Published online 29 July 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 18* **10555–10563**
doi: 10.1093/nar/gkx670

Dynamic stepwise opening of integron *attC* DNA hairpins by SSB prevents toxicity and ensures functionality

Maj Svea Grieb[1], Aleksandra Nivina[2,3,4], Bevan L. Cheeseman[5], Andreas Hartmann[1], Didier Mazel[2,3,*] and Michael Schlierf[1,*]

[1]B CUBE – Center for Molecular Bioengineering, Technische Universität Dresden, Arnoldstraße 18, 01307 Dresden, Germany, [2]Institut Pasteur, Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, 28 rue du Dr. Roux, 75724 Paris, France, [3]CNRS UMR3525, 75724 Paris, France, [4]Paris Descartes University, 75006 Paris, France and [5]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany

Received June 08, 2017; Revised July 19, 2017; Editorial Decision July 20, 2017; Accepted July 22, 2017

ABSTRACT

Biologically functional DNA hairpins are found in archaea, prokaryotes and eukaryotes, playing essential roles in various DNA transactions. However, during DNA replication, hairpin formation can stall the polymerase and is therefore prevented by the single-stranded DNA binding protein (SSB). Here, we address the question how hairpins maintain their functional secondary structure despite SSB's presence. As a model hairpin, we used the recombinogenic form of the *attC* site, essential for capturing antibiotic-resistance genes in the integrons of bacteria. We found that *attC* hairpins have a conserved high GC-content near their apical loop that creates a dynamic equilibrium between *attC* fully opened by SSB and a partially structured attC-6–SSB complex. This complex is recognized by the recombinase IntI, which extrudes the hairpin upon binding while displacing SSB. We anticipate that this intriguing regulation mechanism using a base pair distribution to balance hairpin structure formation and genetic stability is key to the dissemination of antibiotic resistance genes among bacteria and might be conserved among other functional hairpins.

INTRODUCTION

Bacteria have an impressive capability of acquiring antibiotic resistances using the integron recombination platform. The integron is the predominant tool for adaptation in gram-negative bacteria (1,2). It is a genetic platform that assembles, and rearranges promoterless gene cassettes through a unique recombination process in which the integron integrase IntI recognizes and recombines a folded single-stranded DNA hairpin called *attC* (3,4). *attC* sites can form 57 to 141 nucleotide (nt)-long hairpins (3), which display conserved structural features that are recognized by the integrase IntI (5,6). IntI's specificity for folded *attC* hairpins is advantageous for bacteria, as it allows the coupling of recombination with the presence of single-stranded DNA (ssDNA) during replication or DNA exchange among bacteria (2). However, DNA hairpins can also have toxic or deleterious effects, for instance, replication arrest due to polymerase stalling (7). These effects are prevented by the single-stranded DNA binding protein (SSB) which binds to ssDNA with high affinity and low sequence specificity (8,9). Indeed, SSB has been shown to open short double-stranded DNA (dsDNA) stretches of up to 12 bp when migrating along ssDNA (10–12) and to hamper *attC* hairpin folding *in vivo* in the absence of IntI (13). However, it is unknown how SSB removes secondary structure of hairpins up to 141 nt in length and how hairpins maintain their functional folded structure in presence of SSB. Moreover, the recognition dynamics of folded *attC* by the integrase in presence of several thousands of SSB monomers per cell (14,15) is completely unknown. Here, we use an *in vitro* approach based on single-molecule Förster Resonance Energy Transfer (smFRET) (16) to describe how the structure of an *attC* hairpin balances both efficient opening by SSB to avoid toxicity, and reliable recognition by IntI to ensure functionality (Figure 1A). We found that *attC* hairpins have a conserved high GC-content near their apical loop that creates a dynamic equilibrium between *attC* fully opened by SSB and a partially structured attC-6–SSB complex. This complex is recognized by the integrase IntI, which extrudes the full hairpin upon binding while displacing SSB. *In vivo* recombination assays

*To whom correspondence should be addressed. Tel: +49 351 463 43050; Fax: +49 351 463 40322; Email: michael.schlierf@tu-dresden.de
Correspondence may also be addressed to Didier Mazel. Tel: +33 1 40613284; Email: mazel@pasteur.fr

