

Structural heterogeneity of *attC* integron recombination sites revealed by optical tweezers

Ann Mukhortava^{1,†}, Matthias Pöge^{1,†}, Maj Svea Grieb¹, Aleksandra Nivina^{2,3,4},
Celine Loot^{2,3}, Didier Mazel^{2,3} and Michael Schlierf^{1,*}

¹B CUBE – Center for Molecular Bioengineering, TU Dresden, Tatzberg 41, 01307 Dresden, Germany, ²Institut Pasteur, Unité de Plasticité du Génome Bactérien, Département Génomes et Génétique, 28 Rue du Dr Roux, 75015 Paris, France, ³CNRS, UMR3525, 28 Rue du Dr Roux, 75015 Paris, France and ⁴Paris Descartes University, 75006 Paris, France

Received August 22, 2018; Revised December 02, 2018; Editorial Decision December 03, 2018; Accepted December 04, 2018

ABSTRACT

A predominant tool for adaptation in Gram-negative bacteria is the functional genetic platform called integron. Integrations capture and rearrange promoterless gene cassettes in a unique recombination process involving the recognition of folded single-stranded DNA hairpins—so-called *attC* sites—with a strong preference for the *attC* bottom strand. While structural elements have been identified to promote this preference, their mechanistic action remains incomplete. Here, we used high-resolution single-molecule optical tweezers (OT) to characterize secondary structures formed by the *attC* bottom (*attC*_{bs}) and top (*attC*_{ts}) strands of the paradigmatic *attC*_{aadA7} site. We found for both sequences two structures—a straight, canonical hairpin and a kinked hairpin. Remarkably, the recombination-preferred *attC*_{bs} predominantly formed the straight hairpin, while the *attC*_{ts} preferentially adopted the kinked structure, which exposes only one complete recombinase binding box. By a mutational analysis, we identified three bases in the unpaired central spacer, which could invert the preferred conformations and increase the recombination frequency of the *attC*_{ts} *in vivo*. A bioinformatics screen revealed structural bias toward a straight, canonical hairpin conformation in the bottom strand of many antibiotic resistance cassettes *attC* sites. Thus, we anticipate that structural fine tuning could be a mechanism in many biologically active DNA hairpins.

INTRODUCTION

Bacteria possess an impressive capability of adapting to environmental stresses. They are known to develop antibiotic multi-resistance by horizontal gene transfer. The predominant tool for the acquisition and expression of resistance genes in Gram-negative pathogens is a genetic device called integron (1,2). Integrations capture and rearrange promoterless gene cassettes in a unique recombination process involving the excision and integration of single-stranded DNA. Upon SOS response, the integron tyrosine-recombinase IntI specifically recognizes folded, single-stranded DNA hairpins called *attC* sites, which flank the open reading frame in cassettes (3,4) (Figure 1A). IntI excises gene cassettes by recombination between two consecutive *attC* sites and integrates the cassettes predominantly at the *attI* site. To ensure correct expression of the protein-coding sequences from these cassettes, the orientation of the inserted DNA sequence has to be tightly controlled with respect to the cassette promoter *P_c*. Bouvier *et al.* and Nivina *et al.* have identified important structural elements of *attC* sites, which ensure the correct orientation through the strand-selectivity of IntI for the bottom strand (*attC*_{bs}) relative to the top strand (*attC*_{ts}) (5,6). Those structural elements are extra-helical bases (EHBs), an unpaired central spacer (UCS) and a variable terminal structure (VTS, Figure 1B). The role of the EHBs could be inferred from a crystallographic structure due to their prominent interactions with the recombinase IntI (7) and was further shown to be the main factor determining the strand specificity with a contribution of 20 to 300-fold in the strand recognition, depending on the site tested (5,6). The mechanisms imposed by the UCS and the VTS remain unclear. A recent study has reported that a nucleotide skew induced by UCS and VTS stabilizes the folded bottom strand,

*To whom correspondence should be addressed. Tel: +49 351 463 43050; Fax: +49 351 463 40322; Email: michael.schlierf@tu-dresden.de

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present addresses:

Aleksandra Nivina, Department of Chemistry, Stanford University, CA 94305, USA.

Matthias Pöge, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

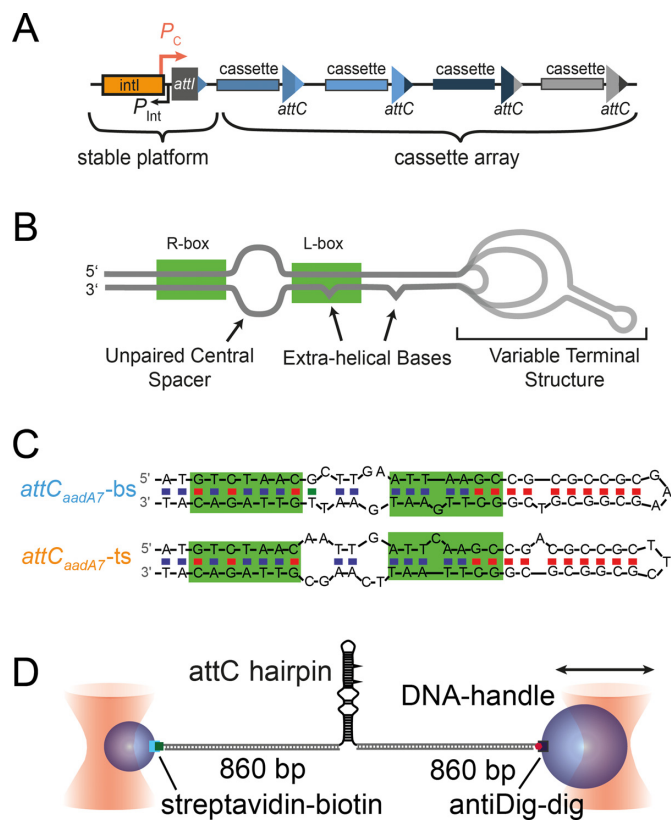


Figure 1. The integron system and schematic structure of *attC* hairpins. (A) Schematic depiction of the integron stable platform and cassette array. (B) Schematic of the structure of *attC* sites. (C) Canonical structure prediction of *attC*_{aadA7} bottom and top strand. (D) Optical tweezers assay to study DNA hairpin secondary structure by mechanical unfolding and refolding.

and by that most likely influences the recombination indirectly (6). An earlier study has suggested that *attC* bottom strands recombine less efficiently if stable parasitic non-recombinogenic secondary structures are formed (8). However, an experimental structural comparison of *attC* bottom strands and top strands has not been carried out to date, despite the apparent importance of the secondary structure of *attC* hairpins for the recombination by IntI.

Here, we used high-resolution single-molecule optical tweezers to characterize secondary structures formed by bottom strand (*attC*_{bs}) and top strand (*attC*_{ts}) of *attC*_{aadA7} site (Figure 1C, D). The force-induced unfolding revealed the existence of two structurally distinct, but energetically similar conformations for both hairpins. The predominant conformation of *attC*_{bs} exposed the IntI binding site, while the predominant conformation of *attC*_{ts} partially buried the binding site. The comparison between wild-type and mutant sequences yielded important insights into the regulation of the exposure of the integrase binding site by a few nucleotides within the *attC*_{aadA7} hairpin. This is directly linked to the efficiency of recombination and thus we speculate that this regulation mechanism is important for strand selectivity as we found a conservation in a bioinformatics analysis of 263 known *attC* sites.

