RecA-T4-wide-4	37
A.27	130905-RecA-T4-wide-5	38
A.28	130905-RecA-T4-wide-6	39
A.29	130905-RecA-T4-wide-7	40
A.30	130905-RecA-T4-wide-8	41
A.31	130924-RecA-lambda-narrow-1	42
A.32	130924-RecA-lambda-narrow-2	43
A.33	130924-RecA-lambda-narrow-3	44
A.34	130924-RecA-lambda-narrow-4	45
A.35	130924-RecA-lambda-narrow-5	46
A.36	130924-RecA-lambda-narrow-6	47
A.37	130924-RecA-lambda-wide-1	48
A.38	130924-RecA-lambda-wide-10	49
A.39	130924-RecA-lambda-wide-2	50
A.40	130924-RecA-lambda-wide-3	51
A.41	130924-RecA-lambda-wide-4	52
A.42	130924-RecA-lambda-wide-5	53
A.43	130924-RecA-lambda-wide-8	54
A.44	130924-RecA-lambda-wide-9	55
B	Classification of conformations and experimental conditions	56
B.1	Definitions	56
B.2	Summary	56
B.2.1	Main findings	56
B.2.2	Notes on experimental conditions	57
B.2.3	How does the initial condition affect the subsequent conformations?	57
B.2.4	Are (experimentally-induced) stuck ends more likely wide to narrow or vice versa?	58
B.2.5	Do folds preferentially form in the direction of the drift, or of the movement from channel to channel?	58
B.2.6	Which is faster or more likely, folding or unfolding?	59
B.2.7	How do the fold rates vary with channel width?	59
B.3	Tables	59
C	Example configurations	70
C.1	Rejected configurations	70
C.2	Accepted configurations	71
D	Extra figures	74
D.1	Plots for T4 DNA only	74
D.2	Plots for λ DNA only	77
	Bibliography	81

1

Introduction

The conformational statistics of DNA have been widely studied, with interest deriving from both theoretical and experimental physics and biology. DNA exhibits a great variety of conformations in biological settings, such as its coiled state in chromatin or its melted state during transcription, but its conformations are also of interest in themselves. Owing to highly-developed experimental techniques and its own experimentally useful properties [1], DNA has been used historically as a model polymer¹.

Confinement of DNA to nanofluidic channels provides the opportunity to study in detail the conformations of individual molecules [2], as well as the potentially exciting biological application of linear sequencing using optical mapping [3]. Here, we investigate the statistics of a DNA-protein complex that is accessible to light microscopy in two dimensions.

The conformation adopted by a confined polymer depends on the relation of physical properties of the polymer to the dimensions of the confinement. The most important physical properties in the context of this study are the persistence length ℓ_P , the contour length L , and the extension X . The extension can be defined relative to any basis vector, but here we are interested in the extension relative to the length of the confining channel: in this context, X is the greatest distance along the channel between two points on the path of the polymer. The contour length is just the total length of the polymer molecule: in terms of DNA, it can be thought of as the number of base pairs times the ‘height’ of each base pair.

The persistence length is the distance along the path of the polymer over which its orientation becomes uncorrelated. If $\hat{\mathbf{t}}(s)$ is the unit tangent vector to the curve of the polymer at distance s along its path, then the persistence length determines the decay of the tangent-tangent correlation function [3]:

$$\langle \hat{\mathbf{t}}(0) \cdot \hat{\mathbf{t}}(s) \rangle = \exp\left(-\frac{s}{\ell_P}\right) \quad (1.1)$$

Where a dimension of the confinement is much smaller than the persistence length, one should expect only small fluctuations along that dimension: intuitively, the stiffness of the

¹There is evidence that the use of (uncoated) DNA as a model polymer is less applicable than previously believed, because it does not exhibit the requisite ‘universality’ of behaviour except at contour lengths well beyond those that are experimentally common [1]. Nonetheless, DNA has many other useful properties, such as its straightforward length selectivity, and well-understood biology. Moreover, it is the subject of the present study.

polymer means that it has not enough space along the dimension to bend back along itself. Notably, we can consider the persistence length as *defining* the flexibility of the polymer, a model which is known as the *wormlike chain model* [4]. A related quantity that will be useful later is the *global persistence length*, g , which measures the distance between hairpin bends along the polymer [5]. A *hairpin* bend is a bend in the contour of the polymer that covers at least the arc of one semicircle.

The persistence length of double-stranded DNA (dsDNA) at 300 K, and in a buffered solution of ionic strength greater than 10 μmol , is approximately 50 nm [3], which is much shorter than the wavelength of any visible light. Consequently, light microscopy is unavailable for studying the conformations of confined plain dsDNA. One effect of coating DNA with the recombinase enzyme RecA is a substantial increase in the stiffness of the polymer, as quantified by the persistence length. In agreement with other estimates [6], Frykholm *et al* [7] report the persistence length of RecA-coated dsDNA to be 1.15 μm , which is therefore accessible to light microscopy. It is this DNA-protein complex that is our experimental subject.

In this study, we analyse results from a confining channel of rectangular cross-section, with length much longer than the contour length of the DNA complex, height $D_H = 140$ nm much less than the persistence length (so as to negate fluctuations in that direction), and varying width D_W . The range of widths spans from approximately half of one persistence length (600 nm) to about three persistence lengths (3000 nm). Assuming the polymer has a constant contour length $L > \ell_P$, then as the channel width increases beyond ℓ_P (that is, as the channel in that dimension leaves the *Odijk regime* [8]), we expect an increase in the folding of the polymer, and therefore a corresponding decrease in the extension.

This expectation is quantified by simulations of a discretized wormlike chain, provided by Kevin Dorfman (private communication). The discretization is made using a touching bead model, with $N + 1$ beads of diameter d connected by rigid bonds. A bending potential determines the stiffness:

$$U(\theta_1, \dots, \theta_N) = k_B T \kappa \sum_{n=1}^{N-1} (1 - \cos \theta_n) \quad (1.2)$$

where n ranges over bonds, θ_n is the angle between bonds n and $n + 1$, and κ is known as the ‘bending constant’.

The persistence length in this model is defined as

$$\ell_P = d \sum_{k=0}^{\infty} \langle \mathbf{t}_n \mathbf{t}_{n+k} \rangle \quad (1.3)$$

where \mathbf{t}_n is the tangent vector to the chain at bond n . The persistence length can then be written in terms of κ and d [1]:

$$\frac{\ell_P}{d} = \frac{\kappa - 1 + \kappa \coth \kappa}{2(\kappa + 1 - \kappa \coth \kappa)} \quad (1.4)$$

Given simulations of this model, we are interested in its predictions for the statistic $\langle X \rangle / L$ as a function of contour length, and hence compare these results with an analysis of microscopy recordings of the experiments; the average is taken over time in the experiments, and, in the simulations, over the simulated ensemble. We find poor agreement, and suggest the following explanation: the experiments analysed were of insufficient duration, and so the

statistics obtained do not represent the simulated ensemble. We corroborate this explanation by considering the folding evidenced in kymographs derived from the microscopy recordings, and compare the experimental hairpin nucleation times with predictions derived by Dorfman (private communication) using a theory from Odijk [5].

2

Methodology

The experimental microscopy recordings were produced by Karolin Frykholm and the worm-like chain simulations by Kevin Dorfman. The corresponding methods are summarised below, and a description of the analysis methods for the microscopy data is given.

2.1 Experimental setting and microscopy

The experiments presented here used lipid-passivated channels, fabricated as described in [2]. dsDNA was coated in fluorescent RecA protein, and the RecA nucleoprotein filaments were funneled from a channel of one width to a channel of another width, in increasing or decreasing order. In this way, it is possible to study the changing conformation of the filament as the channel width changes [7]. The channel dimensions were as given in the Introduction: a height of 140 nm and widths of 600 nm, 900 nm, 1300 nm, 1800 nm, 2400 nm, and 3000 nm.

After moving the filament into a given channel, microscope recordings of the filament were made while constant pressure was maintained in the nanochannel. One ‘video’ of 400 frames was recorded for the different channel widths for each filament; the interval between frames was 0.11 s.

Both T4 and λ DNA were used, with the contour lengths of the T4 DNA complex ranging from 7 microns to 21 microns, and the λ DNA complex ranging from 11 microns to 23 microns; the different DNA types do not produce qualitatively different results under the analysis presented here.

Each experiment on a single nucleoprotein filament is given a codename, such as “130607-RecA-T4-narrow-9”. These codenames will be used below to refer to the individual experiments. In the experiments tagged ‘narrow’ (as in the example here), the molecules start at the narrowest channel, moving to wider channels, whilst in those experiments tagged ‘wide’, the reverse occurs.

A detailed investigation of the conditions of each experiment is presented in Appendix B, in order to elucidate the effects of initial condition and experimental error on the conformations, and to quantify the rate of folding observed.

2.2 Image analysis

To estimate the extensions, X , of the RecA/DNA filaments, the frames from the microscopy videos were first passed through a median filter with a radius of 2, and then the maximum pixel intensity in each column was taken to provide an intensity profile along the channel. Next, to smooth out these profiles and remove noise, a moving average (in time) of window size 3 frames and then (in space) of window size 4 pixels was computed using these intensity profiles. Finally, an error function ‘box’ curve of the form

$$\alpha + \beta (\operatorname{erf}[\gamma(z - \delta)] + \operatorname{erf}[\epsilon(\zeta - z)]) , \quad (2.1)$$

where the Greek letters denote constants to be fitted and z denotes position along the profile, was least-squares fitted to each smoothed profile. This function was then averaged along the channel to compute the extension. Unsmoothed intensity profiles were also used to produce kymographs, by stacking the profiles into a column, so that each row of the column represents a single frame of the experimental recording.

To estimate the contour lengths of the RecA/DNA filaments, a random subset of frames from the microscopy videos corresponding to each filament was presented sequentially on screen. Frames were selected from the recordings of the molecule in all the channels for which data were available. The contour lengths were then computed by tracing a path over the molecule filament on each frame using the computer mouse and then calculating the physical length of this path, until an adequate number of paths (at least 40) was sampled. The mean and standard deviation of these paths were then computed for each molecule in the data set.

There were two rejection criteria for the data. Firstly, if the kymographs suggested experimental error (such as a ‘tangled’ molecule or a molecule becoming folded during the transition between channels of different widths), which was then supported by inspection of the microscope images; the results of this inspection are given in Appendix B. Secondly, data were rejected if the standard deviations of either $\langle X \rangle / L$ or L exceeded 40% of the computed values; this occurred for a total of 10 data points across 6 experiments. To systematize the application of the first rejection criterion, all the microscopy videos were carefully investigated and their initial conditions and conformation changes categorized; see the appendices for these details. Examples of rejected configurations are given in Appendix C.

Finally, to estimate the nucleation times for hairpin bends, we assume that the nucleation can be modelled as a homogeneous Poisson process, over an interval of 100 frames. The maximum-likelihood estimator for the rate parameter of a Poisson distribution is just the mean of the rates observed in the sample. Hence, to estimate the hairpin nucleation times for each channel width, estimates for the rate parameters were first obtained by counting the numbers of nucleations in each set of 100 frames for each channel width. These rate estimates were then converted into time estimates by taking the reciprocal.

2.3 Simulations

Dorfman *et al* made simulations of the wormlike chain model in the regime of interest, using the pruned-enriched Rosenbluth method, as described in [9] and [1]. For each channel size, the simulated data consist of 25 million tours.

3

Results and discussion

Having excluded the clearly experimentally suspicious cases, of 65697 remaining microscopy frames, 61345 were accepted according to the criteria described above (93.4%). Some examples of ‘borderline suspicious’ but nonetheless accepted configurations are discussed below in Appendix C. For the different channel widths, the acceptance rates are: 98.9% (600nm), 99.9% (900nm), 96.1% (1300nm), 91.6% (1800nm), 86.0% (2400nm), and 90.3% (3000nm).

Plots of $\langle X \rangle / L$ against L / ℓ_P were produced, in order to compare with the simulated results. The amount of fluctuation of the filament is quantified by the ratio $\langle X \rangle / L$: where this value is near 1, the filament is mostly straight, and smaller values indicate progressively larger fluctuations or backfolding.

Some results show $\langle X \rangle > L$, and so it is important to note that the quantity by which $\langle X \rangle / L > 1$ is less than one standard deviation of $\langle X \rangle / L$. The standard deviations are shown in Figures 3.3 to 3.8.

The accepted data for both DNA types are plotted in Figure 3.1, which has axes to match the simulated data, plotted in Figure 3.2. Figures 3.3 to 3.8 show the data and simulations plotted on log axes separately for each channel width. The two DNA types are presented separately in the extra figures given in Appendix D.

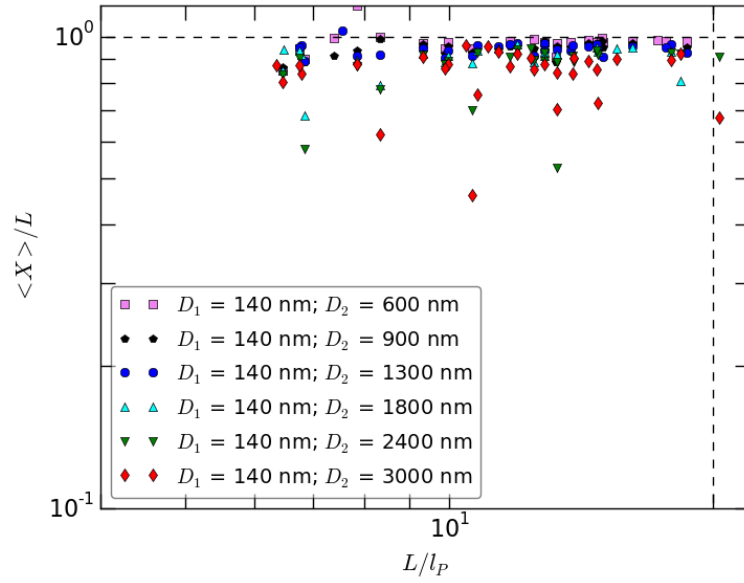


Figure 3.1: Experimental data for both T4 and λ DNA on log axes. The dashed vertical line is the dashed line in 3.2.

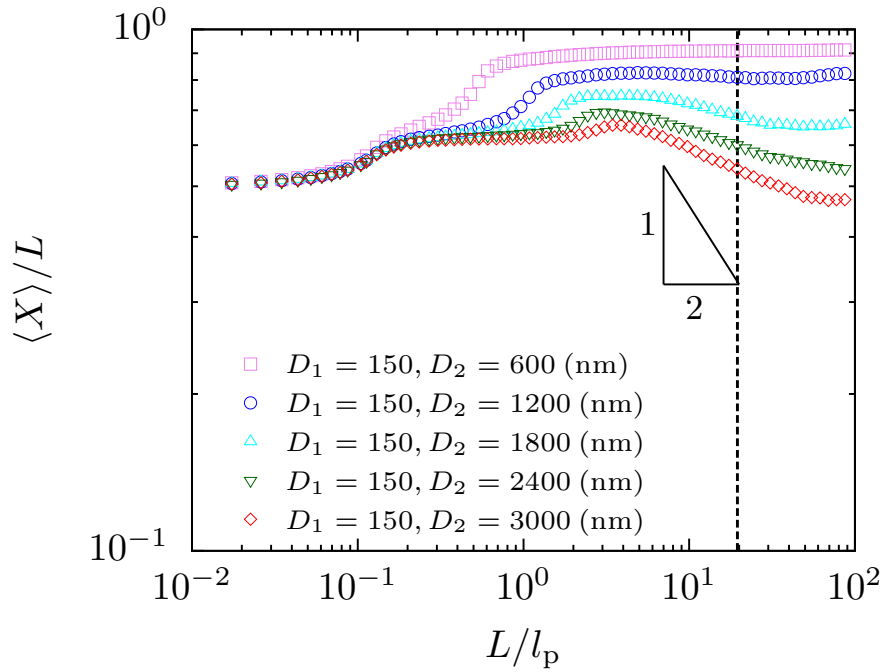


Figure 3.2: Simulated data, from Dorfman.

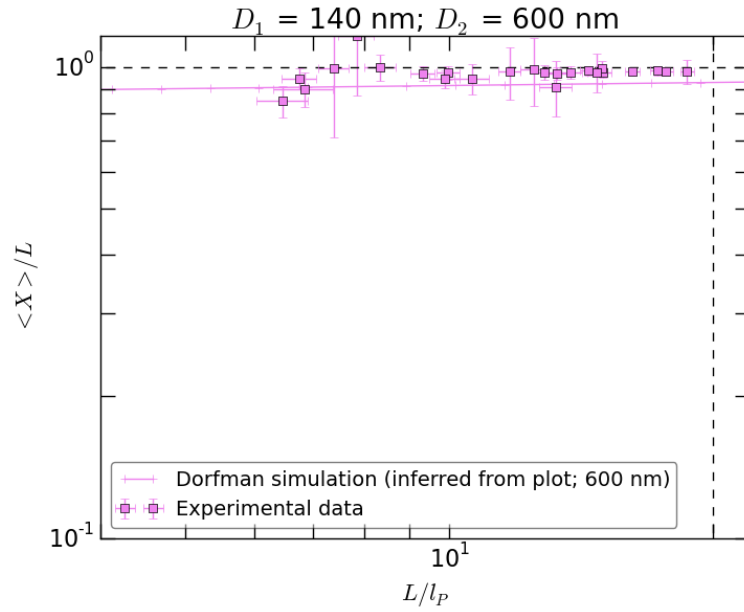


Figure 3.3: 600nm channel; log axes. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

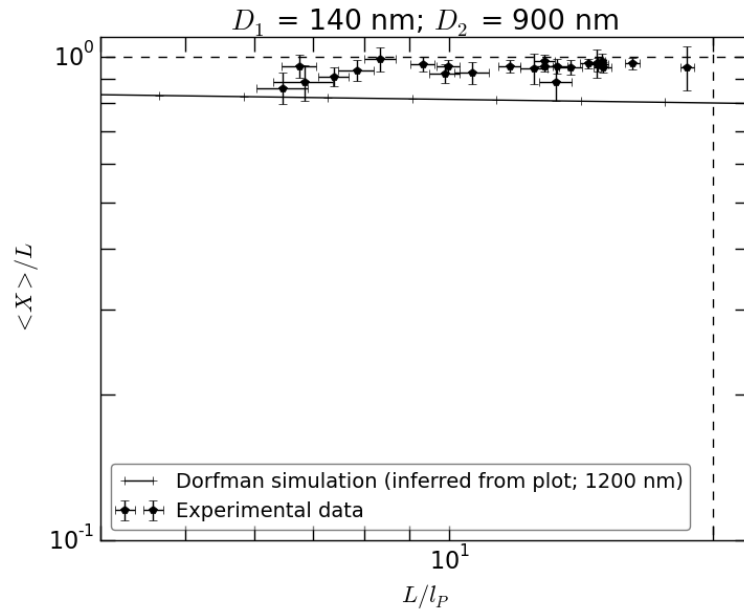


Figure 3.4: 900nm channel; log axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

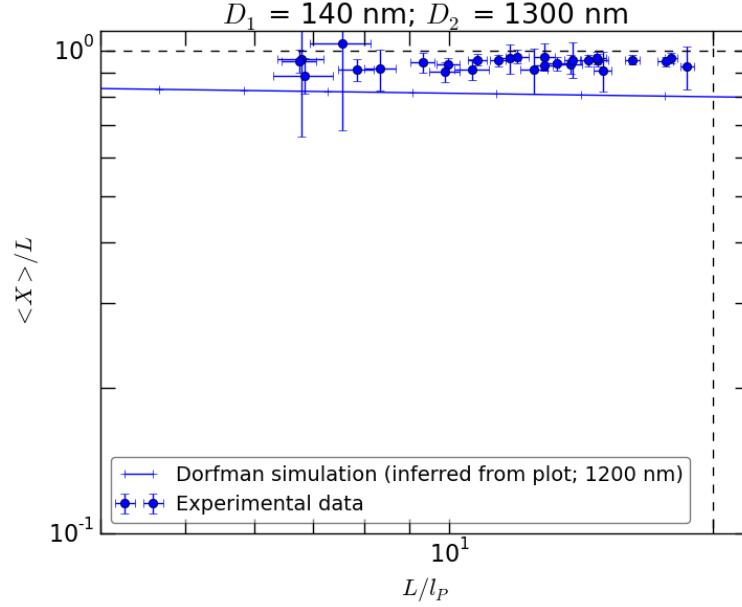


Figure 3.5: 1300nm channel; log axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

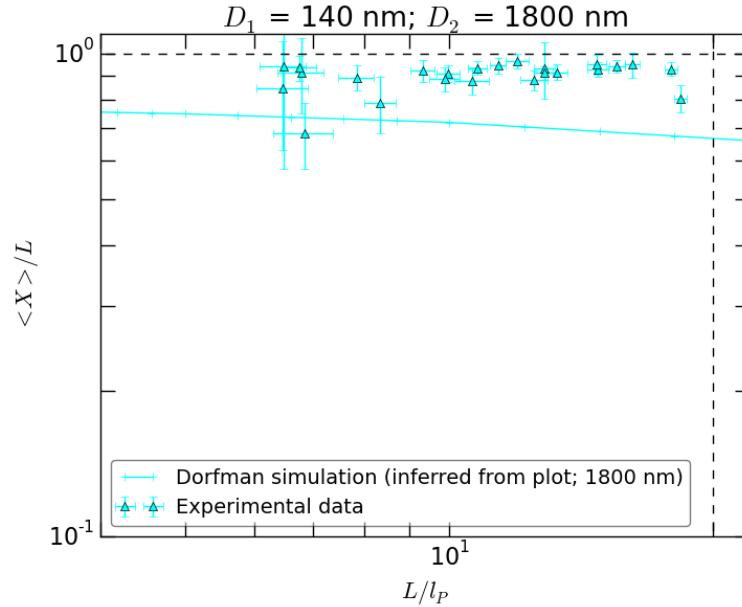


Figure 3.6: 1800nm channel; log axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1800nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The point with lowest $\langle X \rangle / L$ value is 130607-RecA-T4-narrow-9 ($L / \ell_P = 6.83$), which is discussed in Appendix C.