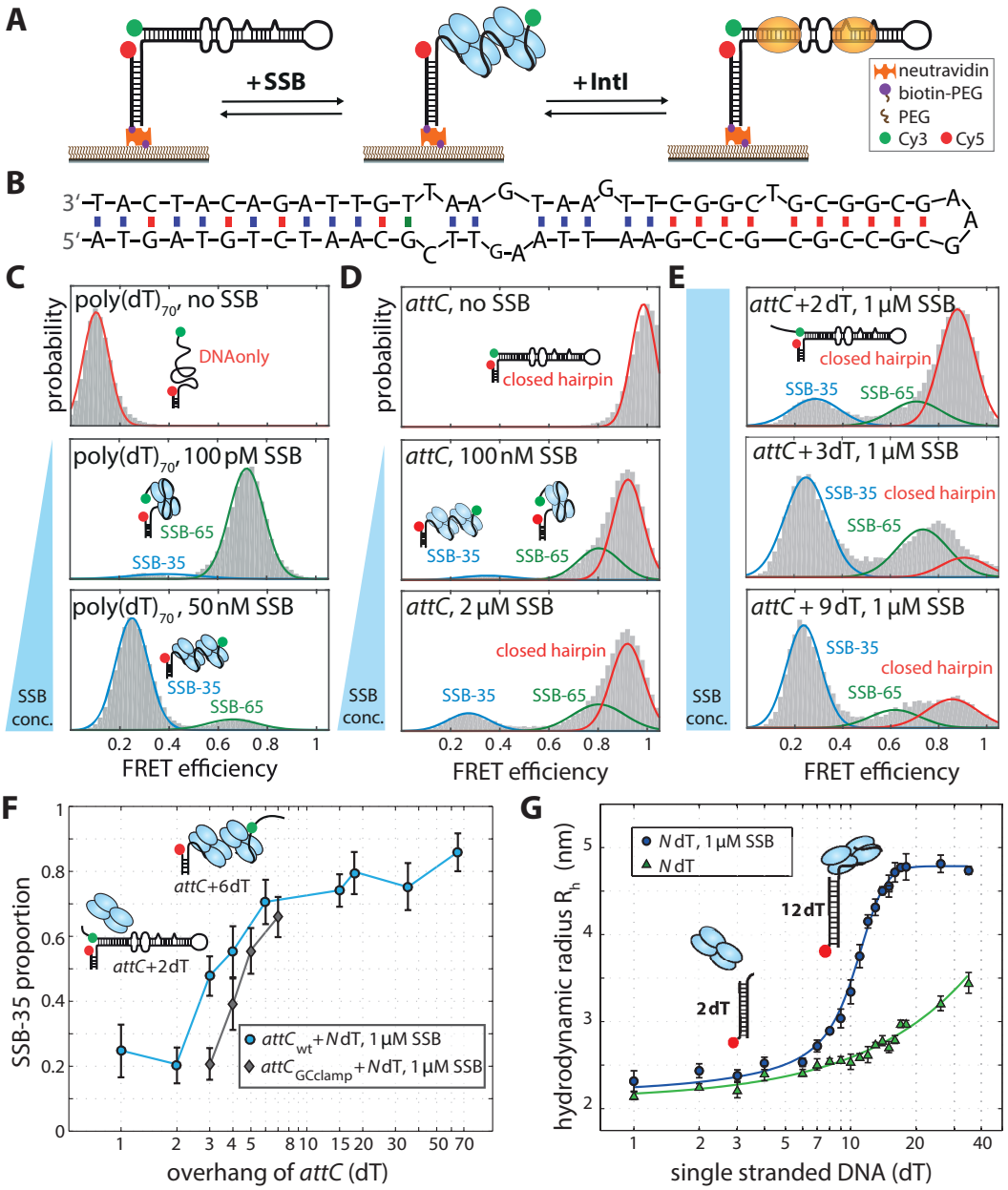

Figure 1. SSB requires a single-stranded DNA overhang to open the *attC* hairpin. (**A**) Schematic illustration of smFRET assay to study the SSB-IntI competition. The average distance between the donor (green) and acceptor (red) serves as a readout of the molecular conformation. (**B**) The secondary structure of the $attC_{aadA7}$ hairpin predicted by RNAfold (43). (**C**) Binding of SSB to poly(dT)$_{70}$. SSB binds in the SSB-65 binding mode at low protein-to-DNA ratios and in the SSB-35 binding mode at higher SSB concentrations (c_{SSB}). (**D**) Inefficient SSB opening of the *attC* hairpin by SSB-65 and SSB-35 binding modes. (**E**) Increased hairpin opening efficiency due to the presence of a ssDNA overhang of 2, 3 or 9 dT nucleotides on the *attC* hairpin. (**F**) Hairpin opening efficiency quantified by SSB-35 proportion with ssDNA overhangs from 1 nt to 65 nt at c_{SSB} = 1μM. DNA breathing was reduced in the $attC_{GCclamp}$ hairpin (grey), which hence requires a longer overhang for opening. (**G**) Hydrodynamic radius of SSB-ssDNA complexes at c_{SSB} = 1 μM obtained by FCS. SSB binds to ssDNA as short as 7 nt and the binding affinity increases with the length of ssDNA until 17 nt. The green line presents the control without SSB. All error bars depict the 95% confidence interval.

of a mutant abolishing the attC-6–SSB complex showed a reduced recombination frequency.

MATERIALS AND METHODS

Proteins

We used the catalytically inactive integron integrase mutant IntI1-Y312F, which is unable to cleave its recombination site but retains full binding ability (17). It was purified using a maltose binding protein (MBP) fusion. The IntI1-MBP fusion protein has been shown to retain full IntI1 functionality *in vitro* and *in vivo* (18). Purification was done in a two-step protocol using first an amylose column (MBPTrap HP, 5 × 1 ml, GE healthcare Life Sciences), followed by an ion exchange column (HiTrap SP HP, 5 × 1 ml, GE Healthcare Life Sciences) to remove contaminants. The *Escherichia coli*

Single-Stranded DNA Binding Protein was obtained from Promega (Madison, USA).

DNA constructs

All oligonucleotides used in this study are listed in Supplementary Table S1. They were obtained from biomers.net (Ulm, Germany), PURIMEX (Göttingen, Germany), eurofins genomics (Ebersberg, Germany), or IBA lifesciences (Göttingen, Germany). Biotin or dye modifications at the ends of the oligonucleotides were attached as NHS esters to an amino-C6-dT or an amino-C6-dA base. Labeled DNA was purified using a reverse phase C18 column (Phenomenex) by High Performance Liquid Chromatography (HPLC).

smFRET

For smFRET measurements the partial duplex DNA constructs carrying one donor (Cy3) and one acceptor (Cy5) were immobilized at the duplex ends on a polyethylene glycol (PEG)-passivated quartz surface via biotin–neutravidin interaction at sample concentrations of 25 pM. The FRET imaging buffer contained 20 mM NaCl, 50 mM Tris–HCl in a saturated, aged Trolox-solution (6-hydroxy-2, 5, 7,8-teramethylchroman-2-carboxylic acid) at pH 8 with an oxygen scavenging system composed of 1 mg/ml glucose oxidase and 0.8 % (w/v) D-glucose, 1 kU/ml catalase. Single molecules were observed at 23°C in a custom-built prism-type total internal reflection microscope (19,20) at a frame rate of 33 s^{-1}. Proteins were added at the concentrations specified and incubated for 10 min prior to data acquisition. The FRET histograms were constructed using FRET time traces of 70 to 150 molecules per condition.

