5082–5097 *Nucleic Acids Research, 2011, Vol. 39, No. 12* *Published online 4 March 2011*
doi:10.1093/nar/gkr115

Segrosome assembly at the pliable *parH* centromere

Meiyi Wu[1], Massimiliano Zampini[1], Malte Bussiek[2], Christian Hoischen[3], Stephan Diekmann[3] and Finbarr Hayes[1,*]

[1]Faculty of Life Sciences and Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK, [2]Department of Genetics, University of Kassel, Heinrich-Plett-Str. 40, D-34132 Kassel and [3]Department of Molecular Biology, Leibniz Institute for Age Research, Fritz-Lipmann-Institute, 07745 Jena, Germany

Received December 16, 2010; Revised January 28, 2011; Accepted February 14, 2011

ABSTRACT

The segrosome of multiresistance plasmid TP228 comprises ParF, which is a member of the ParA ATPase superfamily, and the ParG ribbon–helix–helix factor that assemble jointly on the *parH* centromere. Here we demonstrate that the distinctive *parH* site (~100-bp) consists of an array of degenerate tetramer boxes interspersed by AT-rich spacers. Although numerous consecutive AT-steps are suggestive of inherent curvature, *parH* lacks an intrinsic bend. Sequential deletion of *parH* tetramers progressively reduced centromere function. Nevertheless, the variant subsites could be rearranged in different geometries that accommodated centromere activity effectively revealing that the site is highly elastic *in vivo*. ParG cooperatively coated *parH*: proper centromere binding necessitated the protein's N-terminal flexible tails which modulate the centromere binding affinity of ParG. Interaction of the ParG ribbon–helix–helix domain with major groove bases in the tetramer boxes likely provides direct readout of the centromere. In contrast, the AT-rich spacers may be implicated in indirect readout that mediates cooperativity between ParG dimers assembled on adjacent boxes. ParF alone does not bind *parH* but instead loads into the segrosome interactively with ParG, thereby subtly altering centromere conformation. Assembly of ParF into the complex requires the N-terminal flexible tails in ParG that are contacted by ParF.

INTRODUCTION

The transmission of genetic information from generation-to-generation is a fundamental biological process that must take place with high fidelity. The molecular events that underpin accurate genome segregation in eucaryotes are comparatively well-described (1). In contrast, understanding of the mechanism of procaryotic DNA segregation is more rudimentary. However, the compact genetic modules that mediate the precise partitioning of plasmids are highly informative systems in which to unravel this process in precise detail (2).

Four distinct classes of plasmid segregation cassette have been defined (3). The two most well-studied types each comprise a pair of autoregulated genes and a nearby centromere analogue. The first gene specifies an ATPase that either possesses Walker box motifs (ParA) or is an actin homologue (ParM), whereas the accompanying gene encodes a centromere binding factor (CBF) (4–8). The CBF is a site-specific DNA binding protein that loads onto the centromere to produce a nucleoprotein complex of defined geometry (9–12). The ATPase does not directly contact the centromere, but instead interacts with the CBF to assemble the mature segrosome. In the case of ParM, ATP-induced filamentation from the segrosome propels each member of a plasmid pair in opposite directions to achieve segregation (13). ParA homologues also polymerize in response to ATP binding, a process that is influenced by the CBF and/or by DNA (14–21). The ParA filaments emanating from the segrosome may drive plasmids towards the cell poles, or retraction of the ParA polymers may draw plasmids in opposite directions away from the cytokinetic zone (4,15). Recent *in vivo* studies favour the latter (22,23).

*To whom correspondence should be addressed. Tel: +44 161 3068934; Fax: +44 161 3065201; Email: finbarr.hayes@manchester.ac.uk

The tri-partite segrosome of multiresistance plasmid TP228 comprises the ParA homologue, ParF, and the ParG CBF which assemble on the *parH* centromere (Figure 2). ATP binding promotes the polymerization of ParF into dynamic, extensive multistranded filaments that are implicated in segregation (15). The dimeric ParG protein comprises C-terminal regions that interlock into a ribbon–helix–helix (RHH) fold linked to a pair of flexible N-terminal extensions (24). The folded region harbours the major determinants for dimerization, for binding to the *parH* centromere and to the O_F operator site, as well as for ParF interaction (18,24–26). The ParG mobile tails are also multifunctional. First, arginine fingers stabilize the transition state during nucleotide hydrolysis by their partner proteins. The ParG N-terminal tail includes an arginine finger-like motif that stimulates ATP hydrolysis by ParF (18). This stimulation may be a crucial aspect of the cycle of ParF polymerization and depolymerization during segregation. Second, ParF polymerization is stimulated by the ParG flexible tail (18). The tails either may reorganize or stabilize ParF filaments by tethering ParF monomers within a single protofilament or aligned protofilaments. Alternatively, ParG might cluster at points of polymer growth or disassembly (15,18). In this sense, ParG may play a role similar to formins and related factors that influence the elongation and disassembly of actin filaments in eucaryotes, or may be analogous to microtubule-associated proteins that modulate tubulin dynamics (19). Third, ParG binds to the O_F operator during transcriptional repression of the *parFG* genes (25). O_F comprises eight degenerate 5′-ACTC-3′ boxes arranged in a combination of direct and inverted orientation (26). Each tetramer motif recruits one ParG dimer, implying that the fully bound operator is cooperatively coated by up to eight dimers. The O_F operator apparently has evolved with subsites that bind ParG dissimilarly to produce a nucleoprotein complex fine-tuned for optimal interaction with the transcription machinery (26). A transient β-strand element in the ParG mobile tail associates with the protein's folded RHH domain thereby further modulating the binding of the protein to the operator (25). The mechanism by which this interaction between flexible and folded domains affects DNA binding is elusive.

CBFs have heterogeneous primary sequences that correlate with the diversity in plasmid centromere organization (27). The precise loading of each CBF onto its cognate centromere is a vital early step that is crucial for correct segrosome assembly and the subsequent cascade of events during partitioning. Here, the interaction of ParG with the distinctive *parH* centromere is dissected: *parH* is a complex multisubsite locus that nevertheless can accommodate a variety of synthetic subsite re-arrangements for accurate segregation. Both direct and indirect readout of *parH* potentially are required for correct coating of the centromere by ParG emphasizing that an intricate set of interactions mediate the loading of the protein onto the site. The centromere binding specificity of ParG is enhanced by the protein's flexible N-terminal tails which also are necessary for recruitment of ParF to the mature segrosome.

MATERIALS AND METHODS

Strains, plasmids and molecular biology procedures

Plasmids were propagated and analysed using *Escherichia coli* DH5α (28). Strain BL21 (Novagen) was employed for protein overproduction and plasmid partition assays were performed in the *polA* strain BR825 (29). Recombinant plasmids for overexpression of the *parF* and *parG* genes were described previously (30). ParG derivatives with 9, 19 or 30 amino acid deletions of the N-terminal tail were produced from plasmids constructed elsewhere (25). The partition probe vector pFH450 is a derivative of the bi-replicon plasmid pALA136 (31,32). Plasmid pFH547 comprises the *parFGH* region cloned in pFH450 (33). Plasmid pMW20 was constructed in two steps. First, a promoter-less *parFG* cassette was amplified from pFH547, digested with SacI–XbaI, and inserted between the same sites in the arabinose-inducible expression vector pBAD30 (34) to generate plasmid pMW19. The arabinose-inducible *parFG* cassette then was amplified from pMW19, digested with XhoI, and inserted in the same site in pFH450 to produce pMW20. Derivatives of *parH* possessing a full complement of 5′-ACTC-3′ boxes, but with one or more rearrangements (Figure 1), were constructed by inserting double-stranded oligonucleotides carrying the appropriate sequences and with EcoRV–NsiI compatible ends between the same sites in pMW20. Derivatives of *parH* bearing deletions of 5′-ACTC-3′ boxes were constructed first by amplifying the appropriate regions from pFH547, cleaving the PCR products with BamHI–XhoI, and inserting between the same sites in pFH450. The arabinose-inducible *parFG* cassette from pMW19 was then inserted as an XhoI fragment in the same orientation in each case. The nucleotide sequences of the inserts in all plasmid constructs were verified. DNA cloning and other molecular biology procedures followed standard protocols.

Plasmid segregation assays

Segregation assays were performed using pFH450 or pMW20 derivatives that replicate at low copy number in strain BR825 as detailed elsewhere (33). Briefly, the relevant plasmid-bearing strains were grown for ~25 generations without chloramphenicol selective pressure. Plasmid retention was then determined by replica plating colonies to agar medium with and without the antibiotic. The values presented are the means of at least three independent tests with typical standard deviations (SDs) of ~10%.

Protein production and purification

The hexahistidine-tagged ParF and ParG proteins were overproduced and purified by Ni^{2+} affinity chromatography as described previously (30).

Gel retardation assays

DNA fragments for gel retardation assays were PCR products amplified from appropriate plasmid templates using one primer bearing a 5′ biotin label and a second unlabelled primer, or were generated by annealing

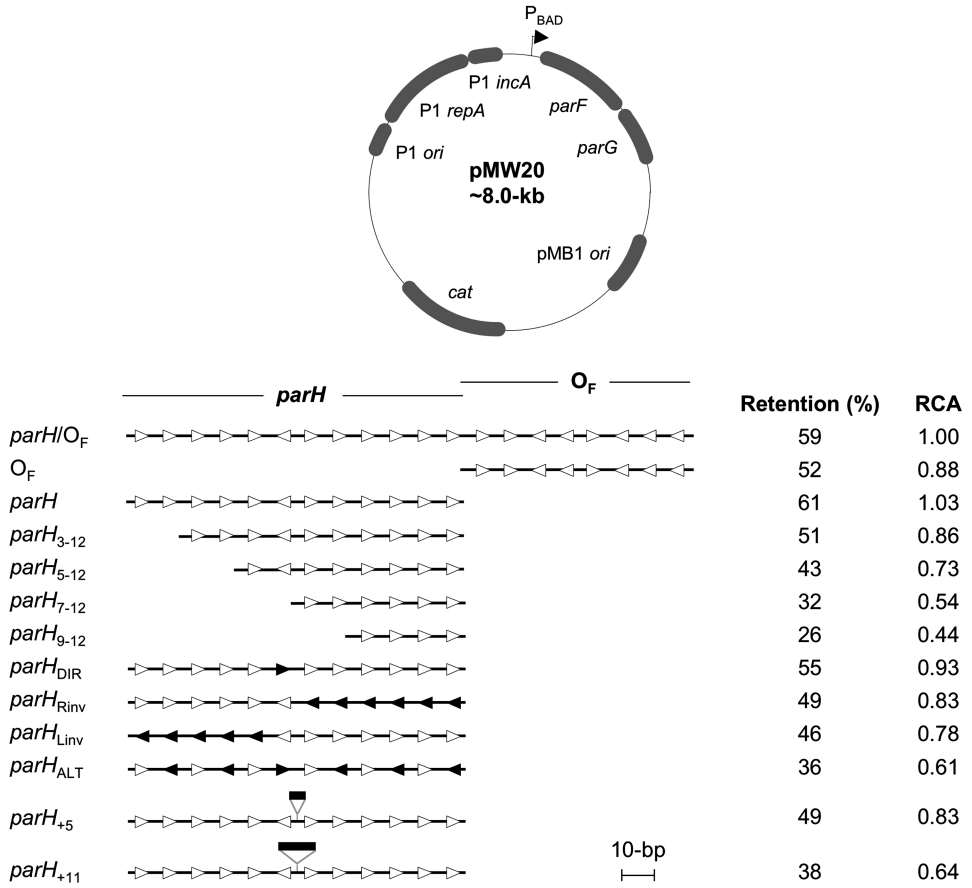

Figure 1. Deletion and mutational analysis of the *parH* centromere. The segregation probe vector pMW20 is illustrated at the top. This plasmid replicates at medium copy number via the pMB1 *ori*. Replication switches to a low copy number via the P1 replicon in a *polA* mutant, but the plasmid is segregationally unstable in this background. The *parFG* genes are expressed from an arabinose-inducible promoter (P_{BAD}) instead of from their native regulatory sequences. The distribution and orientation of the variant 5′-ACTC-3′ motifs in the *parH*-O_F region are indicated by open arrowheads. The segments cloned in pMW20 and tested for segregation activity are shown. Tetramer boxes that were inverted from their normal orientation are denoted with filled arrowheads. The 5- and 11-bp insertions in the $parH_{+5}$ and $parH_{+11}$ sites are indicated by filled bars. Segregational stability assays were conducted as outlined in the text. The relative centromere activities (RCA) associated with the sites are also shown compared to the activity conferred by the intact *parH*-O_F region. The pMW20 vector without *parH*-O_F had an RCA of 0.31. The values presented are the means of at least three independent tests with typical standard deviations of ~10%.