MATERIALS AND METHODS

Hairpin sequences

All hairpin sequences were flanked with overhangs of 20 bases on each side complementary to the DNA handles and a 3 dT spacer. The hairpins were inserted between two double-stranded handles of ≈ 860 base pairs (bp) each (Figure 1D). Biotin- and digoxigenin-functionalized 852 bp (290 nm) and 855 bp (291 nm) DNA handles were synthesized by PCR amplification of λ -DNA. The 5' ends of the forward primers were functionalized with biotin or digoxigenin. The reverse primers introduced nicking enzyme sites to generate a 20-nucleotide ssDNA overhang complementary to the hairpin extensions. Subsequent to PCR amplification, the biotin- and the digoxigenin-handles were purified by ethanol precipitation and then digested with the nicking enzymes Nb.BsmI and Nb.BsrDI, respectively. All oligonucleotides were purchased from Sigma-Aldrich Co. To hybridize the handles with the hairpin, a 1.5:1:1 mixture of the hairpin and DNA handles, respectively, was incubated for 2 min at 90°C. Then the sample was slowly cooled down to 16°C and further ligated 12 h at 16°C with T4 DNA ligase (New England Biolabs). The final product was purified by ethanol precipitation. The full sequences of the DNA hairpins and primers for the handle construction are shown in the Supplementary Table S1A.

Single-molecule optical tweezers experiments

DNA constructs were incubated for 2 h with 2 μm silica beads (Bangs Laboratories, Inc), which were previously covalently functionalized with anti-digoxigenin F_{ab} fragments (Roche). The mixture was diluted in optical tweezers buffer (50 mM sodium phosphate buffer, 250 mM NaCl, 1% (w/v) D-glucose, pH 7.1) and mixed with streptavidin-coated 1 μm silica beads (Bangs Laboratories, Inc.). Measurements were carried out at room temperature in optical tweezers buffer after addition of an oxygen scavenger system (26 U/ml glucose oxidase, 17 000 U/ml catalase). DNA conjugates concentration was adjusted to only sparsely cover the beads leading mainly to single-tether formation. The beads were trapped in the foci of a dual beam optical tweezers platform (JPK NanoTracker) (9). Both trapped beads were brought into close proximity for tether formation. The beads were then separated with a constant velocity yielding force vs. extension traces. Trap stiffness was determined using a previously described calibration protocol (10). The effective trap stiffness was determined with an error of $\sim 10\%$ and varied from 0.056 to 0.062 pN/nm. Data were acquired at a sampling rate of 50 kHz. The signals were corrected for crosstalk.

Force-extension curves and contour length increase

The recorded force–extension curves can be well described with polymer elasticity models. Before unfolding, the DNA elasticity was modeled with the worm-like chain model

(WLC) for the double-stranded DNA handles (11,12):

$$F_{\text{WLC}_{\text{handles}}}(x) = \frac{k_{\text{B}}T}{p_{\text{dsDNA}}} \times \left(\frac{1}{4} \left(1 - \frac{x}{L_{\text{handles}}} \right)^{-2} + \frac{x}{L_{\text{handles}}} - \frac{1}{4} \right),$$

with the persistence length p_{dsDNA} and contour length L_{handles} and extension x . DNA contour length was set to 580 nm and the fit yielded persistence lengths of ~20–27 nm. After unfolding of the hairpin, a combined worm-like chain model was used, modeling the single-stranded DNA elasticity with:

$$F_{\text{WLC}_{\text{ssDNA}}}(\xi_{\text{hairpin}}) = \frac{k_{\text{B}}T}{p_{\text{ssDNA}}} \times \left(\frac{1}{4} \left(1 - \frac{\xi_{\text{hairpin}}}{L_{\text{ssDNA}}} \right)^{-2} + \frac{\xi_{\text{hairpin}}}{L_{\text{ssDNA}}} - \frac{1}{4} \right),$$

with persistence length p_{ssDNA} set to 2 nm (13,14), contour length L_{ssDNA} , and extension ξ_{hairpin} . Upon force-induced unfolding of the hairpin, a flexible ssDNA chain is added to the compliance of the dsDNA construct. Thus, a combined worm-like chain model was used. As the two mechanical parts consisting of dsDNA and unfolded hairpin (ssDNA) were in series, the extension of the full linker consisting of dsDNA and unfolded hairpin was given by:

$$\xi_{\text{construct}}(F) = x(F) + \xi_{\text{hairpin}}(F)$$

where x and ξ_{hairpin} were calculated using the previous equations. This equation was inverted to describe $F_{\text{construct}}(\xi_{\text{hairpin}}, x)$.

In vivo recombination assay

Wild-type and mutant *attC* sites were constructed by annealing of two overlapping phosphorylated oligonucleotides (Supplementary Table S1D), fully complementary except for the overhangs corresponding to EcoRI and BamHI restriction sites. The sites were then ligated into the p4116 plasmid for the delivery of the bottom strand, and into p4117 plasmid for the delivery of the top strand, both digested with EcoRI/BamHI. The plasmids were then transformed into the *Escherichia coli* β 2163 strain (15). The recombination frequency and strand selectivity were determined using a previously developed (16) suicidal conjugation assay described in detail in (6). Briefly, pSW23T vectors were transferred through conjugation from the β 2163 strain into a recipient *E. coli* DH5 α strain, so that the transferred strand carried either the bottom or the top strand of *attC* sites. The recipient strain harbored the pBAD plasmid allowing the expression of IntII integrase, and a pSU plasmid with an *attII* site. Successful *attII* \times *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 six independent experiments. For each construct, the recombination frequencies in each strand (bs or ts) was determined by 48 PCR reactions using primer pairs SWbeg/MFD and

SWend/MFD as previously described (6). The limits of detection correspond to the minimal frequencies of recombination in either strand that could be detected by this method.

Bioinformatics analysis of *attC*_{aadA7} site and *attC* sites from the INTEGRALL database

We predicted *attC*_{aadA7} site conformations using UNAFold (17) under the experimental conditions of our force spectroscopy experiments (0.25 M Na⁺ and 0 M Mg²⁺ ions at 25°C; assumption of a linear DNA molecule). The structures were generated confining the free energies within 5% difference of the minimum free energy. No constraints (like pairing the R' and R'', and L' and L'' sequences) were used. We further used the database containing sequences of 263 mobile integron *attC* sites having less than a 95% identity, obtained from the authors of the INTEGRALL database (18). The simulation conditions were the same, as described above for the *attC*_{aadA7} site. The conformations obtained for each *attC* sequence were categorized into five different groups: (i) straight-complete hairpins with R- and L-boxes fully paired, (ii) straight-incomplete hairpins with paired R-box and not fully paired L-box (e.g. bubble); (iii) kinked-complete hairpins with R- and L-boxes fully paired, (iv) kinked-incomplete hairpins with paired R-box and not fully paired L-box and (v) other structures with unpaired R- and/or L-boxes (see Supplementary Figure S8). Based on the free energy of each conformation, the probability of finding the hairpin in that conformation was calculated using a Boltzmann distribution. *attC* bottom strands formed in 46 \pm 3% (\pm S.D.) of the cases a canonical straight structure, *attC* top strands formed in 36 \pm 3% (\pm S.D.) of the cases a canonical straight structure.