smFRET state finding algorithm

The custom state-finding pipeline consists of three main steps. First, a generic Hidden Markov Model (HMM) was used to find candidate states in the smFRET traces. Second, the found *'HMM states'* are assigned to '*molecular states*' in a user-assisted post-processing step. Third, the rates and state proportions are calculated from the resulting assignment. HMM was performed using the Matlab statistics and machine learning toolbox utilizing the Baum-Welch algorithm (21). First, the FRET efficiency values were quantized into 50 bins between 0 and 1. All traces from one experiment are fit globally by m+n HMM models, differing in their number of hidden states ranging from 1 to m+n, where m is the number of expected states and n the number of over-fitting states. First, the m+n states HMM is run and its states are initialized evenly between 0 and 1. The state number is then reduced by one, taking the previous model's fitted states for initialization and removing the least probable state. This is repeated until all models from 1 to m+n are fit. Here, the maximum number of possible states m was estimated from previous knowledge and preliminary analysis using 'Fast Step Transition and State Identification' algorithm (22) and set to 4. The number of used over-fitting states n was set to 3. The over-fitting of states allows for the correction of misassigned states during the user-assisted post-processing and was found to greatly increase the robustness of the analysis (see Supplementary Figure S4). The user-assisted post-processing is required to map the *'HMM states'*, i.e. the results from the fitting, to the m '*molecular states*', i.e. the actual FRET states characterizing a certain molecular conformation. The user is first presented with m+n FRET histograms summarizing the fits of the different HMM models, chooses the best fit and merges HMM states when necessary. To assist the user in assessing whether the state assignment is correct, the empirical cumulative distribution function (CDF) of the dwell times of each transition in the reassigned data is calculated. The empirical CDF is obtained by the Kaplan–Meier estimator (23) with right censoring provided by the eCDF function of MATLAB. Once the state assignment is performed, the newly *'assigned state model'* is used for all further analysis. For the calculation of the transition rates and the state proportions from the '*assigned state models*', the following parameters and filtering was used. A minimum dwell time of 10 frames was applied to account for the resolution limit and to avoid fitting of noise or blinking. Transition rates $\hat{r}$ between pairs of states A and B were estimated using the maximum likelihood estimator for an exponential distribution, subject to right censoring and calculated by

$$\hat{r} = \frac{N_{A\to B}}{N_A \cdot \hat{\tau}_A}\,, \tag{1}$$

where $N_{A\to B}$ is the number of transitions from A to B observed in all traces, N_A, the total number of transitions out of state A, including those occurring due to the trace ending, and $\hat{\tau}_A$ the average dwell time of state A. The 95 % confidence intervals were estimated using the inverse Fisher information, assuming a normal distribution, as

$$\hat{\sigma} = \frac{\hat{r}}{\sqrt{N_{A\to B}}}\,. \tag{2}$$

State proportions were also calculated from the '*assigned state model*'. The proportion of time spent in each state was estimated from the Viterbi path. Bootstrapping was then performed to obtain an error for the state proportions introduced through state assignment. For all states, FRET traces were re-sampled 10 000 times with replacement. The standard deviation was then estimated from the empirical distribution and used to estimate a 95% confidence interval assuming a normal distribution.

FCS

FCS samples used a partial duplex DNA to attach the ATTO-655 fluorophore. Measurements were performed in 20 mM NaCl, 50 mM Tris–HCl, pH 7.4. The DNA concentration was 1 nM and protein was added at concentrations as indicated. Fluorescence intensity data was acquired using a home-built confocal microscope combined with time-correlated single photon counting (24). Excitation of ATTO-655 was achieved by a pulsed laser source with a wavelength of 640 nm. FCS data analysis was performed using a software correlator and an FCS fitting tool box from PicoQuant (Berlin, Germany).

Free energy profiles

Free energy profile barrier heights were calculated from transition rates using an Arrhenius equation and a 1 ms^{-1} attempt frequency. The closed hairpin structure was set as a zero reference state. The reaction coordinate axis is arbitrarily scaled.

Electrophoretic mobility shift assay

Oligonucleotides were obtained from Sigma-Aldrich, labelled at the 5′ end with radioactive phosphate transferred from [γ-^{32}P]ATP (PerkinElmer) by T4 polynucleotide kinase (ThermoFisher Scientific) and purified with MicroSpin G-25 columns (GE Healthcare). Each reaction contained 500 ng Poly[d(I-C)], 12 mM HEPES NaOH pH 7.7, 12 % glycerol, 4 mM Tris–HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 0.06 mg/ml BSA, 1 mM DTT, 10 % Tween-20, 0.01 nmol of the corresponding ^{32}P-labeled DNA oligonucleotide and the amount of protein indicated, in a final volume of 20 µl. The samples were incubated at 30 °C for 10 min without the probe followed by 20 min with the probe, then loaded to a 5% native polyacrylamide gel (acrylamide/bisacrylamide 37.5:1), with 0.5× TBE as buffer. The gel was run for 2 h at 40 mA with 0.5× TBE as running buffer and visualized using chemo-luminescence film (Amersham HyperfilmTM).

Calculation of base pair compositions of *attC* sites

The list of *attC* sites was obtained from the authors of the INTEGRALL database (25) containing 263 *attC* sites with less than 95 % sequence identity, all originating from mobile integrons. For each site, the GC-content was calculated for the core region (defined as the R-box, the unpaired central spacer, and the L-box) and for the apical region (defined as the rest of the stem and the variable terminal structure).

In vivo recombination assay

Wild-type and mutant *attC* sites were constructed by annealing two overlapping, phosphorylated oligonucleotides (Supplementary Table S1), fully complementary except for the overhangs corresponding to EcoRI and BamHI restriction sites. The sites were then ligated into the pSW23T vector carrying a chloramphenicol (Cm) resistance marker (26) digested with EcoRI/BamHI. The plasmids were then transformed into *Escherichia coli* β2163 strain (3) and their sequence was verified by sequencing. The recombination frequency was measured using a previously developed suicidal conjugation assay (27). Briefly, pSW23T vectors were transferred through conjugation from the β2163 strain into a recipient *E. coli* DH5α strain, so that the transferred strand carried the bottom strand of *attC* sites. The recipient strain harbored the pBAD plasmid allowing the expression of IntI1 integrase (28) and a pSU plasmid with an *attI1* site (27). Successful *attI1* × *attC* recombination in the recipient strain conferred its resistance to Cm, allowing us to measure the rate of recombinant cells (Cm^R) among the total population of recipient cells (Cm^S) by plating. Recombination frequencies correspond to the average of three independent experiments.

