Evolving a new protein-DNA interface via sequential introduction of permissive and specificity-switching mutations

Adam S. B. Jalal[1‡], Ngat T. Tran[1‡], Clare E. Stevenson[2], Xiao Tan[1], David M. Lawson[2], and Tung B. K. Le[1*]

[1]Department of Molecular Microbiology
John Innes Centre, Norwich, NR4 7UH, United Kingdom
[2]Department of Biological Chemistry
John Innes Centre, Norwich, NR4 7UH, United Kingdom

[‡] These authors contribute equally
[*]Corresponding author: tung.le@jic.ac.uk

ABSTRACT

Specific interactions between proteins and DNA are essential to many biological processes. Yet it remains unclear how the diversification in DNA-binding specificity was brought about, and what were the mutational paths that led to changes in specificity. Using a pair of evolutionarily related DNA-binding proteins each with a different DNA preference (ParB and Noc: both having roles in bacterial chromosome maintenance), we show that specificity is encoded by a set of four residues at the protein-DNA interface. Combining X-ray crystallography and deep mutational scanning of the interface, we show that permissive mutations must be introduced before specificity-switching mutations to reprogram specificity, and that mutational paths to a new specificity do not necessarily involve dual-specificity intermediates. Overall, our results provide a glimpse into the possible evolutionary history of ParB and Noc, and in a broader context, might be useful in understanding the evolution of other classes of DNA-binding proteins.

INTRODUCTION

In living organisms, hundreds of DNA-binding proteins carry out a plethora of roles in homeostasis, transcriptional regulation in response to stress, and in maintenance and transmission of genetic information. These DNA-binding proteins do so faithfully due to their distinct DNA-binding specificity towards their cognate DNA sites. Yet it remains unclear how related proteins, sometimes with a very similar DNA-recognition motif, can recognize entirely different DNA sites. What were the changes at the molecular level that brought about the diversification in DNA-binding specificity? As these proteins evolved, did the intermediates in this process drastically switch DNA-binding specificity or did they transit gradually through promiscuous states that recognized multiple DNA sequences? Among many ways to evolve new biological innovations, gene duplication and neo-functionalization has been widely implicated as a major force in evolution[1–5]. In this process, after a gene was duplicated, one copy retained the original function while the other accumulated beneficial and diverging mutations that produced a different protein with a new function. In the case of DNA-binding proteins, a new function could be the recognition of an entirely different DNA site. In this work, we employed a pair of related DNA-binding proteins (ParB and Noc) that are crucial for bacterial chromosome maintenance to better understand factors that might have influenced the evolution of a new DNA-binding specificity.

ParB (Partitioning Protein B) is important for faithful chromosome segregation in two-thirds of bacterial species[6,7]. The centromere-like

parS DNA locus is the first to be segregated following chromosome replication[6–9]. *parS* is bound by ParB, which in turn interacts with ParA and SMC proteins to partition the ParB-*parS* nucleoprotein complex, hence the chromosome, into each daughter cell[7,10–17] (Fig. 1A). ParB specifically recognizes and binds to *parS*, a palindromic sequence (Fig. 1A) that can be present as multiple copies on the bacterial chromosome but almost always locate close to the origin of replication (*oriC*) on each chromosome (Fig. 1A)[6–8,16,18–21]. ParB proteins are widely distributed in bacteria and so must have appeared early in evolution (Fig. 1B)[6]. Noc (Nucleoid Occlusion Factor), a ParB-related protein, was first discovered in *Bacillus subtilis*[22,23]. Like ParB, Noc has a three-domain architecture: an N-terminal domain for protein-protein interaction and for targeting Noc to the cell membrane, a central DNA-binding domain, and a C-terminal dimerization domain[23,24] (Fig. 1A). In contrast to ParB, Noc recognizes a DNA-binding sequence called *NBS* (Noc Binding Site)[24,25] (Fig. 1A). The role of Noc is also different from ParB; Noc functions to prevent the cell division machinery from assembling in the vicinity of the nucleoid, which might be otherwise guillotined, thereby damaging the DNA[23,24] (Fig. 1B). In other words, Noc has a role in preserving the integrity of the chromosome. The genome-wide distribution of *NBS* is also drastically different from that of *parS*. While *parS* sites are restricted in the region around *oriC*, *NBS* are all over the genome except near the terminus of replication (*ter*) [24,25]. The absence of *NBS* near *ter* is crucial to direct the formation of the FtsZ ring and cell division to mid-cell (Fig. 1A). Because of their genomic proximity (Fig. S1) and high sequence similarity, it was suggested that *noc* resulted from a gene duplication event from *parB*[22,26]. A phylogenetic tree showed that *parB* genes are widely distributed in bacteria but *noc* genes are confined to the Firmicutes clade[26] (Fig. 1B). This phylogenetic distribution is most consistent with *parB* appearing early in evolution, possibly before the split between Gram-positive and Gram-negative bacteria, and that the occurrence of *noc* is a later event that happened only in Firmicutes.