RESULTS

Unfolding of *attC*_{aadA7} bottom and top strands exposes a mechanically stable intermediate

To characterize the secondary structure of the *attC*_{bs} DNA hairpin, it was tethered between two glass microspheres of 1 and 2 μ m diameter through two 860-bp-long DNA handles (Figure 1D). By separating the two silica beads from each other, force was applied to the molecular construct, causing the stretching of the DNA handles, which eventually triggered unfolding of the *attC* hairpin. Recording both the force acting on the tether and the positions of beads resulted in force–extension curves (Figure 2). Unfolding events in the force–extension curve contain information about the mechanical stability and the conformation of the DNA hairpin. Figure 2A shows a typical force–extension curve of the *attC*_{bs} hairpin with a constant pulling velocity of 200 nm/s. A sharp transition at a force of approximately 5 pN indicates a structural change, which we attributed to the unfolding of the DNA hairpin structure. The unfolding force of the *attC*_{bs} was determined to be $F_{\text{maj}}(\text{attC}_{\text{bs}}) = 5.7 \pm 0.8$ pN ($k = 215$, mean \pm S.D.). In comparison to previous reports on DNA hairpin structures (19) and our supporting measurements on an artificial perfectly-paired hairpin (Supplementary Figure S1), $F_{\text{maj}}(\text{attC}_{\text{bs}})$ is surprisingly low. A closer inspection of the force–extension curves

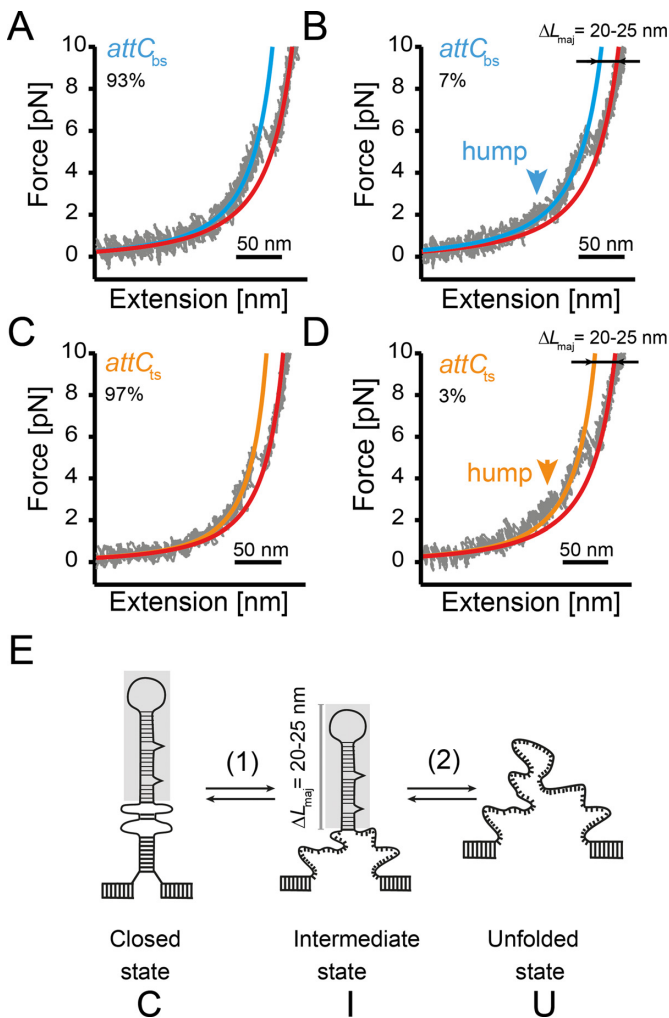


Figure 2. Typical force–extension curves. (A) *attC_{bs}* unfolded in 93% of the force–extension curves with one major unfolding event and (B) in 7% of the cases with a preceding hump-like unfolding. (C) 97% of the *attC_{ts}* force–extension curves show a similar major unfolding event and (D) 3% of the traces exhibit a preceding hump-like unfolding. (E) Two-step unfolding and refolding of the *attC* hairpin via a mechanically stable intermediate state.

of *attC_{bs}* revealed that $\approx 7\%$ of the traces also contained a second, minor, hump-like unfolding transition at an even lower force $F_{\text{min}}(\textit{attC}_{\text{bs}}) \approx 2.8 \pm 0.9$ pN ($k = 15$) (Figure 2B), which was stabilized by the addition of 20 mM MgCl_2 (Supplementary Figure S2). The force at which this minor unfolding event occurred was broadly distributed and often occurred below 1 pN, that prevented this transition to be unambiguously assigned. Our pulling velocity was chosen to be close to equilibrium of the major event, as indicated by multiple un- and refolding events within a single pulling or relaxation cycle (Supplementary Figure S3).

We next measured the contour length change of the major unfolding event (at 5.7 pN) of *attC_{bs}* which allowed us to learn about the underlying structure. The entropic elasticity of dsDNA and ssDNA can be described using a worm-like chain (WLC) polymer elasticity model. Fitting our force–extension traces with a WLC model prior and post unfolding allowed us to extract the contour length

change of unfolding the DNA hairpin $\Delta L_{\text{maj}} \approx 20-25$ nm. As the *attC_{bs}* contains 64 nucleotides (nt) with a length of 0.68 nm/nt (20) and taking into account a 2 nm end-to-end distance between the 5'- and 3'-ends of the closed DNA hairpin, we would expect a contour length change of approximately 41 nm ($64 \text{ nt} \times 0.68 \text{ nm/nt} - 2 \text{ nm} = 41.52 \text{ nm}$) upon full hairpin unfolding. We thus extended the analysis and determined the contour length change occurring during a minor, hump-like unfolding event. The minor unfolding event yielded a contour length change $\Delta L_{\text{min}} \approx 21$ nm (Figure 2B). A rigorous analysis of contour length changes during the minor transition with the WLC model remained impossible due to the instability of the fitting algorithm at forces below 2 pN. The sum of the contour length changes of the minor and the major unfolding event $\Delta L_{\text{tot}} \approx 40-44$ nm agreed well with the expected contour length increase upon unfolding of the entire *attC_{bs}* hairpin. Therefore, we concluded that the *attC_{bs}* unfolds via a mechanically stable intermediate state that corresponds to a contour length change of $\Delta L_{\text{maj}} \approx 20-25$ nm, and, thus, contains $\sim 30-36$ nts (Figure 2E).