RESULTS

SSB requires a single-stranded DNA overhang to open DNA hairpins

In a first step, we characterized the interaction of SSB with the $attC_{aadA7}$ hairpin of 70 nt (3) (Figure 1B). We used the tetrameric SSB from *E. coli* (9), as *E. coli* is a common host of antibiotic resistance integrons. SSB has multiple binding modes, depending on temperature, salt concentration and protein-to-DNA ratio, which are named after the number of nt occluded by SSB: SSB-17, SSB-35, SSB-56 and SSB-65 (29–33). In our low-salt condition (20 mM NaCl, 50 mM Tris–HCl, pH 7.5), SSB complexed unstructured ssDNA of 70 nt ($poly(dT)_{70}$) with its major binding modes SSB-35 and SSB-65 (Figure 1C). The low FRET efficiency population ($E_{FRET} \approx 0.1$) of $poly(dT)_{70}$ vanished upon addition of SSB binding in SSB-65 ($E_{FRET} \approx 0.7$) at 100 pM SSB concentration (c_{SSB}), and gradually transitioned into SSB-35 ($E_{FRET} \approx 0.3$) above 1 nM c_{SSB}, in accordance with previous reports (32). Next, we performed the SSB titration on the *attC* hairpin, where the presence of the SSB-characteristic FRET populations reflects hairpin opening. Indeed, SSB addition decreased the closed hairpin population ($E_{FRET} \approx 0.9$), and SSB-35 and SSB-65-like E_{FRET} populations occurred (Figure 1D). However, the c_{SSB} required for these populations were 1000-fold higher than for $poly(dT)_{70}$, and even at 2 µM c_{SSB}, 75 % of the hairpins remained closed.

The opening efficiency of the *attC* hairpin by SSB was largely increased by the presence of a single-stranded DNA overhang (ssDNA overhang). This effect is demonstrated by the increase of the SSB-35 population with increasing ssDNA overhang (Figure 1E). We quantified this effect by calculating the proportion of SSB-35 at $c_{SSB} = 1$ µM for ssDNA overhang lengths of 1–65 nt (Figure 1F, cyan curve). The observed increase in opening efficiency with the ssDNA overhang suggests that SSB requires binding to the ssDNA overhang prior to hairpin opening and that its minimal interaction site is 3 nt, as this is the apparent minimal ssDNA overhang required. However, this result disagrees with SSB's reported minimal binding site of 17 nt (33,34).

To learn more about the binding of SSB to ssDNA we used Fluorescence Correlation Spectroscopy (FCS) (35). FCS rendered the hydrodynamic radius (R_h) of fluorescently labelled ssDNA of 1 to 35 nt in the absence (green) and the presence (blue) of 1 µM SSB (Figure 1G). SSB binding to the ssDNA increased the R_h and is characterized by the divergence of the two curves. SSB showed binding to ssDNA as short as 7 nt and its binding affinity increased exponentially with increasing ssDNA until the plateau at 17 nt (Figure 1G, Supplementary Figure S1, Supplementary text). We infer that binding of SSB to ssDNA between 7 and 17 nt is governed by a subset of the amino acids involved in the SSB-17 binding mode, composed of Arg-56, Phe-60 and Trp-40, 54 and 88 (31,33,36,37). This hotspot binds to a total of 7 nt of ssDNA according to the crystal structure (Supplementary Figure S2). The binding of 7 nt by SSB does not represent a stable binding mode, but rather a loose association, characterized by a high dissociation constant K_D (SSB-7) ≈ 10 µM (Supplementary Equation S2). It further rationalises why a binding of SSB to 8 nt was re-

ported once, but could have never been confirmed (38). In summary, FCS revealed SSB's minimal association site of 7 nt, but did not explain how an ssDNA overhang of 3 nt on the *attC* hairpin is sufficient for the *attC* hairpin opening by SSB.

The discrepancy between the 3 nt ssDNA overhang required for hairpin opening and the 7 nt required for SSB binding can be explained by DNA breathing, describing the spontaneous opening of double-stranded DNA (dsDNA) at ssDNA–dsDNA junctions (39,40). In the *attC* hairpin, breathing is facilitated by the two AT bp at the end, which expose 4 nt of ssDNA adding to the 3 nt ssDNA overhang to reach 7 nt. We tested this hypothesis by a GC-clamp that reduced breathing by mutating the AT bp at the end to thermodynamically more stable GC bp (Figure 1F, gray line). Indeed, hairpin opening by SSB was shifted to longer ssDNA overhangs (Figure 1F, Supplementary Figure S3), confirming that dsDNA opening requires SSB to bind to ssDNA provided by an ssDNA overhang and by breathing DNA of both strands. We observed opening of *attC* hairpins without a ssDNA overhang with a small percentage (Figure 1D), which can be explained by a low, but non-zero probability that DNA breathing exposes 7 nt. We further found that hairpin opening by SSB is 5-fold more efficient when proceeding in a 3′ to 5′ direction (Supplementary Figure S3, Supplementary text). The directionality was observed only for hairpin opening, and not for binding to ssDNA. Thus, *E. coli* SSB is less efficient in capturing breathing DNA in the 5′ to 3′ direction due to the directional wrapping of ssDNA around the SSB tetramer, as was previously proposed for SSB from *Deinococcus radiodurans* and *E. coli* (10).

SSB–*attC* hairpin complex structures are highly dynamic

Next, we investigated the pathways and kinetics of SSB opening the *attC* hairpin. For this, a state-finding algorithm for smFRET time traces that identifies E_{FRET} states, their lifetimes and transitions was used. However, available algorithms failed to identify states across varying SSB concentrations, as they are biased toward highly populated areas of intrinsically noisy FRET traces. To overcome this limitation, we developed a two-step state-finding pipeline (Supplementary Figure S4) that overfits the data with more states than expected to ensure sparsely populated states to be identified (Materials and Methods). The user-assisted post-processing then allows merging of states if necessary. We confirmed the algorithm by reproducing the results of the well-studied interaction of SSB with poly(dT)$_{70}$ (32) (Supplementary Figure S4).

The analysis of SSB opening the *attC* hairpin shows four distinct FRET states (Figure 2A and B). In addition to the expected closed hairpin (red), SSB-35 (cyan) and SSB-65 (green), a state at $E_{FRET} \approx 0.85$ was found (dark blue). We hypothesized this state to be a partially folded hairpin bound by SSB, as we also observed for artificial hairpins (Supplementary Figure S5). We proposed this hairpin to be composed of six successive GC bp close to the apical loop of *attC* and hence termed the complex attC-6–SSB (inset of Figure 2B). To confirm its existence, we designed a negative control, $attC_{AT}$, in which the GC stretch was replaced by AT bp reducing the probability of attC-6–SSB formation. Indeed, the attC-6–SSB population was significantly reduced and the hairpin was more often completely opened by SSB, indicated by the increased SSB-35 population (Figure 2C). Additionally, a positive control containing only the GC-stretch, while all other bases were replaced by dT ($attC_{GConly}$), showed a prominent attC-6–SSB population (Figure 2D). We conclude that SSB opening of the *attC* hairpin involves a partially closed hairpin–SSB complex.