complementary primers one of which was 5′ biotinylated. Purification of the fragments and conditions for retardation assays were outlined in detail previously (26). Briefly, biotinylated DNA (2 nM) was incubated for 20 min at 25°C in binding buffer [10 mM Tris–HCl, pH 7.5, 50 mM KCl, 1 mM dithiothreitol, 5 mM $MgCl_2$, 0.05 mg/ml poly(dI–dC)] with the ParG concentrations shown in figure legends. Reaction mixtures were electrophoresed on 10% polyacrylamide gels in 0.5× Tris–borate–EDTA (TBE) buffer for 30–90 min at 80 V at 22°C. DNA was transferred by capillary action to positively charged nylon membranes (Roche), and the transferred DNA fragments were immobilized by UV crosslinking. The biotin end-labelled DNA was detected using the LightShift chemiluminescent EMSA kit (Pierce) (30).

DNase I footprinting

The preparation and purification of biotinylated PCR products, conditions for DNase I footprinting reactions with ParG, denaturing gel electrophoresis, and detection of biotinylated DNA followed procedures described in detail recently (26).

Atomic force microscopy and data analysis

For sample preparation in atomic force microscopy (AFM), freshly cleaved mica was functionalized with poly-L-lysine (PL) to support secure immobilization of DNA (35). The mica disc was incubated with 30 µl of a 10 µg/ml aqueous PL solution for 30 s, subsequently washed with 4 ml of Millipore water and dried under a nitrogen stream. A DNA fragment encompassing the *parH* region was amplified using pFH547 as template. Protein–DNA binding reactions (20 µl) included the purified PCR fragment (2.5 nM) and ParG (1–3 µM) in binding buffer (10 mM Tris–HCl, 50 mM KCl, 1 mM EDTA, pH 7.5). After 15–30 min, the mixture was diluted 30-fold in binding buffer and 30 µl of this dilution were immediately placed on the PL-mica. After 60 s of incubation, the mica was rinsed carefully with 2 ml of Millipore water and allowed to dry under a gentle nitrogen stream. Measurements were performed with an

AFM instrument as described (36) and FESP tips (Veeco) in tapping mode. Fields of 1×1 μm were scanned at line rates of 1–2 Hz and resolution of 512×512 pixels. AFM images were plane corrected with SPIP software (Image Metrology, Denmark) and saved in bitmap format for further analysis with ImageJ software (version 1.41o, NIH, USA). The images were scaled to 2048×2048 pixels using bilinear pixel interpolation. The entire contours of the DNA molecules were then traced using the freehand line option and saved as *xy*-coordinates. The traced contours were also marked in the images in order to pinpoint two *xy*-coordinates that define the region along each contour which was occupied by ParG. To this end, the points at which the height begins to increase relative to the free DNA were identified using the height threshold tool and their *xy*-coordinates were defined using the point selection tool. To compare the regions occupied by protein with the putative binding sites in the DNA fragments, the saved *xy*-coordinates were characterized by their distances from the DNA terminus which is nearest the protein complex. These distances, as well as total DNA lengths, were determined by summing the distances between the successive coordinates of the entire contours using Excel.

Bending analysis

Plasmid DNA carrying *parH*, O_F and the 5′-end of *parF* was digested with various restriction endonucleases producing different fragments for curvature analysis. DNA was mixed with loading dye and analysed on equilibrated native polyacrylamide gels as outlined previously (37). Gels were pre-run for ~3 h until current and temperature remained constant. Electrophoresis was carried out in 1× TBE at 150 V (8 mA) for ~4 h (migration distance of bromphenol blue ~14 cm) at 4 or 23°C. Gels were stained for 30 min in an aqueous solution of ethidium bromide (1 mg/l), followed by rinsing in water prior to documentation. The 1 kb Plus DNA ladder (Invitrogen) and the 1 kb DNA ladder (New England Biolabs) served as marker fragments together with plasmid fragments resulting from restriction digests running in the same lane as the *parH* fragment. Migration of all fragments was determined for each gel and calibration curves were plotted using the marker fragments (logarithm of number of base pairs versus distance migrated). The apparent sizes in the acrylamide gel relative to the calibration curve were determined.

RESULTS

Defining and dissecting the *parH* centromere *in vivo*

The O_F operator upstream of *parFG* comprises eight variant 5′-ACTC-3′ motifs arranged in a combination of direct and inverted orientation (Figure 1). The tetramer boxes are separated regularly by 4-bp AT-rich sequences. A single ParG dimer loads onto each 5′-ACTC-3′ box, suggesting that the operator is cooperatively coated by as many as eight dimers during transcriptional repression of *parFG* (26). Inspection of the region further upstream of O_F revealed a second cluster of 12 degenerate 5′-ACTC-3′ motifs (*parH*), also separated by AT-rich spacers (Figure 1). One of the boxes is inverted compared to the remainder. These 12 5′-ACTC-3′ boxes are embedded in a set of longer repeats that were noted previously in this region and which were originally proposed as the putative centromere locus at which segrosome assembly occurs (4,30).

As repeat motifs are characteristic of plasmid centromeres (4), attempts were made to support the partitioning of a segregationally unstable test plasmid harbouring the full complement of 20 5′-ACTC-3′ boxes, i.e. *parH*-O_F, when the ParFG proteins were provided *in trans* from a compatible plasmid. The proteins were produced either from genes under the control of the native *parFG* expression signals or from a lactose-inducible promoter. Expression of *parFG* from a variably inducible arabinose promoter was also tested. In addition, selected synthetic promoters with strengths from weak to high (38) were trialled. None of these approaches elicited improved segregation of the vector possessing the complete set of tetramer boxes compared to the same plasmid lacking the repeat sequences. As an alternative strategy, the *parFG* genes were inserted in the segregation probe vector, pFH450 (31), under the control of an arabinose-inducible promoter (P_{BAD}; 34). This manipulation produced plasmid pMW20 that entirely lacks any of the natural regulatory sequences upstream of *parFG* (Figure 1). With arabinose induction, the plasmid was maintained at a frequency of $18 \pm 3\%$ during non-selective growth for approximately 25 generations in an *E. coli polA* mutant in which the plasmid replicates using the low copy number P1 replicon. However, insertion of the *parH*-O_F region 5′ of the arabinose promoter improved retention to $59 \pm 16\%$, a value close to that observed with the intact partition cassette (33). Similar retention values were obtained when the *parH*-O_F region was cloned elsewhere downstream of *parFG* in pMW20 (data not shown). Thus, the *parH*-O_F region exhibits ectopic, centromere-like activity when located *in cis* to *parFG*. One advantage in expressing *parFG* from P_{BAD} is that any contribution of O_F to centromere activity can be examined independently of its regulatory functions. The conditions required for *in trans* activity of the ParFG proteins at *parH*-O_F have yet to be defined: the appropriate intracellular protein concentrations and/or the temporal pattern of *parFG* expression that is required to support centromere activity *in trans* may be difficult to replicate artificially. Alternatively, a DNA topological requirement or a positional effect of the genes and/or centromere may influence segregation activity. These possibilities require further investigation.

As *parH* and O_F both comprise arrays of 5′-ACTC-3′ boxes, the independent efficacies of the two regions in centromere function were tested in pMW20. The *parH* region displayed centromere activity that was indistinguishable from that of the complete *parH*-O_F region (Figure 1). The operator locus alone also was an effective centromere, albeit with slightly less activity than *parH*: O_F may have dual roles in transcriptional repression of *parFG* (26) and in centromere function. As the activity of *parH* was not enhanced appreciably by O_F, the centromeric

properties of *parH* alone were characterized further. Progressive deletion of pairs of 5′-ACTC-3′ boxes from *parH* was accompanied by concomitant gradual reductions in the retention levels of the test plasmids (Figure 1). Notably, plasmids bearing only four ($parH_{9–12}$) or six ($parH_{7–12}$) tetramer boxes were maintained at approximately half the frequency conferred by full-length *parH*.

The *parH* and O_F sites display comparable centromere activities, but the arrangements of the 5′-ACTC-3′ motifs in the two loci differ markedly: *parH* comprises eleven direct repeats and one inverted repeat (Figure 1), whereas O_F consists of three direct repeats interspersed with five inverted boxes (26). To assess further the malleability of these sequences for partition activity, synthetic *parH* centromeres with reconfigured 5′-ACTC-3′ boxes were cloned in pMW20 and tested in segregation assays (Figure 1). A site in which all of the tetramer boxes were oriented similarly was a proficient centromere ($parH_{DIR}$). Inversion of the six rightward motifs or the five leftmost boxes of *parH* caused only modest decreases in segregation activity ($parH_{Rinv}$ and $parH_{Linv}$, respectively). A site in which alternating tetramers in *parH* were inverted from their canonical orientation was a less effective centromere, although remained partially functional ($parH_{ALT}$). Insertion of one-half helical turn at the centre of *parH* only modestly affected centromere activity ($parH_{+5}$), whereas insertion of a complete turn ($parH_{+11}$) reduced centromere function more severely. Thus, a variety of natural and artificial 5′-ACTC-3′ box dispositions are viable for *parH* centromere action, although a minimum of eight repeats is required for efficient activity. Moreover, altering the relative helical positions of the two halves of *parH* was well-tolerated whereas maintaining these positions, but increasing the distance between the halves, was more deleterious.

Discrete binding of ParG to *parH* and O_F *in vitro*

The ParG protein binds to the O_F operator to achieve transcriptional repression of the *parFG* genes (25,26). DNase I footprinting *in vitro* of the *parH* centromere region revealed that ParG also protects the entire set of 12 5′-ACTC-3′ boxes from digestion on both DNA strands (Figure 2). The AT-rich spacers separating certain boxes were slightly less well protected, most obviously on the top strand. The DNA fragments used in these reactions harboured both *parH* and O_F so as to ascertain whether ParG interacted continuously or discontinuously with the two loci: the zones of protection were separated by a ~5-bp window that remained fully accessible to DNase I (Figure 2). These data correlate with recent observations that delimited the extent of ParG interaction with O_F (26).