In a second set of experiments, we performed the mechanical stretching of the *attC_{ts}* hairpin and observed very similar force–extension curves (Figure 2C, D). The majority of the traces (97%) showed a single unfolding event at $F_{\text{maj}}(\textit{attC}_{\text{ts}}) = 5.5 \pm 1.0$ pN ($k = 232$, mean \pm S.D.) with a contour length change of 20–25 nm, similar to *attC_{bs}*. Only a few curves (3%) exhibited an additional minor unfolding event at low forces of about $F_{\text{min}}(\textit{attC}_{\text{ts}}) = 2.3 \pm 0.9$ pN ($k = 8$) with a ΔL_{min} of about 20 nm. We thus concluded that the *attC_{ts}* unfolds also via a mechanically stable intermediate, like the *attC_{bs}* (Figure 2E).

attC_{bs} and *attC_{ts}* exist in two distinct conformations

Interestingly, both *attC_{bs}* and *attC_{ts}* showed a broad contour length change of 20–25 nm (or 30–36 nt) upon the major unfolding event, which is unexpected for a well-folded DNA hairpin structure (19,21). Therefore, we analyzed the contour length distribution of $n = 414$ major unfolding and refolding events of *attC_{bs}*. For the *attC_{bs}* we found a bimodal distribution of contour length changes with one population (80%) showing a contour length change $\Delta L_{\text{maj},1}(\textit{attC}_{\text{bs}}) = 24.3 \pm 1.3$ nm and a second population (20%) showing a contour length change $\Delta L_{\text{maj},2}(\textit{attC}_{\text{bs}}) = 20.4 \pm 0.9$ nm (Figure 3A). These two contour length changes were also observed within a single tether (Supplementary Figure S3) and correspond to a release of 35 ± 2 nt and 30 ± 1 nt, respectively. In order to correlate the release of nucleotides with the structure of the DNA hairpin, we performed a DNA secondary structure prediction using UNAFold (17). Calculations for the full length *attC_{bs}* yielded the existence of two conformations with nearly identical folding free energies: a straight hairpin and a kinked hairpin structure (Figure 3C).

Considering that the *attC_{bs}* hairpin unfolds in a two-step process via a mechanically stable intermediate structure starting from the 5'- and 3'-ends, we hypothesize that it is the apical stems of the straight and kinked conformations (marked as grey areas, Figure 3C) that form two different mechanically stable intermediate states. For the

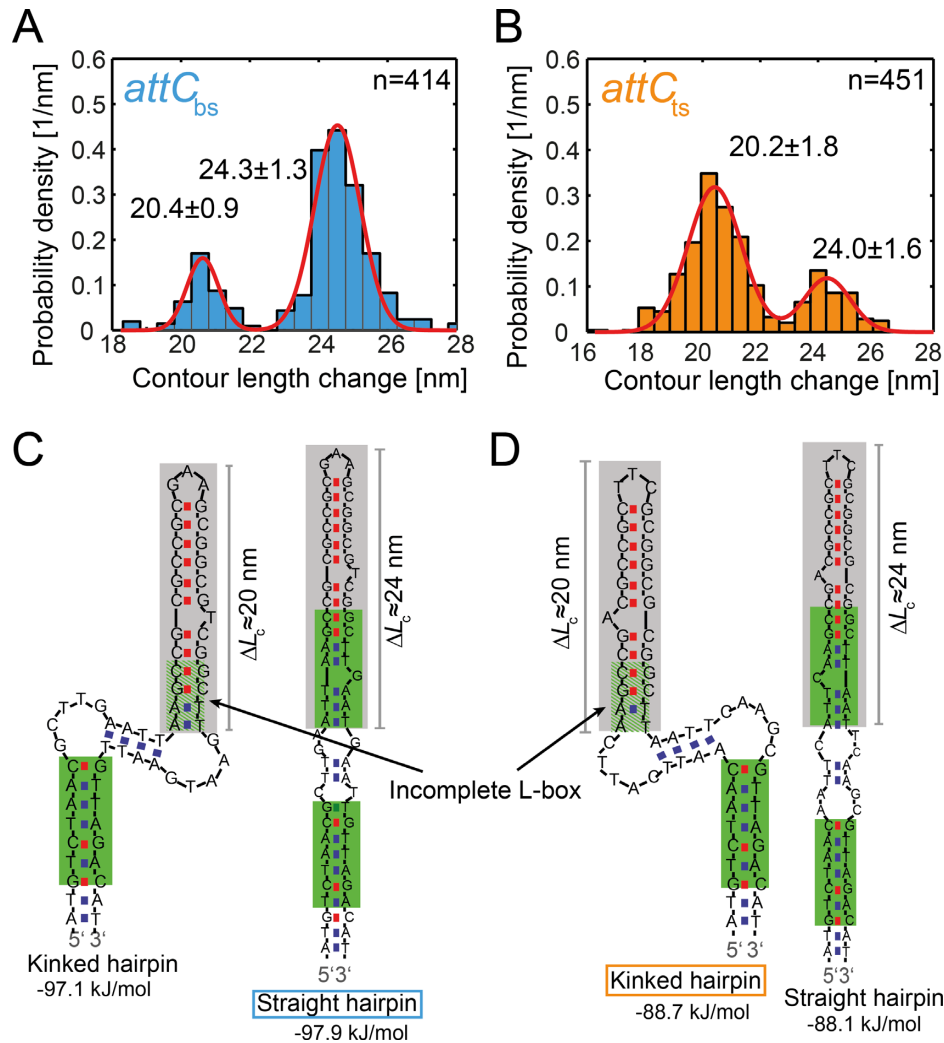


Figure 3. Contour length change of the major unfolding event. Histogram of the contour length change of the major unfolding and refolding event of the *attC_{bs}* and *attC_{ts}* in (A) and (B), respectively. UNAFold predictions (within 5% free energy difference) of the two most stable secondary structures with expected contour length changes of the upper stem for the *attC_{bs}* and *attC_{ts}* in (C) and (D), respectively. The upper stems with the expected contour length changes are indicated in grey. The thermodynamically preferred structure for the *attC_{bs}* is the straight hairpin (blue square) and for the *attC_{ts}* the kinked hairpin (orange square).

kinked hairpin conformation, this intermediate state comprises 28 nt and for the straight hairpin 35 nt, both in excellent agreement with the measured 30 ± 1 nt and 35 ± 2 nt. Thus, we attributed the straight hairpin conformation to the population at ≈ 24 nm contour length change and the kinked conformation to the ≈ 20 nm contour length change. Remarkably, the larger population of $\approx 80\%$ at $\Delta L_{\text{maj},1}(\textit{attC}_{\text{bs}}) \approx 24$ nm correlates well with the lower free energy of the straight hairpin conformation predicted by the UNAFold calculation ($\Delta G_{\text{straight}}(\textit{attC}_{\text{bs}}) = -97.9$ kJ/mol and $\Delta G_{\text{kinked}}(\textit{attC}_{\text{bs}}) = -97.1$ kJ/mol). Therefore, we concluded that the experimentally observed bimodal distribution of the *attC_{bs}* contour length increase reveals the existence of two conformational states of which the straight hairpin is predominant.