The opening pathway of *attC* by SSB was reconstructed using the transition rates (Figure 2E, Supplementary Figure S6) and shows that SSB opens the *attC* hairpin via the attC-6–SSB conformation. The initial transition increases with c_{SSB}, because it depends on SSB binding to available ssDNA at the end of the hairpin. From attC-6–SSB, the complex transitions to the SSB-65 state through opening of the 6 bp with a c_{SSB}-independent rate. The transition to SSB-35, describing the addition of a SSB monomer, is expectedly dependent on c_{SSB}. The transitions back to the hairpin follow the reverse pathway with c_{SSB} independent rates describing either dissociation events, or base pairing of attC-6–SSB. In conclusion, *attC* opening by SSB proceeds in a dynamic and stepwise fashion stabilizing an on-pathway intermediate of a partially closed hairpin termed attC-6–SSB.

IntI recognizes the attC-6–SSB complex and displaces SSB

We investigated the role of the attC-6–SSB state for the biological functionality of *attC* recombination sites, by performing experiments with both SSB and IntI under competition for *attC* (Figure 3A). To minimize IntI-fluorophore interactions, the donor fluorophore was repositioned to the apical loop, without affecting the interaction with SSB (Supplementary Figure S7). When IntI was added to the hairpin and bound it, the low E_{FRET} population of the closed hairpin remained (Figure 3A). However, when SSB was added in the absence of IntI, the closed hairpin population disappeared and was replaced by the characteristic SSB populations, which are highlighted by a blue outline for simplicity (Figure 3A). Keeping the c_{SSB} constant while adding increasing concentration of IntI (c_{IntI}) allowed studying their binding competition, as the appearance of a peak at low E_{FRET} (orange) indicated the displacement of SSB from the hairpin. The results show that even at a 100-fold molar excess of SSB, IntI efficiently displaces SSB from 50% of the hairpins. At equimolar concentration, >90% of the hairpins are closed. This effect is specific to IntI's target site, the *attC* bottom strand, as it was significantly less pronounced on the non-recombinogenic *attC* top strand (Supplementary Figure S8). We concluded that IntI can outcompete SSB for its recombination site.

We hypothesized that IntI's efficient binding of the hairpin despite SSB is due to the recognition of the attC-6–SSB state, which exposes part of the *attC* hairpin. To test this, we used the $attC_{AT}$ hairpin, which has a greatly reduced attC-6–SSB population at $c_{SSB} = 1\mu M$ (Figure 2C). After confirming that this mutation does not affect IntI binding using a gel-shift assay (Supplementary Figure S9), we performed the smFRET competition assay (Figure 3B). Remarkably, at a concentration ratio of 1:100, IntI could only displace SSB in 8% of the cases on $attC_{AT}$ compared to 44% on the

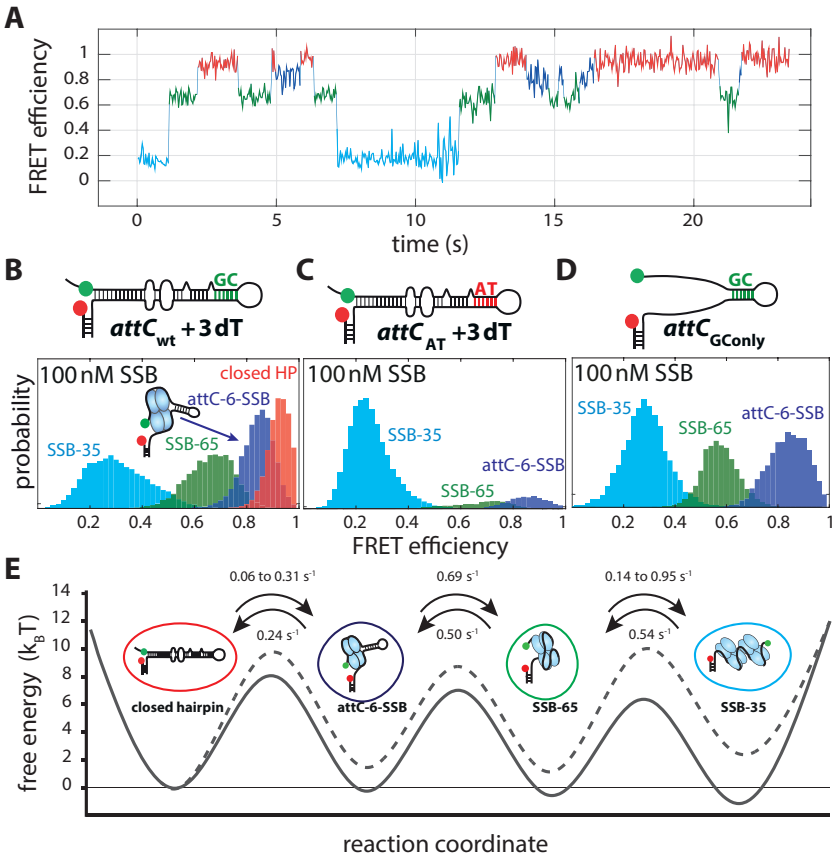

Figure 2. SSB opens the *attC* hairpin in discrete steps. (**A**) Representative FRET time trace of the $attC_{wt}$+3dT hairpin with 100 nM SSB. The state finding algorithm identifies four distinct FRET states highlighted in different colors. (**B**) Histogram of the found states, SSB-35 (cyan), SSB-65 (green), closed hairpin (red) and attC-6–SSB (dark blue) at c_{SSB} = 100 nM. Inset: Schematic of the attC-6–SSB complex. (**C**) Histogram of $attC_{AT}$. The attC-6–SSB and SSB-65 states are rarely populated. (**D**) Histogram of $attC_{GConly}$. A dominant attC-6–SSB state is observed. (**E**) Summary of the extracted transition pathways and rates of SSB opening the $attC_{wt}$+3dT hairpin in a free energy landscape diagram. Dashed line for c_{SSB} = 1 nM, solid line for c_{SSB} = 1 μM.