As a second strategy to probe the interaction of ParG with *parH*, a fragment possessing the centromere (365 bp) was bound by ParG and visualized by AFM (Figure 3A). The images showed protein binding as discrete extended foci. The measured total DNA length was 117 ± 5 nm which was somewhat shorter (~5%) than the 124 nm expected for B-form DNA of this length. Shortening to this extent has been observed previously by AFM of naked DNA (35,39), so the reduced *parH* fragment lengths observed here do not necessarily reflect ParG-induced compaction. Therefore, the approximate DNA helical rise was 0.32 nm/bp for these surface preparations. To examine if the visible ParG binding

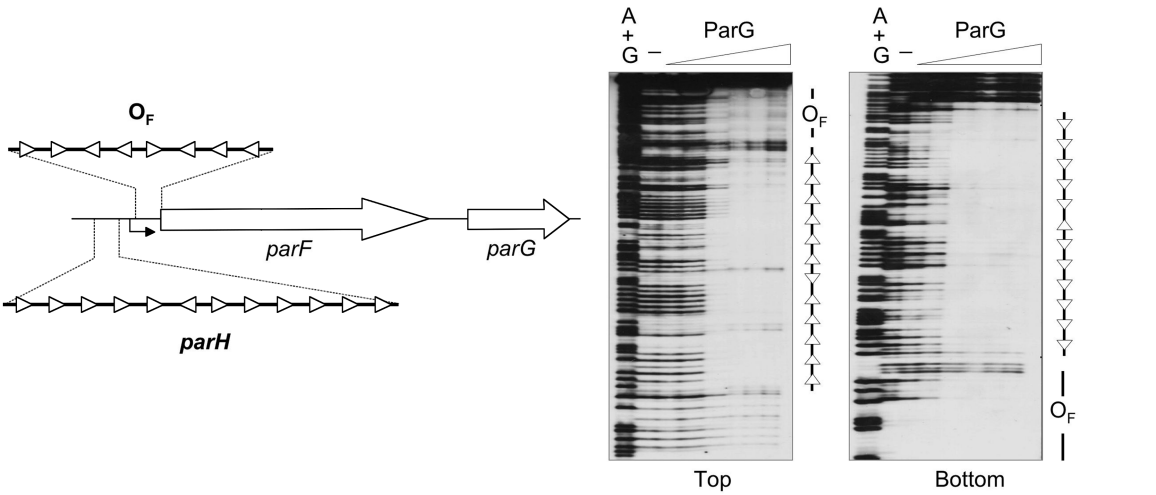

Figure 2. DNase I footprinting of ParG at the *parH* centromere and adjacent O_F operator. The distribution and orientation of the variant 5′-ACTC-3′ motifs in the *parH*-O_F region are indicated by open arrowheads. The bent arrow indicates the putative *parFG* promoter (25,26). Footprinting reactions were performed as outlined in the 'Materials and Methods' section using PCR fragments biotinylated at the 5′-ends of either top or bottom strands. ParG concentrations (μM monomer, left to right): 0, 0.05, 0.1, 0.2, 0.6, 1.0 and 1.5. A + G, Maxam–Gilbert sequencing reactions. The relative dispositions on the top and bottom strands of the *parH* region that are protected from DNase I digestion are shown in the bottom panel. The 5′-ACTC-3′ motifs are boxed.

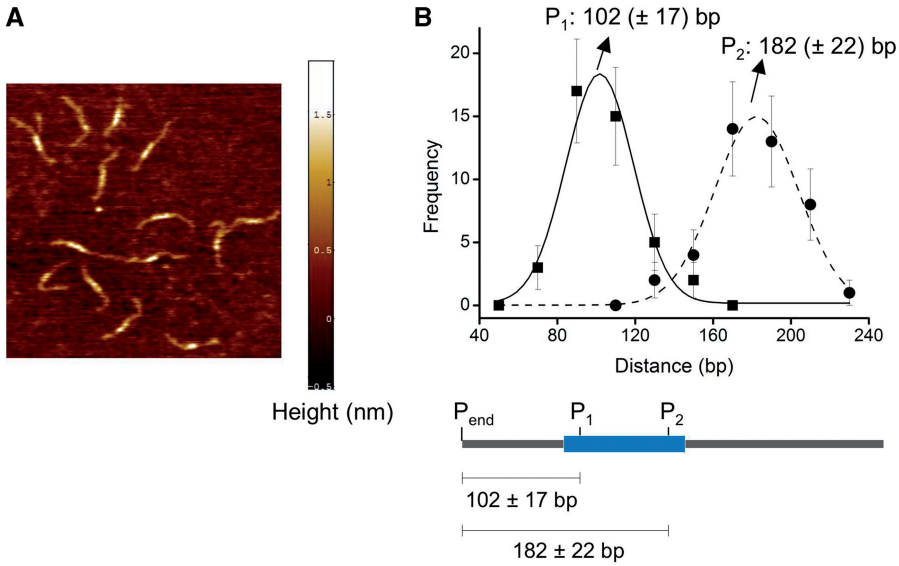

Figure 3. AFM of ParG–DNA complexes. (**A**) Representative AFM images (500 × 500 nm) are shown of ParG bound to a PCR fragment that contained *parH*. Regions with bound ParG are visible as foci of increased height (white regions). (**B**) Histograms of distances between P_{end} and P_1 and P_2 in DNA fragments that contain *parH* after correction for the estimated tip induced lengthening of protein complexes. A traced DNA contour was made whose consecutive *yx*-coordinates were used to calculate total DNA lengths and positions of ParG binding sites where P_{end} is the DNA extremity that is located nearest the protein focus, and P_1 and P_2 are the coordinates that define the initial and end points, respectively, of the protein bound region. Data were fitted with a Gaussian function (curves) to calculate means and SDs. The numbers of base pairs are calculated based on the measured helical rise of 0.32 nm/bp. Squares indicate the distances for the initial point, P_1, and circles mark the distances for the endpoint, P_2. Summaries of the AFM analysis are shown at the bottom. The blue bar denotes the position of *parH* based on DNase I footprinting (Figure 2).

corresponded to the *parH* region, the coordinates P_1 and P_2, defining the foci along digitized traces of the DNA contours, were located as described in the 'Materials and Methods' section. The DNA terminus that was closer to the protein focus was determined and the positions of P_1 and P_2 were specified as distances from this terminus. The distance distributions shown in Figure 3B comprised a correction of the tip induced lengthening of the detected protein region. For an approximation of this effect, it was assumed that the measured DNA width of ~10 nm corresponded to a broadening of 8 nm relative to the known DNA diameter of 2 nm. Therefore, the points determined initially were shifted by 4 nm towards the centres of the foci. The distance maximum in the fragment containing *parH* (Figure 3B) was found at 102 ± 17 bp for position P_1 and 182 ± 22 bp for P_2. These positions roughly define the ends of the *parH* site which is ~100-bp based on DNase I footprinting (Figure 2). In conclusion, the AFM measurements described here confirm the selective binding of ParG to *parH* within the probed fragment at the specified protein concentrations. Although the direction of the fragments is not defined in the AFM images, the observation of single protein foci together with the agreement between the measured distances and the *parH* location established previously (Figure 2) indicate that ParG binding occurred exclusively in this region.

The *parH* centromere lacks intrinsic curvature

The plasmid R1 *parC* centromere displays strong intrinsic curvature and is further distorted into a U-shaped structure by binding of its cognate CBF to the two arms of the site (11,37). In contrast, the *cenE* centromere of plasmid pGENT shows modest inherent curvature which is not altered significantly by the binding of its CBF on a linear DNA fragment (40). A theoretical analysis of the inherent curvature of the *parS* centromere of plasmid P1 suggests that the site is curved (C. Hoischen and S. Diekmann, unpublished data). Moreover, DNA bending by integration host factor (IHF) at the centre of *parS* permits the cognate CBF to span the two arms of the site (41,42). These different topological features may be instrumental in organizing functional segrosomes in different systems.

The sequence of a HindIII–EcoRI fragment encompassing *parH*, O_F and the 5′-end of *parF* (Figure 4) is AT rich (61.2%) and contains blocks of AT nucleotides. These mostly comprise four or five consecutive AT steps suggesting that this region might be curved and show anomalous gel migration. Inherent DNA curvature has clearly identifiable properties. First, curvature can be measured as retarded migration in polyacrylamide gel electrophoresis. Second, curvature is most evident at moderate temperatures and decreases at elevated temperatures (43–45). The quotient *apparent sequence length divided by known sequence length* is termed the *k*-factor with an experimental error in *k*-factor measurement of ±0.03. Variation of curvature values in normal DNA sequences is between *k*-factor values 0.98 and 1.05. Applying this experimental strategy, we recently detected DNA curvature with large *k*-factors for yeast centromeres [*k*-values up to 1.20; (46)] and the plasmid R1 *parC* centromere [$k > 2.0$; (37)]. The *parH* locus and flanking regions were inserted in pUC18 and a set of overlapping fragments were generated by

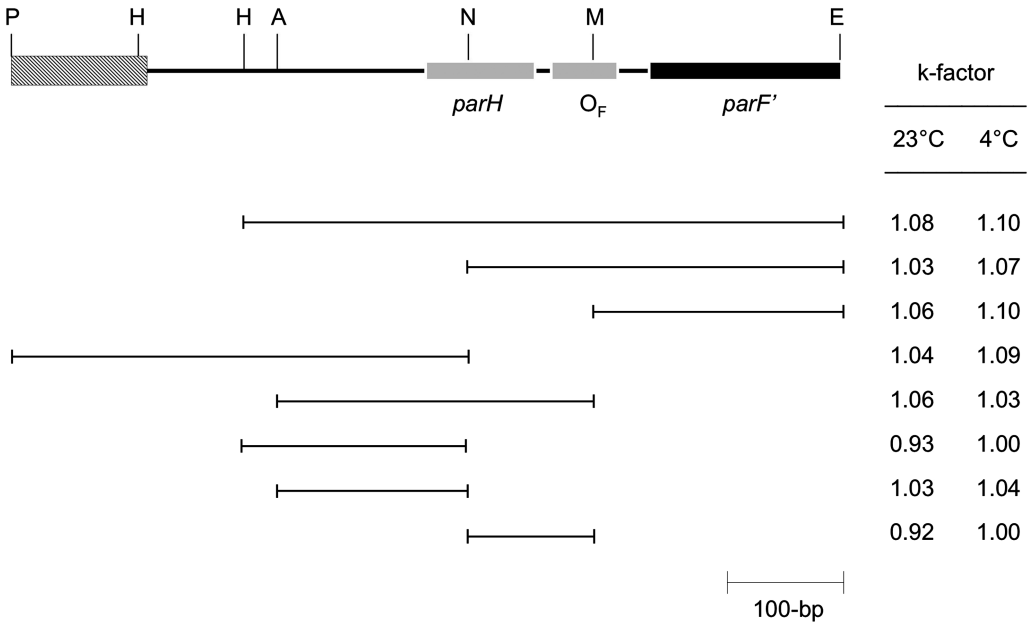

Figure 4. Testing for intrinsic curvature in the *parH*-O_F region. A partial restriction map of the region cloned in pUC18 is illustrated at the top. The cloned fragment spans *parH* and O_F (shown as shaded boxes) and the 5′-end of *parF* (filled box). Vector sequences are shown by the hatched box. A set of subfragments that contain *parH*-O_F at different locations relative to the fragment ends were used in bending analysis. A, AgeI; E, EcoRI; H, HindIII; M, MunI; N, NdeI; P, PstI. The *k*-factors derived from analysis of these restriction fragments under standard electrophoretic conditions (6% polyacrylamide in TBE buffer at 4 or 23°C) are shown.

restriction enzyme digestion (Figure 4). Fragment migration relative to marker fragments in different lanes, as well as to plasmid fragments in the same lane, was analysed in 6% polyacrylamide gels under different temperature conditions, and in agarose gels. Apparent fragment sizes were determined relative to marker fragments with normal migration. Most of the test fragments from the region surrounding *parH* did not show any appreciable gel retardation under standard conditions (23°C, TBE buffer) (Figure 4). Only rather weak retardation of a 520-bp HindIII–EcoRI fragment (*k*-factor = 1.08), a 267-bp AgeI–MunI fragment (*k*-factor = 1.06), and a 219-bp MunI–EcoRI fragment (*k*-factor = 1.06) was measured. The *k*-factors of all tested fragments slightly increased when the temperature was reduced to 4°C. However, a weak moderation in the *k*-factor was evident for the AgeI–MunI fragment at the lower temperature. Although this analysis suggests some secondary structure, perhaps DNA curvature, resulting in a slightly anomalous gel migration, it reveals that the DNA sequence that includes *parH*-O_F is not substantially curved when tested in a linear fragment. This conclusion was supported by the analysis of permuted fragments of identical size and sequence context produced from cloning the *parH*-O_F region in a bending vector: these fragments did not display intrinsic curvature in gel electrophoresis (data not shown).