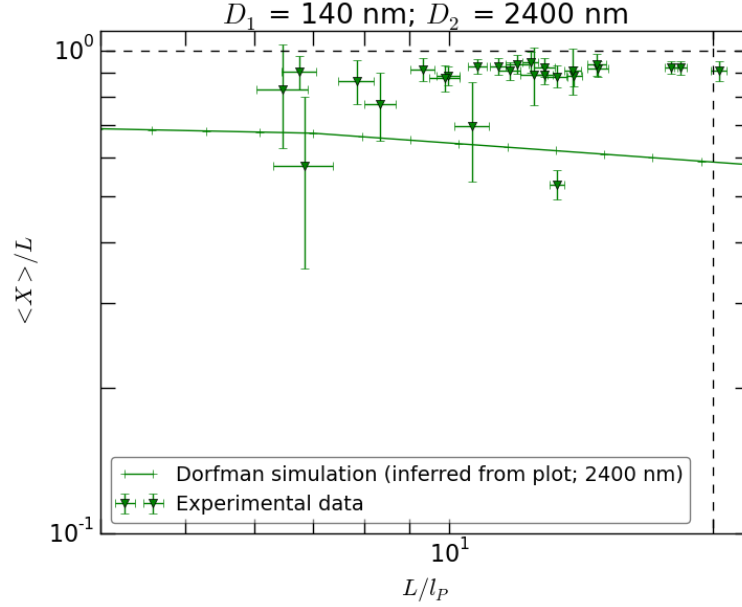


Figure 3.7: 2400nm channel; log axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 2400nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The two lower values are 130607-RecA-T4-narrow-9 ($L/\ell_P = 6.83$) and 130905-RecA-T4-narrow-7 ($L/\ell_P = 13.28$); they are discussed in Appendix C.

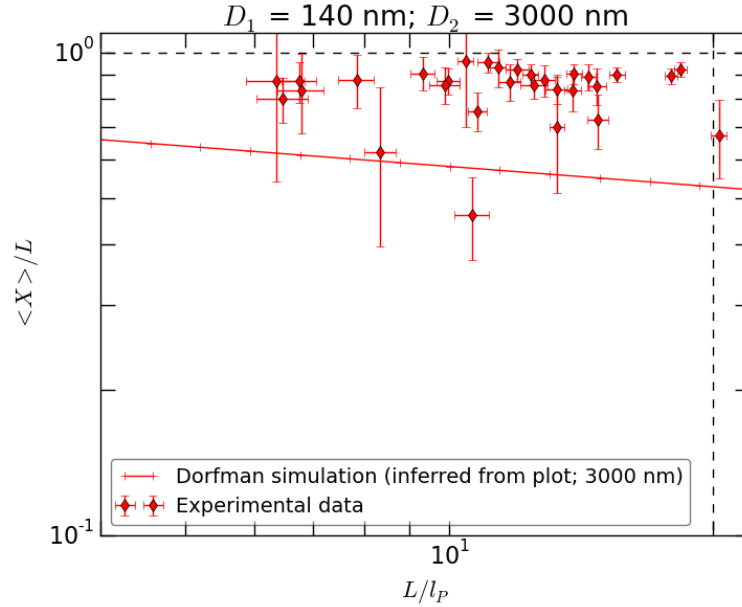


Figure 3.8: 3000nm channel; log axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 3000nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The point with lowest $\langle X \rangle / L$ value is 130607-RecA-T4-narrow-10 ($L/\ell_P = 10.62$), which is discussed in Appendix C.

A slight increase in the size of the fluctuations is visible given increasing channel width, but not of the same magnitude as predicted by simulations. Moreover, the negative slope for the larger channel sizes seen in Figure 3.2 is not visible in the experimental data: this slope represents the increased possibility of backfolding given a length of numerous multiples of the persistence length, and sufficient space in which to make those folds.

3.1 Hairpin nucleation time

We believe that the discrepancy presented here is a consequence of the conformational kinetics. The microscopy videos record at most 45 seconds which, given the frequency of conformational changes seen in the kymographs (Appendix A), is inadequate to obtain statistics in agreement with the equilibrium distributions simulated by Dorfman; moreover, only approximately ten seconds is taken after moving a molecule into a given channel and before starting the data collection, which is likely insufficient for the molecule to reach equilibrium.

The kymographs show few folding events, and to confirm this we can consider the time series of the extension for each experiment. Two such time series are plotted below, for experiments in the 3000 nm channel.

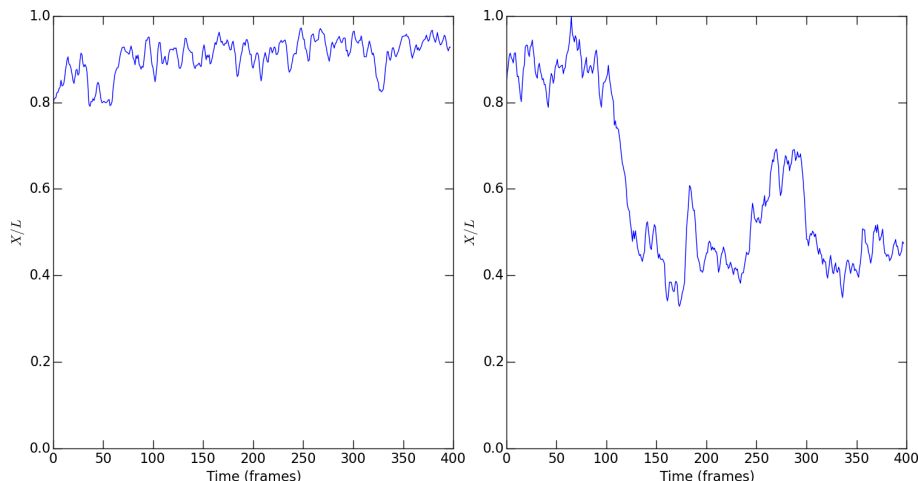


Figure 3.9: Two time series for the ratio $\langle X \rangle / L$, for experiments in the 3000 nm channel. Left: $L = 9\ell_p$. Right: $L = 11\ell_p$. Most time series observed are like the left panel.

In the left panel of Figure 3.9, we see small fluctuations in $\langle X \rangle / L$. Changes in the conformation, such as the nucleation of a hairpin, are seen as large, and usually sudden, changes in $\langle X \rangle / L$; this is seen in the right panel. Most of the experiments observed exhibit time series like that in the left panel.

The hypothesis is further corroborated when considering the simulated distributions that give the averages presented in Figure 3.2. Distributions for channel widths 1800nm, 2400nm and 3000nm are shown below.

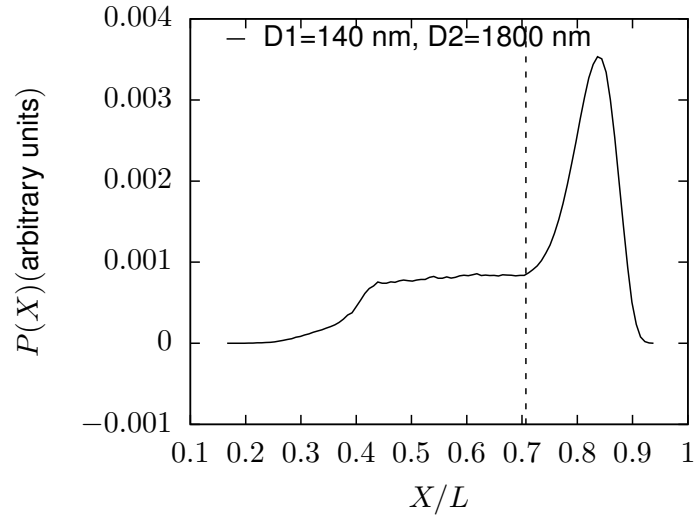


Figure 3.10: Simulated X/L distribution for channel width 1800nm; private communication from Dorfman. The dashed line gives the mean, $\langle X \rangle/L$.

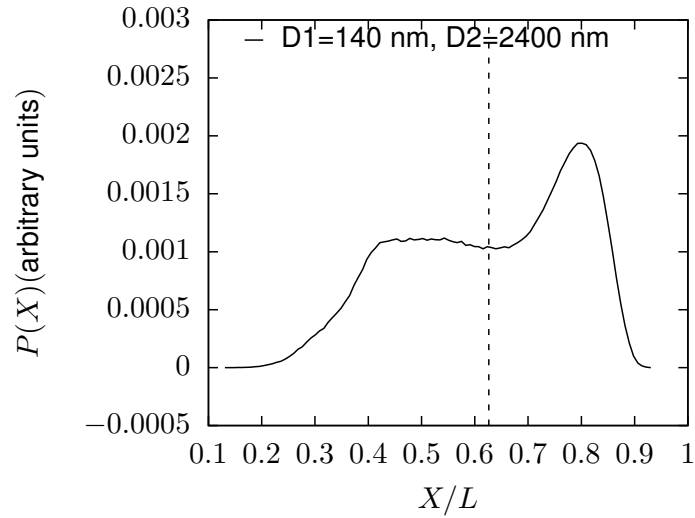


Figure 3.11: Simulated X/L distribution for channel width 2400nm; private communication from Dorfman. The dashed line gives the mean, $\langle X \rangle/L$.

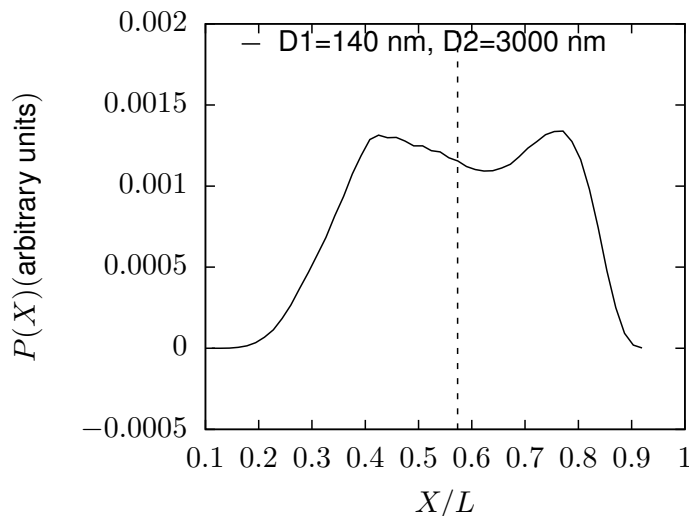


Figure 3.12: Simulated X/L distribution for channel width 3000nm; private communication from Dorfman. The dashed line gives the mean, $\langle X \rangle/L$.

In each of these distributions, we see a mode in the $\langle X \rangle/L$ region between 0.8 and 0.9. The mean is ‘left’ of this mode, as a consequence of the large tails; only the 3000nm distribution has a bimodal form. The long-tailed forms of these distributions, except in the 3000nm case, suggest that there is a large entropic cost associated with the formation of the first hairpin bend. Moreover, if we imagine cutting off some of the density in the tails and sampling from the resulting distributions, we will obtain averages more like the experimental results presented here. Qualitatively, therefore, these distributions appear to corroborate the kinetic hypothesis.

The classification of conformations and experimental conditions given in Appendix B seems to support these conclusions. In particular, we see that, even in the widest (3000nm) channel, folds that significantly change the extension of the molecule occur only 0.53 times every 100 frames on average. Most of the recordings show no folding, and an initial condition that does not indicate significant experimental influence.

More precisely, counting only hairpin nucleations as defined in the Methodology and excluding experimentally suspicious data as before, we count 28 hairpin nucleations in all the experimental time in the 3000nm channel, 7 in the 2400nm channel, 1 each in the 1800nm and 1300nm channels, and none in either the 900nm or 600nm channels. This corresponds to nucleation times of approximately 35 s in the 3000nm channel, 160 s in 2400nm, 1000 s in 1800nm, and 1280 s in 1300nm.

Dorfman (private communication) made preliminary estimates of the nucleation times using global persistence length theory from Odijk [5], and predicted nucleation times of 18 s, 20 s, 21 s, 26 s, 50 s and 210 s for the 3000nm, 2400nm, 1800nm, 1300nm, 900nm and 600nm channels respectively.

Once again, more folding is predicted than is observed. It is not clear at this stage whether the kinetic hypothesis explains this discrepancy, or whether the discrepancy arises for some other reason. Work is on-going to understand this.

3.2 Conclusions

An analysis of light-microscopy recordings of confined filaments of RecA/DNA complex was presented. The accessibility of this complex to light microscopy allows detailed study of the conformations of individual filaments. The average extension of the filaments given varying channel width was compared with simulations of a wormlike chain model, and a disagreement was observed. This disagreement is thought to originate as an experimental artefact: the kinetics of the RecA/DNA complex seem to be such that the experimental recordings were of insufficient duration to collect adequate data. Further evidence for this kinetic hypothesis was presented, in the form of a detailed analysis of the experimental recordings, a qualitative comparison with the simulated distributions, and a quantitative comparison with the hairpin nucleation time predictions of Odijk's global persistence length theory. There remains a further discrepancy between the hairpin nucleation time predictions and the experimental results, which is the subject of on-going investigation.

A

Kymographs

These kymographs were produced by stacking the intensity profiles produced for each sequence of frames as described in the Methodology.

A.1 130603-RecA-T4-wide-2

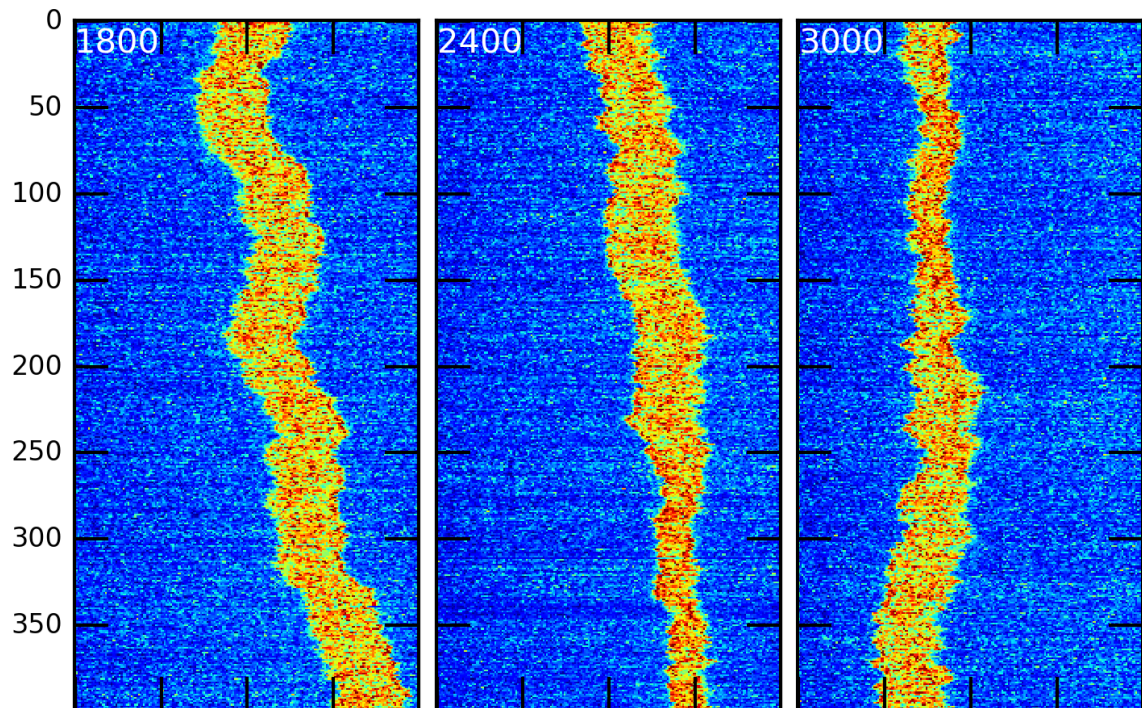


Figure A.1: Vertical axis is time. White annotation gives channel width in nanometres.

A.2 130603-RecA-T4-wide-3

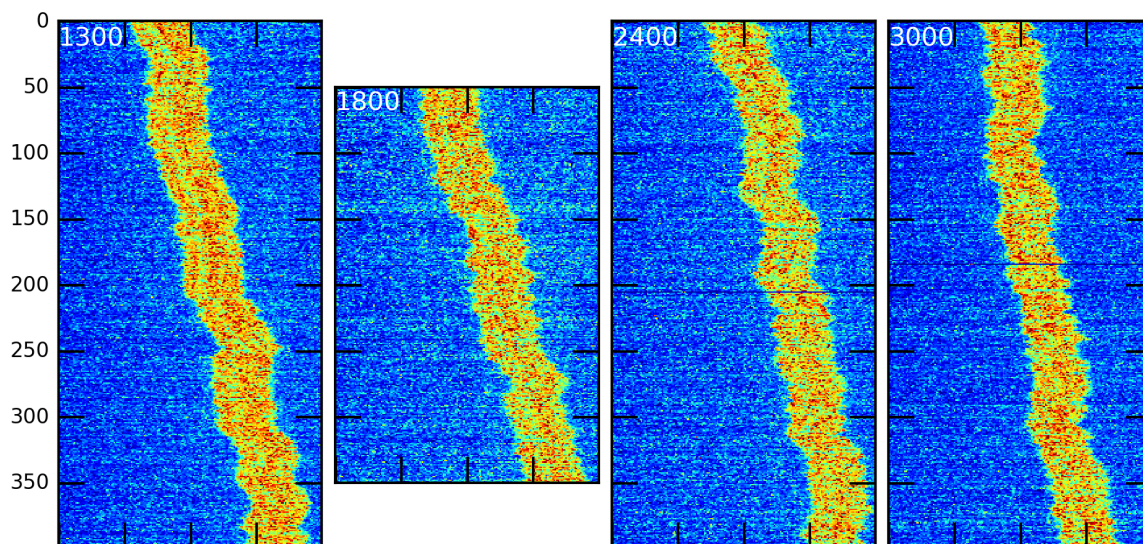


Figure A.2: Vertical axis is time. White annotation gives channel width in nanometres.

A.3 130603-RecA-T4-wide-4

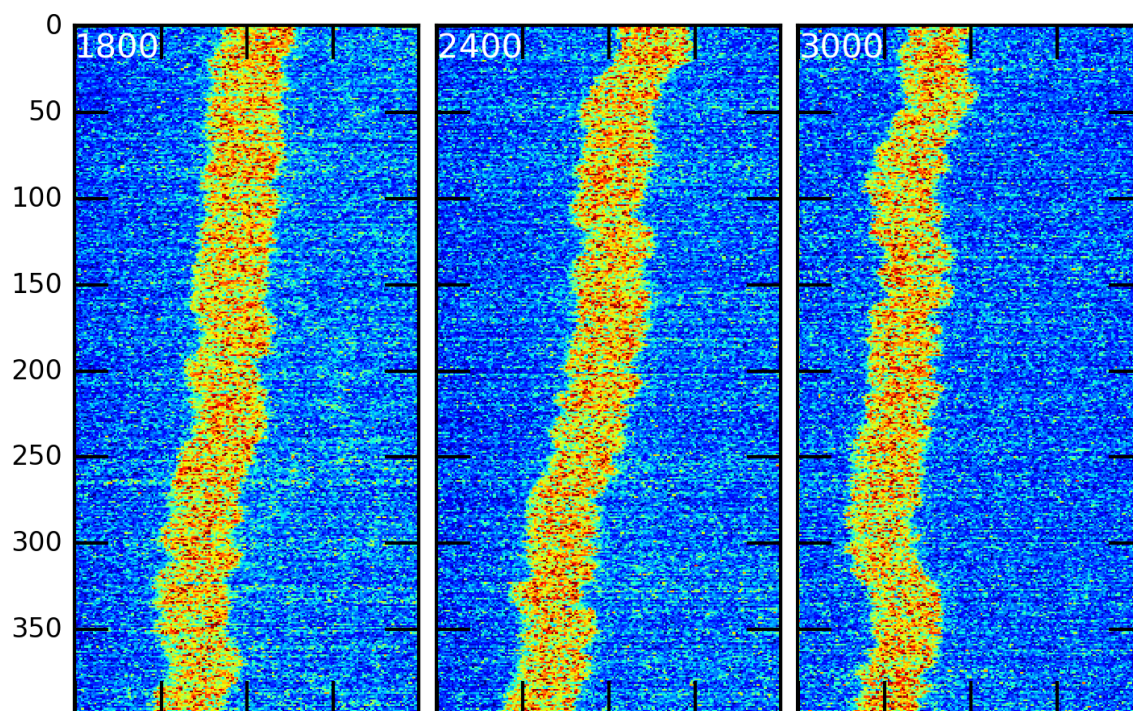


Figure A.3: Vertical axis is time. White annotation gives channel width in nanometres.

A.4 130603-RecA-T4-wide-5

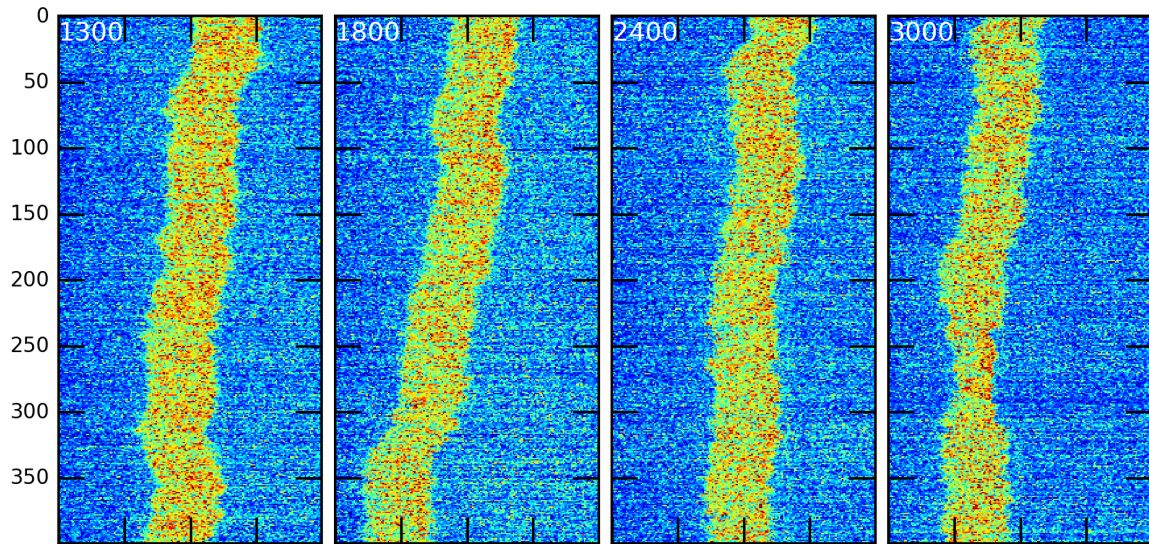


Figure A.4: Vertical axis is time. White annotation gives channel width in nanometres.

