Published online 9 April 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 14* ***6673–6682***
doi:10.1093/nar/gks314

Combinatorial targeting of ribbon–helix–helix artificial transcription factors to chimeric recognition sites

Massimiliano Zampini and Finbarr Hayes*

Faculty of Life Sciences and Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

Received November 4, 2011; Revised March 22, 2012; Accepted March 26, 2012

ABSTRACT

Artificial transcription factors (ATFs) are potent synthetic biology tools for modulating endogenous gene expression and precision genome editing. The ribbon–helix–helix (RHH) superfamily of transcription factors are widespread in bacteria and archaea. The principal DNA binding determinant in this family comprises a two-stranded antiparallel β-sheet (ribbons) in which a pair of eight-residue motifs insert into the major groove. Here, we demonstrate that ribbons of divergent RHH proteins are compact and portable elements that can be grafted into a common α-helical scaffold producing active ATFs. Hybrid proteins cooperatively recognize DNA sites possessing core tetramer boxes whose functional spacing is dictated by interactions between the α-helical backbones. These interactions also promote combinatorial binding of chimeras with different transplanted ribbons, but identical backbones, to synthetic sites bearing cognate boxes for each protein either *in vitro* or *in vivo*. The composite assembly of interacting hybrid proteins offers potential advantages associated with combinatorial approaches to DNA recognition compared with ATFs that involve binding of a single protein. Moreover, the new class of RHH ATFs may be utilized to re-engineer transcriptional circuits, or may be enhanced with affinity tags, fluorescent moieties or other elements for targeted genome marking and manipulation in bacteria and archaea.

INTRODUCTION

The activities of sequence-specific DNA binding proteins are vital to a multitude of cellular processes from transcriptional regulation to genome segregation, and from the movement of transposable elements to site-specific recombination (1). In particular, transcription factors that bind DNA are fundamental for controlling and tuning expression of single genes or multigene regulons in response to diverse environmental and metabolic cues. Transcription factors are equipped with a variety of binding domains that interact with DNA, most frequently with the major groove via α-helical structural elements (2).

There is now considerable interest in designing artificial transcription factors (ATFs) for rational rewiring of gene circuits in synthetic biology, for precision targeting and editing of novel genomic addresses, and for developing regulatable expression systems that control gene clusters introduced into heterologous hosts for metabolic engineering (3–5). Custom ATFs based on Cys_2–His_2–type zinc finger DNA binding proteins have emerged as specially promising candidates with far-reaching and diverse applications (6). Notably, zinc finger proteins fused to nuclease domains can be targeted to pre-defined genomic locations allowing for precise chromosome manipulation in a range of eucaryotes (7). Another class of ATF, transcription activator-like effectors, have been designed for specific binding and transcriptional regulation of endogenous genes (8,9). The rational design of miniature proteins that bind DNA *in vitro* also has met with significant, albeit limited, success (10,11). However, fabrication of miniature protein libraries that specifically recognize arrays of discrete DNA motifs has not yet been described.

Ribbon–helix–helix (RHH) transcription factors are found widely in bacteria and archaea. Although the primary sequences of RHH proteins are diverse, they possess a common homodimeric structure that comprises a pair of antiparallel β-strands (ribbons) framed within an α-helical scaffold (12) (Figure 1A). Typically, numerous RHH dimers assemble cooperatively on an array of regularly spaced DNA sites to exert transcriptional repression of a pertinent gene(s). Binding necessitates the assembly of at least two dimers on adjacent sites via protein–protein interactions mediated by the α-helical backbones. The ribbon is the principal binding determinant, inserting into the major groove of its site, the sequence of which is specific for each protein. The ribbon is a remarkably

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

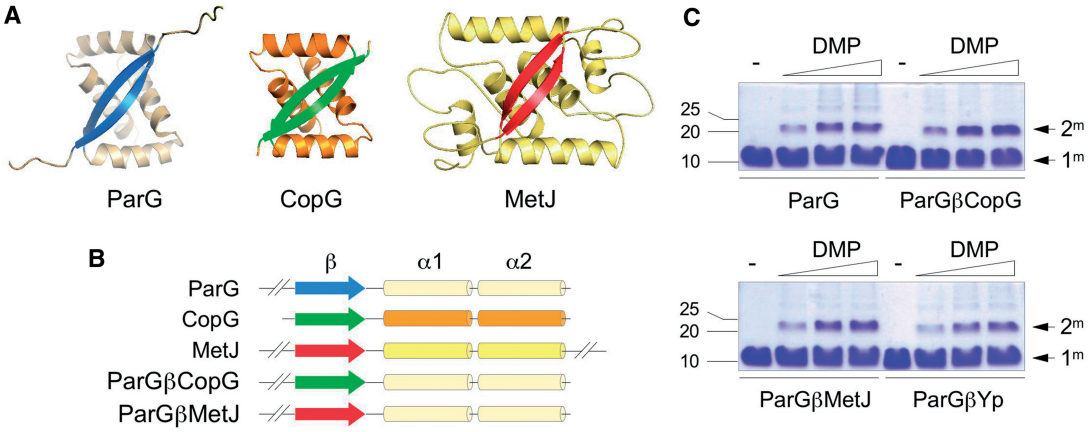

Figure 1. Organization of chimeric ParG:RHH proteins. (**A**) Tertiary structures of ParG (1P94.pdb), MetJ (1CMA.pdb) and CopG (2CPG.pdb) dimers. α-helices are shown in shades of orange and β-strands in blue (ParG), green (CopG) and red (MetJ). (**B**) Linear representations of ParG, CopG, MetJ and chimeras ParGβCopG and ParGβMetJ that comprise ribbons of CopG and MetJ, respectively, transplanted into the ParG α-helical scaffold. (**C**) Chemical cross-linking of ParG, ParGβCopG, ParGβMetJ and ParGβYp with dimethyl pimelimidate (0, 1, 5 and 10 mM, left to right) at 37°C for 60 min followed by SDS–PAGE. Monomeric (1^m) and dimeric (2^m) species, and positions of molecular weight markers (kDa) are shown.

compact binding element as the paired β-strands each comprise only eight identical amino acids. Here, we demonstrate that ribbons with diverse sequences can be exchanged between different α-helical backbones to generate a new class of ATF that can be directed to synthetic DNA sites *in vitro*. Moreover, pairs of RHH ATFs that possess discrete ribbon motifs can be cooperatively targeted via their identical α-helical scaffolds to chimeric sites that possess core motifs for both proteins *in vitro* and *in vivo*. Members of this new type of ATF potentially may be deployed in innovative combinations to engineer novel regulatory circuits, or may be augmented with affinity tags, fluorescent moieties, or other modules that will permit targeted genome identification and manipulation in bacteria and archaea for which ATFs are scarce.

MATERIALS AND METHODS

Bacterial strains and media

Escherichia coli DH5α was used for plasmid construction, and BL21(DE3) was used for protein overproduction. Strains were grown in Luria–Bertani medium with ampicillin (200 μg/ml) and/or spectinomycin (250 μg/ml) as appropriate for plasmid maintenance.