The clustered 5′-ACTC-3′ motifs in *parH* provide a high affinity docking site for ParG

To parallel the *in vivo* deletion analysis of *parH* (Figure 1), DNA fragments with decreasing numbers of 5′-ACTC-3′ boxes (Figure 1) were tested for ParG binding *in vitro*. Full-length *parH* was assembled into a single complex at the lowest ParG concentration that was examined in gel retardation assays (Figure 5A). In contrast, ParG incompletely bound a substrate lacking two 5′-ACTC-3′ boxes even at the highest protein concentration ($parH_{3–12}$). Multiple nucleoprotein complexes with intermediate migrations were also detectable, suggestive of partial occupancy of the motifs by ParG. Target fragments possessing only eight, six or four tetramer boxes each were bound progressively less avidly by the protein ($parH_{5–12}$, $parH_{7–12}$ and $parH_{9–12}$, respectively), and a single pair of 5′-ACTC-3′ motifs was sufficient only for very weak and limited binding ($parH_{11–12}$) (Figure 5A). DNase I footprinting of fragments possessing different numbers of motifs confirmed a positive correlation between ParG binding and increasing motif number (Figure 5B). Footprinting also demonstrated that ParG maintained specificity in protecting the regions harbouring the 5′-ACTC-3′ boxes from digestion even when presented with as few as four boxes.

The variable tetramer motifs in *parH* comprise either canonical 5′-ACTC-3′ boxes or derivatives of this sequence (Figure 2). A single 5′-ACTC-3′ tetramer is insufficient for ParG binding *in vitro* (26). Thus, to assess whether ParG recognizes the assorted tetramer motifs with different avidities, the centromere was subdivided experimentally into pairs of motifs flanked by their natural AT-rich spacers ($parH_{1–2}$ to $parH_{11–12}$) (Figure 6A). Double-stranded oligonucleotides bearing these sequences were tested in gel retardation assays with titrations of ParG. The subsites displayed distinct binding patterns

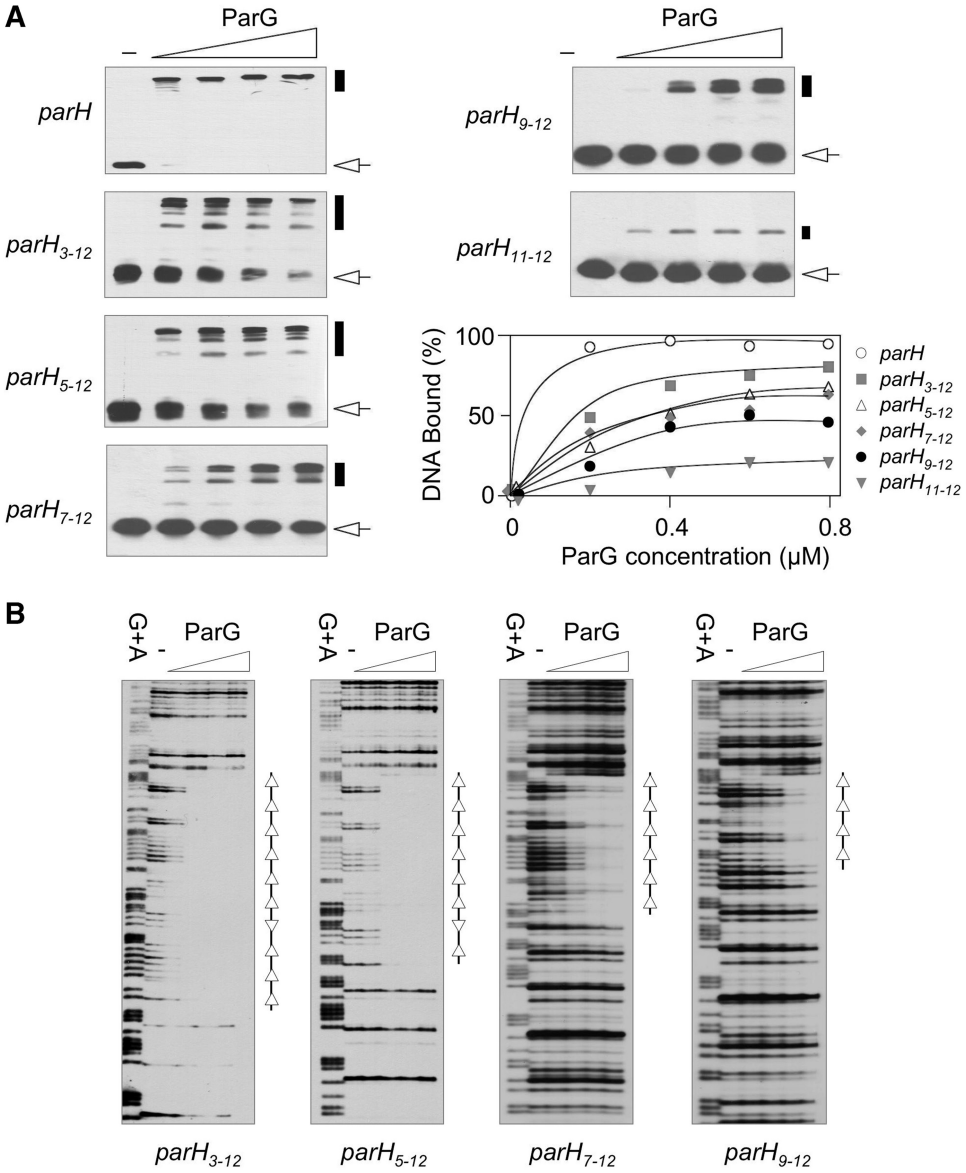

Figure 5. ParG binding to *parH* subsites possessing decreasing numbers of 5′-ACTC-3′ boxes. (**A**) Gel retardation analysis of ParG with the full-length, biotinylated *parH* centromere containing 12 5′-ACTC-3′ motifs, or with fragments possessing 10 ($parH_{3–12}$), 8 ($parH_{5–12}$), 6 ($parH_{7–12}$), 4 ($parH_{9–12}$) or 2 ($parH_{11–12}$) boxes. ParG concentrations (μM monomer, left to right): 0, 0.2, 0.4, 0.6 and 0.8. Arrows and boxes indicate unbound DNA and ParG–DNA complexes, respectively. (**B**) DNase I footprinting of ParG with biotinylated PCR fragments containing 10 ($parH_{3–12}$), 8 ($parH_{5–12}$), 6 ($parH_{7–12}$) or 4 ($parH_{9–12}$) tetramer boxes. Footprints on the top DNA strand (Figure 2) are shown. ParG concentrations (μM monomer, left to right): 0, 0.05, 0.2, 0.6 and 0.8. The distribution and orientation of the variant 5′-ACTC-3′ motifs in the substrates are indicated by open arrowheads.

with ParG: $parH_{1–2}$ and $parH_{5–6}$ were recognized most strongly, whereas ParG bound the $parH_{9–10}$ fragment least well (Figure 6A). Other motif pairs showed intermediate binding properties.

To investigate further the differences in ParG affinities for different *parH* subsites, ParG binding to the biotinylated $parH_{1–2}$ oligonucleotide was challenged with increasing concentrations of each of the unlabelled subsites (Figure 6B). The $parH_{1–2}$ and $parH_{5–6}$ fragments competed most efficiently, whereas $parH_{7–8}$ and $parH_{9–10}$ were weak competitors. Thus, gel retardation assays and competition data revealed that ParG recognizes different centromere subsites with different affinities ($parH_{1–2} \approx parH_{5–6} > parH_{11–12} \approx parH_{7–8} \approx parH_{3–4} > parH_{9–10}$). These differences likely arise both from dissimilarities in the tetramer box sequences as well as from variations in the intervening AT-rich spacers in *parH*, both of which contribute significantly to DNA binding by ParG (26).

The influence of the mobile N-terminus of ParG on centromere interaction

The flexible N-terminal tail of ParG is multifunctional, harbouring an arginine finger-like motif that stimulates

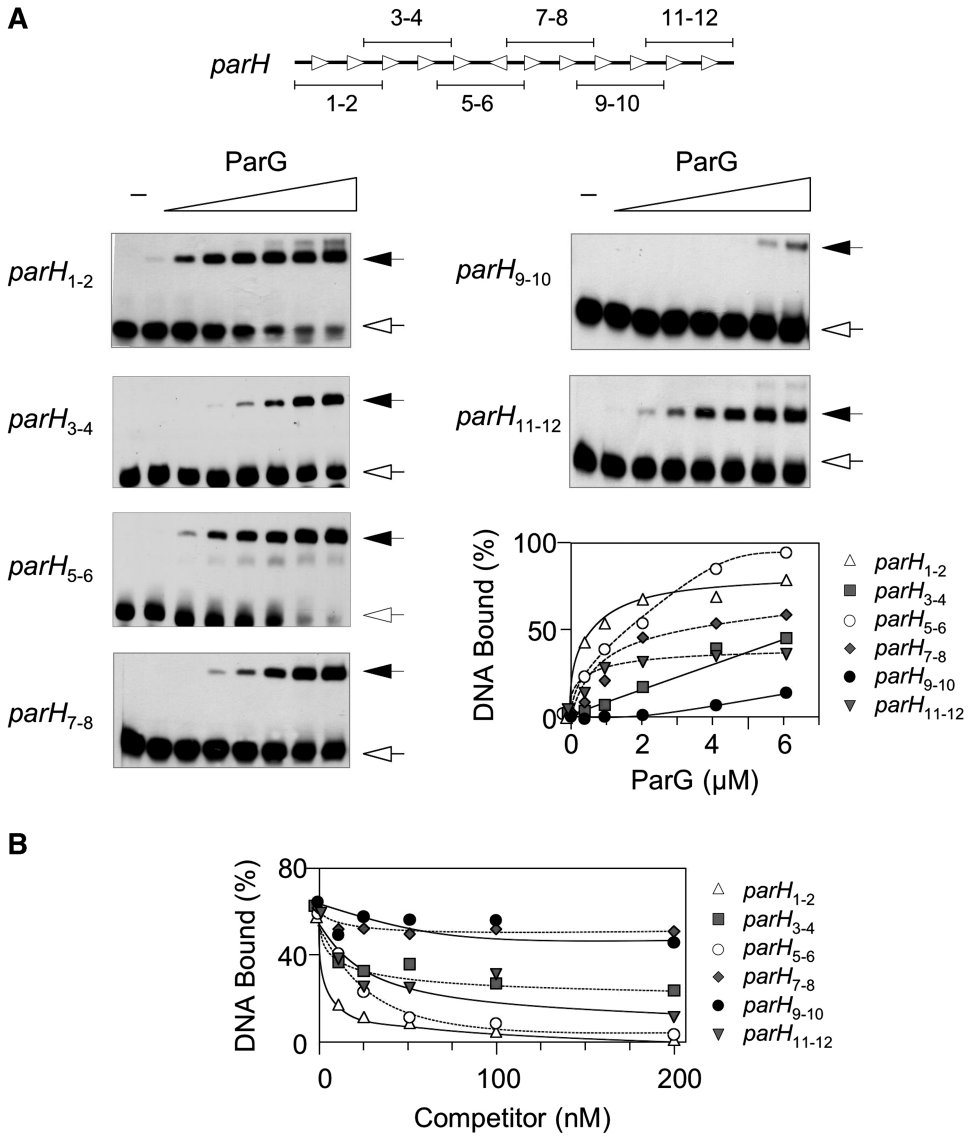

Figure 6. ParG binding to adjacent pairs of 5′-ACTC-3′ boxes from the *parH* centromere. (**A**) Gel retardation assays of ParG with fragments possessing pairs of tetramer boxes from *parH*. The fragments were generated by second-strand synthesis of 49-nt single-strand oligonucleotides bearing the relevant boxes at the centres of the fragments. The single-stranded oligonucleotides each included a common priming site for second-strand synthesis by a 5′-biotinylated primer. Thus, the fragments contained the relevant boxes and immediate flanking regions, but no other *parH* sequences. ParG concentrations (μM monomer, left to right): 0, 0.1, 0.2, 0.5, 1.0, 2.0, 4.0 and 6.0. Open and filled arrows indicate unbound DNA and ParG–DNA complexes, respectively. (**B**) Competition gel retardation assays of ParG binding to the $parH_{1–2}$ fragment. Binding reactions containing ParG (4 μM), a biotinylated fragment (2.5 nM) harbouring the $parH_{1–2}$ pair of 5′-ACTC-3′ boxes that are strongly bound by the protein (panel A), and increasing amounts (up to 200 nM) of unlabelled 19-bp competitor oligonucleotides were incubated at 22°C for 20 min. Reactions were analysed further as described in the 'Materials and Methods' section. Unlabelled competitor DNAs contained the same sequences used in panel A.