A.5 130603-RecA-T4-wide-7

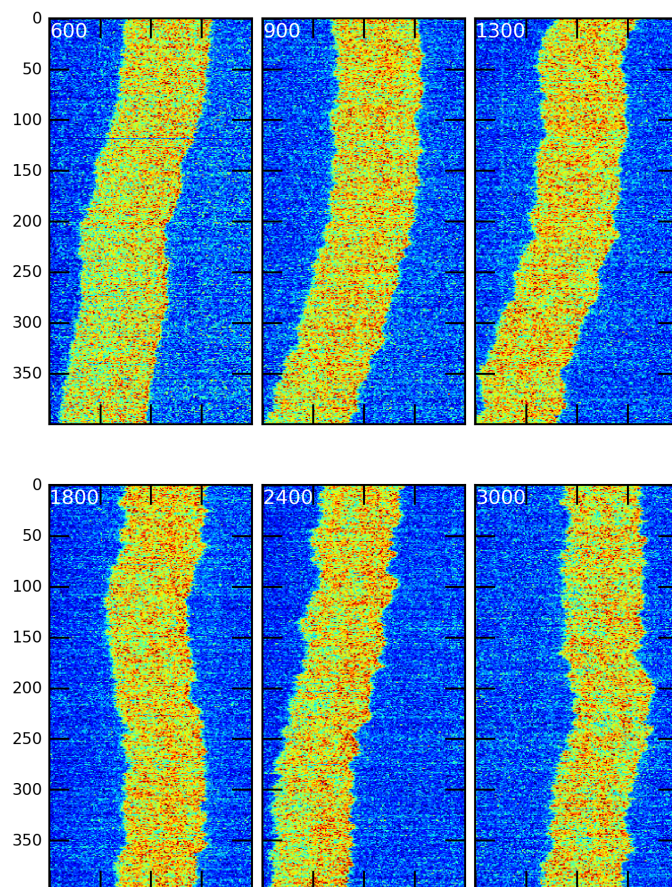


Figure A.5: Vertical axis is time. White annotation gives channel width in nanometres.

A.6 130607-RecA-T4-narrow-10

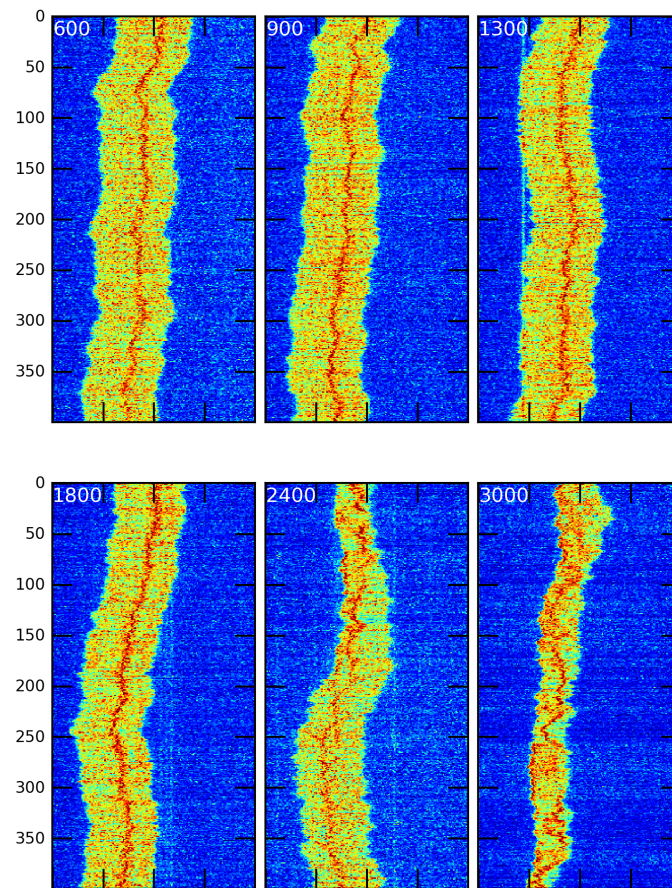


Figure A.6: Vertical axis is time. White annotation gives channel width in nanometres.

A.7 130607-RecA-T4-narrow-7

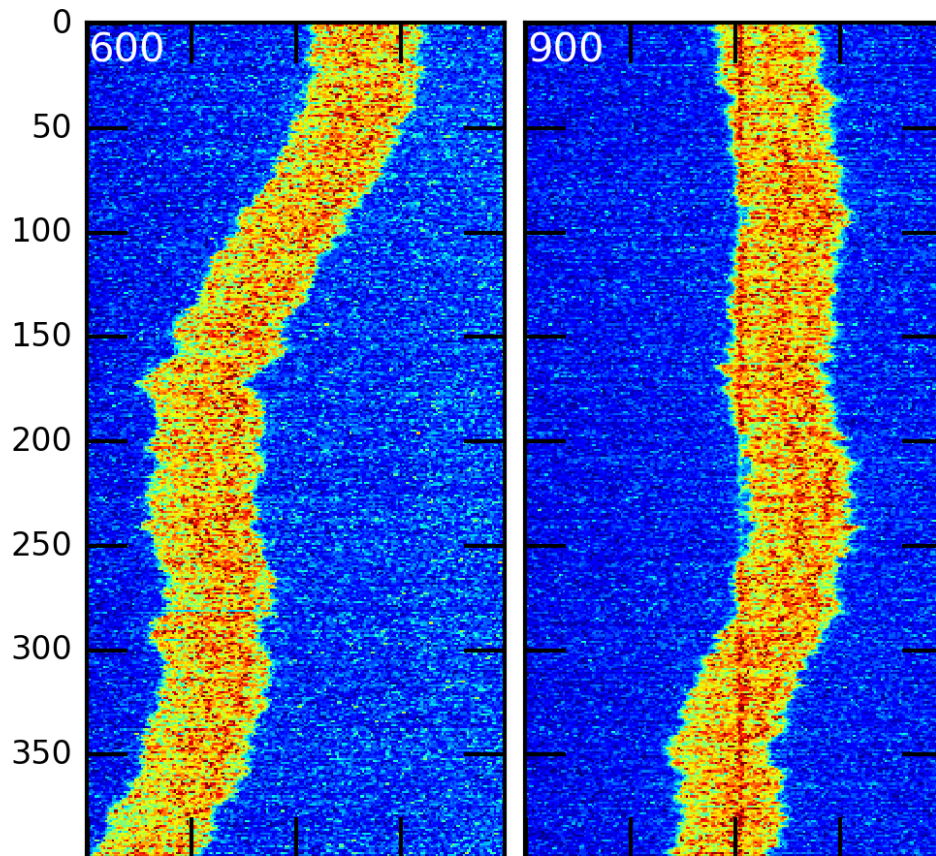


Figure A.7: Vertical axis is time. White annotation gives channel width in nanometres.

A.8 130607-RecA-T4-narrow-8

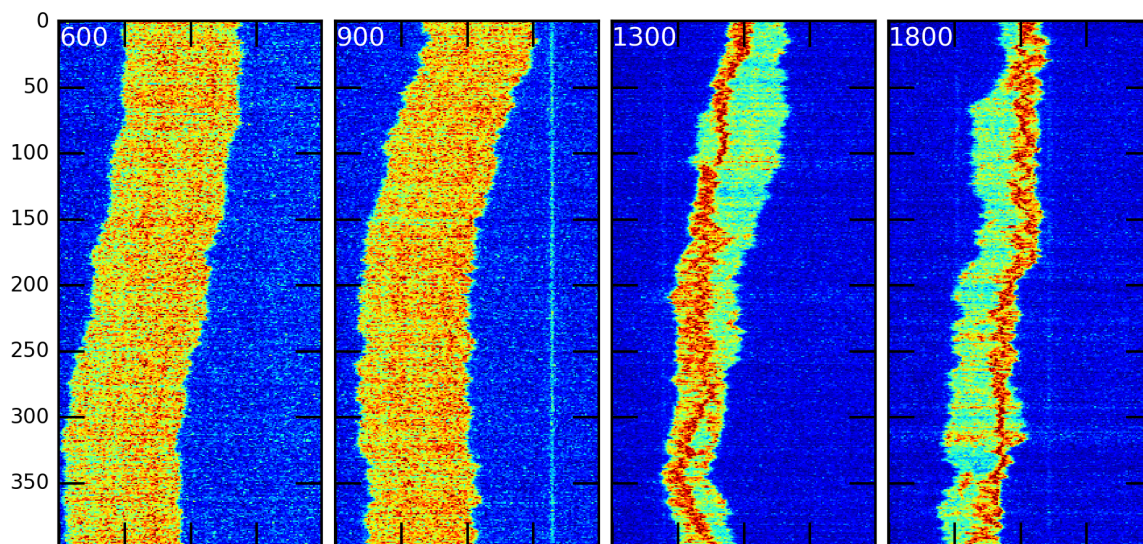


Figure A.8: Vertical axis is time. White annotation gives channel width in nanometres.

A.9 130607-RecA-T4-narrow-9

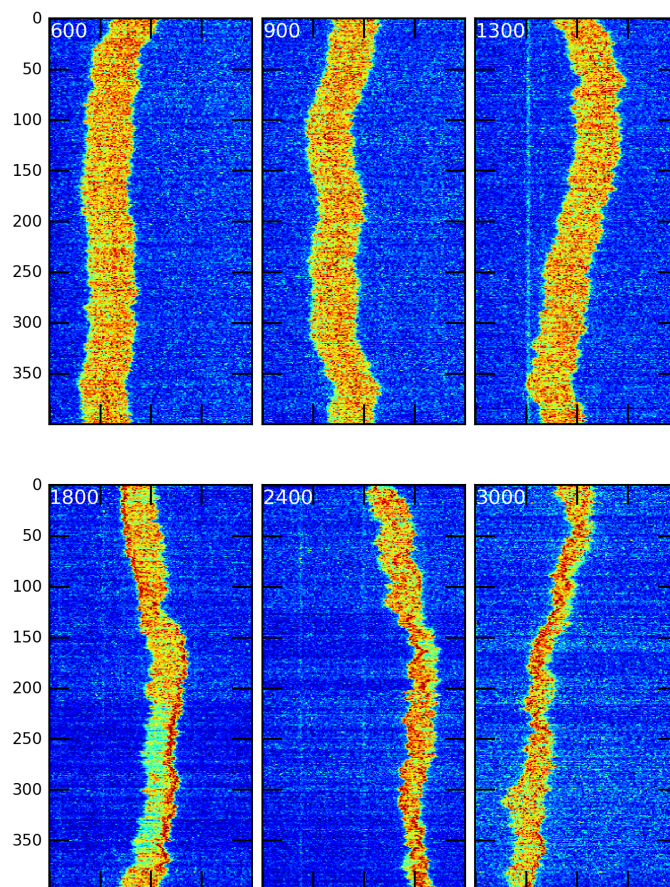


Figure A.9: Vertical axis is time. White annotation gives channel width in nanometres.

A.10 130607-RecA-T4-wide-1

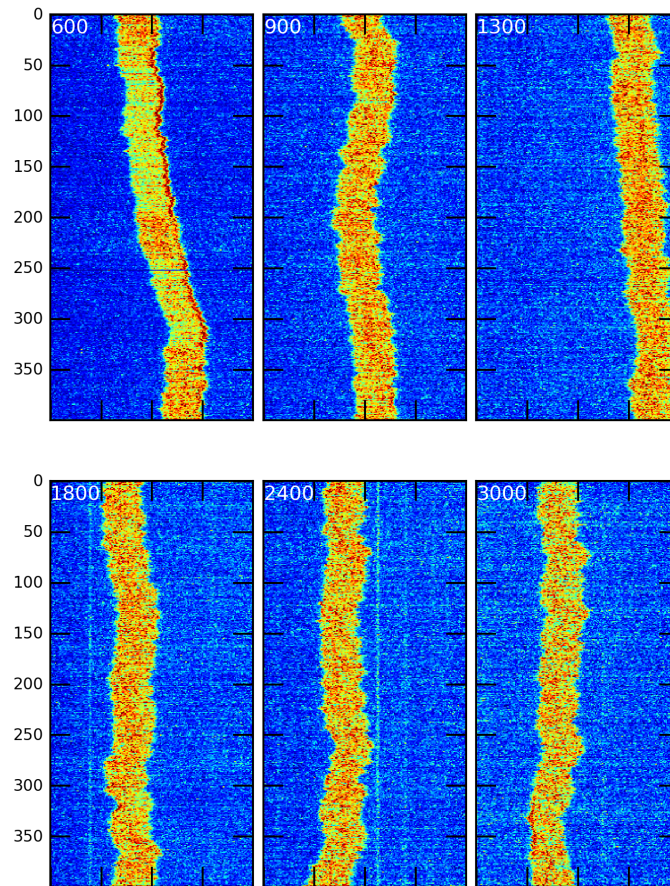


Figure A.10: Vertical axis is time. White annotation gives channel width in nanometres.

A.11 130607-RecA-T4-wide-2

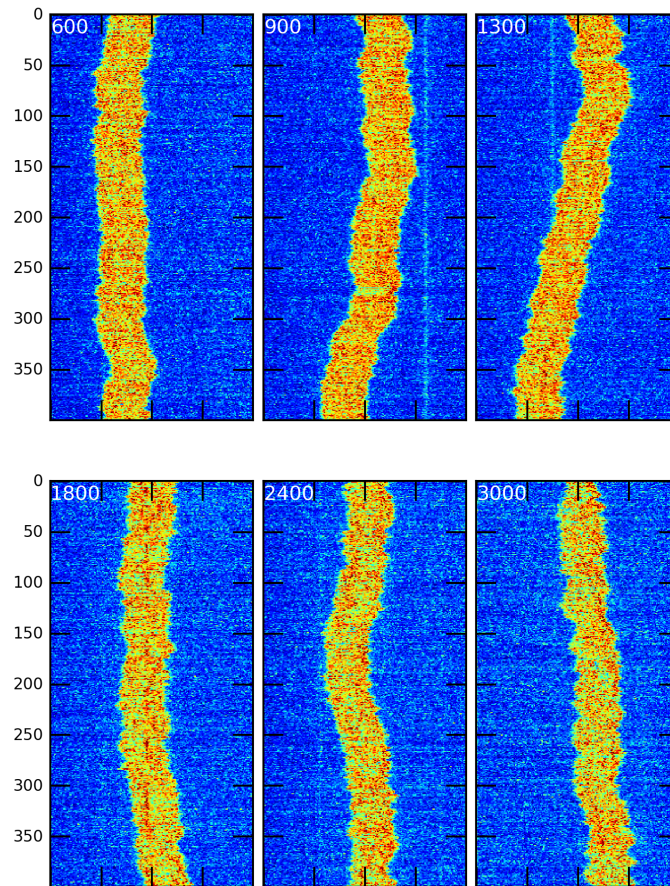


Figure A.11: Vertical axis is time. White annotation gives channel width in nanometres.

A.12 130607-RecA-T4-wide-3

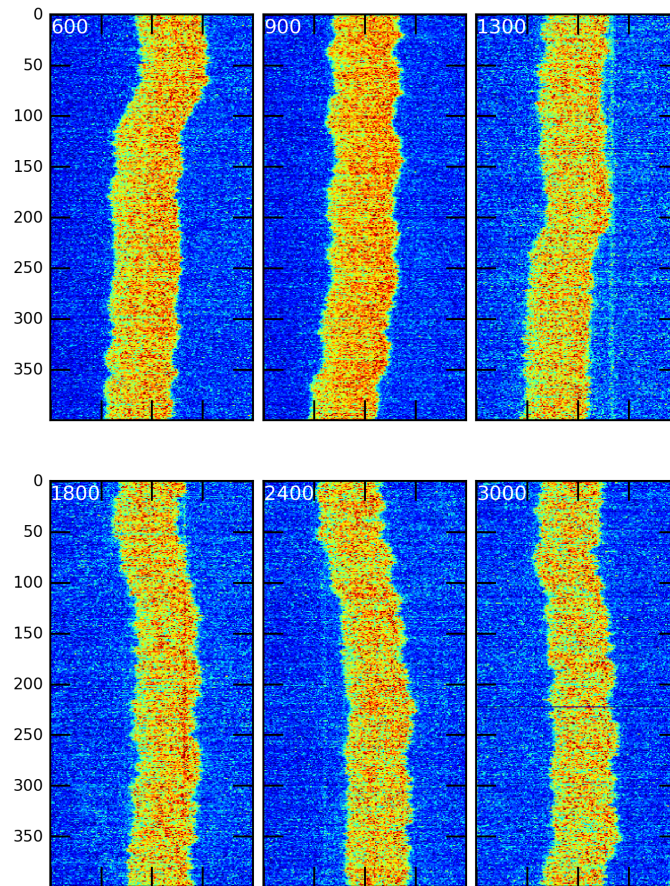


Figure A.12: Vertical axis is time. White annotation gives channel width in nanometres.

A.13 130607-RecA-T4-wide-4

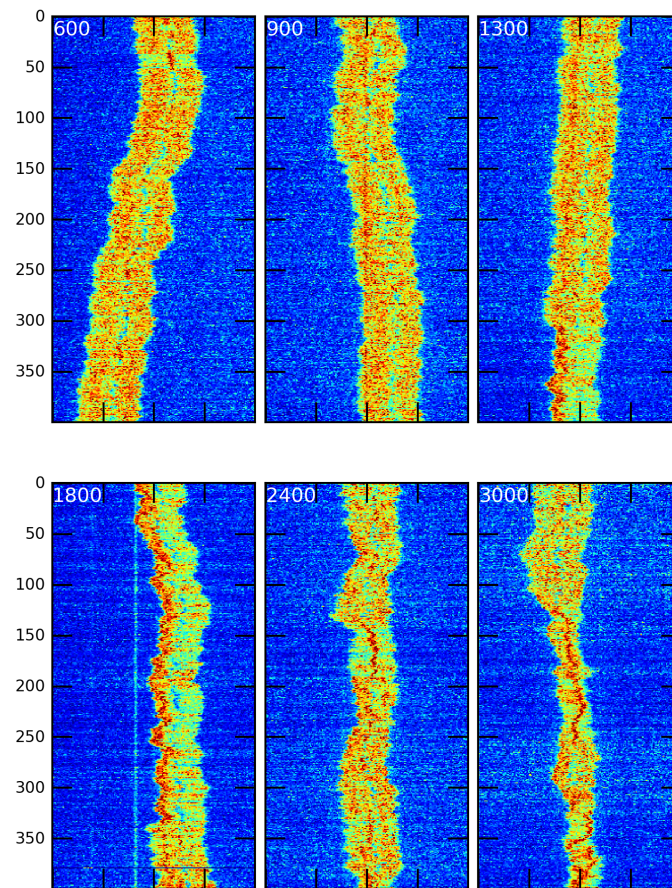


Figure A.13: Vertical axis is time. White annotation gives channel width in nanometres.

A.14 130607-RecA-T4-wide-5

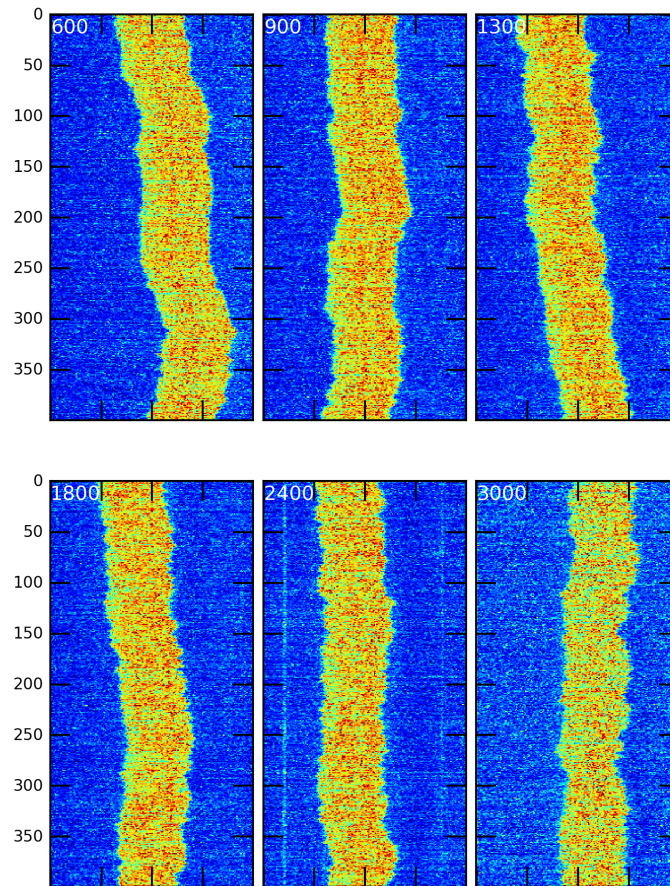


Figure A.14: Vertical axis is time. White annotation gives channel width in nanometres.

A.15 130607-RecA-T4-wide-6

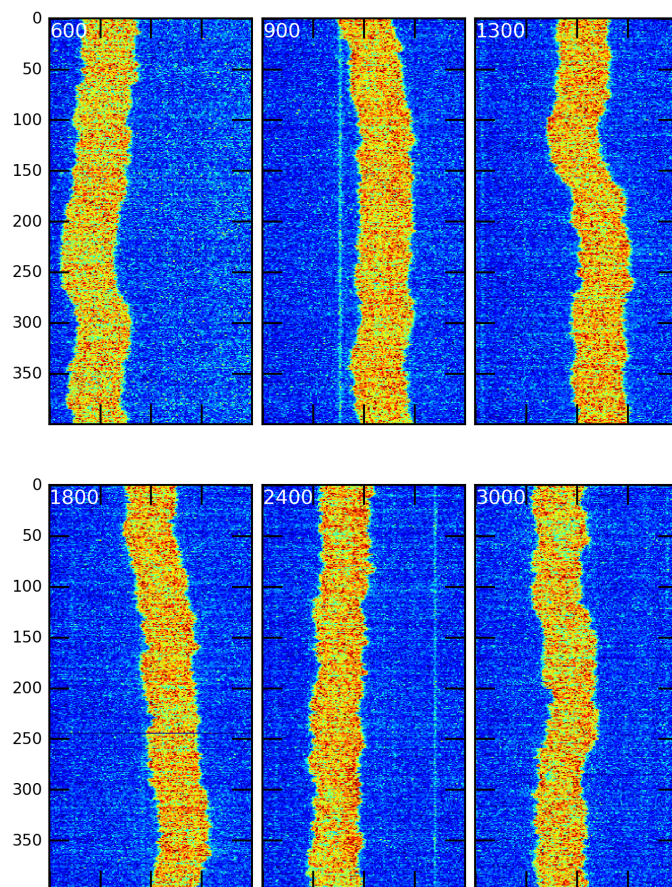


Figure A.15: Vertical axis is time. White annotation gives channel width in nanometres.