Gene synthesis and expression, and protein purification

RHH hybrids were generated by overlap extension PCR (13) using Phusion High-Fidelity DNA Polymerase (New England Biolabs) generating blunt-end amplicons. Briefly, hybrids were generated by splicing two amplicons synthesized using *parG* as template. The first amplicon (PCR-1) contained the ribbon coding sequence of interest at its 3′-end whereas the second amplicon (PCR-2) possessed the same ribbon sequence but at its 5′-end. PCR-3 was performed using equimolar concentrations of purified PCR-1 and PCR-2 as template without primers for the first 10 cycles, followed by an additional 26 cycles in which external primers were added. PCR-3 products containing the spliced recombinant sequence were cloned in pET22b(+) after digestion with NdeI and XhoI. Genes cloned in pET22b(+) also were amplified and subcloned between PfoI and NotI or NdeI and PacI sites in pCDFDuet-1. The gene for $ParG_{Yp}$ was produced using a set of overlapping oligonucleotides that were 5′-phosphorylated, annealed, ligated *in vitro* and cloned in pET22b(+) between NdeI and XhoI sites. The sequences of all cloned inserts were verified. His-tagged proteins were purified as described previously (14).

Gel retardation assays

DNA fragments for gel retardation assays were generated by annealing complementary primers one of which was 5′-biotinylated, or were prepared by second strand synthesis of an unlabeled oligonucleotide using a biotinylated primer. Purification of the fragments and conditions for retardation assays were outlined in detail previously (15). 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 NaCl, 5 mM $MgCl_2$, 0.005–0.05 mg/ml poly(dI-dC)] with the ParG concentrations shown in figure legends. Reaction mixtures were electrophoresed on 7–10% polyacrylamide gels in 0.5 × Tris–borate-EDTA buffer, pH 8.3, for 50–70 min at 90 V at 22°C. DNA was transferred by capillary action or electroblotting to positively charged nylon membranes (Roche), and the transferred DNA fragments were immobilized by UV cross-linking. Biotin end-labeled DNA was detected using chemiluminescent nucleic acid detection reagents (Thermo Scientific Pierce) (14).

DNaseI footprinting and surface plasmon resonance

Preparation and purification of biotinylated PCR products, conditions for DNase I footprinting reactions, denaturing gel electrophoresis, and detection of biotinylated DNA followed procedures described in detail recently (15). Surface plasmon resonance was performed with a Biacore 3000 instrument (Biacore AB). A Streptavidin chip was

primed with immobilization buffer (10 mM Tris–HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA) and then washed with three consecutive 1-min injections (50 µl/min) of activation buffer (1 M NaCl, 50 mM NaOH). A 24-bp biotinylated double-stranded oligonucleotide bearing $copG_L parG_R$ and a second fragment of identical length but unrelated sequence (5′-CCGACGAAGGAGGTGGCTGGTTGG-3′) were prepared at concentrations of 4 nM in immobilization buffer. The fragments were injected at 10 µl/min through separate flow cells until a value of ~160 resonance units was detected. Running buffer [200 mM NaCl, 10 mM Tris–HCl pH 7.5, 1.25 ng/µl poly(dI-dC), 0.05% Tween-20, 1 mM DTT] was used both to prime the chip and to prepare protein dilutions at the desired working concentration (1.6 µM). Binding reactions were performed at 23°C in this buffer by injecting 200 µl protein into each flow cell at 50 µl/min. Binding curves were produced by subtracting the signal corresponding to the flow cell with the unrelated DNA sequence from the signal generated by the test sample. Data were analyzed using BIAevaluation software 3.1.

β-Galactosidase assays

124-bp fragments bearing PO_{H1} or PO_{H2} were cloned in the BamHI site upstream of a promoterless *lac* operon in pRS415 (16) producing transcriptional fusions in which the operon is expressed from the cloned promoter. Strain BL21 harboring the pRS415 derivative was cotransformed with pCDFDuet-1 plasmids encoding the test protein(s). β-galactosidase assays were performed with cells permeabilized with chloroform and SDS (17). Values shown are averages of at least three experiments performed in triplicate.

Kanamycin resistance assays

A promoterless *aph* gene derived from pET30b(+) was PCR amplified and inserted in place of the *lac* operon downstream of PO_{H2} in pRS415. Strain DH5α bearing this plasmid and empty pCDFDuet-1 or pCDFDuet-1 encoding the test protein(s) was grown in ampicillin and spectinomycin for plasmid maintenance to OD_{600} ~0.6. Cultures were reinoculated (2%) in fresh broth containing both antibiotics, as well as with and without kanamycin (1000 µg/ml). Growth was monitored further at OD_{600}. The data shown are representative of experiments performed in triplicate.

RESULTS

Transplanting ribbon DNA recognition determinants into a common α-helical backbone

ParG is both a RHH centromere-binding protein and transcriptional repressor, and also modulates dynamics of the partner ParF segregation protein (14,15,18–23) (Figure 1). The antiparallel β-strands in ParG comprise symmetrical KRVNVNFD motifs (18), whereas the ribbon that contacts DNA in the CopG RHH protein has the sequence KRLTITLS (differences from ParG underlined) (24). The CopG ribbon was substituted in place of the native ribbon in ParG to produce hybrid protein ParGβCopG (Figure 1B and Supplementary Figure S1). This chimera cross-linked into a dimeric species as efficiently as ParG (14) indicating that the hybrid maintains tertiary structure (Figure 1C). The native ParG ribbon recognizes multiple degenerate 5′-ACTC-3′ boxes in both the centromere and O_F operator located upstream of the *parF–parG* cassette, single dimers multimerizing cooperatively thereby coating the two sites. The 5′-ACTC-3′ boxes are separated invariably by 4-bp AT-rich spacers (15,23). Instead, CopG makes multiple contacts in its co-crystal structure with a 13-bp pseudosymmetric element that spans two 5′-TGCA-3′ motifs (24,25). The ParGβCopG hybrid did not bind a ParG site in which a pair of 5′-ACTC-3′ boxes were substituted with 5′-TGCA-3′ motifs for CopG. CopG makes additional base contacts outside of the 5′-TGCA-3′ motifs (24,25) and 5′-GCA$^{C}/_{T}$-3′ tetramers overlap these motifs in certain *copG* sites. Indeed, the co-crystal structure of CopG with DNA (PDB 1b01) reveals that the core base specific recognition contacts span a 5′-GCAC-3′ sequence element (24). The *popG* synthetic locus comprises a ParG site in which the 5′-ACTC-3′ boxes were replaced by 5′-GCA$^{C}/_{T}$-3′ boxes. This site was recognized very efficiently by ParGβCopG in gel retardation assays with an apparent K_d ~1.2 µM (Figure 2A). Binding was specific and cooperative as ParGβCopG did not dock onto sites that possessed a single 5′-GCA$^{C}/_{T}$-3′ box, for example (see Figure 4B). Inserting T residues adjacent to both 5′-GCAC-3′ boxes in *popG* to restore the core 5′-TGCA-3′ motifs for CopG did not enhance binding by ParGβCopG ($popG_1$ in Figure 2A; apparent K_d ~1.2 µM). These results establish that ParGβCopG is a functional chimera with a DNA binding specificity which closely overlaps that of CopG from which the ribbon in this hybrid originates.