nucleotide hydrolyis by the partner ParF ATPase, as well as possessing sequences that promote ParF polymerization (18). Moreover, the interaction of ParG with the O_F operator is modulated by transient associations between the flexible N-terminal and folded C-terminal domains in complex with the target DNA (25). To examine whether the tail also influences interaction with the *parH* centromere, derivatives lacking 9, 19 or 30 amino acids from the ParG N-terminus (25) were tested for centromere assembly (Figure 7). This progressive truncation of the mobile tail was accompanied by increasingly stronger interactions of the ParG deletion derivatives with the centromere both in gel retardation assays and DNase I footprinting. Notably, at protein concentrations at which ParG only weakly shifted *parH* into a series of intermediate nucleoprotein complexes in retardation assays, Δ30ParG assembled the DNA into a major retarded complex (Figure 7A). In addition, Δ30ParG fully protected the *parH* site from DNase I digestion at protein concentrations that were insufficient for protection by the full-length protein (Figure 7B).

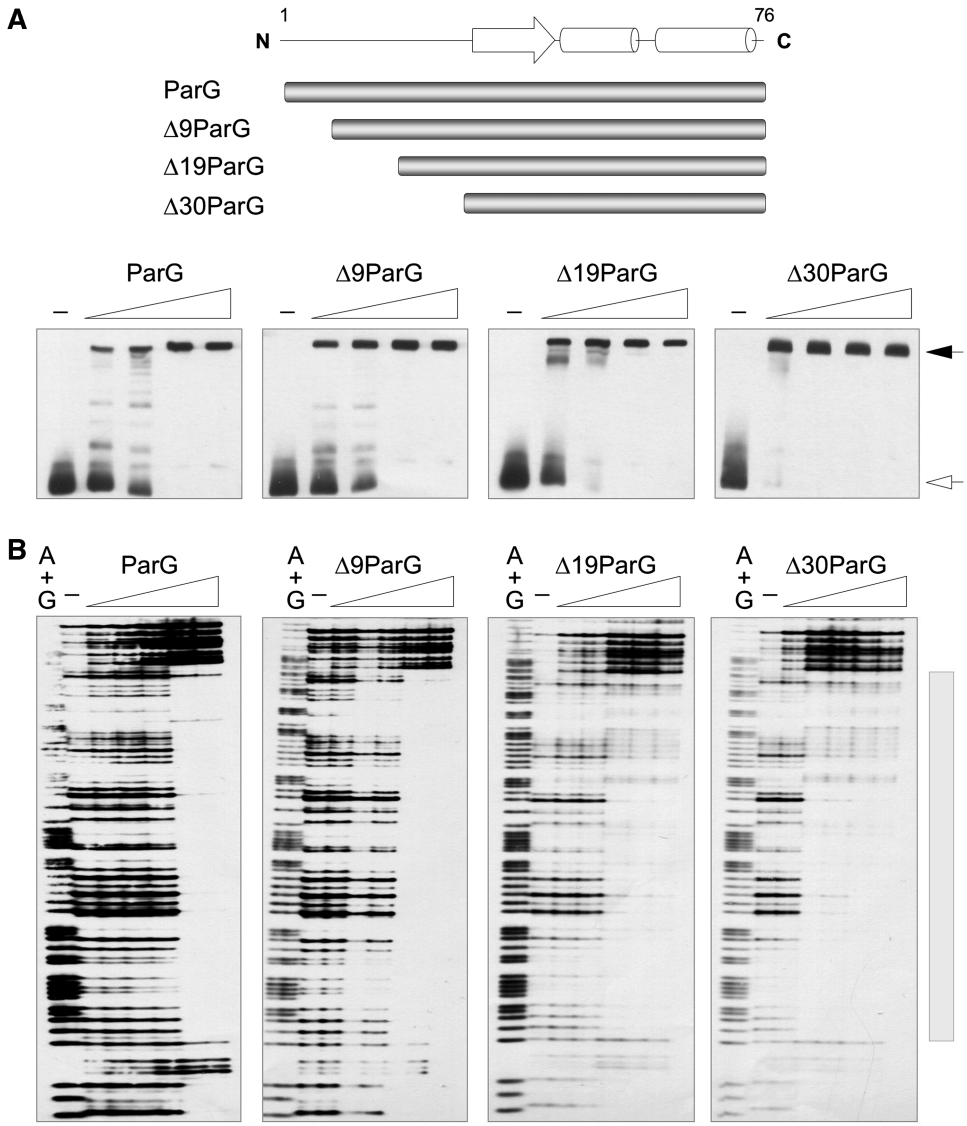

Figure 7. Modulation of ParG-centromere interactions by the protein's mobile N-terminal tail. (**A**) α-helices and β-strands in ParG are indicated by cylinders and arrows, respectively (24). The extent of the N-terminal deletions in Δ9ParG, Δ19ParG and Δ30ParG (25) are shown. These derivatives were used in gel retardation studies with a biotinylated PCR product containing full-length *parH*. Protein concentrations were (μM monomer, left to right): 0, 0.05, 0.1, 0.2 and 0.5. Open and filled arrows indicate unbound DNA and ParG–DNA complexes, respectively. (**B**) DNase I footprinting of ParG, Δ9ParG, Δ19ParG and Δ30ParG at the *parH* centromere. The distribution and orientation of the variant 5′-ACTC-3′ motifs in the *parH* region are indicated by open arrowheads. Footprinting reactions were performed as outlined in the 'Materials and Methods' section using a PCR fragment biotinylated at the 5′-end of the bottom strand. ParG concentrations (μM monomer, left to right): 0, 0.025, 0.05, 0.1, 0.2 and 0.5. A + G, Maxam–Gilbert sequencing reactions. Note that the weak bands within the major zones of protection seen here with Δ19ParG and Δ30ParG were not observed reproducibly.

The preceding observation that deletion of the flexible tail increases *parH* binding by ParG apparently contradicts previous findings that Δ30ParG interacts more weakly at the O_F locus in both gel retardation and surface plasmon resonance assays (25). However, the O_F site subsequently was found to be more extensive than originally described, comprising eight variant 5′-ACTC-3′ boxes instead of the five motifs used in previous experiments (26). The observation that Δ30ParG recognized a substrate containing three tetramer boxes less avidly than did the full-length protein (25) is also intriguing as it suggests that the flexible tail affects DNA binding differently dependent on the number of 5′-ACTC-3′ boxes present in the target site. Therefore, the interaction of the ParG deletion derivatives with pairs of 5′-ACTC-3′ motifs flanked by their natural AT-rich spacers ($parH_{1–2}$ to $parH_{11–12}$) (Figure 6A) was examined (Figure 8A). Unlike the stronger binding that the truncated derivatives displayed to the complete *parH* locus (Figure 7), binding of full-length and deletion versions of ParG was broadly similar to the $parH_{1–2}$, $parH_{3–4}$, $parH_{5–6}$ and $parH_{7–8}$ fragments. However, Δ19ParG generally interacted more weakly with the subsites than either full-length ParG,