Here, we systematically measure the binding preferences of 17 ParB and four Noc family members to *parS* and *NBS* and find that their interactions are specific and conserved among bacterial species. We show that specificity to *parS* or *NBS* is encoded by a small set of four residues at the protein-DNA interface, and mutations in these residues are enough to reprogram DNA-binding specificity. Combining X-ray crystallography and systematic scanning mutagenesis, we show that both permissive and specificity-switching substitutions are required to acquire a new DNA-binding specificity. Guided by these findings, we generate a saturated library with ~10^5 variants of the specificity-defining residues in ParB, and select for mutants that bind to *parS*, *NBS*, or both. We discover several hundred alternative combinations of residues that are capable of binding to *parS* or *NBS*. By analyzing the connectivity of functional variants in the sequence space, we show that permissive and specificity-switching mutations must be introduced in an orderly manner to evolve a new protein-DNA interface. Lastly, we provide evidence that mutational paths that change specificity, at least in our case, tend to be switch-like and do not necessarily involve dual-specificity intermediates.

RESULTS

DNA-binding specificity for *parS* and *NBS* is conserved within ParB and Noc family

To test whether ParB and Noc family members retained their DNA-binding specificity, we selected a group of 17 ParB and four Noc from various bacterial clades for characterization (Fig. 1B-C). We expressed and purified 21 tandemly $(His)_6$- and MBP-tagged ParB and Noc to analyze them in a quantitative bio-layer interferometry assay that directly assess their binding to a *parS*-containing or *NBS*-containing duplex DNA (Fig. 1C). As shown in Fig. 1C, all tested ParB proteins bind preferentially to *parS* over *NBS*, while Noc proteins prefer *NBS* to *parS*. To test DNA-binding specificity *in vivo*, we performed ChIP-seq experiments to quantify the level of ParB or Noc that are bound at a single *parS* or *NBS* site engineered onto the *Escherichia coli* chromosome (Fig. 1D). *E. coli* is a perfect heterologous host for this experiment as it does not possess native ParB/Noc homologs and there are no *parS*/*NBS* site on its genome. Individual ParB or Noc proteins from our collection were expressed from the *lac* promoter and were engineered with

an N-terminal FLAG tag for immunoprecipitation. Consistent with the *in vitro* data, we observed the same trend that ParB members prefer *parS* to *NBS*, and vice versa for Noc members (Fig. 1D). This conservation of DNA preference suggests that there exists a set of conserved residues within each protein family (ParB or Noc) that dictate specificity.

Co-crystal structure of the DNA-binding domain of ParB with *parS* revealed residues that contact DNA

As the first step in identifying specificity residues, we solved a 2.4 Å resolution co-crystal structure of the DNA-binding domain (DBD) of *Caulobacter crescentus* ParB bound to a 20-bp *parS* DNA duplex (Fig. 2A). In the crystallographic asymmetric unit, two very similar ParB DBD monomers (RMSD = 0.1 Å) bind in a two-fold symmetric fashion to a full-size *parS* DNA duplex (Fig. 2A). This structure reveals several regions of each DBD that contact *parS* (Fig. 2B). Firstly, the recognition helix (res. 172-184) of the helix-turn-helix motif inserts into the major grooves of the palindromic *parS* site (Fig. 2B). Second, helix (res. 199-206) and helix (res. 225-235) contribute residues to the protein-DNA interface (Fig. 2B). Lastly, several lysine and arginine residues in the loop (res. 236-254) contact the minor groove side of *parS* in an adjacent complex in the crystal (Fig. S2A), although their contribution to the DNA-binding affinity might not be significant in solution (Fig. 2D). From the structure of the complex, we identified residues that make specific contacts with the DNA bases as well as non-specific contacts with the phosphate backbone (Fig. 2C). We verified the protein-DNA contacts and their contribution to binding affinity by individually mutagenizing each residue to alanine (Fig. 2D). We found that most of the crucial residues for binding to *parS* are within the 162-234 region (Fig. 2D), suggesting their importance in recognizing DNA specifically. We reasoned that specificity residues for *parS* (and *NBS*) must localize within this amino acid region in ParB (and in an equivalent region in Noc).