A.16 130905-RecA-T4-narrow-1

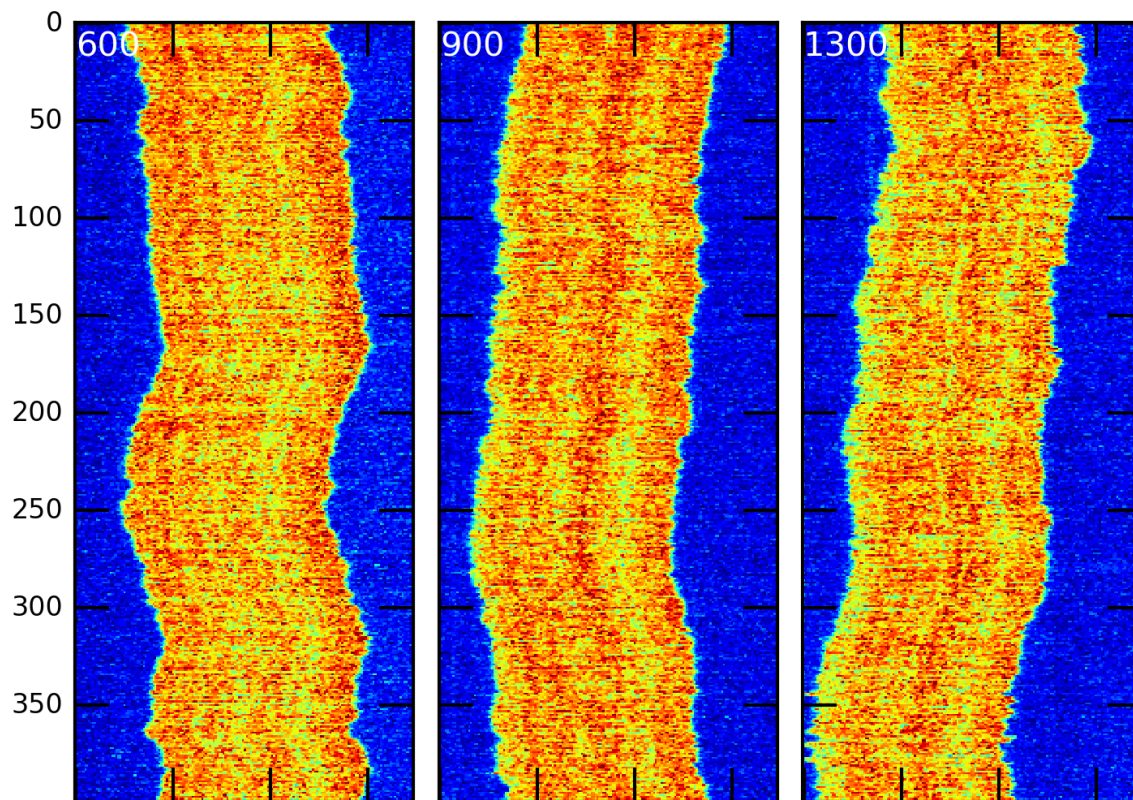


Figure A.16: Vertical axis is time. White annotation gives channel width in nanometres.

A.17 130905-RecA-T4-narrow-2

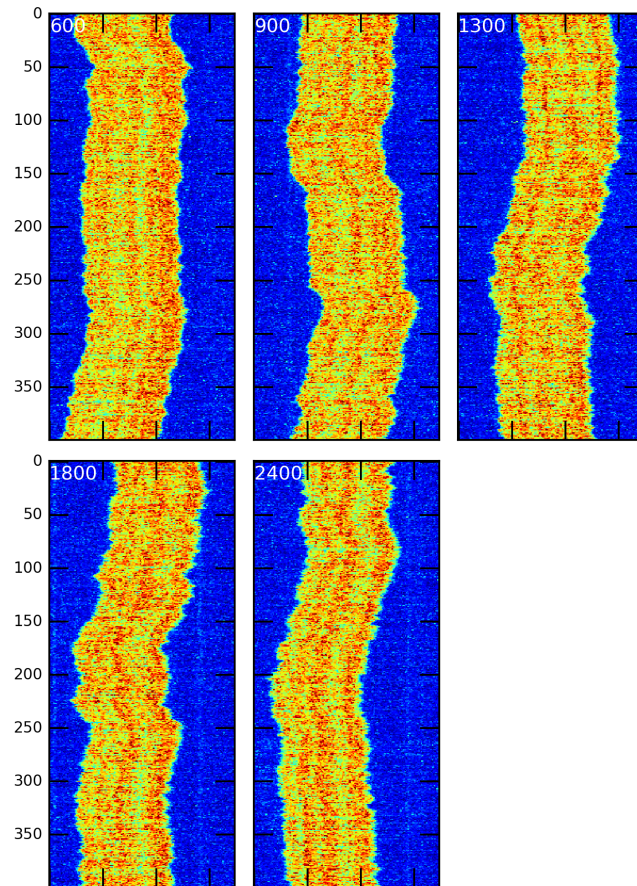


Figure A.17: Vertical axis is time. White annotation gives channel width in nanometres.

A.18 130905-RecA-T4-narrow-3

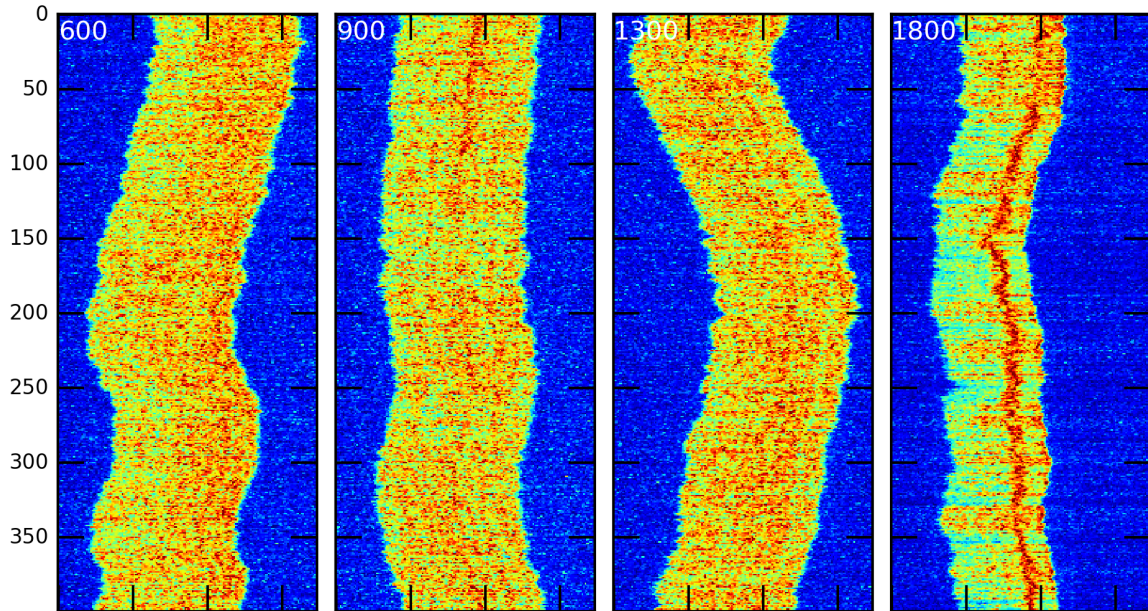


Figure A.18: Vertical axis is time. White annotation gives channel width in nanometres.

A.19 130905-RecA-T4-narrow-4

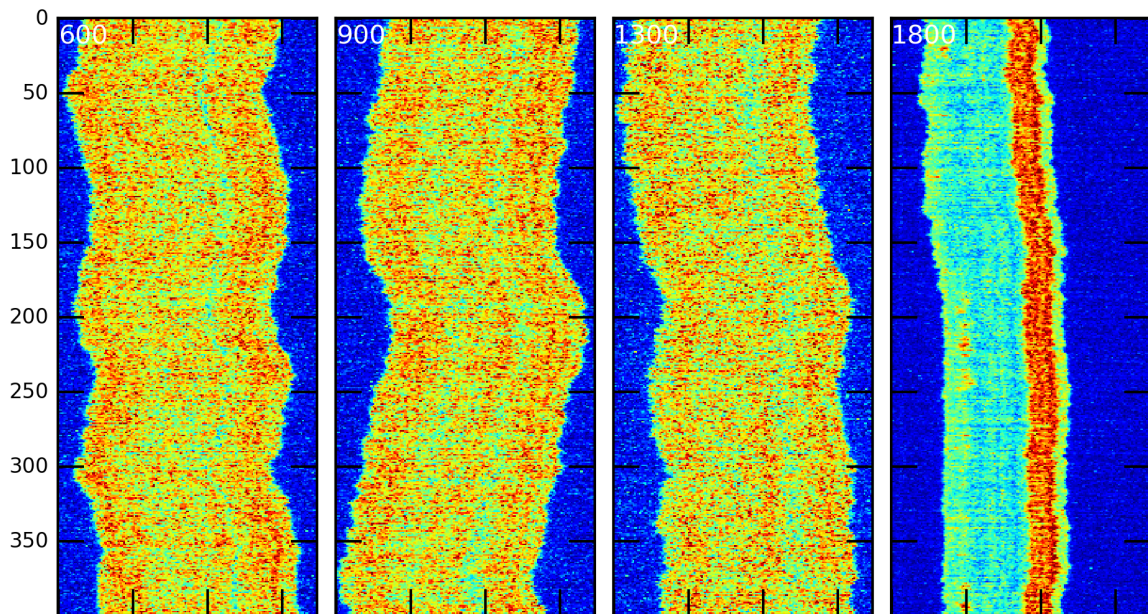


Figure A.19: Vertical axis is time. White annotation gives channel width in nanometres.

A.20 130905-RecA-T4-narrow-5

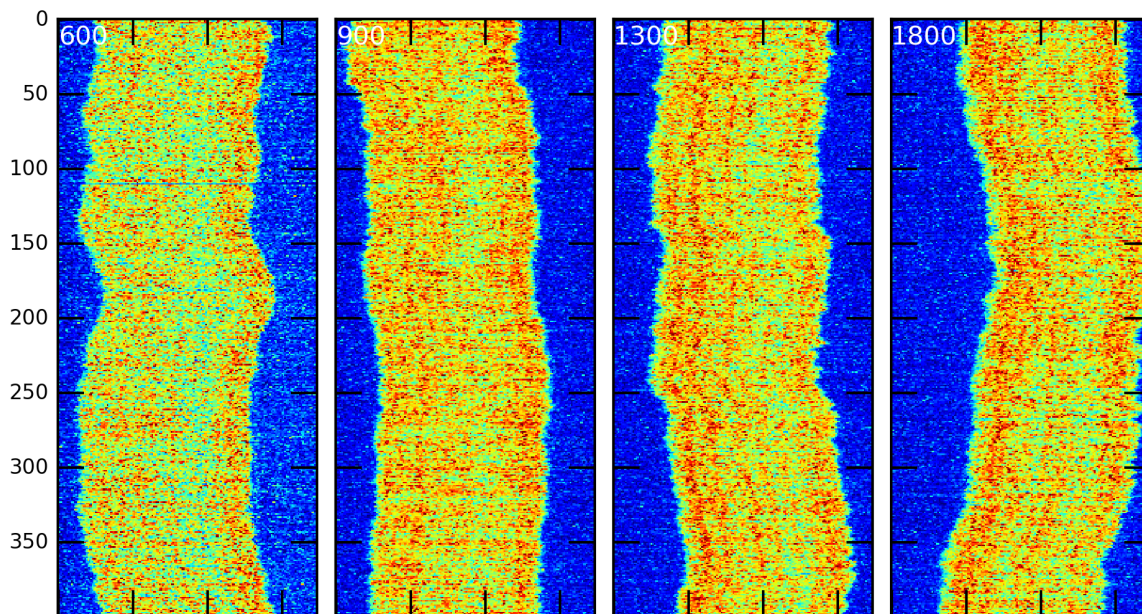


Figure A.20: Vertical axis is time. White annotation gives channel width in nanometres.

A.21 130905-RecA-T4-narrow-6

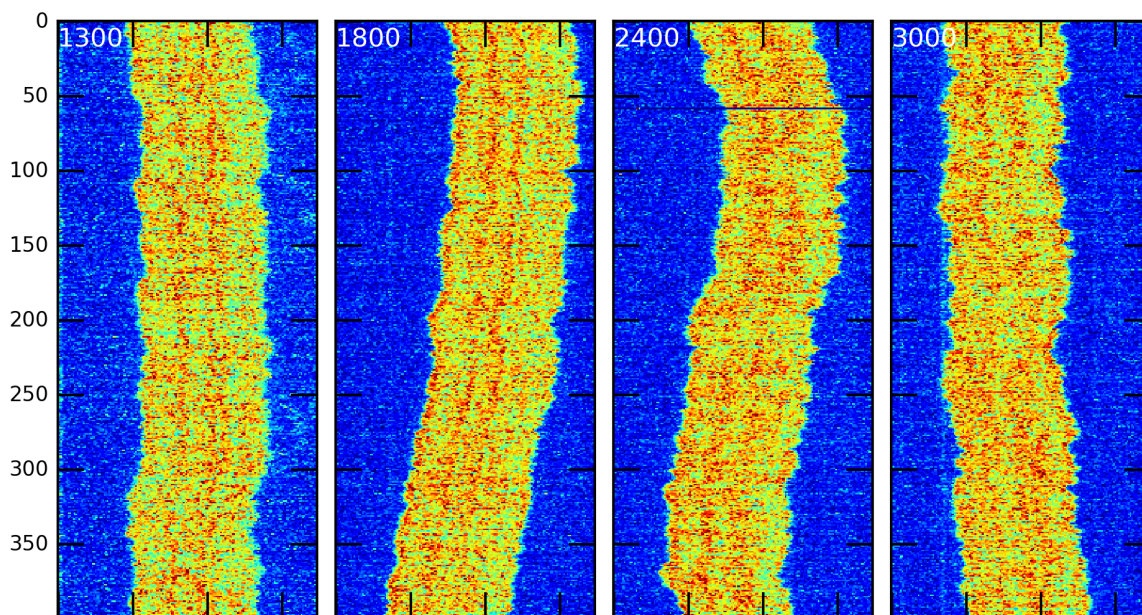


Figure A.21: Vertical axis is time. White annotation gives channel width in nanometres.

A.22 130905-RecA-T4-narrow-7

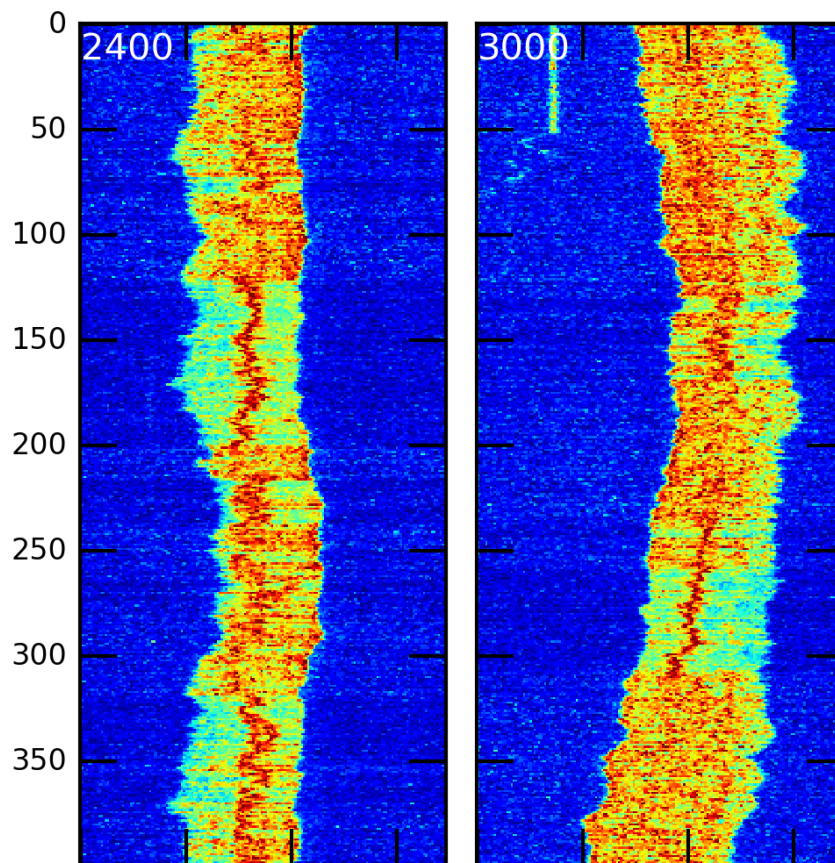


Figure A.22: Vertical axis is time. White annotation gives channel width in nanometres.

A.23 130905-RecA-T4-wide-1

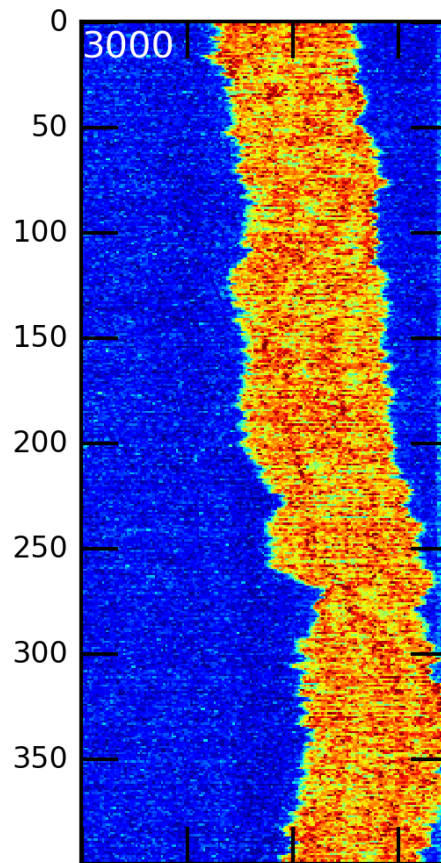


Figure A.23: Vertical axis is time. White annotation gives channel width in nanometres.

A.24 130905-RecA-T4-wide-2

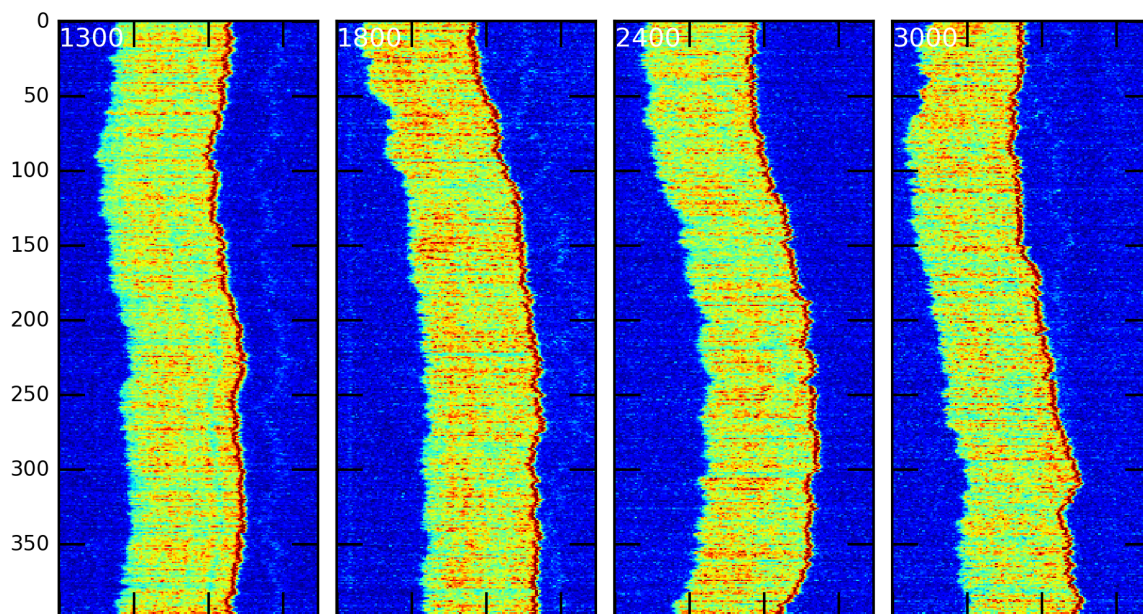


Figure A.24: Vertical axis is time. White annotation gives channel width in nanometres.

A.25 130905-RecA-T4-wide-3

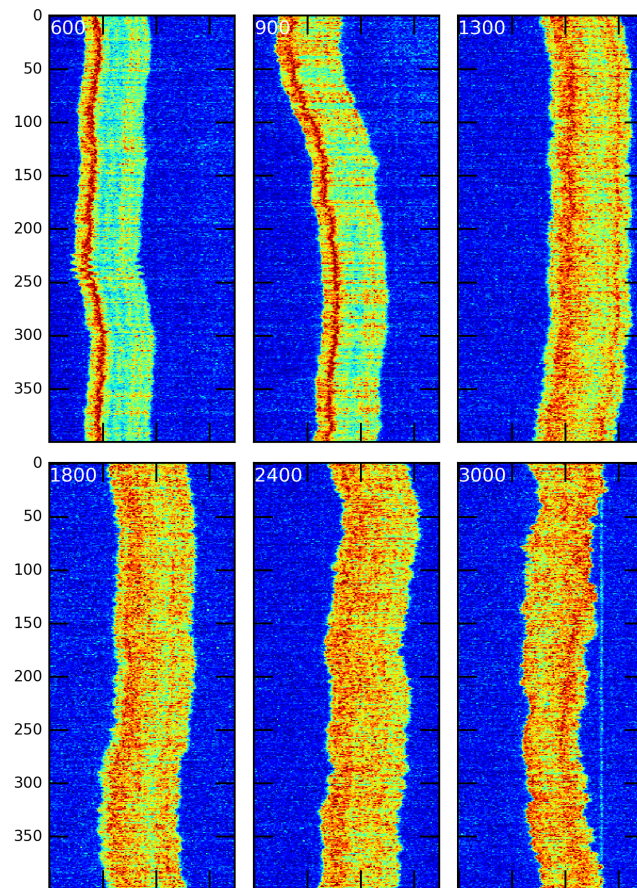


Figure A.25: Vertical axis is time. White annotation gives channel width in nanometres.

A.26 130905-RecA-T4-wide-4

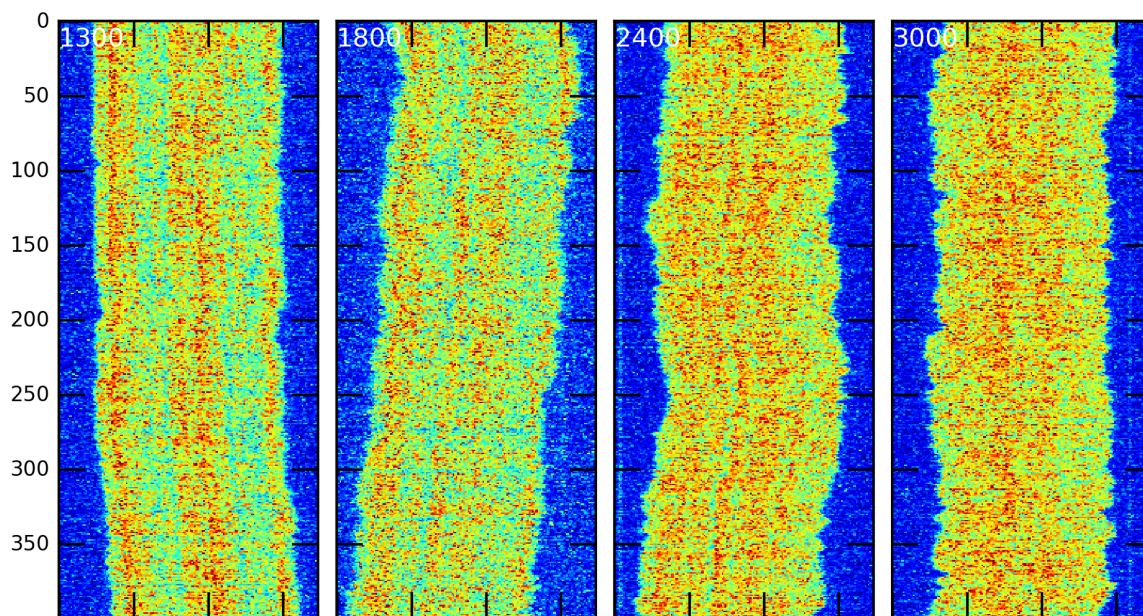


Figure A.26: Vertical axis is time. White annotation gives channel width in nanometres.