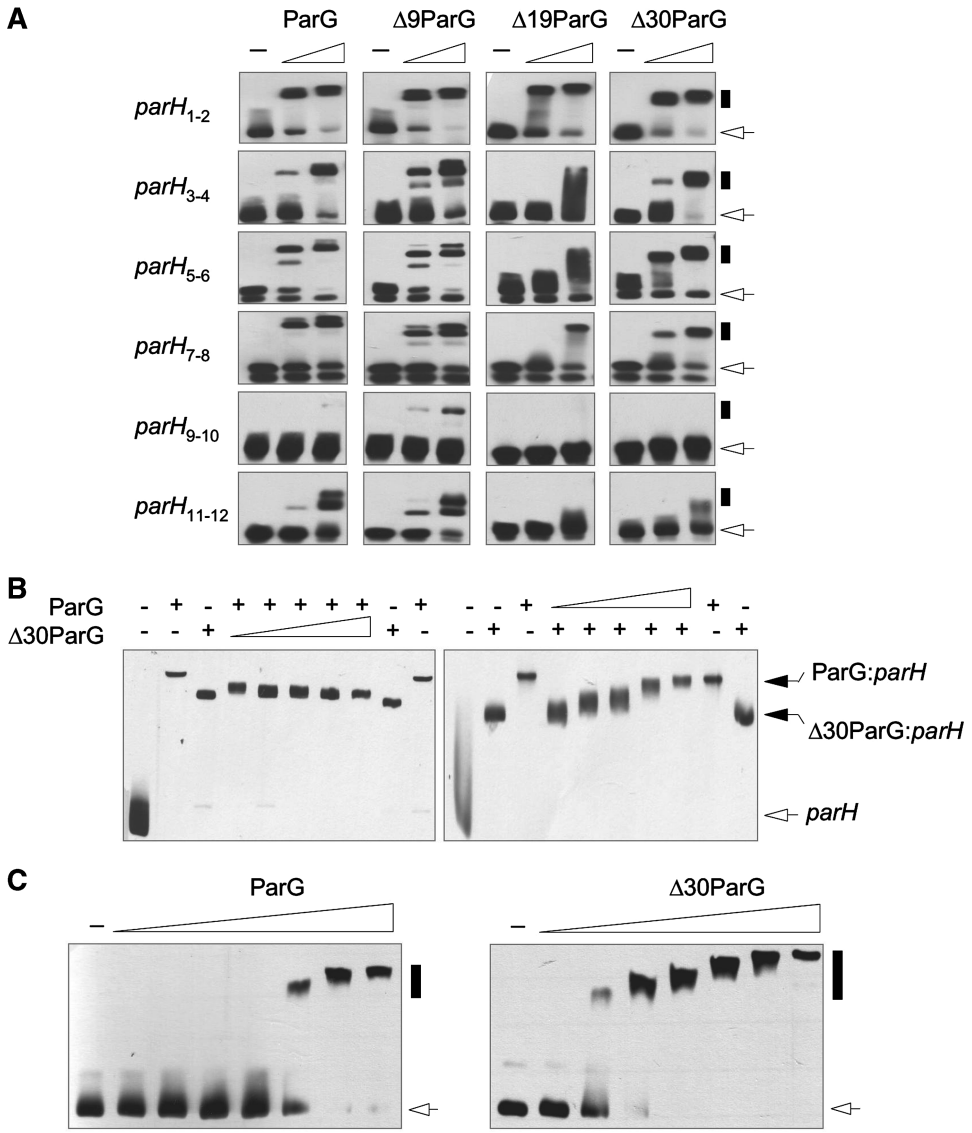

Figure 8. Binding of ParG deletion derivatives to *parH* subsites and competition assays with full-length ParG. (**A**) Gel retardation comparison of ParG, Δ9ParG, Δ19ParG and Δ30ParG binding to adjacent pairs of 5′-ACTC-3′ boxes from the *parH* centromere. The DNA fragments consisted of biotinylated, double-stranded oligonucleotides possessing pairs of tetramer boxes (Figure 6A). Protein concentrations (μM monomer, left to right): 0, 0.5 and 2.0. Arrows and boxes indicate unbound DNA and protein–DNA complexes, respectively. (**B**) Left: competition assays in which a preformed ParG:*parH* complex was challenged with increasing concentrations of Δ30ParG. Where indicated, ParG was present at 0.4 μM. In binding reactions that contained only Δ30ParG, the protein was also included at 0.4 μM. In reactions containing both proteins, Δ30ParG was present at 0.4, 0.8, 2.0, 4.0, or 8.0 μM (left–right). Right, competition assays in which a preformed Δ30ParG:*parH* complex was challenged with increasing concentrations of ParG. Where indicated, Δ30ParG was present at 0.4 μM. In binding reactions that contained only ParG, the protein was also included at 0.4 μM. In reactions containing both proteins, ParG was present at 0.4, 0.8, 2.0, 4.0 or 8.0 μM (left–right). Open and filled arrows indicate unbound DNA and protein–DNA complexes, respectively. (**C**) Deletion of the ParG N-terminal tail increases non-specific DNA binding. ParG and Δ30ParG were incubated with a biotinylated 60-bp fragment bearing the Epstein–Barr virus nuclear antigen binding site. This fragment lacks the 5′-ACTC-3′ boxes recognized by ParG. Protein concentrations (μM monomer, left to right): 0, 2, 2.5, 3, 3.5, 4, 4.5 and 5.0. Arrows and boxes indicate unbound DNA and protein–DNA complexes, respectively.

Δ9ParG or Δ30ParG. Strikingly, the $parH_{9–10}$ and $parH_{11–12}$ oligonucleotides were also bound less well by Δ30ParG than by the wild-type protein. Thus, complete truncation of the N-terminal tail elicited improved binding of ParG to the full-length centromere, but caused weaker binding when certain subsites of *parH* that contain fewer 5′-ACTC-3′ boxes were tested.

As the variant entirely lacking the mobile tail binds *parH* at a lower protein concentration than does full-length ParG, relative centromere binding by ParG and Δ30ParG was assessed by challenging a preformed ParG:*parH* complex with increasing concentrations of Δ30ParG (Figure 8B, left). The complex formed by ParG with *parH* migrates more slowly in gel retardation assays than does the Δ30ParG:*parH* complex allowing ready discrimination between the two species. At a 1:1 ratio, the ParG:*parH* complex was entirely disassembled by Δ30ParG and only nucleoprotein complexes with

intermediate migration were observed. This likely reflects the formation of complexes that include both protein species: either ParG and Δ30ParG may bind simultaneously to the centromere, or monomer exchange between the two species may produce ParG:Δ30ParG heterodimers. Even at a 20:1 ratio of Δ30ParG:ParG, only nucleoprotein species with intermediate mobility were evident indicating that the deletion protein cannot fully displace ParG from the centromere. The converse experiment was also performed. When a preformed Δ30ParG:*parH* complex was challenged with the full-length protein, mixed complexes with migrations intermediate between those produced with ParG or Δ30ParG only were again evident (Figure 8B, right). However, at a high ParG:Δ30ParG ratio, *parH* was assembled entirely into a complex that comigrated with the complex observed only with ParG. Thus, ParG can actively dislodge Δ30ParG from the centromere more effectively than the reverse.

In summary, truncation of the ParG flexible tail permits binding to the full-length *parH* centromere at lower protein concentrations than with the native protein (Figure 7). In contrast, binding of the deletion proteins to subsites possessing only two 5′-ACTC-3′ motifs is only as strong, and in some instances weaker, than binding by full-length ParG (Figure 8A). Moreover, ParG not only competes better than Δ30ParG for the *parH* site, but can displace the deletion protein from the centromere more effectively than Δ30ParG can dislodge ParG (Figure 8B and C). The combined results suggest that the mobile tail modulates the association rate of ParG for the centromere, but that the stability of the ParG-*parH* complex that lacks the tail may be perturbed.

Like all DNA binding proteins, ParG binds to DNA non-specifically *in vitro* when the protein concentration is sufficiently high (24). To ascertain whether the ParG mobile tail contributes to its binding specificity, full-length ParG and Δ30ParG proteins were tested at concentrations up to 5 μM with a biotinylated 60-bp fragment bearing the Epstein–Barr virus nuclear antigen binding site. This fragment lacks the characteristic 5′-ACTC-3′ motifs recognized by ParG and is not bound by the protein at $\leq$3.5 μM, a concentration that is sufficient for full binding of *parH* (Figure 8C, left). The non-specific DNA is shifted into a series of complexes with progressively slower migration at elevated ParG concentrations. In contrast, the non-specific DNA is fully bound in the presence of Δ30ParG at a concentration of 3 μM (Figure 8C, right). Equivalent results were obtained with a second fragment of non-specific DNA (data not shown). Thus, removal of the flexible N-terminal tail from ParG also causes the protein to bind non-specific DNA more avidly.

Loading of ParF into the segrosome

The influence of ParF on the binding of ParG to the region upstream of *parFG* was examined previously, both in the absence and presence of ATP (30). However, this analysis was performed before the boundaries of the *parH* and O_F loci were delineated accurately [Figure 1; (26)], and before the dynamic effects of ATP on ParF behaviour had been investigated (15,18). In view of this, the assembly of ParF into the segrosome was re-assessed using *parH* centromeric DNA defined here. First, the ParF protein does not bind the *parH* centromere in gel retardation assays either in the absence (30) or presence of ATP (data not shown). Second, the ParG:*parH* complex was supershifted progressively into a more slowly migrating complex that remained in the gel loading well as ParF was titrated into the reactions. At a ParF:ParG ratio of 32:1, the centromere was supershifted entirely (Figure 9A, left). Third, ParF slightly modified the DNase I footprinting patterns that ParG produced at *parH* (Figure 9B). ParG was included in footprinting reactions at three concentrations: 0.1 (low), 0.25 (medium) and 0.5 (high) μM. In each case, ParF was omitted or was included at 4 or 20 μM. At low ParG concentration, no protection of the centromere was observable on top or bottom DNA strands either in the absence or presence of ParF. Incomplete protection of the site was apparent at medium ParG concentration. ParF modestly enhanced this protection. At high ParG concentration, ParG fully protected both the centromere and O_F operator from DNase I digestion. The two loci were separated by the distinctive window of DNase I accessibility noted earlier (Figure 2). However, DNase I access to this zone was hindered by the presence of ParF which also slightly ameliorated protection at the other boundary of *parH*. The alterations in DNase I protection patterns induced by ParF were subtle, but reproducible, and demonstrate that centromeric DNA assembled into the segrosome may be organized differently than when bound only by ParG.

ParG binds progressively more weakly to *parH* subsites that comprise decreasing numbers of 5′-ACTC-3′ boxes (Figure 5A). The capacity of ParF to co-assemble with ParG into these subsites was tested. Supershifted complexes were weakly detectable with as few as six tetramer motifs derived from the centromere ($parH_{7–12}$). In contrast, supershifted species were not evident with four ($parH_{9–12}$) or two ($parH_{11–12}$) 5′-ACTC-3′ boxes (data not shown). Indeed, the ParG-only complexes disappeared at the highest ParF concentrations used in these experiments indicating that the complexes are sufficiently unstable that ParF can strip weakly bound ParG from the DNA via protein–protein interactions (30).

The N-terminal tail of ParG is required for segrosome assembly

The ParFGH segrosome was detectable in gel retardation assays when increasing concentrations of ParF were incubated with ParG and a *parH* fragment (Figure 9A, left). To examine whether segrosome assembly necessitated the N-terminal mobile tails of ParG, deletion versions of the protein were employed in retardation assays with ParF and the *parH* site. A biotinylated fragment (25 nM) bearing the *parH* locus was incubated with 0.5 μM of Δ9ParG, Δ19ParG or Δ30ParG (Figure 7A) and titrated with ParF (0.2–20 μM). Segrosome formation was not detectable when Δ9ParG (Figure 9A, right), Δ19ParG or Δ30ParG were present.