$attC_{wt}$. Even at equimolar concentration, only 45% of the hairpins were bound by IntI. This supports the hypothesis that IntI recognizes the attC-6–SSB state and subsequently displaces SSB from the hairpin and that the lack of the attC-6–SSB state significantly reduces the recognition by IntI.

To test whether the reduced recognition of the $attC_{AT}$ hairpin by IntI due to the action of SSB is also visible *in vivo*, we compared the recombination frequencies of the $attC_{wt}$ and the $attC_{AT}$ hairpins using a well-established *in vivo* recombination assay (27). Indeed, the results show that IntI recombines the $attC_{AT}$ hairpin approximately 50% less efficient than the $attC_{wt}$ (Figure 3C). These results show that the attC-6–SSB state plays an important role for the recognition by IntI and therefore for the functionality of the integron both *in vitro* and *in vivo*.

DISCUSSION

Is this regulation mechanism by a high GC content close to the apical loop found only in the $attC_{aadA7}$ hairpin or a common motif in Integron recombination sites? For this, we performed a bioinformatic analysis in which we compared the GC-content of the apical and the core region of 263 different *attC* sites occurring *in vivo* (Figure 4A). This analysis revealed that the apical part of the hairpins, close to the loop, has significantly more GC base pairs than the core region. In conclusion, we propose that an attC-6–SSB like state is the key to maintaining *attC* functionality in the presence of SSB. This state is generated by a conserved, weighted base pair distribution along the hairpin.

Based on our results, we propose the following model of the interaction of SSB and IntI with *attC* hairpins both *in vivo* and *in vitro* (Figure 4B). SSB can open the *attC* hairpin, if it can bind to a single-stranded DNA overhang of at least 3 nt. However, the conserved base pair distribution of *attC* hairpins, featuring a high GC-content in the apical tip of the hairpin, does not allow SSB to open the hairpin irreversibly. Instead, the high GC-content creates a dynamic equilibrium transitioning on the seconds scale between the fully opened hairpin (SSB-65 and SSB-35) and a partially folded *attC* hairpin (attC-6–SSB). We show that IntI recognizes the short hairpin of the attC-6–SSB complex, binds to it as a monomer and then recruits a second monomer, while displacing SSB (Supplementary Figure S8). Assuming an SSB concentration of 1 μM, the $attC_{aadA7}$ will transition on average every 4–5 s to the attC-6–SSB state and thus allowing IntI binding. This model can also explain the regulation of *attC* hairpins in living bacteria, where SSB is constitutively expressed, while IntI expression is suppressed

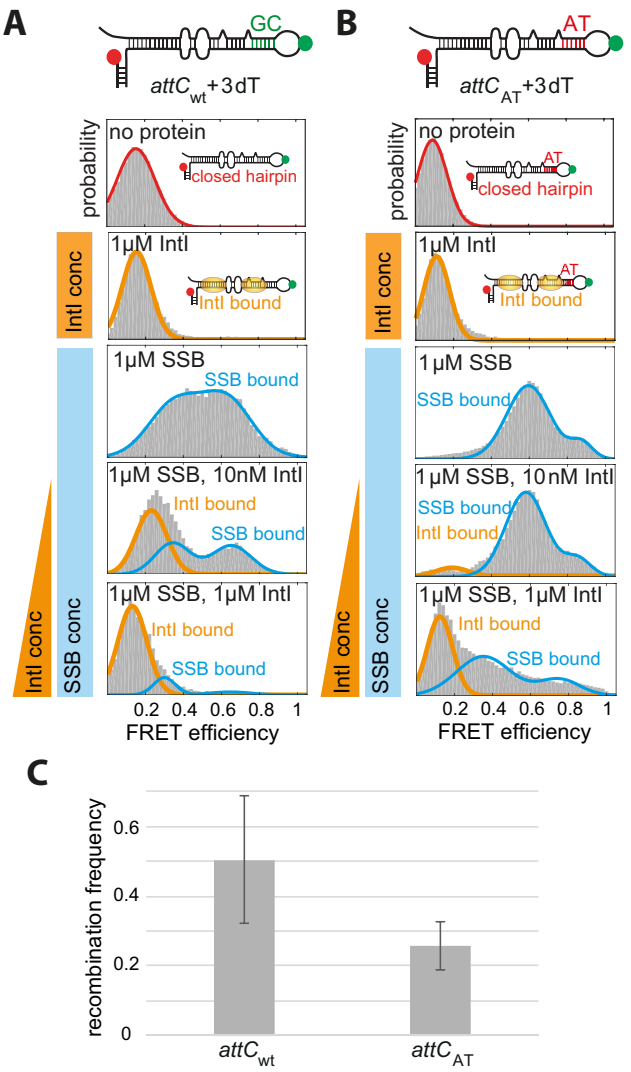

Figure 3. IntI outcompetes SSB for the *attC* hairpin. (**A**) The *attC* hairpin is labeled at the apical loop to minimize IntI-fluorophore interactions. When IntI binds to the hairpin, the E_{FRET} shows the closed hairpin. The SSB-characteristic populations form in the presence of 1 μM SSB, summarized by the blue outline. Upon addition of IntI to the SSB-*attC* complex, the low E_{FRET} peak appears, representing the displacement of SSB by IntI. (**B**) The same competition experiment with $attC_{AT}$, which does not stabilize attC-6–SSB, characterized by the difference in the histograms at 1 μM SSB compared to the wild type. Here, SSB displacement by IntI is much less efficient. (**C**) *In vivo* recombination frequencies show that $attC_{AT}$ recombination is approximately 50% less efficient than the wildtype *attC* site. Error bars show the standard deviation ($P = 0.093$).