A.27 130905-RecA-T4-wide-5

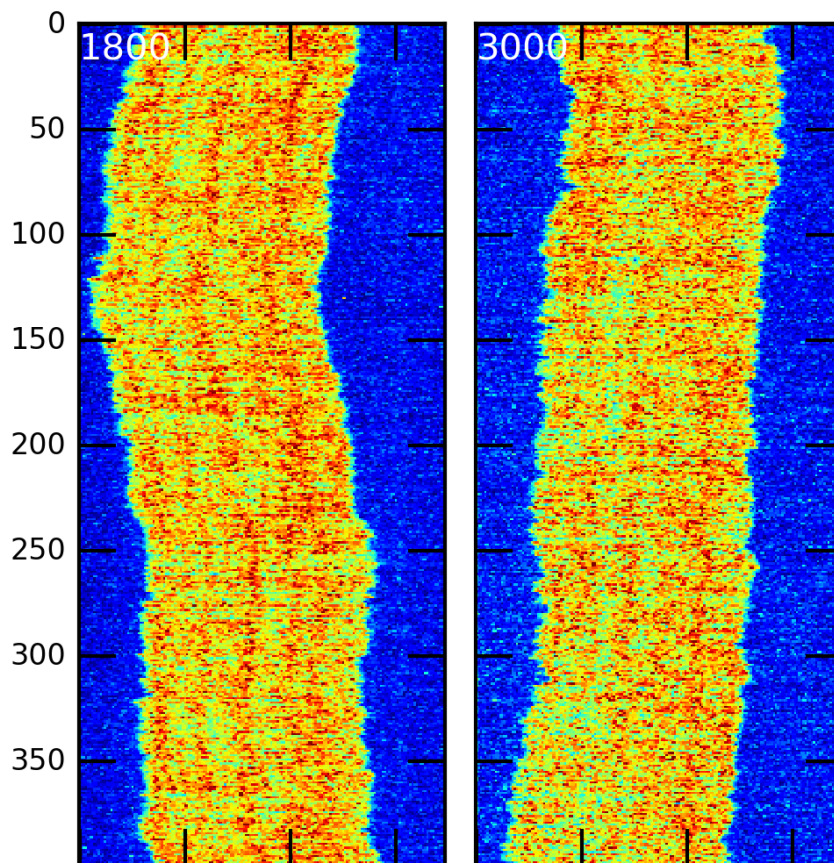


Figure A.27: Vertical axis is time. White annotation gives channel width in nanometres.

A.28 130905-RecA-T4-wide-6

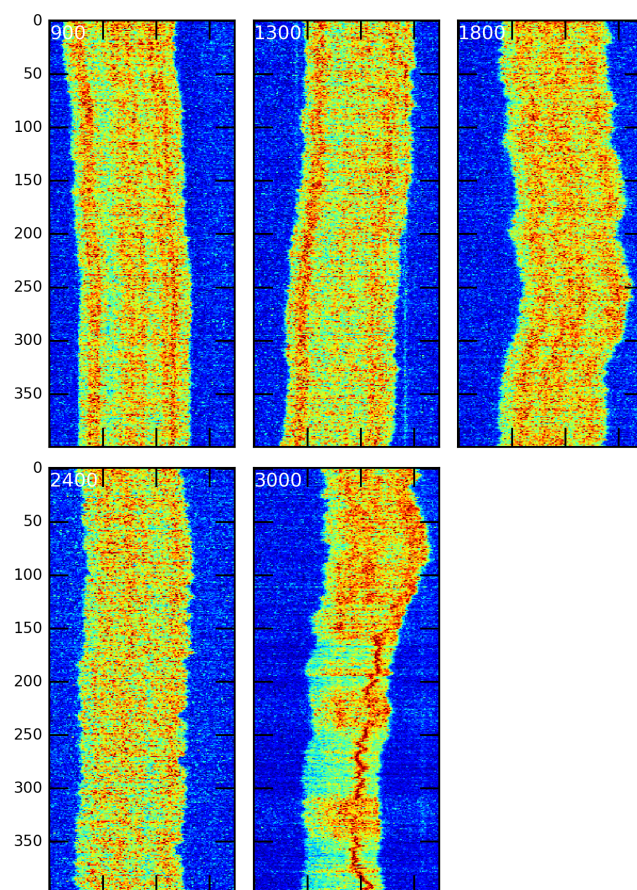


Figure A.28: Vertical axis is time. White annotation gives channel width in nanometres.

A.29 130905-RecA-T4-wide-7

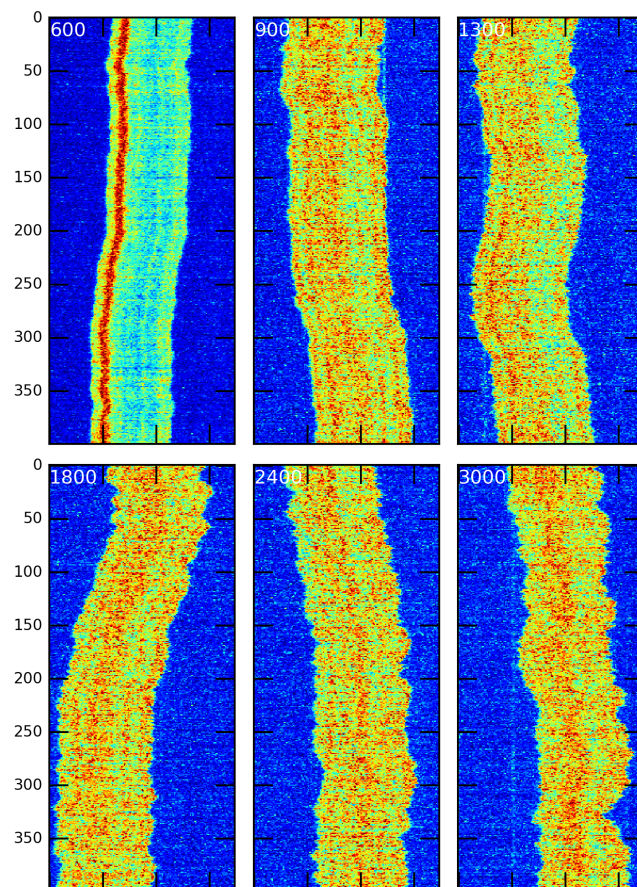


Figure A.29: Vertical axis is time. White annotation gives channel width in nanometres.

A.30 130905-RecA-T4-wide-8

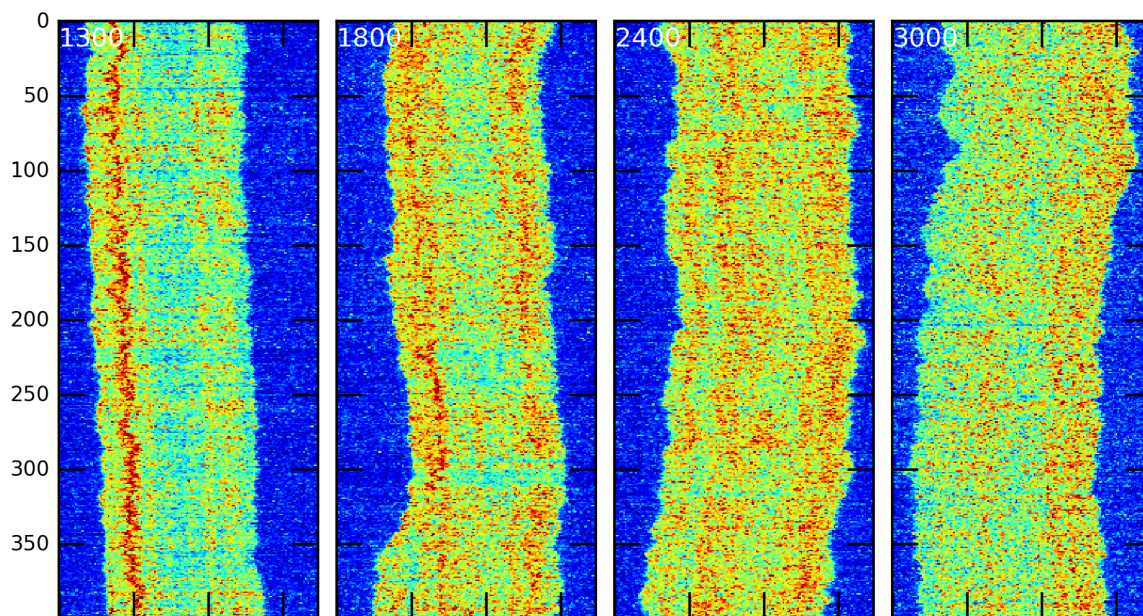


Figure A.30: Vertical axis is time. White annotation gives channel width in nanometres.

A.31 130924-RecA-lambda-narrow-1

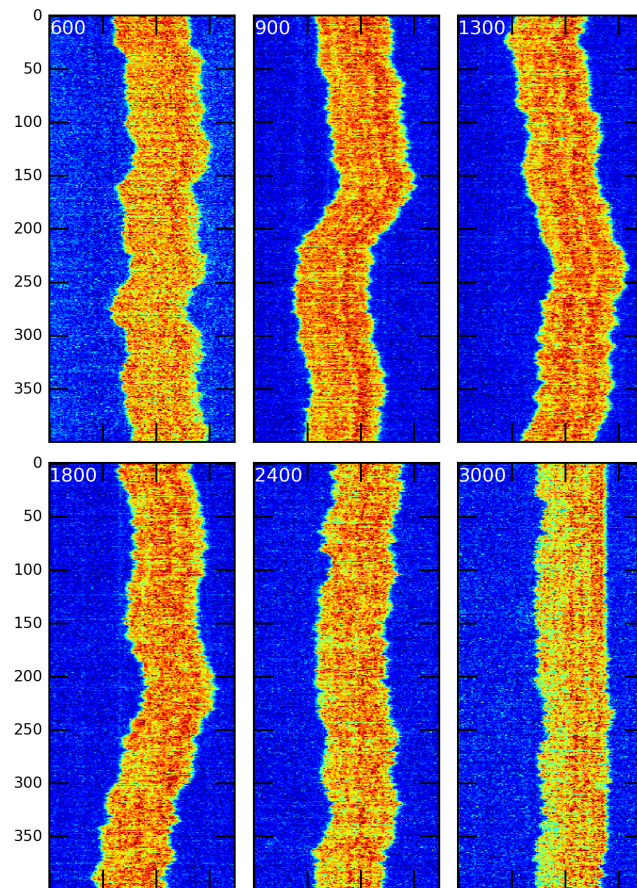


Figure A.31: Vertical axis is time. White annotation gives channel width in nanometres.

A.32 130924-RecA-lambda-narrow-2

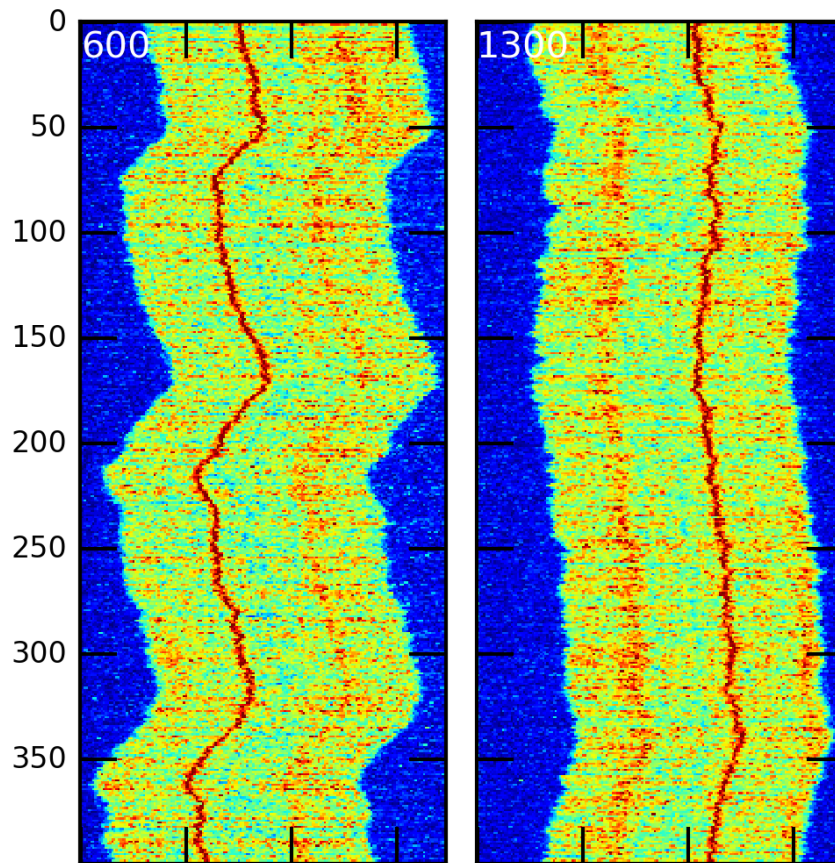


Figure A.32: Vertical axis is time. White annotation gives channel width in nanometres.

A.33 130924-RecA-lambda-narrow-3

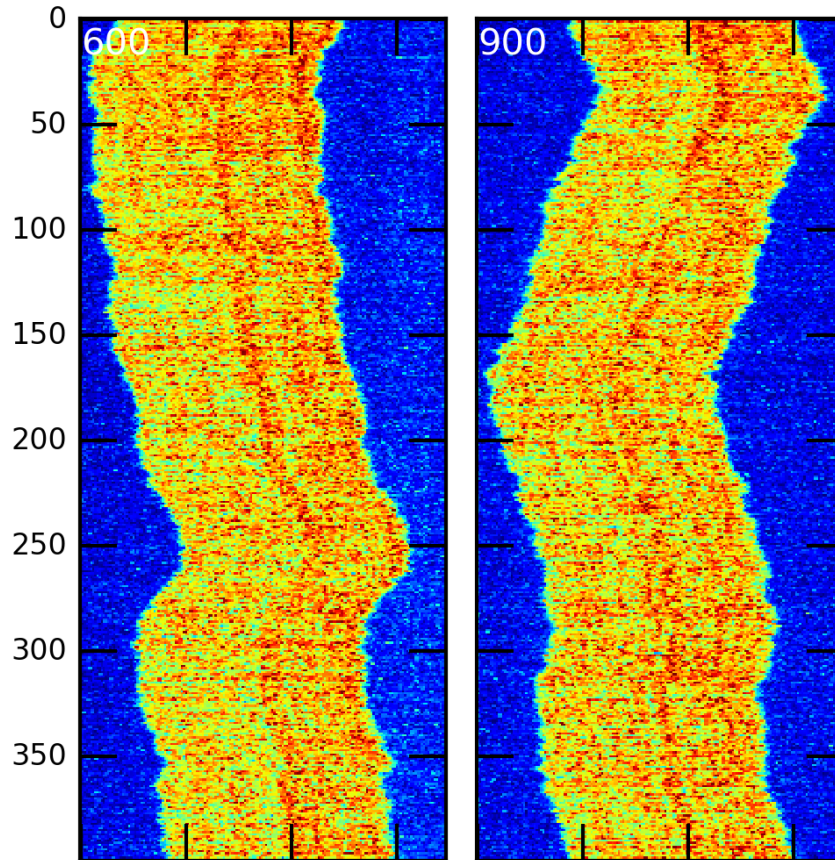


Figure A.33: Vertical axis is time. White annotation gives channel width in nanometres.

A.34 130924-RecA-lambda-narrow-4

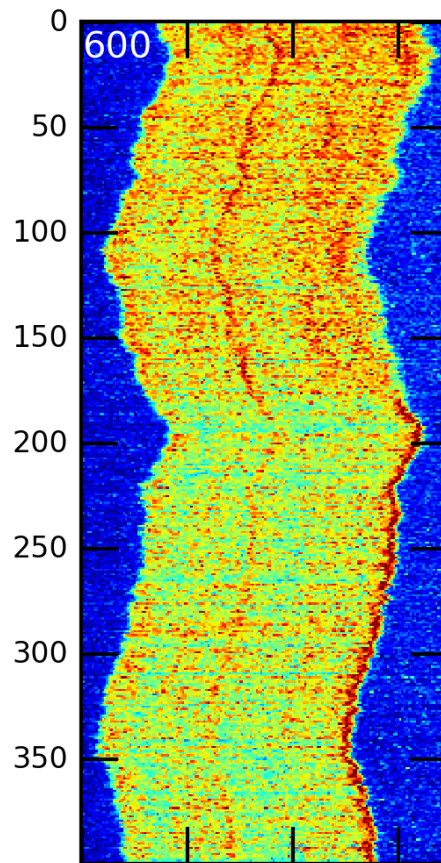


Figure A.34: Vertical axis is time. White annotation gives channel width in nanometres.

A.35 130924-RecA-lambda-narrow-5

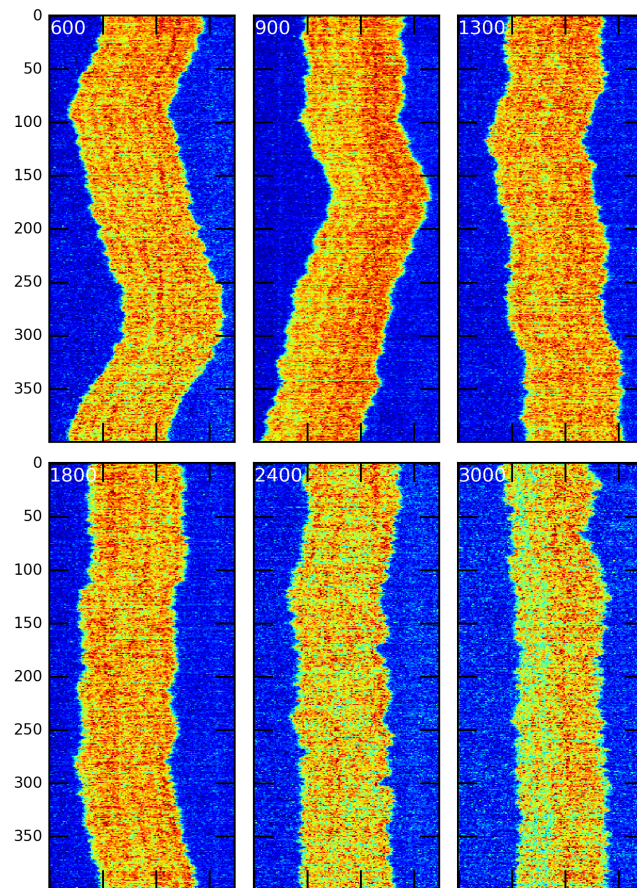


Figure A.35: Vertical axis is time. White annotation gives channel width in nanometres.

A.36 130924-RecA-lambda-narrow-6

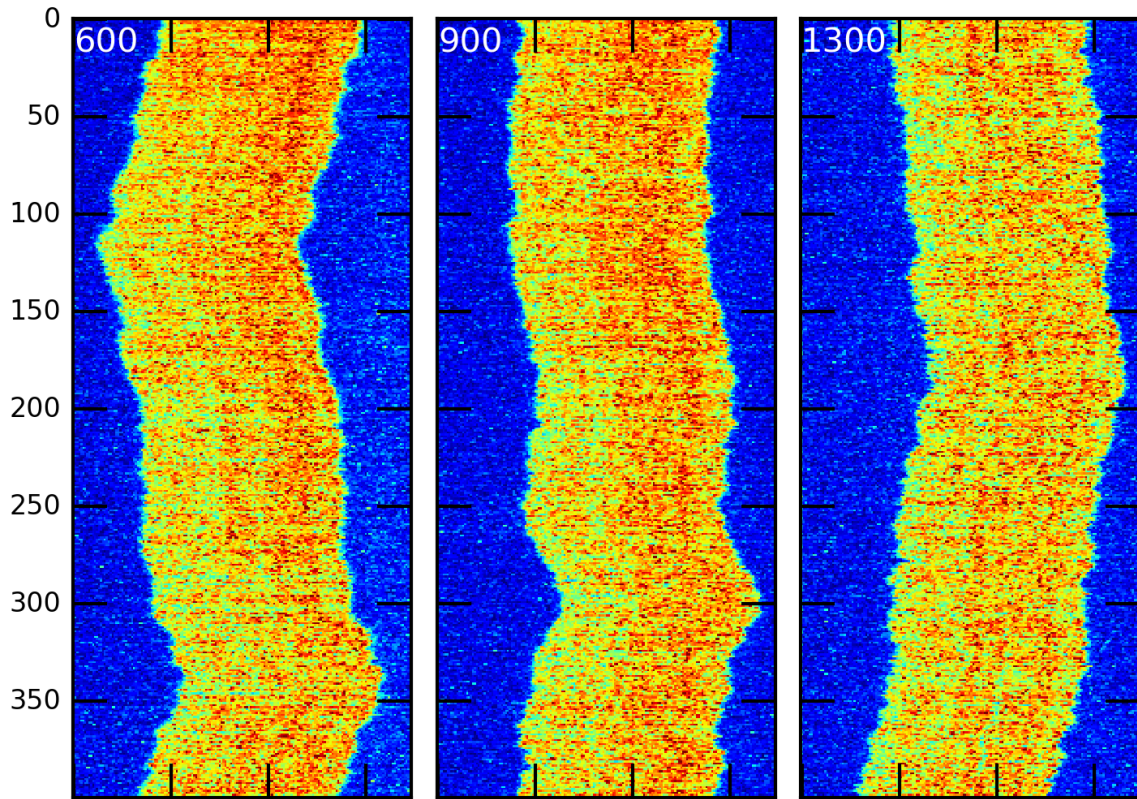


Figure A.36: Vertical axis is time. White annotation gives channel width in nanometres.