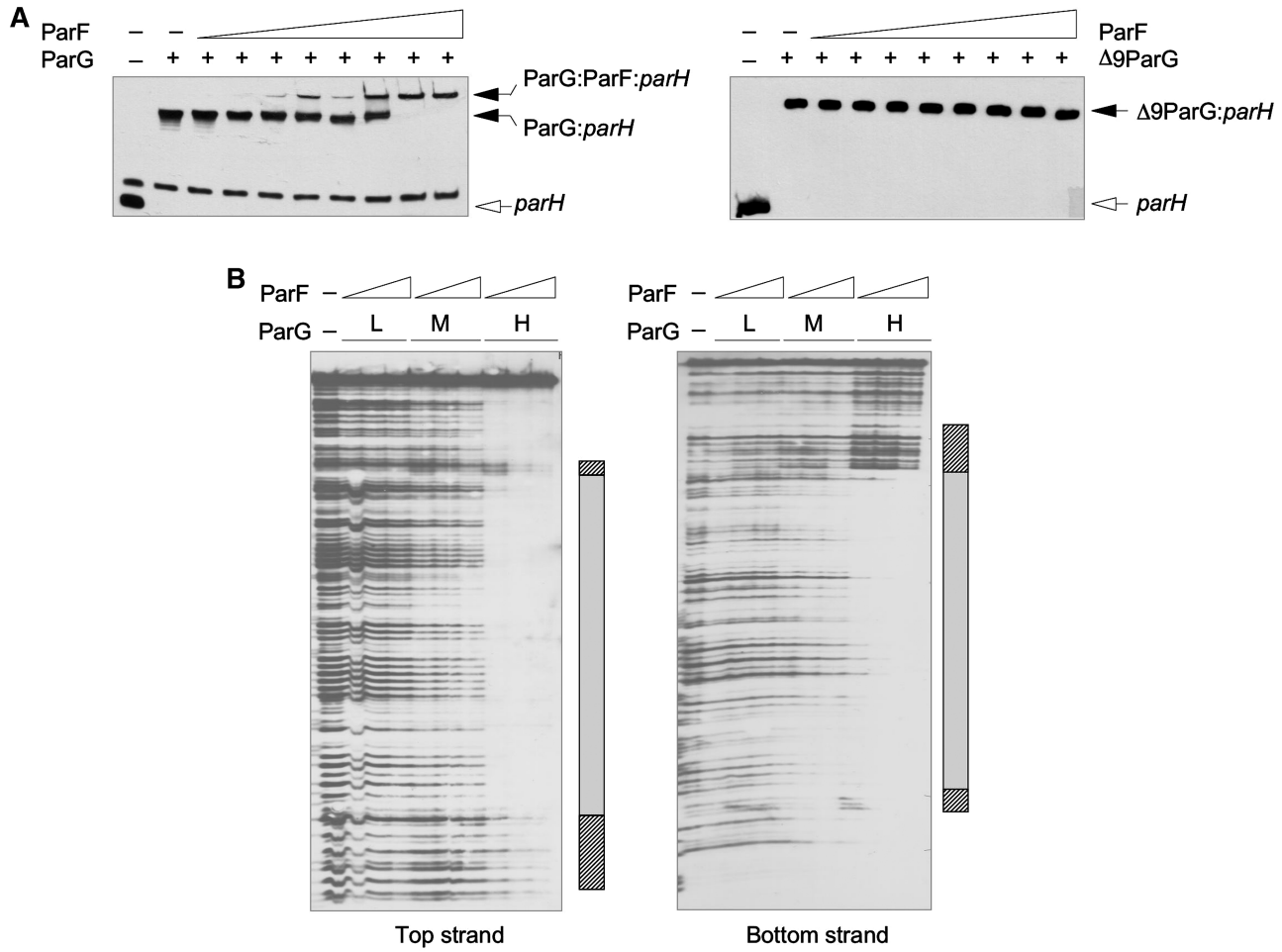

Figure 9. Assembly of ParF into the segrosome. (**A**) Supershifting of the ParG:*parH* complex in the presence of ParF. A 320-bp biotinylated PCR product containing *parH* was incubated simultaneously with ParG (0.5 μM, left) or Δ9ParG (0.5 μM, right) and increasing concentrations of ParF and analysed by gel retardation assays. ParF concentrations (μM monomer, left to right): 0.2, 0.5, 1.0, 2.0, 4.0, 8.0, 16 and 20. Open and filled arrows indicate unbound DNA and protein–DNA complexes, respectively. The band above free *parH* in the left panel is likely to be the same fragment with an atypical secondary structure that is commonly observed in DNA preparations of the centromere and O_F locus (26,30). (**B**) Influence of ParF on the DNase I protection pattern by ParG at the *parH* centromere. Footprinting reactions were performed as described in the 'Materials and Methods' section using a 244-bp PCR fragment encompassing the *parH*-O_F region. Reactions contained ParG at low (L; 0.1 μM), medium (M; 0.2 μM) or high (H; 0.5 μM) concentrations, and ParF at 0, 4 or 20 μM (left–right in each case). Reactions were performed on both top and bottom DNA strands. The major zones of ParG protection are indicated by the shaded rectangles, and the regions of altered pattern observed in the presence of ParF are denoted by the adjoining hatched boxes.

Thus, removal of the flexible tail abolishes segrosome assembly in gel retardation assays.

DISCUSSION

Plasmid centromeres are diverse. They typically comprise tandem repeats whose numbers, lengths, sequences, orientations and spacing are highly variable from replicon-to-replicon (4,5). Although characteristically positioned either upstream or downstream of the corresponding segregation genes, atypical centromeres that are situated more remotely have been described (47,48). Moreover, there are examples in which more than one plasmid locus functions as an independent centromere, or in which sites may act synergistically (49,50), suggesting that cross-talk between segrosomes assembled at discrete positions may contribute to partitioning in certain cases. Among the most well-studied examples, the ParB protein uses separate domains to associate with two different types of motif in the *parS* centromere of the P1 plasmid, as well as to bridge sites located on different molecules. ParB binding is facilitated by the DNA bending protein IHF, producing nucleoprotein complexes with distinctive topologies (9,51). In the case of the R1 plasmid, the interaction of the cognate CBF with the *parC* centromere generates ring-like, superhelical complexes (10–12). These characteristic protein–DNA superstructures are recognized in turn by the polymerizing ATPase (52) to form the mature segrosome. The *parH* centromere of plasmid TP228 was proposed to comprise a short set of ~20-bp iterated sequences based on bioinformatics analysis (30). However, in light of recent work that defined the core binding site of the ParG protein as a

5′-ACTC-3′ box flanked by AT-rich sequences (26), we have reappraised the organization of the *parH* locus here (Figure 1). The region 5′ of the *parFG* genes harbours 20 degenerate 5′-ACTC-3′ motifs separated by AT-rich spacers. The complete array of tetramer boxes, a subset of eight motifs that comprise the O_F operator (26), and the 12 distal boxes (*parH*) each were proficient centromeres, although the independent activity of O_F was modestly weaker than that of *parH* alone. Remarkably, synthetic *parH* sites in which multiple tetramer boxes were inverted relative to their normal orientation also were effective centromeres, albeit none were as competent as the wild-type locus. This observation parallels previous findings that inversion of tetramer boxes did not diminish ParG binding at the O_F locus (26). ParG dimers may be able to load onto the native tetramer boxes in the artificial *parH* sites and subsequently tether additional dimers non-specifically to neighbouring, mutated boxes through protein–protein interactions. This may produce a complex that mimics the wild-type configuration sufficiently that the ParF factor can still be recruited to assemble a functional segrosome. Thus, segregation may require only that a sufficient number of ParG dimers assemble on the centromere, permitting ParF to interact effectively with ParG displayed on a variety of promiscuous tetramer box arrangements. The absence of intrinsic curvature in *parH* fits with this concept as it reveals that the centromere is not locked into a single topology. Accordingly, insertion of a half-helical turn at the centre of the site was more well-tolerated than insertion of a full turn indicating that distance constraints between the halves of the site may be more crucial than their spatial positioning relative to helical phase. Thus, the site may be a tractable element that potentially can be flexed, if required, when complexed with ParG and ParF in the segrosome. Analogously, the ParB protein of the P1 plasmid can interact with multiple binding motifs in the *parS* centromere to generate a variety of nucleoprotein configurations (51). How the proposed plasticity of the ParG-*parH* interaction integrates with loading of ParF and the subsequent segregation process that requires directional plasmid movement remains to be resolved.

The antiparallel β-strands within RHH factors such as ParG comprise their principal DNA binding determinants, directing the protein dimers to the major groove (53). However, the binding of ParG at *parH* also is influenced profoundly by the N-terminal flexible tail (~30 amino acids): derivatives with increasing truncations of the ParG tail shifted the centromere progessively more efficiently in gel retardation assays, as well as more readily protecting the full-length site from DNase I digestion (Figure 7). Nevertheless, subsites of *parH* are bound equally well by ParG and its deletion derivatives, or are bound less proficiently by the latter in certain instances (Figure 8A). Moreover, a version of ParG that entirely lacks the mobile tail competes less well for centromeric DNA when challenged with the full-length protein. Previous chemical shift mapping of ParG-O_F interactions highlighted the improved discrimination between the operator and non-specific DNA that the ParG mobile tail confers (25). Correspondingly, deletion of the tail increases non-specific DNA binding by ParG (Figure 8C). Bases in the 5′-ACTC-3′ boxes in *parH* are likely contacted by residues in the anti-parallel β-strands of ParG, whereas the mobile tail may provide phosphate backbone contacts that enhance the strength of the interaction. In addition, a transient β-strand previously detected in the mobile tail (25) may be implicated in enhancing the specificity of the interaction with centromeric DNA by a mechanism that has yet to be fully revealed. It is noteworthy that other RHH factors similarly possess disordered extensions that modulate their DNA binding properties (54–58) indicating that this may be a widespread feature among this class of proteins. The weaker interaction of ParG deletion derivatives with certain *parH* subsites also may reflect compromised dimer–dimer interactions in these shorter substrates compared to the full-length sequence on which an array of dimers can cooperatively assemble.

Direct readout of sequences by DNA binding proteins involves interaction with specific bases in the binding site. In contrast, indirect sequence readout entails modulation of protein–DNA complex formation by bases that are not specifically contacted by the protein. These non-contacted bases may contribute, for example, to distortion of the DNA thereby promoting stable nucleoprotein complex formation by assisting in the precise alignment of amino acid residues and bases contacted during direct readout (59). We speculate that the 5′-ACTC-3′ boxes in *parH* are read out directly by ParG and that the intervening AT-rich spacers participate indirectly and that these combined contributions explain the dissimilar binding patterns of ParG to different *parH* subsites. For example, the $parH_{1–2}$ and $parH_{7–8}$ subsites both possess identical pairs of 5′-ACTC-3′ boxes, but different spacer and flanking sequences whose indirect read out by ParG elicits different affinities for the subsites in gel retardation assays (Figure 6). Although the *parH* site does not exhibit global curvature (Figure 4), the AT-rich flanking regions may display localized flexibility that promotes the association of ParG dimers assembled on adjacent tetramer boxes. Accordingly, increasing the AT content of the spacer in a natural binding site ameliorates the interaction with ParG, whereas this interaction is disimproved by other mutations (26). Thus, direct and indirect readout, combined with cross-talk between protein dimers loaded onto neighbouring tetramer boxes, may cooperatively promote coating of the *parH* site with an array of interacting ParG dimers. The flexible tails of ParG provide another layer of complexity to the centromere interaction: the tails modulate the DNA binding affinity of ParG. This modulation may be via a direct interaction with the the *parH* site, or by cooperative ParG dimer–dimer interactions mediated through the N-terminal tails. Elucidation of the tertiary structures of ParG complexed with the centromere will clarify further the nucleoprotein interactions that underpin segrosome assembly.

Like ParG, the ParR and ω proteins encoded by plasmids R1/pSK41 and pSM19035, respectively, are RHH factors that bind their cognate centromeres.

However, the sequences of the ribbon motifs that are involved in DNA recognition differ between the various proteins, and the DNA repeats that the proteins contact also are dissimilar (10,12,21,26,52). ParR assembles on DNA as a dimer-of-dimers, each dimer binding to one half-site of a 20-bp repeat that occurs four times in the pSK41 *parC* centromere (10). The ω protein recognizes multiple sets of contiguous heptad repeats that are arranged in different numbers and combinations of direct and indirect orientation on pSM19035: each set of repeats displays centromere-like properties (50,60). Like the ParR-*parC* and ParG-*parH* interactions, ω cooperatively binds a pair of contiguous DNA motifs as a dimer-of-dimers. Strikingly, both ParR and ω assemble into nucleoprotein superstructures that enwrap the centromeric DNA and which may be required for recruitment of the polymerizing NTPase (10,12). Further analysis is needed to determine whether the extensive repeats that comprise the *parH* site (Figures 1 and 2) also act as a scaffold for assembly of a nucleoprotein superstructure onto which ParF docks.

The ParF component alone does not interact directly with the *parH* site either in the presence or absence of ATP. Instead the ATPase assembles into the segrosome via interactions with ParG to generate the mature complex. Both the RHH and flexible domains of ParF contact ParG, the mobile regions being required to stimulate ATP hydrolysis by ParF as well as promoting ParF polymerization (18). The unstructured tails are also essential for recruitment of ParF to the segrosome (Figure 9A, right). Assembly of the intact complex with full-length ParG was detectable as a higher order, supershifted species that failed to enter the polyacrylamide matrix in gel retardation assays (Figure 9A, left). DNase I footprinting patterns at the edges of the *parH* site were altered modestly by the addition of ParF into reactions that contained ParG, suggestive of perturbations of the centromere boundaries when the tripartite complex is assembled. These perturbations may parallel observations with other segrosomes in which the centromere can be extensively distorted or wrapped about the CBF in higher order nucleoprotein superstructures (10–12). A plethora of questions remain unanswered about the functioning of centromeric DNA in tripartite segregation complexes that also involve a CBF and a ParA-type ATPase. What is the evolutionary force that drives the extensive sequence and organization diversity among plasmid centromeres? Does this diversity result in profound variations in the architecture of segrosomes assembled on different episomes, or are there unifying themes in the organization of these complexes? An evolutionarily conserved segregation complex seems to exist in eucaryotes, although it is modulated between different organisms. Does the same pertain among procaryotic segrosome structures? How are homologous ParA proteins captured by apparently dissimilar CBF–centromere complexes? The solutions to these and other conundrums will illuminate further how centromere sequences underpin the segregation process.

ACKNOWLEDGEMENTS

We thank Daniela Barillà and Marisa Caccamo for their critical reading of the manuscript. AFM studies were performed at the University of Twente, The Netherlands.