Two further characteristics of ParGβCopG and ParG binding were compared. First, ParG recognizes natural and synthetic sites possessing various arrangements of direct and inverted 5′-ACTC-3′ motifs (15,23). Similarly, as well as the *popG* site containing directly-oriented boxes, ParGβCopG bound substrates bearing inverted 5′-GCAC-3′ boxes ($popG_2$ and $popG_3$, apparent K_d ~2.5 and 2.1 µM, respectively) (Figure 2A). Second, the 4-bp AT-rich spacers between 5′-ACTC-3′ boxes are read out indirectly by ParG (23). Artificial sites with more GC-rich spacers are bound poorly by the protein partly due to less effective interactions between dimers bound on neighboring boxes (15). Similarly, the $popG_4$ site bearing flanking regions with reduced AT content was bound only weakly by ParGβCopG with an apparent $K_d > 4$ µM (Figure 2A). Thus, ParGβCopG retains the capacity of ParG dimers to bind to directly- and inversely-oriented core tetramer motifs and, like ParG, is sensitive to the AT-richness of the bases flanking these boxes.

The ribbon of MetJ, another archetypal RHH protein (26), was selected as a second example for designing hybrid proteins. MetJ binds a tandemly repeated 8 bp palindrome, or met-box, with the optimal sequence 5′-AGAC GTCT-3′ (27). The MetJ ribbon that recognizes this box is KKITVSIP which differs at six positions (underlined)

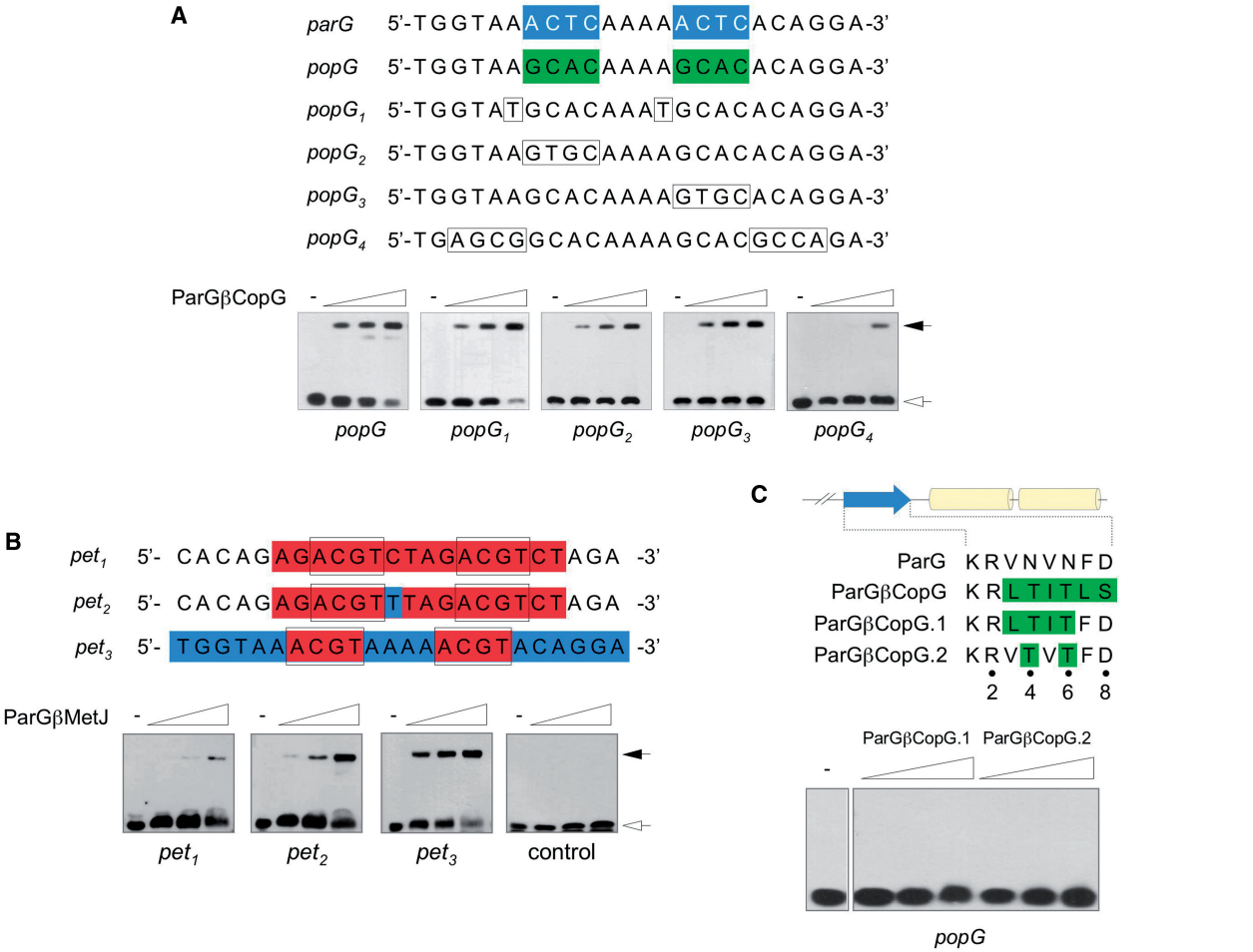

Figure 2. Docking of ParGβCopG and ParGβMetJ onto hybrid sites. (**A**) The *popG* locus comprises a site for ParG binding in which 5′-ACTC-3′ boxes were replaced by 5′-GCAC-3′ core motifs recognized by ParGβCopG. The hybrid was tested in gel retardation assays with double-stranded oligonucleotides bearing *popG*, and with variant sites in which T residues were introduced 5′ of the tetramer motifs (*popG*$_1$), left and right 5′-GCAC-3′ tetramers were inverted (*popG*$_2$ and *popG*$_3$, respectively), and AT contents of regions flanking the 5′-GCAC-3′ boxes were decreased (*popG*$_4$). Open and filled arrows mark unbound DNA and nucleoprotein complexes, respectively. (**B**) Sequences derived from sites recognized by ParG and MetJ are represented by blue and red boxes, respectively. The *pet*$_1$ site bears the sequence recognized by MetJ in the crystal structure; *pet*$_2$ is similar to *pet*$_1$ but carries a C-to-T mutation in the spacer between the two 5′-ACGT-3′ boxes; *pet*$_3$ consists of a ParG site in which core 5′-ACTC-3′ motifs recognized by the protein are replaced with 5′-ACGT-3′ sequences for ParGβMetJ binding. The control is a 24-bp double-stranded oligonucleotide with a randomized sequence. Open and filled arrows mark unbound DNA and nucleoprotein complexes, respectively. (**C**) Testing ribbon sequence requirements for ParGβCopG binding. The ParG ribbon comprises KRVNVNFD that is altered to KRLTITLS in ParGβCopG which possesses the CopG ribbon (green). Amino acids one and two are common to parental and hybrid proteins. In ParGβCopG.1, residues 3–6 correspond to those in CopG, whereas residues 4 and 6 in ParGβCopG.2 are derived from CopG. Other residues originate from ParG. ParGβCopG.1 and ParGβCopG.2 were tested in gel retardation assays with the *popG* site. Proteins in panels A–C were used at 0, 0.5, 1.0 and 1.5 μM (left to right).

compared to the ParG β-strands. The ParG ribbon was replaced with the equivalent MetJ sequence producing ParGβMetJ (Figure 1B and Supplementary Figure S1). The protein formed dimers in solution (Figure 1C) and bound weakly to a pair of tandem met-boxes with apparent $K_d > 4\,\mu M$ (*pet*$_1$ in Figure 2B). As outlined earlier, ParG binding to its native site requires a high AT content between core 5′-ACTC-3′ motifs (19). A single C-to-T transition between met-boxes in *pet*$_1$ produced *pet*$_2$ which was bound more avidly by ParGβMetJ (apparent $K_d \sim 1.6\,\mu M$). Alignment of MetJ dimers on met-boxes is centered on the 5′-ACGT-3′ sequence (27). Embedding a pair of 5′-ACGT-3′ boxes in a *parG* site (*pet*$_3$) improved binding by ParGβMetJ further still (apparent $K_d \sim 1.2\,\mu M$) (Figure 2B) demonstrating that this core motif from the met-box is sufficient for docking of this hybrid protein.