The same analysis of the distribution of the contour length change was performed for the *attC_{ts}*. A total of $n = 451$ major unfolding and refolding events of the *attC_{ts}*

revealed again a bimodal distribution of contour length changes with a population $\Delta L_{\text{maj},1}(\textit{attC}_{\text{ts}}) = 24.0 \pm 1.6$ nm (35 ± 2 nt, 24% of the events) and a second population at $\Delta L_{\text{maj},2}(\textit{attC}_{\text{ts}}) = 20.2 \pm 1.8$ nm (30 ± 1 nt, 76% of the events, see Figure 3B). A UNAFold secondary structure prediction for *attC_{ts}* exhibited two energetically close conformations—again, a kinked hairpin and a straight hairpin (Figure 3D). In accordance with the *attC_{bs}*, the upper stems comprised of 28 and 35 nt for the kinked and straight hairpins, respectively, and were assumed to represent the mechanically stable intermediate states for the *attC_{ts}* (grey highlight, Figure 3D). In contrast to the *attC_{bs}*, for the *attC_{ts}*, the occupancy of the ‘20 nm’ population, corresponding to the kinked hairpin conformation, is higher, which correlates well with the lower free energy of the kinked hairpin conformation of the *attC_{ts}* predicted by UNAFold ($\Delta G_{\text{straight}}(\textit{attC}_{\text{ts}}) = -88.1$ kJ/mol and $\Delta G_{\text{kinked}}(\textit{attC}_{\text{ts}}) = -88.7$ kJ/mol). Therefore, the predom-

inant conformation for $attC_{ts}$ hairpin upon folding is a kinked hairpin conformation. Notably, the analyses of contour length changes upon the major unfolding events during the first pulling cycle after tether formation yielded similar bimodal distributions for both $attC_{bs}$ and $attC_{ts}$ (Supplementary Figure S3). This corroborates the preceding analysis and renders the existence of force-induced structural heterogeneity unlikely. We therefore concluded that despite their complementary sequence, $attC_{bs}$ and $attC_{ts}$ preferentially fold into different structures.

The unpaired central spacer of the $attC$ strands influences the hairpin secondary structure

As the bottom and top strands, although being complementary, assume different predominant conformations, the preference for a certain shape must be encoded in the hairpin sequence. We hypothesized that the structural elements—the variable terminal structure (VTS) and the unpaired central spacer (UCS) region—influence the thermodynamic stability and by that shift the equilibrium between the kinked and the straight hairpin. In order to test our hypothesis, we constructed hybrids of $attC_{bs}$ and $attC_{ts}$. In a first set of experiments, we exchanged the VTS sequence of the $attC_{bs}$ with the VTS sequence of the $attC_{ts}$ and vice versa to create $attC_{bs}^{VTS-ts}$ and $attC_{ts}^{VTS-bs}$, respectively (Supplementary Table S1B). Single-molecule force spectroscopy experiments on the hybrids revealed a similar two-step unfolding with an unfolding force of the major unfolding event of 6.5 ± 1.0 pN ($k = 249$, mean \pm S.D.) for the $attC_{bs}^{VTS-ts}$ and 5.8 ± 0.8 pN ($k = 241$, mean \pm S.D.) for the $attC_{ts}^{VTS-bs}$. Again, most of the traces showed only one, major unfolding event and only a few traces both the minor and the major unfolding events. For both $attC_{bs}^{VTS-ts}$ and $attC_{ts}^{VTS-bs}$ constructs, we observed a bimodal contour length change distribution, with the nearly-identical ratio between populations to the one observed for the wild-type $attC_{bs}$ and $attC_{ts}$, respectively (Figure 4A, B). UNAFold predicted for the hybrids $attC_{bs}^{VTS-ts}$ and $attC_{ts}^{VTS-bs}$ straight and kinked hairpins as possible conformations, with the straight hairpin being thermodynamically preferred for $attC_{bs}^{VTS-ts}$ and the kinked hairpin—for $attC_{ts}^{VTS-bs}$ (Supplementary Figure S5). UNAFold predictions were in excellent agreement with the results of our structural analysis (Figure 4A, B). It is important to note, that both the ΔL_c distributions and the predominant conformations for VTS-modified mutants were analogous to the corresponding ones for wild-type strands. Thus, we concluded that the VTS region does not determine the preferred conformation of the hairpin.

As the next step, we constructed $attC_{bs}^{UCS-ts}$ and $attC_{ts}^{UCS-bs}$ hybrids by swapping 3 UCS bases between $attC_{bs}$ and $attC_{ts}$ (Supplementary Table S1B). Single-molecule optical tweezers experiments revealed the unfolding forces of 5.6 ± 0.8 pN ($k = 229$, mean \pm S.D.) for $attC_{bs}^{UCS-ts}$ and 6.2 ± 0.8 pN ($k = 212$, mean \pm S.D.) for the $attC_{ts}^{UCS-bs}$. Again, we observed two step unfolding patterns for both hybrids, which were similar to the wild-type $attC_{bs}$ and $attC_{ts}$. Both hybrids, $attC_{bs}^{UCS-ts}$ and $attC_{ts}^{UCS-bs}$, showed a bimodal distribution of contour

length changes of their major unfolding event (Figure 4C, D). Surprisingly, the populations of the distributions were inverted in comparison to the wild-type strands (see Figure 3A, B). We correlated this change with structural predictions by UNAFold and found that for the $attC_{ts}^{UCS-bs}$ the energetically preferred conformation is a straight hairpin (in contrast to the kinked hairpin for $attC_{ts}$, Supplementary Figure S5). The UNAFold calculations for $attC_{bs}^{UCS-ts}$ revealed three energetically-similar structural predictions—a kinked and a straight hairpin configuration, and a hairpin with less base pairing in the UCS region ('big bubble', Supplementary Figure S6). Both upper stems of the 'kinked' and 'big-bubble' conformations contain 28 nt which corresponds to the contour length change of ≈ 20 nm. Thus, the enlarged population of $attC_{bs}^{UCS-ts}$ at ≈ 20 nm resembles a mixture of these two conformations (Figure 4C) with the prevailing 'big-bubble' sub-population over the 'kinked' according to the ratio of free energies (Supplementary Information). It is important to note, that the conformation of the hairpin with the big bubble is structurally closely related to the straight configuration. Indeed, e.g. two sides of the bubble could collapse on themselves turning a big-bubble hairpin into a straight one—a process which could occur during IntI binding (Supplementary Figure S7). The correlation of the mechanical structural analysis and the UNAFold calculations for UCS-hybrids sequences suggests that the UCS is an important regulator of the conformation of the $attC$ DNA hairpin, and thus might be involved in the strand selectivity upon recombination.

In vivo recombination is affected by a changed UCS sequence

Taking together the selectivity of IntI for the $attC_{bs}$ during recombination (5–7) and the existence of different predominant conformations for the $attC_{bs}$ and $attC_{ts}$, we hypothesized that the strand selectivity of IntI is to a partially conditioned by the structural heterogeneity of the $attC$ hairpin, e.g. the existence of kinked and straight conformations, which hinder or allow IntI binding, respectively. To test this hypothesis *in vivo*, we used a well-established assay (22) to characterize the recombination efficiency of wild-type $attC_{bs}$ and $attC_{ts}$, and the hybrid $attC$ hairpins $attC_{bs}^{VTS-ts}$, $attC_{ts}^{VTS-bs}$, $attC_{bs}^{UCS-ts}$ and $attC_{ts}^{UCS-bs}$ (Figure 5).