FUNDING

Biotechnology and Biological Sciences Research Council (grant G003114 to F.H.). Overseas Research Student Awards Scheme, scholarship (to M.W., partial). Funding for open access charge: Biotechnology and Biological Sciences Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Bloom,K. and Joglekar,A. (2010) Towards building a chromosome segregation machine. *Nature*, **463**, 446–456.
2. Hayes,F. and Barillà,D. (2010) Extrachromosomal components of the nucleoid: recent developments in deciphering the molecular basis of plasmid segregation. In Dorman,C.J. and Dame,R.T. (eds), *Bacterial Chromatin*. Springer Publishing, Dordrecht, The Netherlands, pp. 49–70.
3. Schumacher,M.A. (2008) Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochem. J.*, **412**, 1–18.
4. Hayes,F. and Barillà,D. (2006) The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nature Rev. Microbiol.*, **4**, 133–143.
5. Hayes,F. and Barillà,D. (2006) Assembling the bacterial segrosome. *Trends Biochem. Sci.*, **31**, 247–250.
6. Schumacher,M.A. (2007) Structural biology of plasmid segregation proteins. *Curr. Opin. Struct. Biol.*, **17**, 103–109.
7. Salje,J. (2010) Plasmid segregation: how to survive as an extra piece of DNA. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 296–317.
8. Gerdes,K., Howard,M. and Szardenings,F. (2010) Pushing and pulling in prokaryotic DNA segregation. *Cell*, **141**, 927–942.
9. Schumacher,M.A. and Funnell,B.E. (2005) Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature*, **438**, 516–519.
10. Schumacher,M.A., Glover,T.C., Brzoska,A.J., Jensen,S.O., Dunham,T.D., Skurray,R.A. and Firth,N. (2007) Segrosome structure revealed by a complex of ParR with centromere DNA. *Nature*, **450**, 1268–1271.
11. Hoischen,C., Bussiek,M., Langowski,J. and Diekmann,S. (2008) *Escherichia coli* low-copy-number plasmid R1 centromere *parC* forms a U-shaped complex with its binding protein ParR. *Nucleic Acids Res.*, **36**, 607–615.
12. Moller-Jensen,J., Ringgaard,S., Mercogliano,C.P., Gerdes,K. and Löwe,J. (2007) Structural analysis of the ParR/*parC* plasmid partition complex. *EMBO J.*, **26**, 4413–4422.
13. Garner,E.C., Campbell,C.S., Weibel,D.B. and Mullins,R.D. (2007) Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science*, **315**, 1270–1274.
14. Ebersbach,G. and Gerdes,K. (2004) Bacterial mitosis: partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell. *Mol. Microbiol.*, **52**, 385–398.
15. Barillà,D., Rosenberg,M.F., Nobbmann,U. and Hayes,F. (2005) Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. *EMBO J.*, **24**, 1453–1464.
16. Lim,G.E., Derman,A.I. and Pogliano,J. (2005) Bacterial DNA segregation by dynamic SopA polymers. *Proc. Natl Acad. Sci. USA*, **102**, 17658–17663.
17. Adachi,S., Hori,K. and Hiraga,S. (2006) Subcellular positioning of F plasmid mediated by dynamic localization of SopA and SopB. *J. Mol. Biol.*, **356**, 850–863.
18. Barillà,D., Carmelo,E. and Hayes,F. (2007) The tail of the ParG DNA segregation protein remodels ParF polymers and enhances

ATP hydrolysis via an arginine finger-like motif. *Proc. Natl Acad. Sci. USA*, **104**, 1811–1816.
19. Machón,C., Fothergill,T.J.G., Barillà,D. and Hayes,F. (2007) Promiscuous stimulation of ParF protein polymerization by heterogeneous centromere binding factors. *J. Mol. Biol.*, **374**, 1–8.
20. Bouet,J.Y., Ah-Seng,Y., Benmeradi,N. and Lane,D. (2007) Polymerization of SopA partition ATPase: regulation by DNA binding and SopB. *Mol. Microbiol.*, **63**, 468–481.
21. Pratto,F., Cicek,A., Weihofen,W.A., Lurz,R., Saenger,W. and Alonso,J.C. (2008) *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on *parS* DNA during plasmid segregation. *Nucleic Acids Res.*, **36**, 3676–3689.
22. Fogel,M.A. and Waldor,M.K. (2006) A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev.*, **20**, 3269–3282.
23. Ringaard,S., van Zon,J., Howard,M. and Gerdes,K. (2009) Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl Acad. Sci. USA*, **106**, 19369–19374.
24. Golovanov,A.P., Barillà,D., Golovanova,M., Hayes,F. and Lian,L.Y. (2003) ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. *Mol. Microbiol.*, **50**, 1141–1153.
25. Carmelo,E., Barillà,D., Golovanov,A.P., Lian,L.Y., Derome,A. and Hayes,F. (2005) The unstructured N-terminal tail of ParG modulates assembly of a quaternary nucleoprotein complex in transcription repression. *J. Biol. Chem.*, **280**, 28683–28691.
26. Zampini,M., Derome,A., Bailey,S.E.S., Barillà,D. and Hayes,F. (2009) Recruitment of the ParG segregation protein to different affinity DNA sites. *J. Bacteriol.*, **191**, 3832–3841.
27. Fothergill,T.J.G., Barillà,D. and Hayes,F. (2005) Protein diversity confers specificity in plasmid segregation. *J. Bacteriol.*, **187**, 2651–2661.
28. Woodcock,D.M., Crowther,P.J., Doherty,J., Jefferson,S., DeCruz,E., Noyer-Weidner,M., Smith,S.S., Michael,M.Z. and Graham,M.W. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.*, **17**, 3469–3478.
29. Ludtke,D.N., Eichorn,B.G. and Austin,S.J. (1989) Plasmid-partition functions of the P7 prophage. *J. Mol. Biol.*, **209**, 393–406.
30. Barillà,D. and Hayes,F. (2003) Architecture of the ParF-ParG protein complex involved in procaryotic DNA segregation. *Mol. Microbiol.*, **49**, 487–499.
31. Hayes,F. (1998) A family of stability determinants in pathogenic bacteria. *J. Bacteriol.*, **180**, 6415–6418.
32. Martin,K.A., Friedman,S.A. and Austin,S.J. (1987) Partition site of the P1 plasmid. *Proc. Natl Acad. Sci. USA*, **84**, 8544–8547.
33. Hayes,F. (2000) The partition system of multidrug resistance plasmid TP228 includes a novel protein that epitomizes an evolutionarily-distinct subgroup of the ParA superfamily. *Mol. Microbiol.*, **37**, 528–541.
34. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121–4130.
35. Bussiek,M., Mucke,N. and Langowski,J. (2003) Polylysine-coated mica can be used to observe systematic changes in the supercoiled DNA conformation by scanning force microscopy in solution. *Nucleic Acids Res.*, **32**, e137.
36. Van der Werf,K.O., Putman,C.A.J., De Grooth,B.G., Segerink,F.B., Schipper,E.H., van Hulst,N.F. and Greve,J. (1993) Compact stand-alone atomic force microscope. *Rev. Sci. Instrument*, **64**, 2892–2897.
37. Hoischen,C., Bolshoy,A., Gerdes,K. and Diekmann,S. (2004) Centromere *parC* of plasmid R1 is curved. *Nucleic Acids Res.*, **32**, 5907–5915.
38. De Mey,M., Maertens,J., Lequeux,G.J., Soetaert,W.K. and Vandamme,E.J. (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol.*, **7**, 34.
39. Rivetti,C., Guthold,M. and Bustamante,C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, **264**, 919–932.
40. Derome,A., Hoischen,C., Bussiek,M., Grady,R., Adamczyk,M., Kedzierska,B., Diekmann,S., Barillà,D. and Hayes,F. (2008) Centromere anatomy in the multidrug resistant pathogen *Enterococcus faecium*. *Proc. Natl Acad. Sci. USA*, **105**, 2151–2156.
41. Funnell,B.E. (1991) The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. *J. Biol. Chem.*, **266**, 14328–14337.
42. Hayes,F. and Austin,S. (1994) Topological scanning of the P1 plasmid partition site. *J. Mol. Biol.*, **243**, 190–198.
43. Diekmann,S. and Wang,J.C. (1985) On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. *J. Mol. Biol.*, **186**, 1–11.
44. Diekmann,S. (1987) Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. *Nucleic Acids Res.*, **15**, 247–265.
45. Crothers,D.M., Haran,T.E. and Nadeau,J.G. (1990) Intrinsically bent DNA. *J. Biol. Chem.*, **265**, 7093–7096.
46. Bechert,T., Heck,S., Fleig,U., Diekmann,S. and Hegemann,J.H. (1999) All 16 centromere DNAs from *Saccharomyces cerevisiae* show DNA curvature. *Nucleic Acids Res.*, **27**, 1444–1449.
47. Williams,D.R., Macartney,D.P. and Thomas,C.M. (1998) The partitioning activity of the RK2 central control region requires only *incC, korB* and KorB-binding site O_B3 but other KorB-binding sites form destabilizing complexes in the absence of O_B3. *Microbiology*, **144**, 3369–3378.
48. Grigoriev,P.S. and Lobocka,M.B. (2001) Determinants of segregational stability of the linear plasmid-prophage N15 of *Escherichia coli*. *Mol. Microbiol.*, **42**, 355–368.
49. Ebersbach,G. and Gerdes,K. (2001) The double *par* locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. *Proc. Natl Acad. Sci. USA*, **98**, 15078–15083.
50. Dmowski,M., Sitkiewicz,I. and Ceglowski,P. (2006) Characterization of a novel partition system encoded by the δ and ω genes from the streptococcal plasmid pSM19035. *J. Bacteriol.*, **188**, 4362–4372.
51. Vecchiarelli,A.G., Schumacher,M.A. and Funnell,B.E. (2007) P1 partition complex assembly involves several modes of protein-DNA recognition. *J. Biol. Chem.*, **282**, 10944–10952.
52. Pratto,F., Suzuki,Y., Takeyasu,K. and Alonso,J.C. (2009) Single-molecule analysis of protein-DNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. *J. Biol. Chem.*, **284**, 30298–30306.
53. Schreiter,E.R. and Drennan,C.L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nature Rev. Microbiol.*, **5**, 710–720.
54. Somers,W.S. and Phillips,S.E. (1992) Crystal structure of the Met repressor-operator complex at 2.8Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.
55. Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature*, **367**, 754–757.
56. Phillips,K. and Phillips,S.E.V. (1994) Electrostatic activation of *Escherichia coli* methionine repressor. *Structure*, **2**, 309–316.
57. Benanti,E.L. and Chivers,P.T. (2007) The N-terminal arm of the *Helicobacter pylori* Ni^{2+}-dependent transcription factor NikR is required for specific DNA binding. *J. Biol. Chem.*, **282**, 20365–20375.
58. Ni,L., Jensen,S.O., Tonthat,N.K., Berg,T., Kwong,S.M., Guan,F.H., Brown,M.H., Skurray,R.A., Firth,N. and Schumacher,M.A. (2009) The *Staphylococcus aureus* pSK41 plasmid-encoded ArtA protein is a master regulator of plasmid transmission genes and contains a RHH motif used in alternate DNA-binding modes. *Nucleic Acids Res.*, **37**, 6970–6983.
59. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 143–177.
60. de la Hoz,A.B., Pratto,F., Misselwitz,R., Speck,C., Weihofen,W., Welfle,K., Saenger,W., Welfle,H. and Alonso,J.C. (2004) Recognition of DNA by ω protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: analysis of binding to operator DNA with one to four heptad repeats. *Nucleic Acids Res.*, **32**, 3136–3147.