The combined data with ParGβCopG and ParGβMetJ illustrate that transplantation of diverse ribbons between distantly-related RHH proteins produces functional RHH hybrids. Second, the chimeras possess specific DNA binding activity and recognize 4-bp core motifs that are closely-related to motifs bound by the proteins from which the ribbons originate. Third, the chimeric sites with which the hybrid proteins interact retain features of sites bound by ParG—specifically 4-bp spacers that are AT-rich—that reflect interactions between the α-helical scaffolds of dimers bound at adjacent core sites.

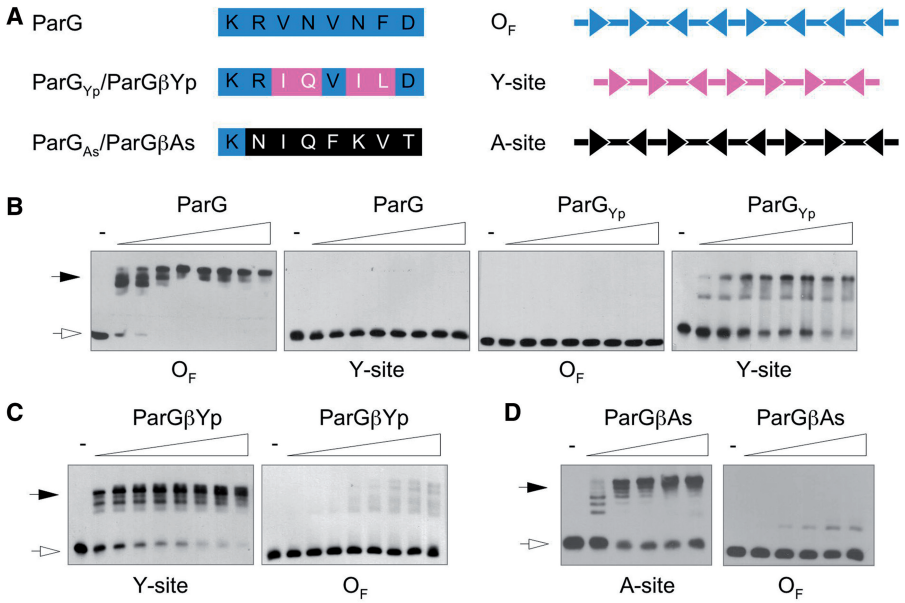

Figure 3. Design and binding of hybrids with ribbon motifs from putative RHH proteins. (**A**) Left: Sequences of the ParG β-strand and putative ribbons from homologues in *Y. pseudotuberculosis* and *A. salmonicida* (GenBank accession numbers CAQ76574 and CAQ81938, respectively). Differences from ParG are highlighted in magenta and black. These motifs were swapped into ParG to generate hybrids ParGβYp and ParGβAs. Right: Organization of 5′-ACTC-3′ boxes in the O_F site bound by ParG (15), and 5′-GA$^{C}/_{T}$A-3′ and 5′-CAT$^{T}/_{G}$-3′ motifs in the Y- and A-sites of *Y. pseudotuberculosis* and *A. salmonicida*, respectively. Arrow orientation indicates direct or inverted boxes. (**B**) Gel retardation assays with O_F, the Y-site, ParG and $ParG_{Yp}$. (**C**) Gel retardation assays with O_F, the Y-site, and ParGβYp. Assays in panels (B) and (C) used 0, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0 and 1.5 μM (left to right) of the indicated proteins. (**D**) Gel retardation assays with O_F, the A-site, and ParGβAs. Assays used 0, 0.1, 0.2, 0.3, 0.4 and 0.5 μM (left to right) ParGβAs. Open and filled arrows mark unbound DNA and nucleoprotein complexes, respectively, in panels A–D.

Engineering a functional RHH hybrid necessitates splicing an entire ribbon sequence

Hydrophilic side-chains at positions 2, 4 and 6 within the eight amino acid β-strands of RHH factors make specific nucleotide base contacts during DNA binding (12). To ascertain whether replacing specifically these residues generated an active chimera, a hybrid was designed that contained residues 4 and 6 from CopG with other residues from ParG. The lysine and arginine residues at positions 1 and 2, respectively, are common to parental and hybrid proteins. A second variant possessed residues 3–6 from CopG and residues 7 and 8 from ParG. Neither the ParGβCopG.1 or ParGβCopG.2 variants bound the *popG* site in gel retardation assays (Figure 2C). ParG, like all DNA binding proteins, binds non-specifically at high concentrations *in vitro* (18). Similarly, ParGβCopG.1 and ParGβCopG.2 both bound non-specific DNA at concentrations similar to ParG, revealing that these hybrids retain some affinity for DNA, albeit of wider or unknown specificity (Supplementary Figure S2). Grafting specifically the hydrophilic amino acids from the CopG ribbon into ParG may produce hybrids that recognize different DNA sequences compared to the chimera possessing a complete β-strand exchange. However, transplanting an entire ribbon is more likely to produce RHH hybrids with predictable binding specificity than exchanging subsets of β-strand residues.

Database mining and fabrication of hybrids from uncharacterized RHH candidates

We previously have highlighted several ParG homologues from diverse bacterial species (18,21,28). In view of the continued expansion in genome sequencing, reiterative BLAST searching of the NCBI database was used here to identify further ParG homologues. Redundant entries with identical sequences were removed, thereby revealing > 250 unique sequence homologues. Although the tertiary structures of these proteins have not been elucidated, they are predicted RHH factors, like ParG. More than 90 of the homologues possess unique ribbon sequences that are distinct either from that found in prototypical ParG or in other homologues. The differences range from single amino acid variations to predicted ribbons in which all eight residues are dissimilar to those in ParG. Two homologues were selected for further analysis, $ParG_{Yp}$ from *Yersinia pseudotuberculosis* and $ParG_{As}$ from *Aliivibrio salmonicida*, which share moderate (46%) and low (18%) sequence identity, respectively, with archetypal ParG. The putative ribbons in $ParG_{Yp}$ and $ParG_{As}$ differ from ParG at four and seven positions, respectively (Figure 3A and Supplementary Figure S1).

The degenerate 5′-ACTC-3′ motifs separated by 4-bp AT-rich spacers that comprise ParG binding sites (15,23) provide clues for *in silico* identification of equivalent sites for ParG homologues. Accordingly, examination of the

regions upstream of *parF* homologues that precede genes for $ParG_{Yp}$ and $ParG_{As}$ pinpointed arrays of tetramer repeats separated by 4-bp AT-rich patches: the core motifs are 5′-GA$^{C}/_{T}$ A-3′ and 5′-CAT$^{T}/_{G}$-3′, respectively, forming the Y- and A-sites (Figure 3A). Purified $ParG_{Yp}$ bound efficiently to a Y-site fragment, but did not interact with the O_F site recognized by ParG. Conversely, ParG did not bind the Y-site demonstrating that the native proteins interact with the cognate sites, but not with non-cognate fragments (Figure 3B).