A.37 130924-RecA-lambda-wide-1

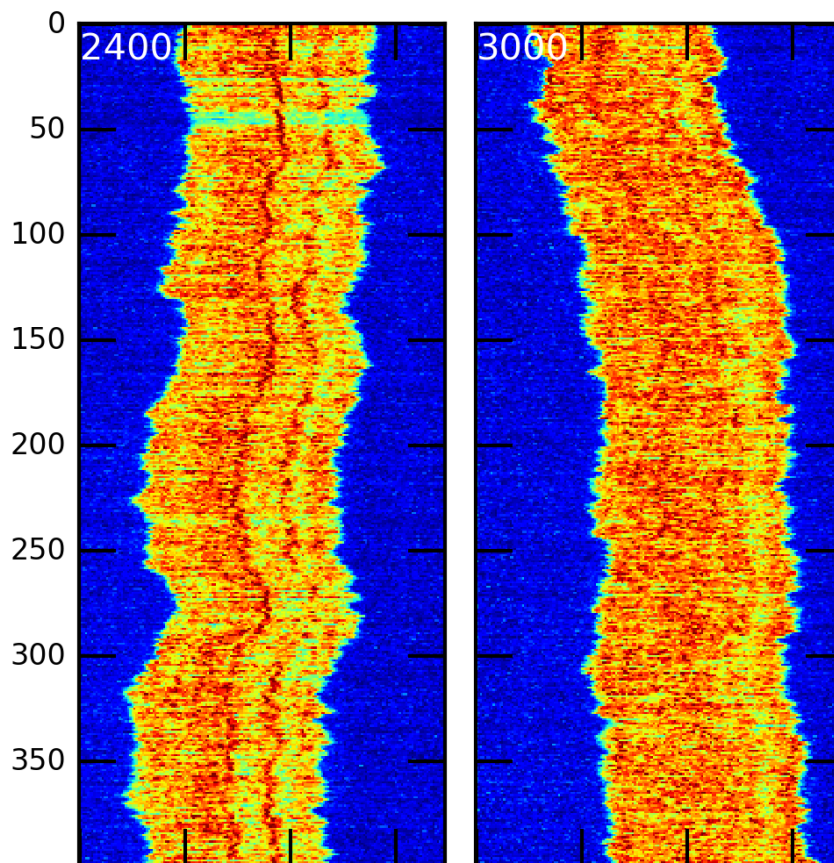


Figure A.37: Vertical axis is time. White annotation gives channel width in nanometres.

A.38 130924-RecA-lambda-wide-10

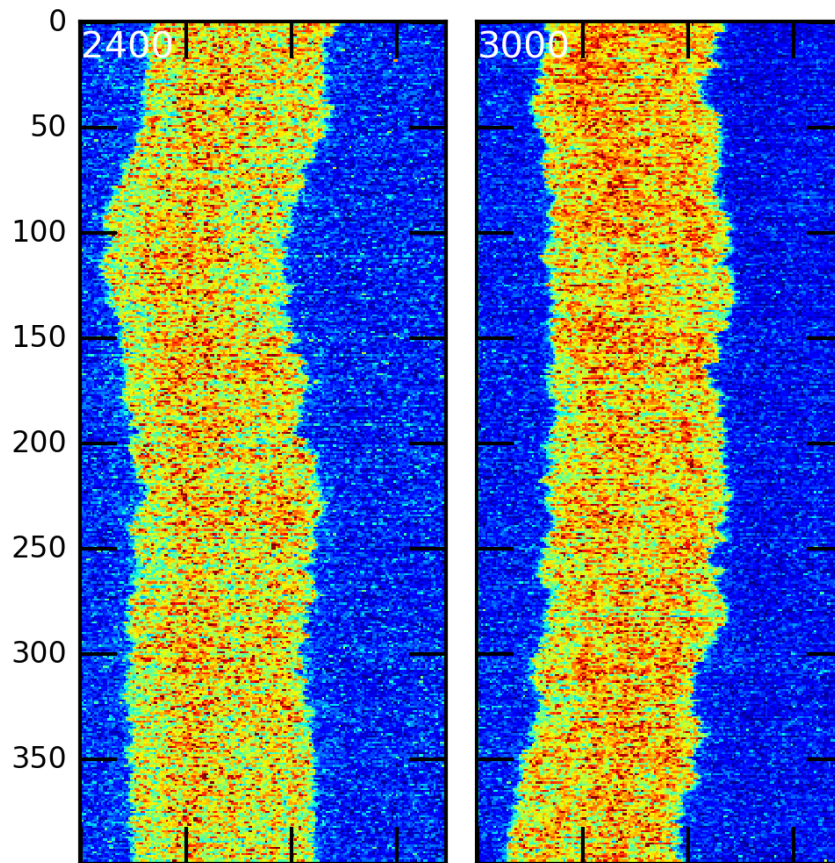


Figure A.38: Vertical axis is time. White annotation gives channel width in nanometres.

A.39 130924-RecA-lambda-wide-2

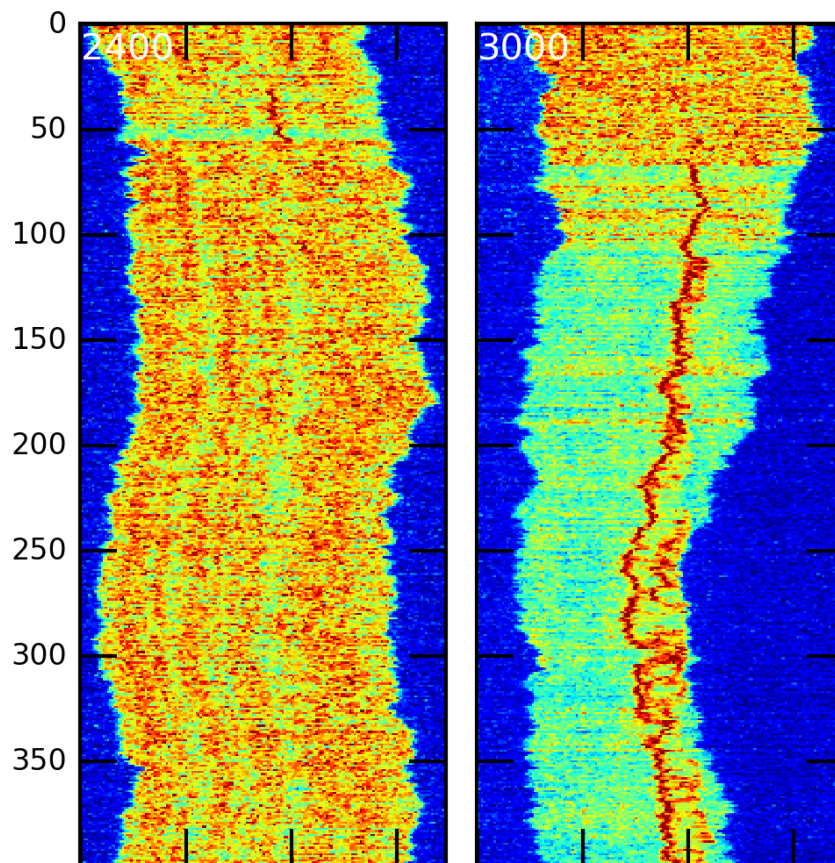


Figure A.39: Vertical axis is time. White annotation gives channel width in nanometres.

A.40 130924-RecA-lambda-wide-3

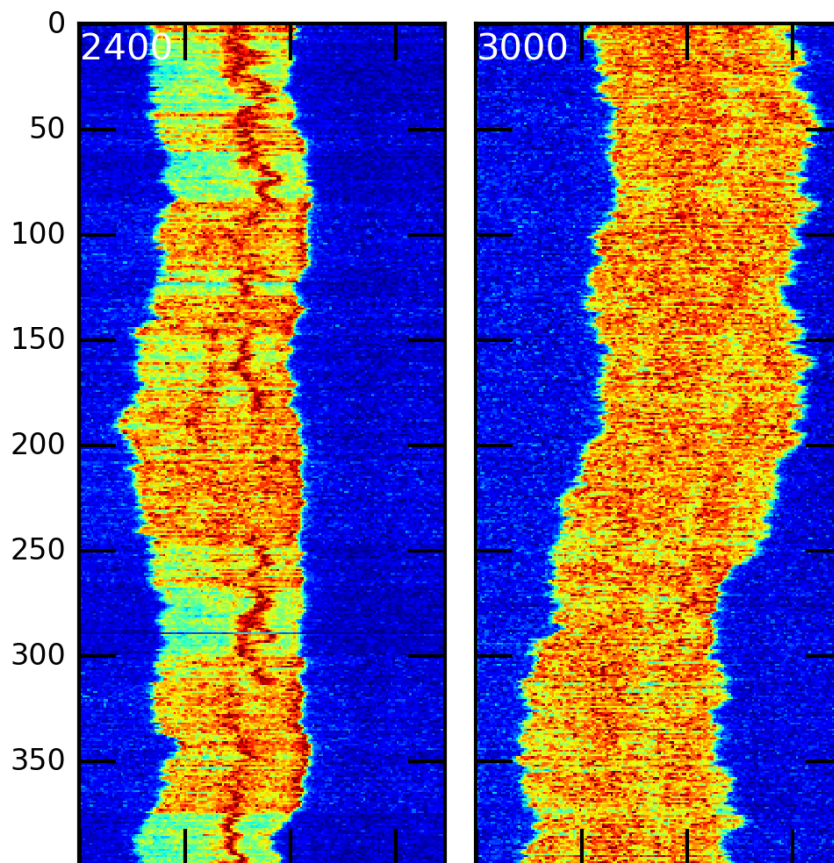


Figure A.40: Vertical axis is time. White annotation gives channel width in nanometres.

A.41 130924-RecA-lambda-wide-4

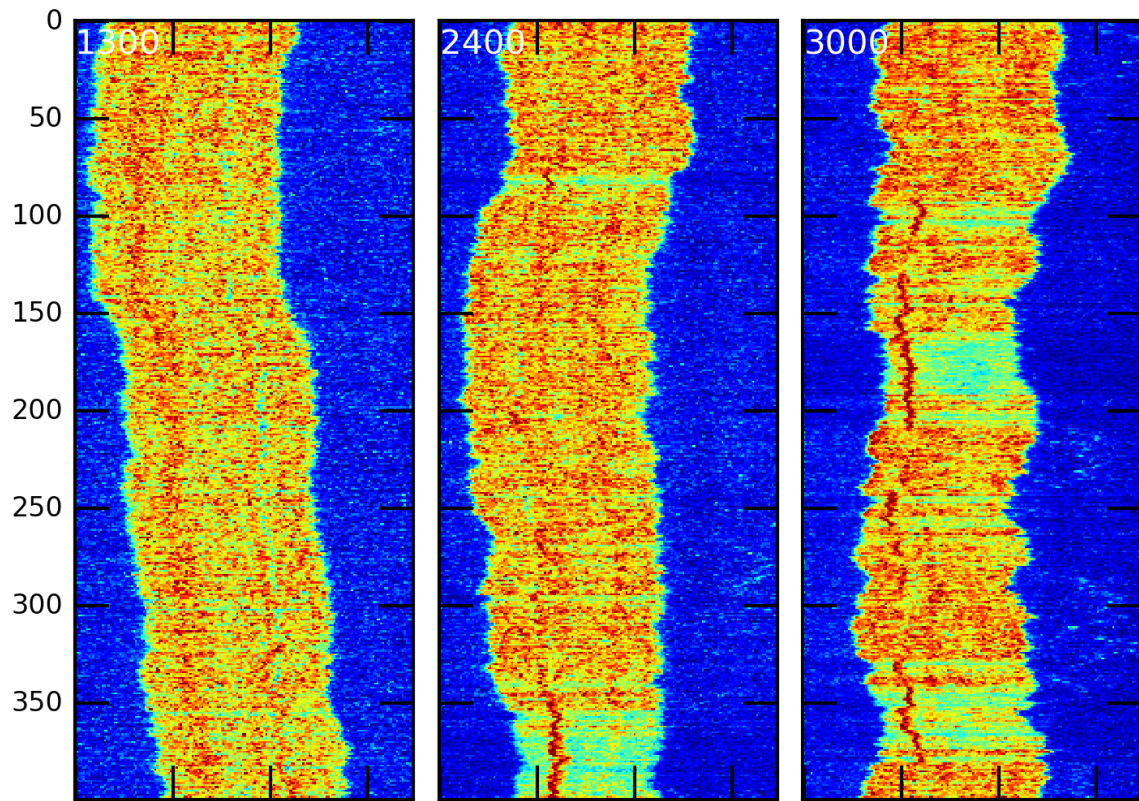


Figure A.41: Vertical axis is time. White annotation gives channel width in nanometres.

A.42 130924-RecA-lambda-wide-5

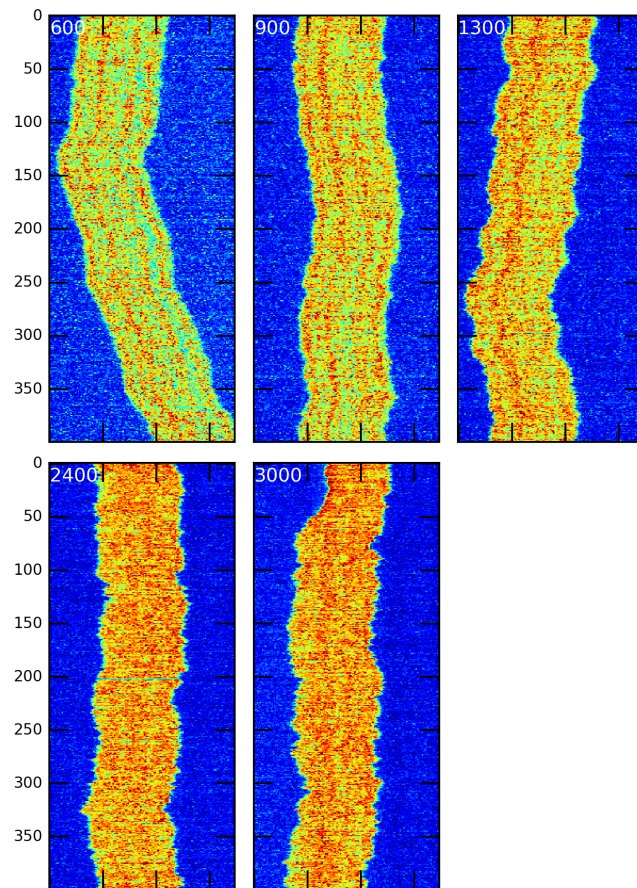


Figure A.42: Vertical axis is time. White annotation gives channel width in nanometres.

A.43 130924-RecA-lambda-wide-8

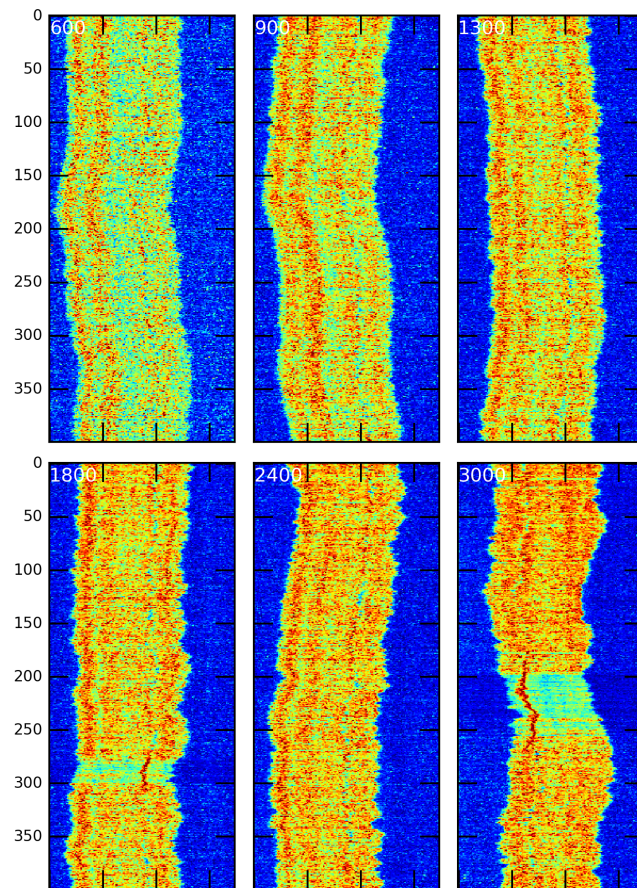


Figure A.43: Vertical axis is time. White annotation gives channel width in nanometres.

A.44 130924-RecA-lambda-wide-9

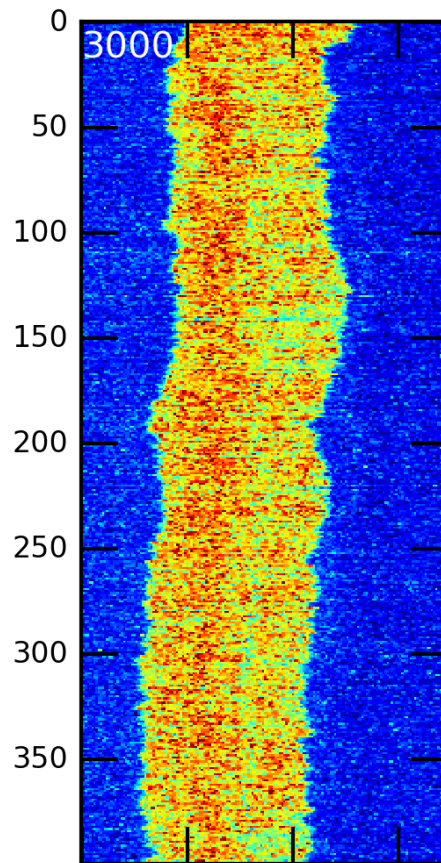


Figure A.44: Vertical axis is time. White annotation gives channel width in nanometres.

B

Classification of conformations and experimental conditions

B.1 Definitions

A conformation exhibiting a *fold* is a conformation on which there is at least one point with tangent vector

Fold rate means the number of changes in conformation, per 100 frames, that cause a significant change in the extension of the molecule.

Hairpin nucleation time is computed as defined in the Methodology.

The *drift* and *fold direction* give the direction of any drift of the filament in the channel, and the alignment of folded conformations (such as ‘U’) with the movement from channel to channel. Therefore, in both cases, ‘forwards’ means with the direction of travel from channel to channel, and ‘backwards’ means against the direction of travel. ‘None’ means there is no drift or folding. ‘Neither’ means that there are folded conformations but that they are not aligned with the movement in either direction. ‘Both’ means there is drift or there are folds in both directions.

Finally, the symbols describing the conformations of the filament are explained in Table B.1.

B.2 Summary

B.2.1 Main findings

- The initial conditions can have a large effect on the conformations adopted, but in the majority of cases, the initial condition is a filament with only small fluctuations along its length, with the filament aligned with the channel (denoted ‘~’; see Table B.1).
- There is a preference for unfolded conformations, which do not form folds followed by equal numbers of instances of folding and unfolding, followed by more unfolding than folding (where the initial condition was folded but not stuck).

- Most of the folding occurs in the widest channels, but even at 3000 nm, only 0.53 extension-changing folds every 100 frames on average.
- Moving from a wide channel to a narrow channel is more likely to produce an experimental error than vice versa, although “stuck” ends are not significantly more likely in either direction.
- In most cases, there is some drift of the molecules in the channel, probably resulting from unequal pressure. But also in most cases, there is no folding.
- Where there is folding, there is an effect of drift on folding, particularly when the drift is aligned with the direction of movement from channel to channel, such that the conformations produced tend to align with the drift and channel movement.

B.2.2 Notes on experimental conditions

Firstly, there was no systematic timing in the experiments, and in particular, no systematic equilibration time before starting recording, after moving each molecule into its channel.

Secondly, there was on some unknown occasions “pressure leakage” in the channels, and attempts were made at compensating for this by applying pressure in the opposing direction in the channel; therefore, the pressure in the channels was not necessarily in equilibrium, and sometimes drift of the molecule is recorded in the videos. It is not clear what proportion of the drift is explained in this way.

Thirdly, on some unknown occasions, where “irreversible folds” (suspected to be experimentally induced) were noticed in the molecules, it was attempted to unfold the molecule by moving it “back and forth” in the channel. It is not clear where this occurred, or whether it occurred during or prior to recording.

B.2.3 How does the initial condition affect the subsequent conformations?

Most of the initial conditions are \sim (see Table B.1 for explanation of these conformation symbols). Where the initial conditions are not \sim , the initial condition only seems to affect the subsequent conformation if it involves “stuck” folded ends (ie, those that somehow seem to have become glued to the bulk of the filament).

The following cases show stuck ends:

- 130607-RecA-T4-narrow-8 (at 1300 nm and subsequently 1800nm)
- 130905-RecA-T4-narrow-3 (at 1800nm – the last width for which we have data)
- 130905-RecA-T4-narrow-4 (at 1800nm – the last width for which we have data)
- 130905-RecA-T4-narrow-7 (at 3000nm)
- 130905-RecA-T4-wide-3 (at 900 nm and subsequently 600 nm – but not at the “front” end)
- 130905-RecA-T4-wide-7 (at 600 nm – again not at the “front” end)
- 130924-RecA-lambda-wide-3 (at 2400 nm – the last width for which we have data)

The following cases show unusual initial conditions that do not include stuck ends:

- 130607-RecA-T4-narrow-9 (initially folded front at 1800 and 2400nm, which later unfolds)
- 130607-RecA-T4-wide-4 (an initially folded front end at 1800 nm that later unfolds)
- 130905-RecA-T4-wide-2 (a bright “clump” visible at the front end throughout all the recordings)
- 130905-RecA-T4-wide-8 (a point on filament around which bending is easier; perhaps a gap in the protein coating?)
- 130924-RecA-lambda-wide-4 (a loop is visible from the start at 3000 nm)
- 130924-RecA-lambda-wide-8 (a loop is visible from 2400 nm through 1800 nm, but not thereafter)

B.2.4 Are (experimentally-induced) stuck ends more likely wide to narrow or vice versa?

Neither direction of movement is more likely to produce “stuck” folded ends. Looking again at the cases detailed above, we have 4 cases of stuck ends when moving from narrow to wide, and 3 cases of stuck ends when moving from wide to narrow. But there are 5 other wide-to-narrow experimental oddities, against just 1 for narrow-to-wide, which suggests that wide-to-narrow is nonetheless more likely to produce experimental error during the progression of the experiment.

B.2.5 Do folds preferentially form in the direction of the drift, or of the movement from channel to channel?

In most cases, there is some drift of the molecule in the channel, and no folding. Sometimes, the drift seems obviously to affect folding, as in the case of 130607-RecA-T4-narrow-10 (unfolding at 2400 nm due to channel pressure change; initial condition is folded in direction of movement) or 130603-RecA-T4-wide-5 (folds occur when drift is 0, preferentially in direction of movement). Where the initial condition is folded, it is more often the case that this folding is in the direction of movement (9 vs 6 cases), but it is hard to say if this is a significant difference.