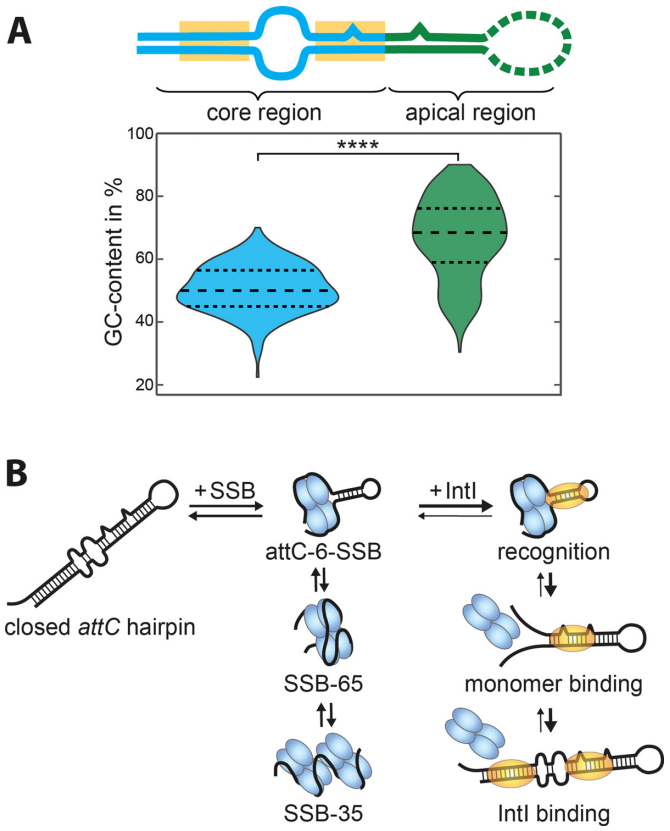

Figure 4. A conserved weighted basepair distribution allows for the dynamic SSB-hairpin equilibrium. (**A**) Analysis of the GC-content of 263 different *attC* hairpins obtained from the INTEGRALL database (12). *attC* sites were divided into a core region and apical region as depicted in cyan and green at the top. While the core region has a median GC-content (dashed line) of ≈50%, the apical region shows a median GC-content of ≈ 70%. The asterisks indicate a significant difference between the populations ($P < 10^{-4}$). (**B**) Model of how SSB opens the *attC* hairpin *via* the attC-6–SSB state recognized by IntI.

in steady growth conditions and only activated by the SOS response, an indicator for the need of adaptation (41). In steady growth conditions, SSB assists replication by destabilizing *attC* hairpins. The frequent formation of the attC-6–SSB state should not present an obstacle for the replication machinery (42). However, as soon as IntI expression is induced, it recognizes the attC-6–SSB complex and displaces SSB even at great molar excess of SSB. As long as IntI is expressed, recombination is favoured over replication and the integron is active. This is crucial to bacteria, because the integron is their major tool for adapting to environmental stresses, for example, by acquiring antibiotic resistance genes.

In summary, we present here a regulation mechanism how *attC* DNA hairpins can be both genetically stable, which requires easy opening of the hairpins, and biologically functional, which requires stable folding of the hairpins. The key to this regulation is most likely an evolutionary optimized GC-base pair distribution along the hairpin, which allows SSB to open the hairpin, while providing a recognition site for the recombining integrase IntI. In the case of *attC* hairpins, this regulation ensures antibiotic resistance acquisition and transmission in bacteria. There are many functional DNA hairpins in bacteria, which play critical roles during the recognition of the origin of replication, transcription, and recombination (4). We therefore anticipate that this elegant regulation mechanism is also used as a general approach by other hairpins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Céline Loot and all members of the Schlierf group for discussions, in particular Georg Krainer and Varsha Natarajan for assistance in the experimental design and Philip Gröger and Marko Swoboda for assistance in the data analysis. We thank Bénédicte Michel for her helpful comments on the manuscript. We gratefully acknowledge Iain Patten and the participants of the 'writing for publication' workshop for their support in writing this manuscript.

FUNDING

German Federal Ministry of Education and Research [BMBF 03Z2EN11 to M.S.]; Franco-German PROCOPE program of the German Academic exchange service [DAAD 55931827 to M.S.]; French Ministry of Foreign Affairs [PHC 28351PC to D.M.]; Dresden International Graduate School for Biomedicine and Bioengineering [DFG GS97 to M.S.G. and B.L.C.]; Institut Pasteur, Centre National de la Recherche Scientifique, Fondation pour la Recherche Médicale [FDT20150532465 to A.N.]; Doctoral School 'Frontières du Vivant' funded by the University Paris Descartes. Funding for open access charge: Institutional funds (to M.S.).

Conflict of interest statement. None declared.