Hybrids were designed in which the putative ribbons of $ParG_{Yp}$ and $ParG_{As}$ were stitched into the ParG backbone, producing ParGβYp (Figure 1C) and ParGβAs (Supplementary Figure S1). The specificity of ParGβYp resembled that of $ParG_{Yp}$: the chimera bound the Y-site efficiently, did not bind a randomized DNA fragment (data not shown), and recognized the O_F fragment only weakly (Figure 3C). The last may reflect indirect readout of the intervening AT-rich patches between the core motifs (23). Analogously, the ParGβAs hybrid assembled the A-site into a ladder of progressively more retarded complexes and ultimately into a major retarded species (Figure 3D). At low concentrations, ParGβAs bound the O_F fragment only weakly (Figure 3D) which may reflect a broader binding specificity. In conclusion, the results establish that putative ribbon sequences from RHH factors whose tertiary structures are unelucidated can be substituted into the ParG α-helical scaffold producing functional ATFs whose specificity differs from that provided by the ParG ribbon. Importantly, this strategy also is a powerful approach to assign uncharacterized proteins as RHH factors.

Simultaneous binding of native and chimeric RHH factors to hybrid DNA sites

RHH proteins typically do not bind a single core DNA motif, but identical RHH dimers instead assemble cooperatively on adjacent motifs. For ParG, cooperativity is promoted by interactions between the α-helical backbones across AT-rich spacers of appropriate length that traverse the core motifs. The backbone sequences differ between disparate RHH proteins implying that dimer–dimer interactions are not viable between these proteins. By contrast, chimeras based on ParG possess diverse ribbons but the α-helical scaffold sequences are identical (Figure 1B and Supplementary Figure S1), potentially permitting cooperative docking of RHH proteins with different binding specificities on artificial DNA sites bearing the corresponding motifs. A synthetic site was designed that included a core 5′-ACTC-3′ box from the O_F operator bound by ParG (15) fused to a segment derived from a CopG binding site that includes the 5′-GCA$^{C}/_{T}$-3′ motif bound by ParGβCopG ($copG_LparG_R$ in Figure 4A). Neither ParG or ParGβCopG alone bound this site in gel retardation or surface plasmon resonance assays verifying that a fragment bearing different subsites is not recognized by the proteins individually (Figure 4B and C). However, when ParG and ParGβCopG were incubated in equimolar ratios with $copG_LparG_R$ a single retarded species was observed in gel retardation and a strong combined response was evident in surface plasmon resonance. Moreover, DNase I footprinting demonstrated that ParG and ParGβCopG protected a cloned $copG_LparG_R$ locus specifically (Figure 4D). By contrast with $copG_LparG_R$, the $parG_LcopG_R$ fragment comprises the left and right segments, respectively, of ParG and CopG sites (Figure 4A). This fragment was bound more weakly by ParG and ParGβCopG than the $copG_LparG_R$ site which may reflect the dispositions of flanking AT bases in the two sites.

The $copG_LparG_R$ and $parG_LcopG_R$ sites were designed so that core motifs for ParG and ParGβCopG binding were separated by 4-bp spacers found invariably in natural ParG binding sites. This spacing accommodates cooperative interactions between the proteins' common α-helical backbones. Altering the interbox spacing in $copG_LparG_R$ or $parG_LcopG_R$ by deletion or insertion of a single base pair reduced simultaneous binding by the proteins, emphasizing that the spacer length between core motifs is crucial for optimal binding of ParG-based dimers that possess distinct ribbon sequences (Figure 4B). In summary, the combined data illustrate that RHH proteins with identical α-helical backbones, but heterologous ribbon sequences, can be guided concurrently to a chimeric site designed with subsites for both proteins.

Transcriptional repression by RHH hybrids *in vivo*

RHH proteins are transcriptional repressors that control expression of single genes or more complex regulons (12). The potential of RHH hybrids to act as transcriptional repressors *in vivo* was tested by replacing the variant 5′-ACTC-3′ motifs recognized by ParG in the O_F operator with 5′-GCAC-3′ boxes for ParGβCopG (Figure 5A). The resulting hybrid promoter-operator (PO_{H1}) was inserted as a transcriptional fusion upstream of the *lac* operon in vector pRS415 (16) driving expression of > 2000 U of β-galactosidase activity in *E. coli*. Provision of ParGβCopG *in trans* from a compatible vector reduced expression ~10-fold (Figure 5B) whereas native ParG repressed expression only ~2-fold. Thus, the engineered ParGβCopG protein is a functional transcriptional repressor that downregulates a synthetic promoter-operator *in vivo*.

As RHH proteins with identical backbones, but disparate ribbons, can interact *in vitro* as described earlier, the potential of hybrid pairs to act coordinately on an artificial promoter-operator *in vivo* was examined. The PO_{H2} sequence includes the O_F regulatory region except that alternating 5′-ACTC-3′ motifs were substituted with 5′-GCAC-3′ and 5′-ACGT-3′ boxes for ParGβCopG and ParGβMetJ, respectively (Figure 5A). Unregulated PO_{H2} expressed ~1500 U of β-galactosidase activity when placed upstream of the *lac* operon (Figure 5B). ParGβMetJ alone repressed expression from PO_{H2} modestly, whereas ParGβCopG reduced β-galactosidase values ~10-fold. Interestingly, concomitant production of ParGβMetJ and ParGβCopG repressed expression from PO_{H2} to an intermediate level compared to repression by the separate proteins (Figure 5B). This may result from partial occupation of the hybrid operator by the proteins, may reflect