The remaining, spontaneous, cases are detailed in Table B.3. We see that, where there is no drift, there is no preference for folding to occur in either direction. Where there is a forwards drift (ie, in the direction of movement from channel to channel), there is a preference for folding also in the forwards direction. Finally, where there is a backwards drift (against the direction of movement from channel to channel), there is still a slight preference for folding in the forwards direction (opposed to the drift). This suggests that there is only an effect of drift on the direction of spontaneous folding when the drift is in the direction of channel movement.

B.2.6 Which is faster or more likely, folding or unfolding?

There is clearly a preference for unfolded configurations, except where folding is apparently induced by experimental conditions. Most recordings show no folding at all. Of the remainder, there is a preference for equal numbers of folding and unfolding (15 cases), followed by more unfolding (7 cases), and finally, there are only 4 cases where the recordings show more folding than unfolding. The cases of more unfolding are explained by their having a folded initial condition. There is no difference between the speed at which conformations change from folded to unfolded or vice versa, except in the cases where the initial condition has “stuck” folded ends, in which cases unfolding at the ends does not occur.

B.2.7 How do the fold rates vary with channel width?

Ignoring the effect of contour length on the fold rate, the average fold rates for the different channel widths are: 0.00 at 600 and 900 nm; 0.01 at 1300 nm; 0.05 at 1800 nm; 0.19 at 2400 nm; and 0.53 at 3000 nm. Clearly, as expected, most of the folding happens in the largest channels.

B.3 Tables

Table B.1: Legend for conformation symbols

Symbol	Explanation
~	Small fluctuations along the length of the filament, with the filament aligned with the channel.
∩	Fewer, larger fluctuations along the length of the filament, like an ‘S’ on its side.
S	An ‘S’ shape, perpendicular to the length of the channel.
U	A single U bend, perpendicular to the length of the channel.
∩	A single ∩ bend, perpendicular to the length of the channel.
Ω	A bend in the form of an ‘Ω’.
⊃	A single ⊃ bend, aligned with the length of the channel.
⊂	A single ⊂ bend, aligned with the length of the channel.
α	Like the α symbol: a bend aligned with the length of the channel, but with ends crossed over.
∞	Like α, but with ends not quite crossed over, and of asymmetric lengths.
σ	A curl in the left end of the filament, of the form of the symbol.
⊖	A curl in the right end of the filament, of the form of the symbol.
α	A curl in the left end of the filament, of the form of the symbol.
⊖	A curl in the right end of the filament, of the form of the symbol.
⌘	A loop in the middle of the filament, apparently formed by the molecule crossing itself.

Table B.2: Hairpin nucleation times

Channel width (nm)	Average nucleation time (s)
3000	34.6

2400	163.4
1800	1001.0
1300	1276.0
900	∞
600	∞

Table B.3: Average fold rates for different directions of drift and channel movement, excluding cases of experimental oddity detailed above

Drift	Fold direction	Mean fold rate
None	Forwards	0.60
None	Backwards	0.70
None	Neither	0.25
Forwards	Forwards	1.25
Forwards	Backwards	0.25
Forwards	Neither	0.75
Backwards	Forwards	0.50
Backwards	Backwards	0.00
Backwards	Neither	0.81

Table B.4: 130603-RecA-T4-wide-2. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 6.47$

D	Initial condition	Fold rate	Drift	Fold direction
3000	U	0.75	None	Forwards
2400	~	0.25	Forwards	Backwards
1800	~	0	Forwards	None

Table B.5: 130603-RecA-T4-wide-3. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 6.79$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0	Forwards	None
2400	~	0	Forwards	None
1800	~	0	Forwards	None
1300	~	0	Forwards	None

Table B.6: 130603-RecA-T4-wide-4. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 6.34$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0	None	None
2400	~	0	Backwards	None

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

1800 ~ 0 Backwards None

Table B.7: 130603-RecA-T4-wide-5. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 7.54$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	0.50	Backwards	Forwards	Folds occur when drift is 0
2400	~	0	Backwards	None	
1800	~	0	Backwards	None	
1300	~	0	Backwards	None	

Table B.8: 130603-RecA-T4-wide-7. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 12.50$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	0	None	None	
2400	~	0	Backwards	None	
1800	~	0	Backwards	None	
1300	~	0	Backwards	None	
900	~	0	Backwards	None	
600	~	0	Backwards	None	Drift not constant; pressure leakage?

Table B.9: 130607-RecA-T4-narrow-10. T4 DNA. Started at 600 nm channel. $L/\ell_P = 10.62$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
600	~	0	Forwards	None	
900	~	0	Both	None	
1300	~	0	None	None	
1800	~	0	Forwards	None	
2400	∞	1.00	Forwards	Forwards	Final unfolding occurs once drift begins. Suggests unfolding due to channel pressure change.
3000	σ	1.50	Forwards	Forwards	See previous comment.

Table B.10: 130607-RecA-T4-narrow-7. T4 DNA. Started at 600 nm channel. $L/\ell_P = 7.38$

D	Initial condition	Fold rate	Drift	Fold direction
600	~	0	Forwards	None
900	~	0	Both	None

Table B.11: 130607-RecA-T4-narrow-8. T4 DNA. Started at 600 nm channel. $L/\ell_P = 13.22$

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

D	Initial condition	Fold rate	Drift	Fold direction	Comments
600	~	0	Forwards	None	
900	~	0	Forwards	None	
1300	α	0	Forwards	Forwards	Suggests experimental error
1800	α	0	Forwards	Forwards	Suggests experimental error

Table B.12: 130607-RecA-T4-narrow-9. T4 DNA. Started at 600 nm channel. $L/\ell_P = 6.83$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
600	~	0	Forwards	None	
900	~	0	Both	None	
1300	~	0	Forwards	None	
1800	α	1.00	None	Neither	Initial fold in direction of movement.
2400	σ	1.25	Backwards	Neither	Initial fold in direction of movement.
3000	~	0.50	Forwards	Forwards	But final configuration (\supset) does not point forwards.

Table B.13: 130607-RecA-T4-wide-1. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 6.46$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	S	0.25	None	Neither	Initial S becomes ~ by frame 5.
2400	~	0	None	None	
1800	~	0	None	None	
1300	~	0	Forwards	None	
900	~	0	None	None	
600	~	0	Forwards	None	

Table B.14: 130607-RecA-T4-wide-2. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 6.74$

D	Initial condition	Fold rate	Drift	Fold direction
3000	\cap	0.75	None	Forwards
2400	~	0	None	None
1800	~	0	None	None
1300	~	0	Backwards	None
900	~	0	Backwards	None
600	~	0	None	None

Table B.15: 130607-RecA-T4-wide-3. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 9.32$

D	Initial condition	Fold rate	Drift	Fold direction
---	-------------------	-----------	-------	----------------

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

3000	~		0	None	None
2400	~		0	None	None
1800	~		0	None	None
1300	~		0	None	None
900	~		0	None	None
600	~		0	Backwards	None

Table B.16: 130607-RecA-T4-wide-4. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 8.35$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	0.25	None	Backwards	
2400	~	1.50	None	Backwards	
1800	a	0.25	None	Backwards	Only fold is an unfolding
1300	~	0.25	None	Backwards	
900	~	0	None	None	
600	~	0	Backwards	None	

Table B.17: 130607-RecA-T4-wide-5. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 9.89$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0	None	None
2400	~	0	None	None
1800	~	0	None	None
1300	~	0	None	None
900	~	0	None	None
600	~	0	Forwards	None

Table B.18: 130607-RecA-T4-wide-6. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 7.84$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0	None	None
2400	~	0	None	None
1800	~	0	None	None
1300	~	0	Forwards	None
900	~	0	None	None
600	~	0	None	None

Table B.19: 130905-RecA-T4-narrow-1. T4 DNA. Started at 600 nm channel. $L/\ell_P = 14.97$

D	Initial condition	Fold rate	Drift	Fold direction
---	-------------------	-----------	-------	----------------

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

600	~		0	Both	None
900	~		0	None	None
1300	~		0	Forwards	None

Table B.20: 130905-RecA-T4-narrow-2. T4 DNA. Started at 600 nm channel. $L/\ell_P = 12.82$

D	Initial condition	Fold rate	Drift	Fold direction	
600	~		0	Forwards	None
900	~		0	None	None
1300	~		0	None	None
1800	~		0	Forwards	None
2400	~		0	None	None

Table B.21: 130905-RecA-T4-narrow-3. T4 DNA. Started at 600 nm channel. $L/\ell_P = 13.75$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
600	~	0	Both	None	
900	~	0	None	None	
1300	~	0	Both	None	
1800	o	0	None	Backwards	Remains in initial condition throughout

Table B.22: 130905-RecA-T4-narrow-4. T4 DNA. Started at 600 nm channel. $L/\ell_P = 18.68$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
600	~	0	None	None	
900	~	0	None	None	
1300	~	0	None	None	
1800	o_o	0	Backwards	Both	Remains o_o throughout. Bright region near rightmost loop suggests stuck to itself.

Table B.23: 130905-RecA-T4-narrow-5. T4 DNA. Started at 600 nm channel. $L/\ell_P = 16.17$

D	Initial condition	Fold rate	Drift	Fold direction	
600	~		0	None	None
900	~		0	None	None
1300	~		0	None	None
1800	~		0	Both	None

Table B.24: 130905-RecA-T4-narrow-6. T4 DNA. Started at 600 nm channel. $L/\ell_P = 11.96$

D	Initial condition	Fold rate	Drift	Fold direction
1300	~		0 None	None
1800	~		0 None	None
2400	~		0 Forwards	None
3000	~		0.50 None	Backwards

Table B.25: 130905-RecA-T4-narrow-7. T4 DNA. Started at 600 nm channel. $L/\ell_P = 13.28$

D	Initial condition	Fold rate	Drift	Fold direction
2400	\supset		0.75 None	Backwards
3000	∞		0 Both	Forwards

Table B.26: 130905-RecA-T4-wide-1. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 10.43$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~		1.50 Forwards	Forwards

Table B.27: 130905-RecA-T4-wide-2. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 11.36$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	\sim	0.75	Forwards	Neither	Bright clump on right end
2400	~	0	Both	None	Bright clump on right end
1800	~	0	Forwards	None	Bright clump on right end
1300	~	0	None	None	Bright clump on right end

Table B.28: 130905-RecA-T4-wide-3. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 10.78$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	1.00	None	Backwards	
2400	~	0	None	None	
1800	~	0	None	None	
1300	~	0	None	None	
900	σ (tight loop)	0	Forwards	Backwards	Bend suggestive of experimental error, but at wrong end..
600	σ (tight loop)	0	Both	Backwards	Bend seemingly residual of that introduced at 900 nm.

Table B.29: 130905-RecA-T4-wide-4. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 17.90$

D	Initial condition	Fold rate	Drift	Fold direction
---	-------------------	-----------	-------	----------------

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

3000	~		0	None	None
2400	~		0	None	None
1800	~		0	None	None
1300	~		0	None	None

Table B.30: 130905-RecA-T4-wide-5. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 15.54$

D	Initial condition		Fold rate	Drift	Fold direction
3000	~		0	None	None
1800	~		0	None	None

Table B.31: 130905-RecA-T4-wide-6. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 14.79$

D	Initial condition		Fold rate	Drift	Fold direction
3000	Ω (longer tails)		0.50	None	Forwards
2400	~		0	None	None
1800	~		0	None	None
1300	~		0	None	None
900	~		0	None	None

Table B.32: 130905-RecA-T4-wide-7. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 12.85$

D	Initial condition		Fold rate	Drift	Fold direction	Comments
3000	~		1.00	Forwards	Forwards	
2400	~		0	None	None	
1800	~		0	Backwards	None	
1300	~		0	None	None	
900	~		0	None	None	
600	σ		0	None	Backwards	Remains in initial condition throughout. Initial condition suggestive of experimental error, but at wrong end..

Table B.33: 130905-RecA-T4-wide-8. T4 DNA. Started at 3000 nm channel. $L/\ell_P = 18.36$

D	Initial condition		Fold rate	Drift	Fold direction	Comments
3000	~		0	None	None	
2400	~		0.50	None	None	Filament seems to have a point around which bending is easier.
1800	σ		0	None	Backwards	See previous comment. This point is the origin of the σ bend.

Table B.39: 130924-RecA-lambda-narrow-6. λ DNA. Started at 600 nm channel. $L/\ell_P = 14.40$

D	Initial condition	Fold rate	Drift	Fold direction
600	~		0 Both	None
900	~		0 Both	None
1300	~		0 Forwards	None

Table B.40: 130924-RecA-lambda-wide-1. λ DNA. Started at 3000 nm channel. $L/\ell_P = 13.88$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~		0 Forwards	None
2400	~		0 Both	None

Table B.41: 130924-RecA-lambda-wide-10. λ DNA. Started at 3000 nm channel. $L/\ell_P = 12.37$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0.75	Forwards	Neither
2400	~	0.25	Backwards	Neither

Table B.42: 130924-RecA-lambda-wide-2. λ DNA. Started at 3000 nm channel. $L/\ell_P = 20.29$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0.50	None	Forwards
2400	~	0	None	None

Table B.43: 130924-RecA-lambda-wide-3. λ DNA. Started at 3000 nm channel. $L/\ell_P = 14.42$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	0.50	Backwards	Forwards	
2400	⊖	0	None	Forwards	Suggests experimental error

Table B.44: 130924-RecA-lambda-wide-4. λ DNA. Started at 3000 nm channel. $L/\ell_P = 13.82$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	1.50	None	Neither	Suspicious bend like ⊖ (with bright spot)
2400	~	0.25	Backwards	Neither	See previous comment
1300	~	0	Forwards	None	

Table B.45: 130924-RecA-lambda-wide-5. λ DNA. Started at 3000 nm channel. $L/\ell_P = 11.72$

APPENDIX B. CLASSIFICATION OF CONFORMATIONS AND EXPERIMENTAL
 B.3. TABLES CONDITIONS

D	Initial condition	Fold rate	Drift	Fold direction
3000	⊂	1.00	None	Neither
2400	~	0	None	None
1300	~	0	Both	None
900	~	0	None	None
600	~	0	Both	None

Table B.46: 130924-RecA-lambda-wide-8. λ DNA. Started at 3000 nm channel. $L/\ell_P = 14.74$

D	Initial condition	Fold rate	Drift	Fold direction	Comments
3000	~	1.25	None	Neither	
2400	~	0	None	None	Sometimes seems to want to form τ
1800	~	0.25	None	Neither	See previous comment; this is the cause of the folding here (one fold, one unfold).
1300	~	0	None	None	
900	~	0	Forwards	None	
600	~	0	Forwards	None	

Table B.47: 130924-RecA-lambda-wide-9. λ DNA. Started at 3000 nm channel. $L/\ell_P = 11.08$

D	Initial condition	Fold rate	Drift	Fold direction
3000	~	0	None	None

C

Example configurations

Below, I detail some rejected configurations, and some accepted, but nonetheless experimentally suspicious, configurations.

C.1 Rejected configurations

130607-RecA-T4-narrow-8 ($L/\ell_P = 13.22$, not shown) is folded over or knotted at the beginning of the 1300nm sequence, and never unfolds throughout the rest of the experiment. It seems highly likely that this is an experimental error, as there is little change in the conformation from this point on. The error can arise when the molecule transitions between channels of different widths, getting tangled in the process. Once a molecule becomes tangled in this way, it often does not untangle at any point in the rest of the experiment.

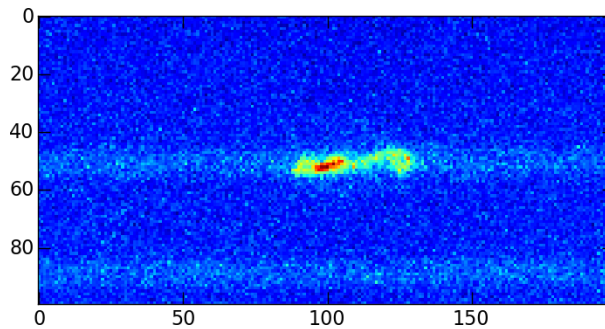


Figure C.1: 130607-RecA-T4-narrow-8 at the start of the 1300nm channel sequence

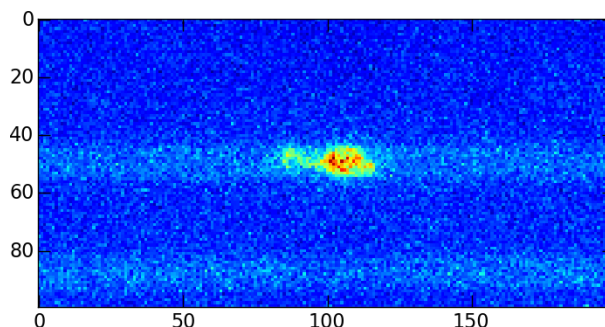


Figure C.2: 130607-RecA-T4-narrow-8 at the start of the 1800nm channel sequence

Similarly suspicious experimental conditions are visible in the following cases: 130905-RecA-T4-narrow-3 (1800nm), 130905-RecA-T4-narrow-4 (1800nm), 130905-RecA-T4-wide-3 (600nm and 900nm), 130905-RecA-T4-wide-7 (600nm), 130905-RecA-T4-wide-8 (1300nm), and 130924-RecA-lambda-wide-3 (2400nm).

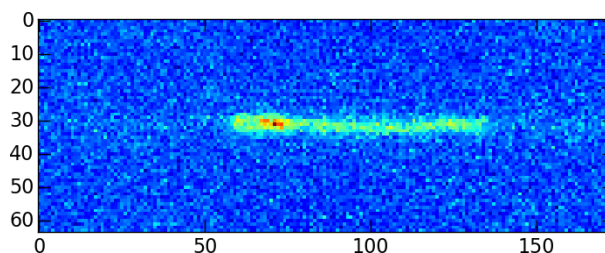


Figure C.3: 130605-RecA-T4-wide-7 at the start of the 600nm channel sequence

The following points were rejected because their relative $\langle X \rangle / L$ errors exceeded the 0.4 threshold: 130603-RecA-T4-wide-4 (0.48 at 1800 nm, 0.59 at 2400 nm); 130603-RecA-T4-wide-3 (0.65 at 2400 nm); 130603-RecA-T4-wide-2 (0.53 at 2400 nm, 0.73 at 3000 nm); 130607-RecA-T4-wide-1 (0.67 at 1300 nm); 130607-RecA-T4-narrow-9 (0.61 at 3000 nm); and 130603-RecA-T4-wide-5 (0.55 at 1800 nm, 0.40 at 2400 nm, 0.53 at 3000 nm).

C.2 Accepted configurations

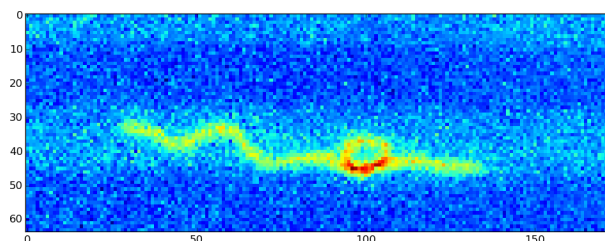


Figure C.4: 130924-RecA-lambda-wide-2 ($L/\ell_P = 20.30$) at frame 131 in the 3000nm channel

Evident in the wider channels, there are three molecules whose $\langle X \rangle / L$ values fall below the simulated curve, unlike the rest of the data; please refer back to Section 3. These cases are worth investigating further, since it is not immediately clear whether they exhibit experimental

errors. The cases in question are 130607-RecA-T4-narrow-9 ($L/\ell_P = 6.83$), 130607-RecA-T4-narrow-10 ($L/\ell_P = 10.62$), and 130905-RecA-T4-narrow-7 ($L/\ell_P = 13.28$); please refer also to the corresponding kymographs in Appendix A. Moreover, 130924-RecA-lambda-wide-2 ($L/\ell_P = 20.30$), shown in the figure above, appears interesting, since it is a long molecule which seems naturally to adopt a much more folded configuration in the 3000nm channel than the 2400nm channel; however, it appears to display a 'knot' (as shown above) in many of the microscopy frames.

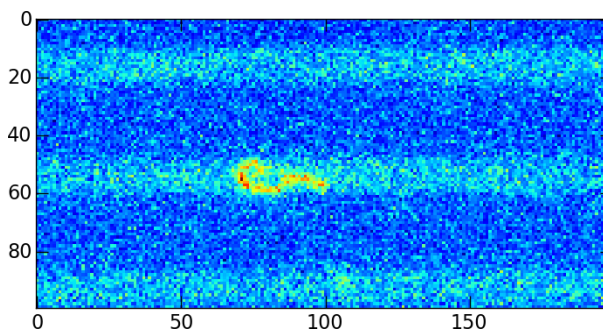


Figure C.5: 130607-RecA-T4-narrow-9 ($L/\ell_P = 6.83$) at the start of the 1800nm channel sequence

130607-RecA-T4-narrow-9 ($L/\ell_P = 6.83$) is folded at the beginning of the 1800nm sequence, but it is not clear if this is an experimental error, or a natural back-fold. The molecule does seem to unfold around frame 138 of that sequence, with the folded end tumbling over the rest of the molecule until the left and right ends have switched and the molecule is extended. However, at around frame 210, the originally folded end once more folds back up against the rest of the molecule. At the beginning of the 2400nm sequence, the molecule appears quite tangled, and similar patterns are displayed for the rest of the experiment.

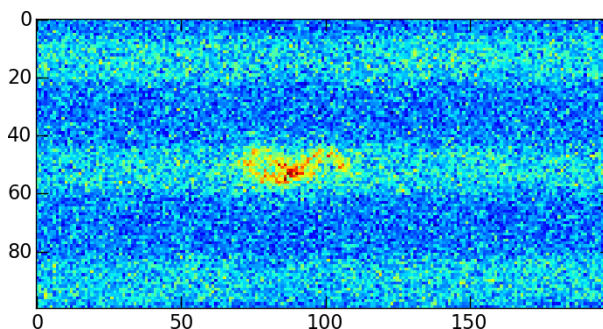


Figure C.6: 130607-RecA-T4-narrow-10 ($L/\ell_P = 10.62$) at the start of the 2400nm channel sequence

130607-RecA-T4-narrow-10 ($L/\ell_P = 10.62$) is folded at the beginning of the 2400nm sequence, but unfurls beginning at around frame 175, and remains without back-folds for the rest of the sequence. At the beginning of the 3000nm sequence, it is again completely folded over, and remains at least slightly folded for the rest of the sequence. It is not clear if this is due to experimental error.