As expected, for the $attC_{aadA7}$ site the delivery of the bottom strand lead to recombination at high efficiency (2.7×10^{-1} , Figure 5, left panel, light green) and the recombination took place exclusively in the delivered bottom strand. The delivery of the top strand mostly led to recombination of $attC_{ts}$ (Figure 5, right panel, deep green), even though we could also observe recombination of $attC_{bs}$ (Figure 5, right panel, light green), which occurred after re-synthesis of the corresponding strand. However, the recombination frequency obtained in the case of top strand delivery was 100-fold lower than the one obtained upon bottom strand delivery (2.0×10^{-3} for $attC_{ts}$ versus 2.7×10^{-1} for $attC_{bs}$, Figure 5, ts and bs delivered), which originates from the lower affinity of IntI towards the $attC_{ts}$ hairpin (5,6). The inversion of the VTS between $attC_{bs}$ and $attC_{ts}$

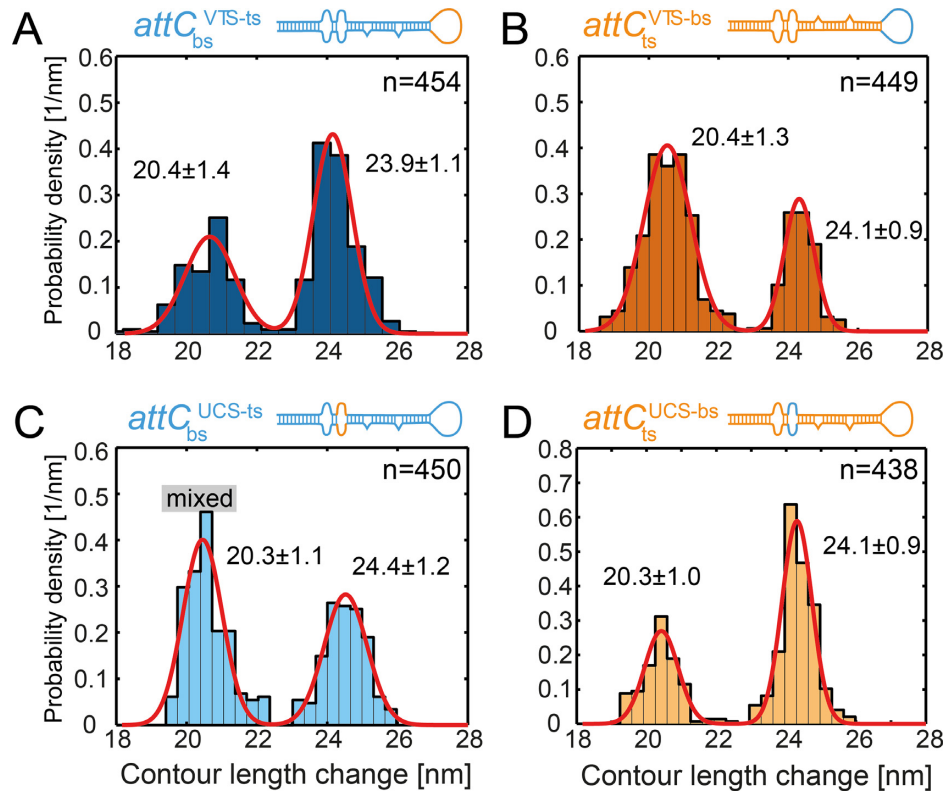


Figure 4. Contour length change of the major unfolding event of the *attC* mutants. Histograms of the contour length change of the major unfolding and refolding event of (A) *attC*_{bs}^{VTS-ts}, (B) *attC*_{ts}^{VTS-bs}, (C) *attC*_{bs}^{UCS-ts} and (D) *attC*_{ts}^{UCS-bs}. Mutated regions in the hairpins are illustrated on top in color code (blue originally bottom strand, orange originally top strand). Predicted UNAFold structures are shown in Supplementary Figure S4.

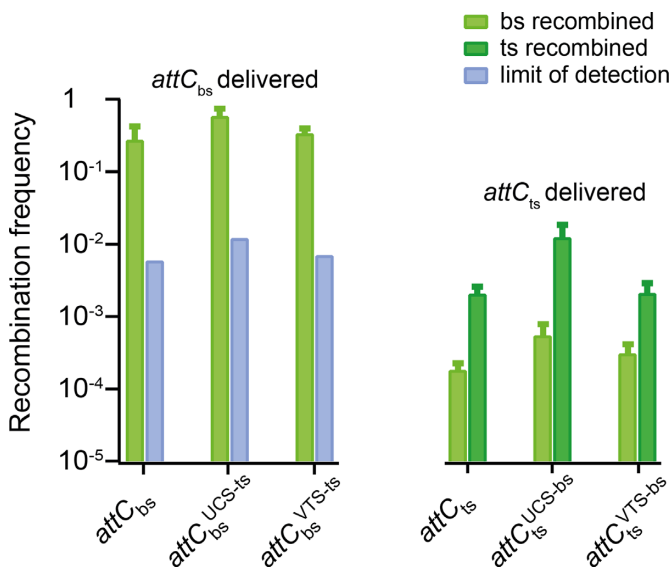


Figure 5. Recombination frequencies of wild-type and mutant *attC* sites *in vivo*. *attC*_{bs} or *attC*_{ts} was delivered. Light green indicates bottom strand recombination, dark green indicates top strand recombination. Blue indicates the limit of detection (see Supplementary information). Values represent the mean of six independent experiments and error bars correspond to mean deviations.

did not lead to any changes in recombination frequency and strand selectivity of IntI (Figure 5).

The mutation in the UCS region led to a 6-fold increase in recombination efficiency of *attC*_{ts}^{UCS-bs} (1.21×10^{-2}) in comparison to *attC*_{ts} (2.03×10^{-3}), which supports our hypothesis that a straight hairpin conformation assists the efficient recombination (Figure 5). The introduction of the top strand UCS in the *attC*_{bs} (*attC*_{bs}^{UCS-ts} mutant) did not notably affect the high recombination efficiency of the bottom strand (Figure 5) despite the inverted populations observed during optical-tweezers structural analysis. However, in this case we did not expect a significant drop in recombination efficiency, since the EHBs are major determinants of strand selectivity in *attC* sites, and their correct positioning in *attC*_{bs} has a predominant effect on IntI binding and recombination (Supplementary Figure S7) (5,6). Despite the fact that according to UNAFold calculations the preferred structure of *attC*_{bs}^{UCS-ts} is the bubble-like hairpin that corresponds to a population with a short contour length increase upon unfolding, this structure is most likely recombinogenic due to its simple conversion to the straight hairpin upon IntI binding (Supplementary Figure S5).

Taken together, we showed that the exchange of three bases in UCS of one strand to the corresponding UCS bases of the opposite strand perturbs the conformational heterogeneity, favouring opposite conformations of the hairpin. By that, it may alter the *in vivo* recombination efficiency, as one (kinked) conformation exposes only one IntI binding box. The *in vivo* results support the hypothesis that the binding of IntI is influenced by the conformational hetero-

geneity of the *attC* hairpin. We conclude, that this intriguing mechanism of structural fine tuning is determined by the unpaired *attC* secondary structure elements—mainly by the UCS.