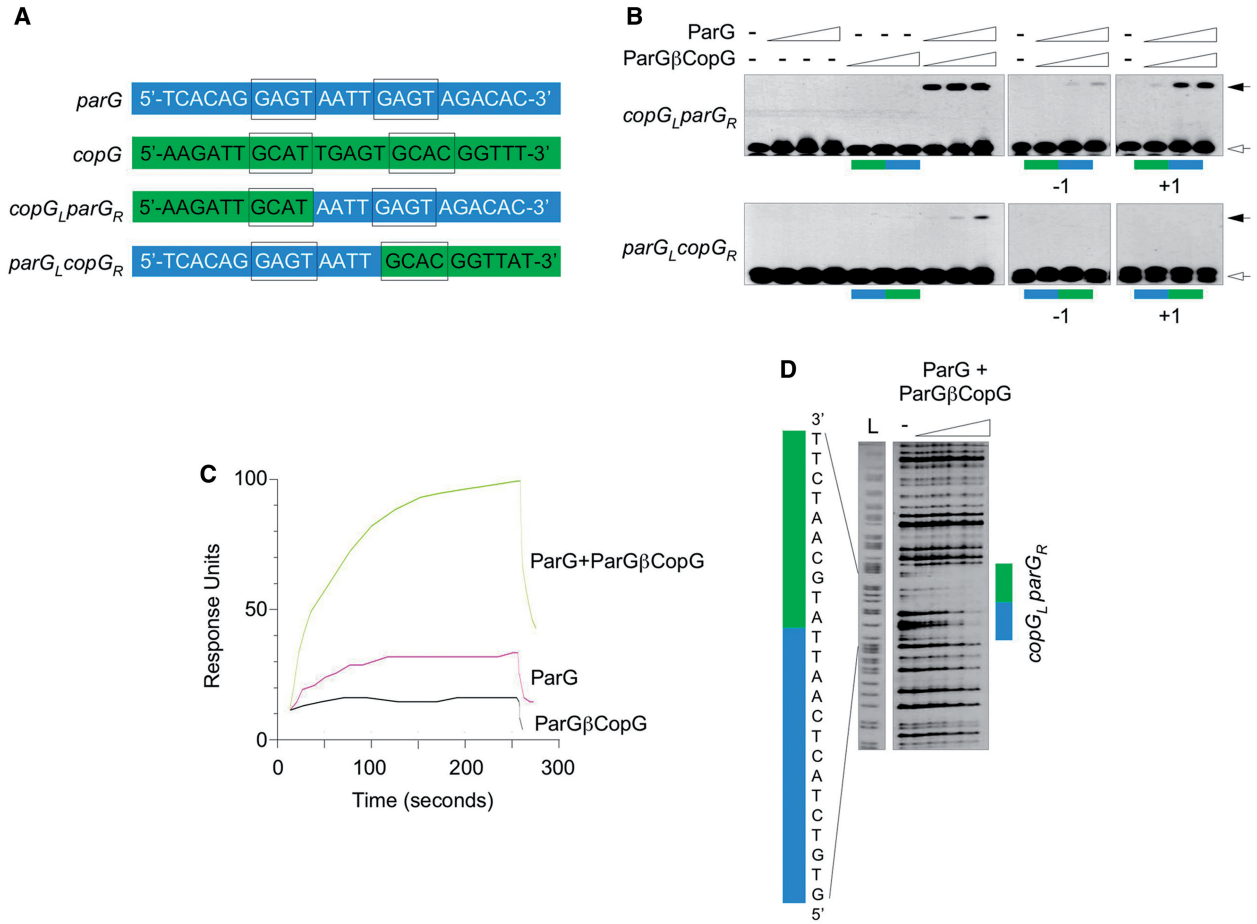

Figure 4. Simultaneous binding of ParG and hybrid ParGβCopG to chimeric sites. (**A**) Sequences of binding sites for ParG, CopG and hybrid sites tested for binding by ParG and/or ParGβCopG. Sequences from sites recognized by ParG and CopG are highlighted in blue and green, respectively. Core recognition motifs for ParG and ParGβCopG—an inverted 5′-ACTC-3′ sequence and 5′-GCA$^{C}/_{T}$-3′, respectively—are boxed. (**B**) Gel retardation assays used proteins mixed in equimolar ratios to give final concentrations of 0, 0.5, 1.0 and 1.5 μM (left to right) that were analyzed for binding to the $copG_LparG_R$ and $parG_LcopG_R$ sites (left panels). In addition, sites harboring single base-pair deletions or insertions in the central AT-rich regions were tested (middle and right panels, respectively). Open and filled arrows mark unbound DNA and nucleoprotein complexes, respectively. (**C**) Surface plasmon resonance analysis of ParG and ParGβCopG binding individually and simultaneously to the $copG_LparG_R$ site. (**D**) DNase I footprinting of ParG and ParGβCopG on $copG_LparG_R$. Proteins were mixed in equimolar ratios to give final concentrations of 0, 3, 9, 16 and 20 μM (left to right). The location of the $copG_LparG_R$ site is highlighted. L, Maxam–Gilbert A + G sequencing ladder.

formation of heterodimeric hybrids or differences in protein half-lives may elicit the intermediate repression pattern. Similar repression profiles were noted with other combinations of RHH protein, including wild-type ParG and ParGβCopG on a hybrid operator that possessed binding sites for both proteins (M. Zampini and F. Hayes, unpublished data). Pairs of Arc RHH variants with surface patches of excess positive or negative charge have been described. Some of these variants folded preferentially as heterodimers (29). Analogous charge swaps at the ParG dimer-dimer interface may assist in enforcing tight combinatorial control of ParG chimeras *in vivo*.

Repression of a second reporter gene by RHH hybrids was examined by placing PO_{H2} upstream of a promoterless kanamycin resistance determinant. *Escherichia coli* bearing the transcriptional fusion was resistant to >1000 μg/ml of the antibiotic compared to a strain without the fusion that was sensitive to <50 μg/ml kanamycin. ParGβCopG provided *in trans* repressed expression from PO_{H2}, dramatically inhibiting growth in the presence of the antibiotic (Figure 5C). Production of ParGβCopG in the absence of kanamycin did not block growth (Figure 5C, *inset*). As noted with the PO_{H2}-*lac* fusion, ParGβMetJ alone or with ParGβCopG repressed expression of the PO_{H2}-*aph* fusion to intermediate levels. Overall, the data establish that RHH hybrids can be targeted to synthetic operator sites *in vivo*. Moreover, combinations of chimeric RHH proteins and operators potentially can be employed to achieve gradations in gene expression.

DISCUSSION

The most successful strides in designing ATFs have been made by engineering extant proteins. Random or site-directed mutagenesis of DNA recognition elements can generate functional derivatives with altered binding properties, although developing variants with new