REFERENCES

1. Mazel,D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, **4**, 608–620.
2. Escudero,J.A., Loot,C., Nivina,A. and Mazel,D. (2015) The integron: adaptation on demand. *Microbiol. Spectr.*, **3**, doi:10.1128/microbiolspec.MDNA3-0019-2014.
3. Bouvier,M., Demarre,G. and Mazel,D. (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J.*, **24**, 4356–4367.
4. Bikard,D., Loot,C., Baharoglu,Z. and Mazel,D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol. Mol. Biol. Rev.*, **74**, 570–588.
5. MacDonald,D., Demarre,G., Bouvier,M., Mazel,D. and Gopaul,D.N. (2006) Structural basis for broad DNA-specificity in integron recombination. *Nature*, **440**, 1157–1162.
6. Nivina,A., Escudero,J.A., Vit,C., Mazel,D. and Loot,C. (2016) Efficiency of integron cassette insertion in correct orientation is ensured by the interplay of the three unpaired features of attC recombination sites. *Nucleic Acids Res.*, **44**, 7792–7803.
7. Voineagu,I., Narayanan,V., Lobachev,K.S. and Mirkin,S.M. (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9936–9941.
8. Shereda,R.D., Kozlov,A.G., Lohman,T.M., Cox,M.M. and Keck,J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 289–318.
9. Lohman,T.M. and Ferrari,M.E. (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.*, **63**, 527–570.
10. Eggington,J.M., Kozlov,A.G., Cox,M.M. and Lohman,T.M. (2006) Polar destabilization of DNA duplexes with single-stranded overhangs by the Deinococcus radiodurans SSB protein. *Biochemistry*, **45**, 14490–14502.
11. Roy,R., Kozlov,A.G., Lohman,T.M. and Ha,T. (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature*, **461**, 1092–1097.
12. Reddy,M.S., Vaze,M.B., Madhusudan,K. and Muniyappa,K. (2000) Binding of SSB and RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA polymerization on single-stranded DNA. *Biochemistry*, **39**, 14250–14262.
13. Loot,C., Parissi,V., Escudero,J.A., Amarir-Bouhram,J., Bikard,D. and Mazel,D. (2014) The integron integrase efficiently prevents the melting effect of Escherichia coli single-stranded DNA-binding protein on folded attC sites. *J. Bacteriol.*, **196**, 762–771.
14. Villani,G., Pierre,A. and Salles,B. (1984) Quantification of SSB protein in E. coli and its variation during RECA protein induction. *Biochimie*, **66**, 471–476.
15. Reyes-Lamothe,R., Sherratt,D.J. and Leake,M.C. (2010) Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science (New York, NY)*, **328**, 498–501.
16. Roy,R., Hohng,S. and Ha,T. (2008) A practical guide to single-molecule FRET. *Nat. Methods*, **5**, 507–516.
17. Frumerie,C., Ducos-Galand,M., Gopaul,D.N. and Mazel,D. (2010) The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity. *Nucleic Acids Res.*, **38**, 559–569.
18. Gravel,A., Messier,N. and Roy,P.H. (1998) Point mutations in the integron integrase IntI1 that affect recombination and/or substrate recognition. *J. Bacteriol.*, **180**, 5437–5442.
19. Swoboda,M., Henig,J., Cheng,H.-M., Brugger,D., Haltrich,D., Plumeré,N. and Schlierf,M. (2012) Enzymatic oxygen scavenging for photostability without pH drop in single-molecule Exeperiments. *ACS Nano*, **6**, 6364–6369.
20. Cheng,H.-M., Gröger,P., Hartmann,A. and Schlierf,M. (2015) Bacterial initiators form dynamic filaments on single-stranded DNA monomer by monomer. *Nucleic Acids Res.*, **43**, 396–405.
21. Durbin,R., Eddy,S.R., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis; Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press.
22. Shuang,B., Cooper,D.R., Taylor,J.N., Kisley,L., Chen,J., Wang,W., Li,C.B., Komatsuzaki,T. and Landes,C.F. (2014) Fast step transition and state identification (STaSI) for discrete single-molecule data analysis. *J. Phys. Chem. Lett.*, **5**, 3157–3161.
23. Lawless,J.F. (2011) *Statistical Models and Methods for Lifetime Data*, 2nd edn. John Wiley and Sons. Inc.
24. Hartmann,A., Krainer,G. and Schlierf,M. (2014) Different fluorophore labeling strategies and designs affect millisecond kinetics of DNA hairpins. *Molecules*, **19**, 13735.
25. Moura,A., Soares,M., Pereira,C., Leitão,N., Henriques,I. and Correia,A. (2009) INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*, **25**, 1096–1098.
26. Demarre,G., Guérout,A.-M., Matsumoto-Mashimo,C., Rowe-Magnus,D.A., Marlière,P. and Mazel,D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate Escherichia coli host strains. *Res. Microbiol.*, **156**, 245–255.
27. Biskri,L., Bouvier,M., Guerout,A.-M., Boisnard,S. and Mazel,D. (2005) Comparative study of class 1 integron and Vibrio cholerae superintegron integrase activities. *J. Bacteriol.*, **187**, 1750.
28. Demarre,G., Frumerie,C., Gopaul,D.N. and Mazel,D. (2007) Identification of key structural determinants of the IntI1 integron integrase that influence attC× attI1 recombination efficiency. *Nucleic Acids Res.*, **35**, 6475–6489.
29. Bujalowski,W. and Lohman,T.M. (1986) Escherichia coli single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry*, **25**, 7799–7802.
30. Overman,L.B., Bujalowski,W. and Lohman,T.M. (1988) Equilibrium binding of Escherichia coli single-strand binding protein to single-stranded nucleic acids in the (SSB)65 binding mode. Cation and anion effects and polynucleotide specificity. *Biochemistry*, **27**, 456–471.
31. Raghunathan,S., Kozlov,A.G., Lohman,T.M. and Waksman,G. (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat. Struct. Biol.*, **7**, 648–652.
32. Roy,R., Kozlov,A.G., Lohman,T.M. and Ha,T. (2007) Dynamic structural rearrangements between DNA binding modes of E. coli SSB protein. *J. Mol. Biol.*, **369**, 1244–1257.
33. Suksombat,S., Khafizov,R., Kozlov,A.G., Lohman,T.M. and Chemla,Y.R. (2015) Structural dynamics of E. coli single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *Elife*, **4**, e08193.

34. Bujalowski,W. and Lohman,T.M. (1989) Negative co-operativity in Escherichia coli single strand binding protein-oligonucleotide interactions. I. Evidence and a quantitative model. *J. Mol. Biol.*, **207**, 249–268.
35. Ries,J. and Schwille,P. (2012) Fluorescence correlation spectroscopy. *Bioessays*, **34**, 361–368.
36. Khamis,M.I., Casas-Finet,J.R., Maki,A.H., Murphy,J.B. and Chase,J.W. (1987) Role of tryptophan 54 in the binding of E. coli single-stranded DNA-binding protein to single-stranded polynucleotides. *FEBS Lett.*, **211**, 155–159.
37. Casas-Finet,J.R., Khamis,M.I., Maki,A.H. and Chase,J.W. (1987) Tryptophan 54 and phenylalanine 60 are involved synergistically in the binding of E. coli SSB protein to single-stranded polynucleotides. *FEBS Lett.*, **220**, 347–352.
38. Krauss,G., Sindermann,H., Schomburg,U. and Maass,G. (1981) Escherichia coli single-strand deoxyribonucleic acid binding protein: stability, specificity, and kinetics of complexes with oligonucleotides and deoxyribonucleic acid. *Biochemistry*, **20**, 5346–5352.
39. Jose,D., Datta,K., Johnson,N.P. and von Hippel,P.H. (2009) Spectroscopic studies of position-specific DNA 'breathing' fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4231–4236.
40. Phelps,C., Lee,W., Jose,D., von Hippel,P.H. and Marcus,A.H. (2013) Single-molecule FRET and linear dichroism studies of DNA breathing and helicase binding at replication fork junctions. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 17320–17325.
41. Guerin,E., Cambray,G., Sanchez-Alberola,N., Campoy,S., Erill,I., Da Re,S., Gonzalez-Zorn,B., Barbé,J., Ploy,M.-C. and Mazel,D. (2009) The SOS response controls integron recombination. *Science (New York, NY)*, **324**, 1034.
42. Jergic,S., Horan,N.P., Elshenawy,M.M., Mason,C.E., Urathamakul,T., Ozawa,K., Robinson,A., Goudsmits,J.M.H., Wang,Y., Pan,X. *et al.* (2013) A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *EMBO J.*, **32**, 1322–1333.
43. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.