DISCUSSION

Low mechanical stability of imperfect DNA hairpin structures

DNA secondary structures such as hairpins are formed in the context of many biological processes (23). While they appear at first glance as simple structures consisting of an inverted repeat sequence, numerous structural features of natural hairpins remain elusive. In this study, we analyzed the DNA hairpins formed by the *attC_{aadA7}* site, which is involved in DNA recombination of the bacterial integron system. According to previous single-molecule mechanical studies, simple stem-loop structures typically unfold and re-fold as a two-state system (19,21,24) along a single pathway (25,26) and exhibit mechanical stability on the order of 10–13 pN (19,27). In contrast, *attC_{bs}* and *attC_{ts}* hairpins unfold at lower forces (<6 pN) via two consecutive events and show evidence of complex folding dynamics involving on-pathway intermediates. The partial unfolding of the hairpins into intermediate states (step 1, Figure 2E) occurred mostly at forces lower than the detection limit, similar to a recently reported case for the HIV-1 RNA hairpin with an unstable lower stem (28,29) or DNA hairpins with single base damages like 8-oxoguanine (30). The major unfolding event for *attC* hairpins was typically at forces around 5.5 pN. This drop in unfolding force from 10 pN (typical for a simple stem-loop structure) most likely originates from the imperfections in the *attC* hairpin—namely, extra-helical bases and the unpaired central spacer, which prevent perfect base stacking and, by extension, high mechanical resistance. Such a low unfolding force may have functional implications like allowing an easy opening of DNA hairpins *in vivo* by polymerases or SSB (31,32) to minimize negative effects of unintended hairpin formation, e.g. during RNA transcription or DNA replication.

attC_{bs} and *attC_{ts}* show distinct conformations

Analysing the contour length changes in detail allowed us to find a heterogeneous ensemble of two conformations for both *attC_{bs}* and *attC_{ts}*. We correlated the measured contour length changes with secondary structure predictions by UNAFold to identify a kinked hairpin structure in addition to the canonical straight DNA hairpin. While secondary structure heterogeneity for nucleic acids is long known, here using our optical tweezers assay we directly observed two competing structures. Importantly, both structures are energetically very similar and differ less than 1 kJ/mol ($\approx 0.4 k_B T$) in their predicted stability, which allowed direct observation of a population of both structural states. We further revealed that, the *attC_{bs}* predominantly adopts a straight-hairpin structure, while the *attC_{ts}* favours a kinked hairpin in about 75% of the analysed molecules. In Table 1 we compare the experimental ratio of straight canonical

vs kinked structures. Notably, a discrepancy in the absolute numbers exist between the experimentally observed populations and those expected according to the Boltzmann distribution based on the folding free energies. This could indicate that in our experimental test, the refolded structures are biased by the remaining folding force. We therefore analysed also the first unfolding events of DNA tethers and observed a similar ratio as in our cumulative distribution. Thus, we assume that the refolding under external force does not skew the structural distributions. This might indicate that the calculations of the folding free energies are still lacking some unknown contributions. It is important to note, that already minor contributions of 1 kJ/mol or less affect the populations.

Based on the correlation of the recombination efficiency with a preferred secondary structure, we speculate that due to the absence of one complete IntI binding box, it is likely that the kinked conformation would lead to a significantly reduced IntI binding and, by extension, contribute to strand selectivity. Indeed, an earlier study already reported a correlation between a calculated probability to form a canonical straight hairpin structure and the measured recombination efficiency (8). Here, we observed directly the duality of a straight canonical structure and a kinked structure for the same *attC* site. We speculate that not only recognition of the kinked hairpin structure by IntI could be impaired, but also the recombination due to a structural offset, which would disrupt the interaction site between both integrase monomers or generate a steric hindrance during synaptic complex formation during *attC* × *attC* or *attC* × *attI* recombination.

The conformational switch can be induced by three bases

Previous studies reported that a switch of the entire UCS of the *attC_{bs}* to the *attC_{ts}* can increase the recombination efficiency of the *attC_{ts}* (5,6). Here, we found that changes in the UCS affect the structure adopted by the *attC* site significantly. The exchange of only three nucleotides in the UCS region inverts the populations of kinked and straight conformations (Figure 4C, D). Indeed, *attC_{ts}^{UCS-bs}* adopted a canonical straight hairpin and showed a 6-fold increase in recombination efficiency. This effect of three exchanged nucleotides can be explained by analysing the free energy of the loop formed by these nucleotides based on a previous model (33). This model considers an energetic penalty of a mismatched base pair next to a Watson-Crick base pair, i.e. loss of base stacking interactions. The purine-containing mismatches show in calculations a lower energetic penalty (*attC_{bs}* and *attC_{ts}^{UCS-bs}*) than the pyrimidine-containing mismatches (*attC_{ts}* and *attC_{bs}^{UCS-ts}*). Thus, the purine-containing UCS region (A, G) is more stabilized than the pyrimidine one (C, T). Noteworthy, changes in the VTS did neither affect the structural ensemble compared to the wild-type nor the recombination efficiency. However, the VTS of the *attC_{aadA7}* site contains only a very small 3 nucleotide-long loop and most likely exerts only little influence on the structure formation process. Longer loops or even extended complicated VTS sequences—as present in

Table 1. Summary of the experimentally determined contour length changes and populations as well as the predictions (grey shading) based on UNAFold calculations

	# of unfolding-refolding events $n/\#$ of tethers m	Measured L_{maj} , nm (nt) occurrence/occurrence in first unfolding		Predicted L_{maj} , nm (nt) Predicted occurrence	
		Kinked hairpin conformation	Straight hairpin conformation	Kinked hairpin conformation	Straight hairpin conformation
$attC_{bs}$	$n = 414/m = 65$	20.4±0.9 nm (~30 nt) 20%/15%	24.3±1.3 nm (~35 nt) 80%/85%	19 nm (28 nt) 42%	23.8 nm (35 nt) 58%
$attC_{ts}$	$n = 451/m = 65$	20.2±1.8 nm (~30 nt) 76%/83%	24.0±1.6 nm (~35 nt) 24%/17%	19 nm (28 nt) 56%	23.8 nm (35 nt) 44%
$attC_{bs}^{VTS-ts}$	$n = 454/m = 41$	20.4±1.4 nm (~30 nt) 38%/27%	23.9±1.1 nm (~35 nt) 62%/73%	19 nm (28 nt) 42%	23.8 nm (35 nt) 58%
$attC_{ts}^{VTS-bs}$	$n = 449/m = 29$	20.4±1.3 nm (~30 nt) 69%/59%	24.1±0.9 nm (~35 nt) 31%/41%	19 nm (28 nt) 56%	23.8 nm (35 nt) 44%
$attC_{bs}^{UCS-ts}$	$n = 450/m = 29$	20.3±1.1 nm (~30 nt) 55%/41%	24.4±1.2 nm (~35 nt) 45%/59%	19 nm (28 nt) 14%	23.8 nm (35 nt) 33% bubble: 19nm (28nt) 53%
$attC_{ts}^{UCS-bs}$	$n = 438/m = 30$	20.3±1.0 nm (~30 nt) 33%/10%	24.1±0.9 nm (~35 nt) 67%/90%	19 nm (28 nt) 33%	23.8 nm (35 nt) 67%

e.g. the VCR_{2/1} site—might show a stronger contribution to a multidimensional conformational space (8).