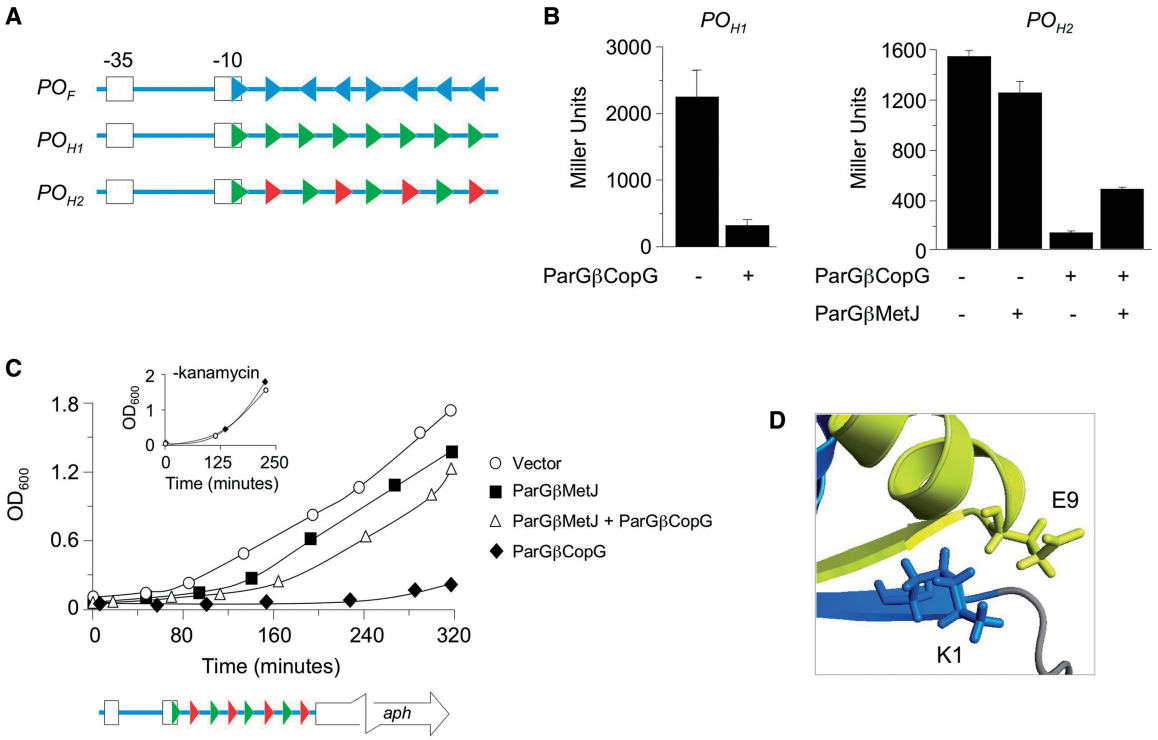

Figure 5. Chimera ParGβCopG is a functional transcriptional repressor *in vivo*. (**A**) The regulatory region (PO_F) for *parF-parG* expression comprises a characteristic Gram-negative promoter in which an array of eight 5′-ACTC-3′ boxes (blue) recognized by ParG is situated downstream of the putative −10 motif (15). These tetramer boxes were replaced by 5′-GCAC-3′ motifs (green) derived from a *copG* site (Figure 1) in the synthetic PO_{H1} promoter-operator, and by alternating 5′-GCAC-3′ and 5′-ACGT-3′ boxes (red) in PO_{H2}—the latter are recognized by ParGβMetJ. (**B**) Synthetic promoter-operators were cloned as transcriptional fusions to a *lacZ* gene in pRS415. Intrinsic promoter activities when empty pCDFDuet-1 was present *in trans* and repression of these activities by ParGβCopG and/or ParGβMetJ produced from pCDFDuet-1 were tested. (**C**) Resensitization of *E. coli* to kanamycin. Native regulatory signals controlling expression of *aph* were supplanted by the PO_{H1} promoter-operator. ParGβCopG provided *in trans* from a compatible pCDFDuet-1 based vector repressed expression from PO_{H1} and ablated kanamycin resistance (filled diamonds). ParGβMetJ alone (filled squares) and ParGβCopG and ParGβMetJ simultaneously (open triangles) repressed *aph* expression to intermediate levels. Open circles show growth of a strain bearing the PO_{H1}-*aph* fusion and empty pCDFDuet-1. The inset shows that strains carrying the PO_{H1}-*aph* fusion and either pCDFDuet-1 (open circles) or pCDFDuet-1 expressing ParGβCopG (filled diamonds) grew similarly. (**D**) Interaction between positively charged lysine at β-strand position one and negatively charged glutamate approaching from a nearby α-helix may assist in stabilizing the ribbon in ParG and functional hybrids. A similar interaction will occur at the other ribbon extremity.

specificities by this route is challenging (30,31). Fusing multiple DNA binding domains with distinct recognition patterns in tandem to generate multidomain proteins that bind extended DNA sites has been fruitful in the case of artificial zinc finger proteins (6,7). Swapping integral DNA recognition determinants between protein scaffolds has proven more taxing: active hybrids have been produced when similar α-helical DNA binding elements have been transplanted between closely related proteins, but exchange between more disparate proteins often generates disrupted or non-functional variants (32,33). By contrast, we have demonstrated here that the ribbon of the RHH superfamily is a compact and highly portable element that can be grafted in its entirety into a universal α-helical scaffold to produce functional chimeras. Among eight residues that comprise the ribbon, the designed hybrids contained four to seven amino acids that differ from those in the supplanted ParG ribbon (Supplementary Figure S1). The spliced antiparallel β-strand structures apparently can be stably accommodated within the common α-helical backbone of ParG and make appropriate DNA major groove contacts. Thus, the interactions that underpin the RHH fold in native proteins are maintained in the hybrids. To our knowledge, the only comparable study with RHH proteins involved swapping residues between ribbons of Arc and Mnt repressors with simultaneous exchange of additional residues outside of the ribbons. These multiple changes were accompanied by alterations in binding specificity (34,35).

Non-specific binding is characteristic of all sequence specific DNA binding proteins when tested at elevated concentrations *in vitro*. Not all RHH hybrids that we designed were active: chimeras bearing the β-strand motifs from the Arc, ArtA and ParD (36–38) proteins swapped into the ParG backbone did not bind DNA, even non-specifically at high protein concentrations (Supplementary Figures S1 and S3). The tertiary structures of these inactive variants may be disrupted. Proper packing of the ribbon within the α-helical scaffold will influence the stability and functionality of the chimeric proteins. Thus, packing incompatibilities could be associated with the failure of the Arc ribbon to be spliced on the ParG backbone, as it contains

phenylalanine as one of the ribbon–helix interface residues (36). Interestingly, each of the ribbons that produced a functional chimera when transplanted into ParG possesses a positively-charged lysine residue at position one of the β-strand, like ParG itself (Figures 2 and 3). By contrast, the Arc, ArtA and ParD ribbons have neutral residues at this position (Supplementary Figure S1). Similarly, the ω and CcdA RHH repressors have non-charged residues at the first positions in their β-strands (39,40) and these ribbons also produced inactive chimeras when grafted onto the ParG backbone (Supplementary Figures S1 and S3). However, when the first positions in the β-strands of these hybrids were altered to lysine, the resulting proteins bound non-specific DNA avidly in the case of the modified ω ribbon and showed a preference for RYRY core motifs in the case of the CcdA ribbon (unpublished data). Further analysis is required to determine the binding specificities of these hybrids and whether they match those associated with the parental ribbons. The interaction of lysine residues at the N-termini of the β-strands with negatively-charged glutamate residues that approach from nearby α-helices may assist in clamping the ribbon into a stable configuration in ParG and in active hybrids (Figure 5D). Determination of the tertiary structures of selected hybrids, both functional and non-functional, will clarify this hypothesis.

ParG recognizes degenerate 5′-ACTC-3′ boxes that are separated consistently by AT-rich spacers. Cooperative dimer–dimer interactions mediated by ParG α-helical scaffolds across the AT-rich regions dictate that cores invariably possess centre-to-centre distances of 4-bp resulting in regular positioning of ParG dimers along the DNA (15,23). Similarly, other RHH proteins generally bind non-palindromic motifs, although often these are extended sequences > 4-bp whose spacing varies from protein to protein (12). Combining several RHH hybrids with different specificities to form composite ATFs and the use of longer binding sites will allow the recognition of extended, more specific sequences in whole bacterial genomes. Moreover, each functional hybrid described here required only substitution of tetramer boxes within a ParG site for efficient binding while adhering to the centre-to-centre distance between these motifs observed with ParG. As illustrated by the $ParG_{Yp}$ and $ParG_{As}$ examples (Figure 3), a very powerful strategy ensues for identification of tetrameric core motifs recognized by hybrid RHH repressors, as well as for pinpointing putative proteins as *bona fide* RHH factors. This approach also may clarify whether it is feasible to decipher a code that relates ribbon sequences to the DNA sites that they recognize. Elucidation of this code may be simplified using ribbons spliced into a universal α-helical scaffold, as outlined here, so that any contribution of the scaffold to binding is normalized.