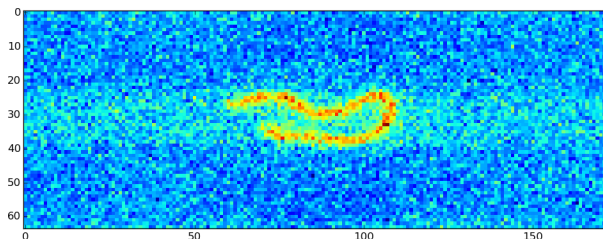


Figure C.7: 130905-RecA-T4-narrow-7 ($L/\ell_P = 13.28$) at the start of the 2400nm channel sequence

Similarly, 130905-RecA-T4-narrow-7 ($L/\ell_P = 13.28$) is folded over at the beginning of the 2400nm sequence (the smallest channel for which we have data for this molecule), but it never seems to unfurl completely for the rest of the experiment.

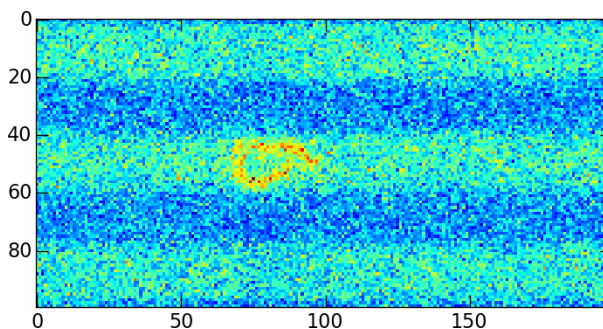


Figure C.8: 130607-RecA-T4-wide-4 ($L/\ell_P = 8.35$) at frame 141 in the 3000nm channel

Finally, 130607-RecA-T4-wide-4 ($L/\ell_P = 8.35$; and which, like all experiments tagged 'wide', starts from the 3000nm channel, and moves to progressively narrower ones) folds over beginning around frame 110, unfolds substantially around frame 300, and then curls back up for the remainder of the time in that channel. A similar bimodal pattern is evident in the 2400nm channel. However, at the beginning of the 1800nm channel, one end is folded back slightly, which may be due to experimental error in the transition between channels. The molecule nonetheless unfolds at around frame 335.

130607-RecA-T4-wide-4 is the one of few molecules which seem to show clearly, particularly here in the two widest channels, a coherent folded configuration that mostly does not seem to be the result of an external experimental condition. Since it shows at least two modes of configuration – folded and unfolded – the average $\langle X \rangle$ is skewed accordingly, and hence this molecule is one of few that is consistent with the simulations. Note that its mean $\langle X \rangle / L$ values fall close to the simulated data (despite the large standard error).

D

Extra figures

D.1 Plots for T4 DNA only

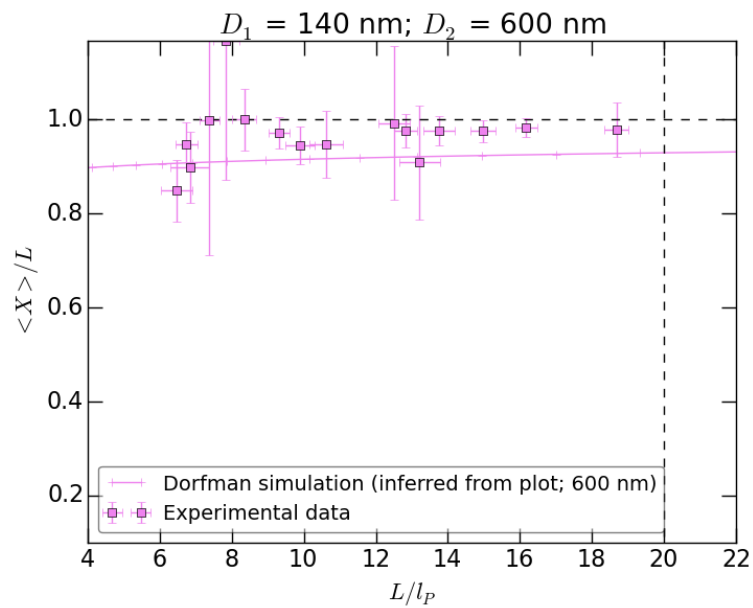


Figure D.1: 600nm channel; linear axes. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

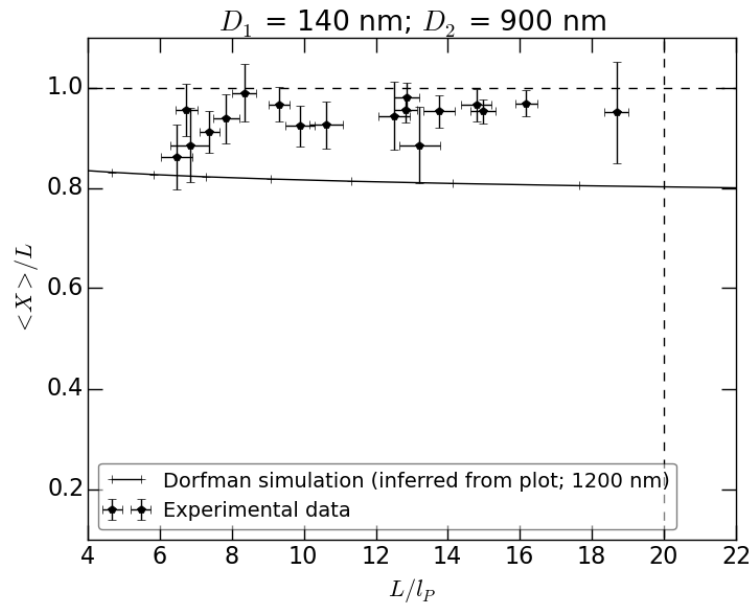


Figure D.2: 900nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

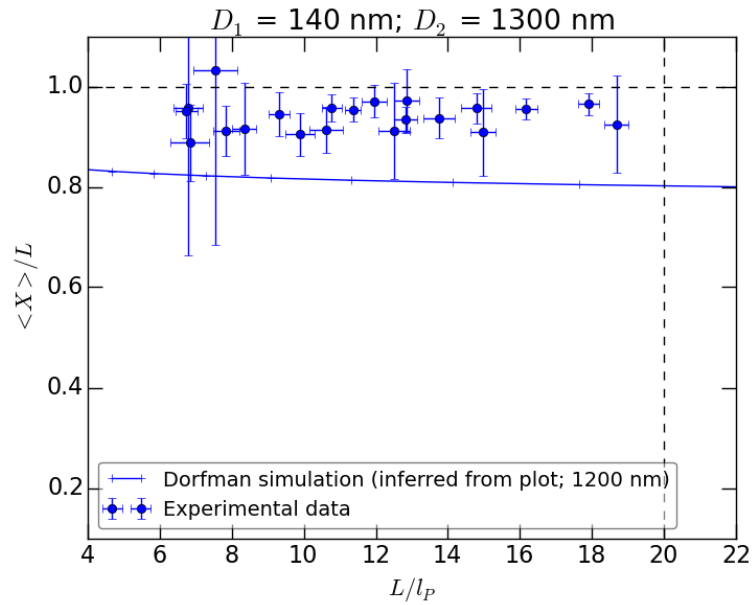


Figure D.3: 1300nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

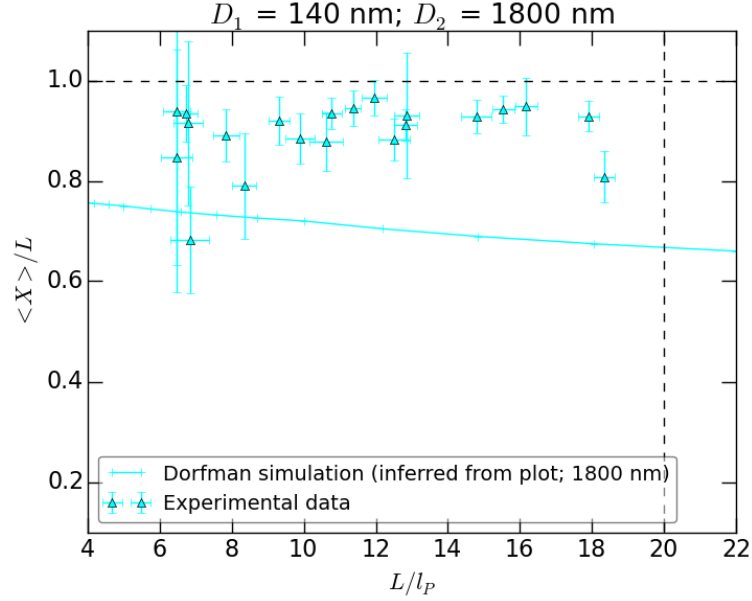


Figure D.4: 1800nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1800nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The point with lowest $\langle X \rangle / L$ value is 130607-RecA-T4-narrow-9 ($L/l_P = 6.83$), which was discussed in Appendix C.

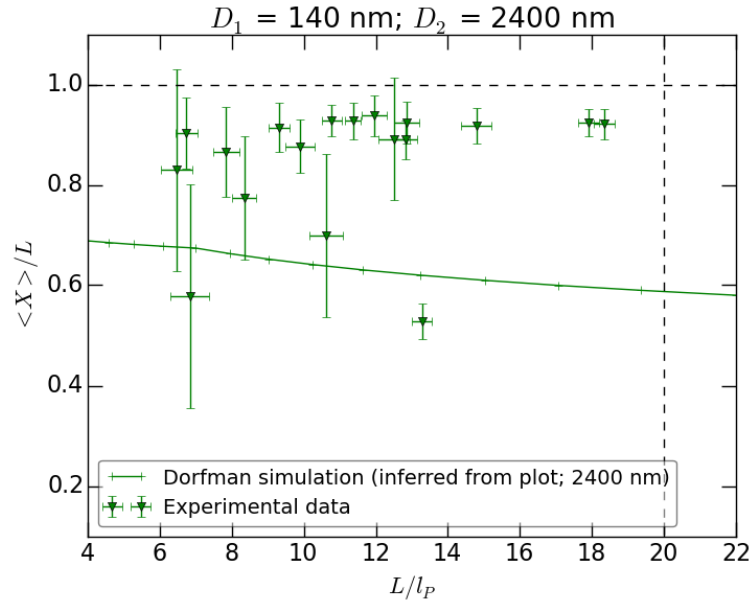


Figure D.5: 2400nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 2400nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The two lower values are 130607-RecA-T4-narrow-9 ($L/l_P = 6.83$) and 130905-RecA-T4-narrow-7 ($L/l_P = 13.28$); they were discussed in Appendix C.

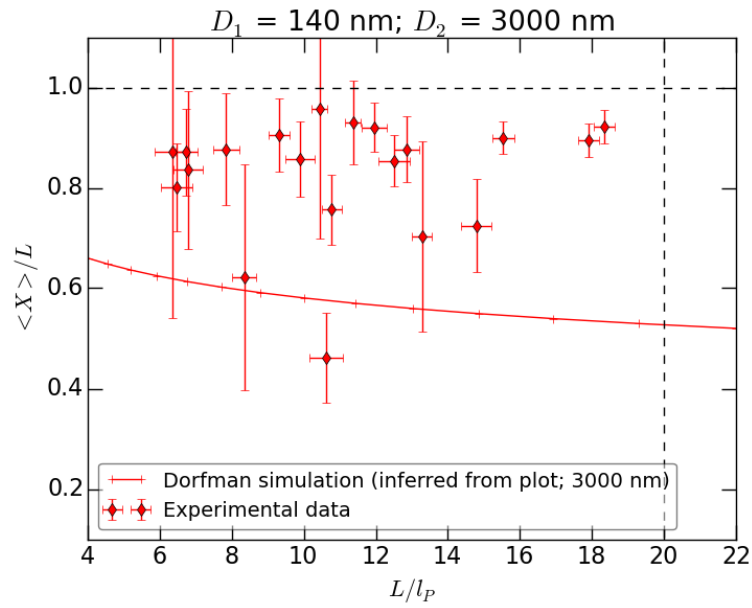


Figure D.6: 3000nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 3000nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation. The point with lowest $\langle X \rangle / L$ value is 130607-RecA-T4-narrow-10 ($L/\ell_P = 10.62$), which was discussed in Appendix C.

D.2 Plots for λ DNA only

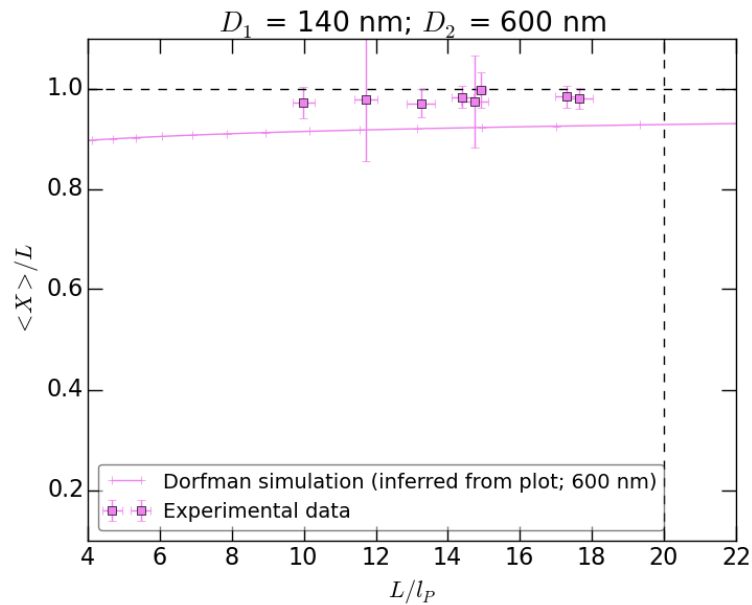


Figure D.7: 600nm channel; linear axes. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

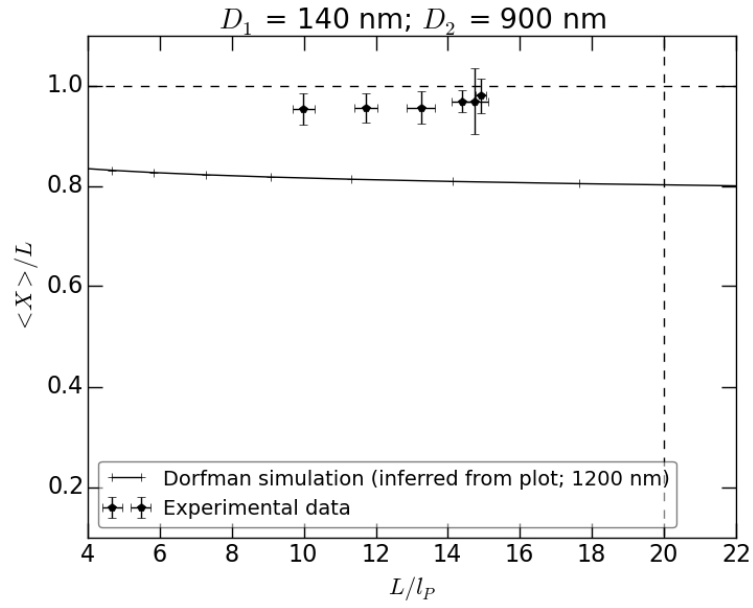


Figure D.8: 900nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

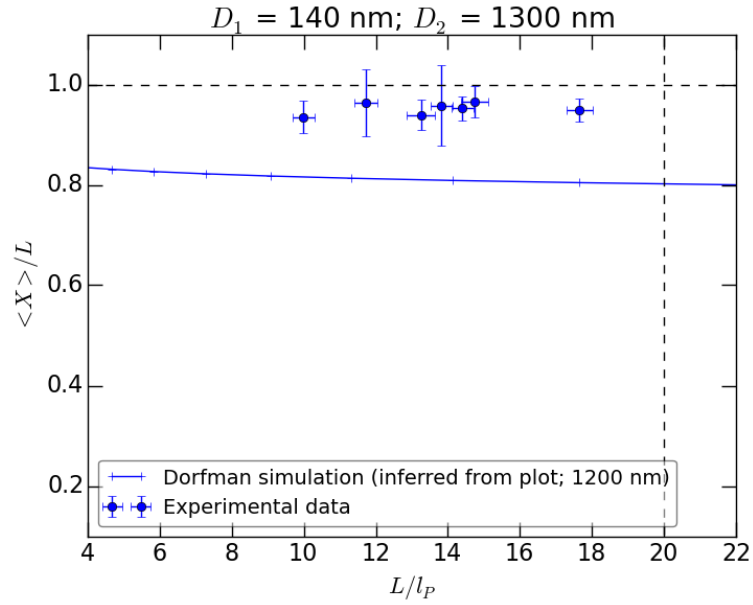


Figure D.9: 1300nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1200nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

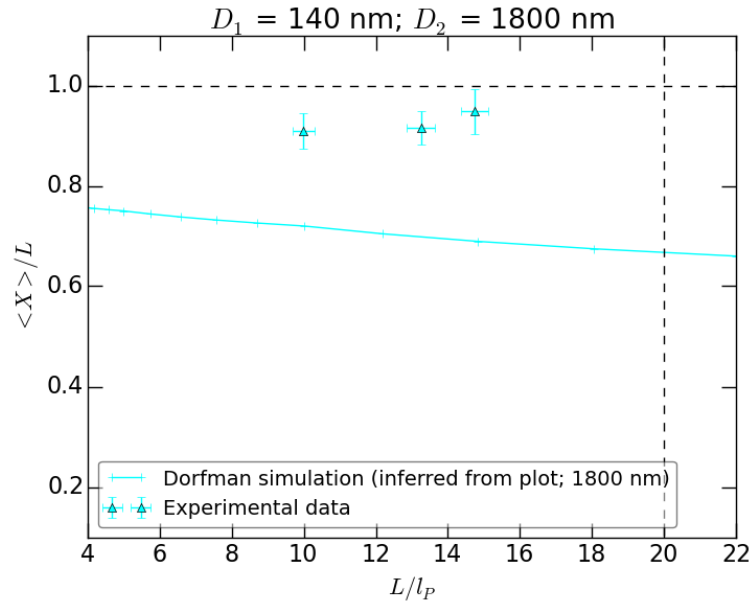


Figure D.10: 1800nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 1800nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

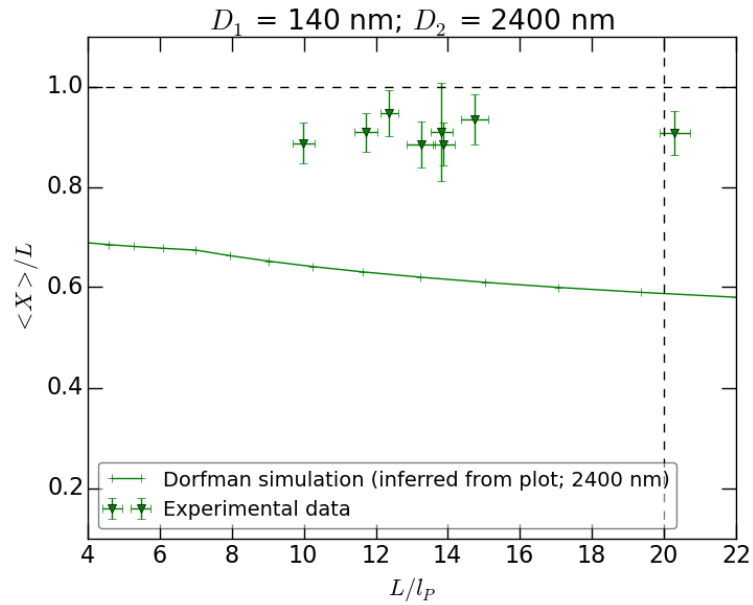


Figure D.11: 2400nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 2400nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

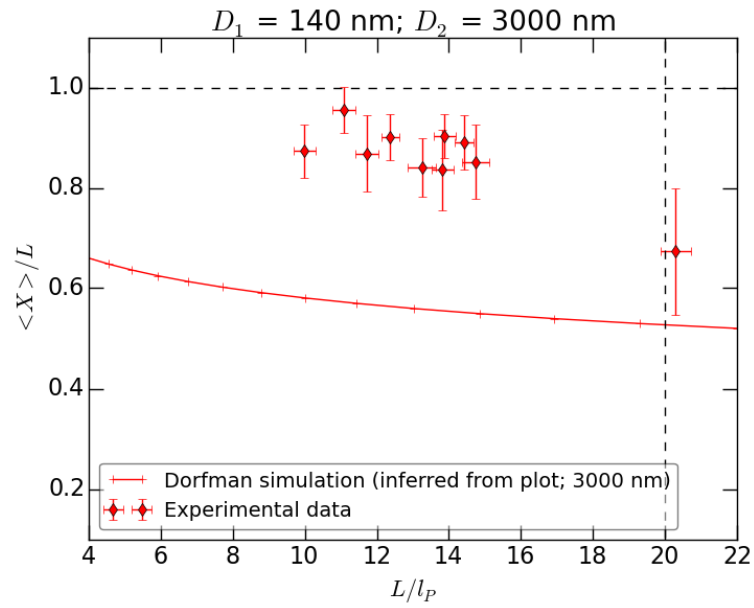


Figure D.12: 3000nm channel; linear axes. Points represent experimental data, whilst the line represents the rescaled simulated data for the 3000nm channel. The dashed vertical line is the dashed line in 3.2. Error bars represent one standard deviation.

Bibliography

- [1] D. R. Tree, A. Muralidhar, P. S. Doyle, K. D. Dorfman, Is DNA a good model polymer?, *Macromolecules* 46 (20) (2013) 8369–8382.
- [2] F. Persson, J. Fritzsche, K. U. Mir, M. Modesti, F. Westerlund, J. O. Tegenfeldt, Lipid-based passivation in nanofluidics, *Nano letters* 12 (5) (2012) 2260–2265.
- [3] W. Reisner, J. N. Pedersen, R. H. Austin, DNA confinement in nanochannels: physics and biological applications, *Reports on Progress in Physics* 75 (10) (2012) 106601.
- [4] M. Doi, S. F. Edwards, *The theory of polymer dynamics*, Vol. 73, Oxford University Press, 1988.
- [5] T. Odijk, DNA confined in nanochannels: hairpin tightening by entropic depletion, *The Journal of chemical physics* 125 (20) (2006) 204904.
- [6] M. Hegner, S. B. Smith, C. Bustamante, Polymerization and mechanical properties of single RecA–DNA filaments, *Proceedings of the National Academy of Sciences* 96 (18) (1999) 10109–10114.
- [7] K. Frykholm, M. Alizadehheidari, J. Fritzsche, J. Wiggenius, M. Modesti, F. Persson, F. Westerlund, Probing physical properties of a DNA-protein complex using nanofluidic channels, *Small* 10 (5) (2014) 884–887.
- [8] T. Odijk, The statistics and dynamics of confined or entangled stiff polymers, *Macromolecules* 16 (8) (1983) 1340–1344.
- [9] A. Muralidhar, D. R. Tree, Y. Wang, K. D. Dorfman, Interplay between chain stiffness and excluded volume of semiflexible polymers confined in nanochannels, *The Journal of chemical physics* 140 (8) (2014) 084905.