Straight and kinked conformations in *attC* sites

Is the observed conformational preference to straight and kinked structures a common structural motif in *attC* sites? To answer this question, we performed a bioinformatics analysis of 263 *attC* sites provided by the authors of the INTEGRALL database (18). We predicted the structures of these *attC* sites using UNAFold and often found multiple structural models within the interval of 5% folding energy difference. We classified these models in kinked or straight conformations upon IntI binding site pairing (see Materials and Methods). Using a Boltzmann equation, we calculated the probabilities of conformational occupancy and found that the *attC* top strands are 30% less likely to form a canonical straight conformation than *attC* bottom strands.

In summary, using a single-molecule optical tweezers assay we found a heterogeneous ensemble of *attC* DNA hairpin conformations. We resolved a conformational bias of the *attC_{aadA7}* bottom strand for the canonical straight stem-loop conformation, while the *attC_{aadA7}* top strand was found to be structurally biased toward a kinked DNA structure. Key for the bias of the *attC_{bs}* to the straight conformation is a fine-tuned sequence element, the unpaired central spacer. Upon a sequence inversion of only three nucleotides, the population of conformations was inverted and the recombination efficiency of the *attC_{ts}* was increased 6-fold *in vivo*. This may originate from a steric hindrance for the integrase binding and synaptic complex formation during recombination imposed by the kinked conformation of the DNA hairpin. We anticipate that this fine structural regulation of conformations is also used in other DNA hairpin systems involved in DNA replication, transcription and recombination to ensure not only strand selectivity, but also stability and functionality.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all members of the Schlierf and Mazel group for discussions.

Author contributions: M.S., M.S.G., A.N. and A.M. conceived the experiment, A.M., M.P., M.S.G. and M.S. designed the experiments. A.M., M.P., M.S.G., A.N. and C.L. performed the experiments, A.M. wrote the analysis software and M.P., A.M., A.N., D.M. and M.S. analysed the data. A.M. and M.S. wrote the manuscript. All authors discussed the results and commented on the 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 N° 28351PC to D.M.]; Institut Pasteur, the Centre National de la Recherche Scientifique; the 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: SLUB/TU Dresden.

Conflict of interest statement. None declared.

REFERENCES

- Mazel, D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, **4**, 608–620.
- 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.
- Francia, M.V., Zabala, J.C., de la Cruz, F. and Lobo, J.M.G. (1999) The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J. Bacteriol.*, **181**, 6844–6849.
- Johansson, C., Kamali-Moghaddam, M. and Sundström, L. (2004) Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.*, **32**, 4033–4043.
- Bouvier, M., Ducos-Galand, M., Loot, C., Bikard, D. and Mazel, D. (2009) Structural features of Single-Stranded integron cassette *attC* sites and their role in strand selection. *PLoS Genet.*, **5**, e1000632.
- 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. 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.
 8. Loot, C., Bikard, D., Rachlin, A. and Mazel, D. (2010) Cellular pathways controlling integron cassette site folding. *EMBO J.*, **29**, 2623–2634.
 9. Mukhortava, A. and Schlierf, M. (2016) Efficient formation of Site-Specific Protein-DNA hybrids using Copper-Free click chemistry. *Bioconjug. Chem.*, **27**, 1559–1563.
 10. Tolić-Nørrelykke, S.F., Schäffer, E., Howard, J., Pavone, F.S., Jülicher, F. and Flyvbjerg, H. (2006) Calibration of optical tweezers with positional detection in the back focal plane. *Rev. Sci. Instrum.*, **77**, 103101.
 11. Ceconi, C., Shank, E.A., Bustamante, C. and Marqusee, S. (2005) Direct observation of the three-state folding of a single protein molecule. *Science*, **309**, 2057–2060.
 12. Bustamante, C., Marko, J.F., Siggia, E.D. and Smith, S. (1994) Entropic elasticity of lambda-phage DNA. *Science*, **265**, 1599–1600.
 13. Chi, Q., Wang, G. and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Phys. A Stat. Mech. Appl.*, **392**, 1072–1079.
 14. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing Single-Stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, **86**, 2530–2537.
 15. 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 (IncP α) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.*, **156**, 245–255.
 16. 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.
 17. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.
 18. Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., Correia, A., Leitao, N., Henriques, I., Correia, A., Leitão, N. *et al.* (2009) INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*, **25**, 1096–1098.
 19. Woodside, M.T., Behnke-Parks, W.M., Larizadeh, K., Travers, K., Herschlag, D. and Block, S.M. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 6190–6195.
 20. Bosco, A., Camunas-Soler, J. and Ritort, F. (2014) Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.*, **42**, 2064–2074.
 21. Woodside, M.T., Anthony, P.C., Behnke-Parks, W.M., Larizadeh, K., Herschlag, D. and Block, S.M. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, **314**, 1001–1004.
 22. Biskri, L., Bouvier, M., Guérout, A.-M., Boisnard, S. and Mazel, D. (2005) Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J. Bacteriol.*, **187**, 1750.
 23. 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.
 24. Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I. and Bustamante, C. (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science*, **292**, 733–737.
 25. Neupane, K., Foster, D.A.N., Dee, D.R., Yu, H., Wang, F. and Woodside, M.T. (2016) Direct observation of transition paths during the folding of proteins and nucleic acids. *Science*, **352**, 239–242.
 26. Neupane, K., Manuel, A.P., Lambert, J. and Woodside, M.T. (2015) Transition-Path probability as a test of Reaction-Coordinate quality reveals DNA hairpin folding is a One-Dimensional diffusive process. *J. Phys. Chem. Lett.*, **6**, 1005–1010.
 27. Alemany, A. and Ritort, F. (2014) Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins. *Biopolymers*, **101**, 1193–1199.
 28. Mazauric, M.-H., Seol, Y., Yoshizawa, S., Visscher, K. and Fourmy, D. (2009) Interaction of the HIV-1 frameshift signal with the ribosome. *Nucleic Acids Res.*, **37**, 7654–7664.
 29. Ritchie, D.B., Cappellano, T.R., Tittle, C., Rezajooei, N., Rouleau, L., Sikkema, W.K.A. and Woodside, M.T. (2017) Conformational dynamics of the frameshift stimulatory structure in HIV-1. *RNA*, **23**, 1376–1384.
 30. McCauley, M.J., Furman, L., Dietrich, C.A., Rouzina, I., Núñez, M.E. and Williams, M.C. (2018) Quantifying the stability of oxidatively damaged DNA by single-molecule DNA stretching. *Nucleic Acids Res.*, **46**, 4033–4043.
 31. 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.
 32. Grieb, M.S., Nivina, A., Cheeseman, B.L., Hartmann, A., Mazel, D. and Schlierf, M. (2017) Dynamic stepwise opening of integron *attC* DNA hairpins by SSB prevents toxicity and ensures functionality. *Nucleic Acids Res.*, **45**, 10555–10563.
 33. SantaLucia, J. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.