ParG multimerizes cooperatively on native DNA sites via backbone interactions between dimers bound at neighboring core motifs (15,23). Athough the RHH hybrids described here possess different ribbons, they are constructed with identical ParG backbones. These common α-helical scaffolds can be used to target pairs of distinct hybrids to synthetic sites that are designed with core motifs for both ribbons *in vitro* (Figure 4). An intriguing picture also emerged with analogous experiments *in vivo*. A single hybrid efficiently repressed an artificial promoter-operator in which core motifs for ParG were replaced by motifs for the hybrid (Figure 5B). Thus, the artificial RHH proteins are functional transcriptional repressors that retain the capacity to interfere with the action of RNA polymerase (25,41). Moreover, a promoter-operator designed with alternating core motifs for two different RHH hybrids was repressed strongly by one hybrid, weakly by the second hybrid, and to an intermediate level by both chimeras simutaneously (Figure 5B and C). These results are specially tantalizing as they establish that combinatorial use of RHH hybrids potentially can be used to modulate differential expression of target genes. Achieving gradations of expression is useful, for example, when analyzing toxic genes or when over- and under-expression of a gene of interest is sought to examine effects on cellular pathways. Combinatorial use of RHH ATFs as part of a molecular toolbox also has promise, for example, in targeting specific genomic loci with fluorescent fusion proteins to track movement of these sites during procaryotic chromosome replication and segregation. The hybrids also could be deployed as molecular roadblocks at specific loci to inhibit bacteriophage propagation, may be supplemented with peptide tags or developed into a novel two-hybrid system to study protein–protein interactions *in vivo*. The composite assembly of interacting RHH ATFs offers advantages associated with combinatorial approaches to DNA recognition compared with existing classes of ATFs that involve binding of a single protein (42). Moreover, the ParG C-terminus accommodates both short peptide affinity tags and > 200 amino acid domains without disrupting DNA binding or dimerization (14), indicating the suitability of the RHH fold for adding effector modules to innovate novel ATFs in procaryotes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3.

ACKNOWLEDGEMENTS

We thank Daniela Barillà and Maria Schumacher for incisive reading of the manuscript.

FUNDING

Biotechnology and Biological Sciences Research Council (to F.H.). Funding for open access charge: The University of Manchester.

Conflict of interest statement. None declared.

REFERENCES

1. Stormo,G.D. and Zhao,Y. (2010) Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.*, **11**, 751–760.
2. Garvie,C.W. and Wolberger,C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, **8**, 937–946.

3. Ansari,A.Z. and Mapp,A.K. (2002) Modular design of artificial transcription factors. *Curr. Opin. Chem. Biol.*, **6**, 765–772.
4. Collins,C.H., Yokobayashi,Y., Umeno,D. and Arnold,F.H. (2003) Engineering proteins that bind, move, make and break DNA. *Curr. Opin. Biotechnol.*, **14**, 371–378.
5. Khalil,A.S. and Collins,J.J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.*, **11**, 367–379.
6. Beerli,R.R. and Barbas,C.F. III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.*, **20**, 135–141.
7. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
8. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
9. Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. and Arlotta,P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.
10. Chin,J.W. and Schepartz,A. (2001) Concerted evolution of structure and function in a miniature protein. *J. Am. Chem. Soc.*, **123**, 2929–2930.
11. Stewart,A.L., Park,J.H. and Waters,M.L. (2011) Redesign of a WW domain peptide for selective recognition of single-stranded DNA. *Biochemistry*, **50**, 2575–2784.
12. Schreiter,E.R. and Drennan,C.L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.*, **5**, 710–720.
13. Higuchi,R., Krummel,B. and Saiki,R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.*, **16**, 7351–7367.
14. Barillà,D. and Hayes,F. (2003) Architecture of the ParF-ParG protein complex involved in procaryotic DNA segregation. *Mol. Microbiol.*, **49**, 487–499.
15. 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.
16. Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.
17. Miller,J.H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 72–74.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. Wu,M., Zampini,M., Bussiek,M., Hoischen,C., Diekmann,S. and Hayes,F. (2011) Segrosome assembly at the pliable *parH* centromere. *Nucleic Acids Res.*, **39**, 5082–5097.
24. Gomis-Rüth,F.X., Solá,M., Acebo,P., Párraga,A., Guasch,A., Eritja,R., González,A., Espinosa,M., del Solar,G. and Coll,M. (1998) The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.*, **17**, 7404–7415.
25. Hernández-Arriaga,A.M., Rubio-Lepe,T.S., Espinosa,M. and del Solar,G. (2009) Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. *Nucleic Acids Res.*, **37**, 4799–4811.
26. Augustus,A.M., Reardon,P.N., Heller,W.T. and Spicer,L.D. (2006) Structural basis for the differential regulation of DNA by the methionine repressor MetJ. *J. Biol. Chem.*, **281**, 34269–34276.
27. He,Y.Y., Stockley,P.G. and Gold,L. (1996) *In vitro* evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ. *J. Mol. Biol.*, **255**, 55–66.
28. 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.
29. Nohaile,M.J., Hendsch,Z.S., Tidor,B. and Sauer,R.T. (2001) Altering dimerization specificity by changes in surface electrostatics. *Proc. Natl Acad. Sci. USA*, **98**, 3109–3114.
30. Salinas,R.K., Folkers,G.E., Bonvin,A.M.J.J., Das,D., Boelen,R. and Kaptein,R. (2005) Altered specificity in DNA binding by the lac repressor: a mutant lac headpiece that mimics the gal repressor. *Chembiochem*, **6**, 1628–1637.
31. Desai,T.A., Rodionov,D.A., Gelfand,M.S., Alm,E.J. and Rao,C.V. (2009) Engineering transcription factors with novel DNA-binding specificity using comparative genomics. *Nucleic Acids Res.*, **37**, 2493–2503.
32. Wharton,R.P., Brown,E.L. and Ptashne,M. (1984) Substituting an α-helix switches the sequence-specific DNA interactions of a repressor. *Cell*, **38**, 361–369.
33. Kolkhof,P., Teichmann,D., Kisters-Woike,B., von Wilcken-Bergmann,B. and Müller-Hill,B. (1992) Lac repressor with the helix-turn-helix motif of λ cro binds to *lac* operator. *EMBO J.*, **11**, 3031–3038.
34. Knight,K.L. and Sauer,R.T. (1989) DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues. *Proc. Natl Acad. Sci. USA*, **86**, 797–801.
35. Raumann,B.E., Knight,K.L. and Sauer,R.T. (1995) Dramatic changes in DNA-binding specificity caused by single residue substitutions in an Arc/Mnt hybrid repressor. *Nat. Struct. Biol.*, **2**, 1115–1122.
36. 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.
37. Ni,L., Jensen,S.O., Ky Tonthat,N., 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.
38. Oberer,M., Zangger,K., Gruber,K. and Keller,W. (2007) The solution structure of ParD, the antidote of the ParDE toxin antitoxin module, provides the structural basis for DNA and toxin binding. *Protein Sci.*, **16**, 1676–1688.
39. Murayama,K., Orth,P., de la Hoz,A.B., Alonso,J.C. and Saenger,W. (2001) Crystal structure of ω transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 Å resolution. *J. Mol. Biol.*, **314**, 789–796.
40. Madl,T., Van Melderen,L., Mine,N., Respondek,M., Oberer,M., Keller,W., Khatai,L. and Zangger,K. (2006) Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *J. Mol. Biol.*, **364**, 170–185.
41. Smith,T.L. and Sauer,R.T. (1996) Role of operator subsites in Arc repression. *J. Mol. Biol.*, **264**, 233–242.
42. Moretti,R. and Ansari,A.Z. (2008) Expanding the specificity of DNA targeting by harnessing cooperative assembly. *Biochimie*, **90**, 1015–1025.