Four amino acid residues dictate interaction specificity between Noc and *NBS*

To discover the region of Noc that determines specificity for *NBS*, we constructed a series of chimeric proteins in which different regions of *Caulobacter* ParB were replaced with the corresponding regions of *B. subtilis* Noc (Fig. 3A). Replacing the entire region (res. 162-230) containing the helix-turn-helix motif, helix (res. 199-206), and part of helix (res. 225-235) with the corresponding region of *B. subtilis* Noc produced a chimera that binds to both *parS* and *NBS*, but with a clear preference for *NBS* (Chimera 1, Fig. 3A). Swapping a smaller region (res. 162-207) containing just the helix-turn-helix motif and an adjacent helix (res. 199-206) created a chimera that has an improved specificity for *NBS* (Chimera 4, Fig. 3A). These results indicate that the region (res. 162-207) contains most of the specificity residues for *NBS*.

To better understand the high degree of specificity conserved within the ParB and Noc families, we mapped a sequence alignment of ~1800 ParB and ~400 Noc orthologs onto the ParB(DBD)-*parS* crystal structure to determine amino acid sequence preferences for those residues required for interaction specificity (Fig. 3B). We focused our attention on the region between residues 162 and 207, which was shown above to contain the specificity residues (Fig. 3B). Of those amino acids that contact *parS* (Fig. 2B-C), six residues (Q162, G170, K171, S172, N178, and R204) are conserved between ParB and Noc family members (Fig. 3B). Two residues (R173 and G201) in ParB that contact *parS* but are changed to Q173 and R201, respectively, in Noc homologs (Fig. 3B). Other residues at positions 179 and 184 vary among ParB homologs but are almost invariably a lysine in Noc family members (Fig. 3B). We hypothesized that these amino acids (Q173, K179, K184, and R201) (Fig. 3B) are specificity residues that dictate Noc preference for *NBS*. To test this hypothesis, we generated a variant of *Caulobacter* ParB in which these four residues were introduced at the structurally equivalent positions (i.e. R173Q, T179K, A184K and G201R). We purified and tested this variant in a bio-layer interferometry assay with *parS* and *NBS*. As shown in Fig. 3A, this *Caulobacter* ParB (RTAG→QKKR) (PtoN15) variant completely switched its binding preference to a non-cognate *NBS* site. We also identified additional highly conserved residues (E200 and K227) that might contribute to the *NBS*-binding preference (Fig. 3A-B), however, a minimal set

of four residues are enough to reprogram specificity.

Systematic dissection of ParB-*parS* and Noc-*NBS* interfaces reveals the contribution of each specificity residue to the DNA-binding preference

To systematically dissect the role of each specificity residue, we constructed a complete set of ParB mutants that have either a single, double, or triple amino acid changes between the four specificity positions, from a *parS*-preferred wild-type *Caulobacter* ParB ($R^{173}T^{179}A^{184}G^{201}$) to an *NBS*-preferred variant ($Q^{173}K^{179}K^{184}R^{201}$). We named them ParB-to-Noc intermediates (PtoN, 15 variants in total). To simplify the nomenclature, we named the mutants based on the specificity residues being considered, for example, an *NBS*-preferred variant ($Q^{173}K^{179}K^{184}R^{201}$) is shortened to PtoN15 (QKKR). PtoN variants were C-terminally $(His)_6$ tagged, expressed, and purified to homogeneity. Subsequently, we tested ParB and 15 PtoN variants with a series of 16 different DNA sites, each representing a transitional state from *parS* to *NBS* with each of the two variable positions (1 and 6) changed to any of other four DNA bases (Fig. 3C). We visualized 16x16 interactions as a heatmap where each matrix position reflects a response value from our bio-layer interferometry assays.

This systematic pairwise interaction screen led to several notable observations (Fig. 3C). First, there are two non-functional variants (PtoN1: QTAG and PtoN7: QTAR) that were unable to interact with any of the 16 DNA sites (Fig. 3C). Second, three variants (PtoN4: RTAR, PtoN5: QKAG, and PtoN6: QTKG) switched their specificity to a DNA site that has features borrowed from both *parS* and *NBS*. Meanwhile, several variants (PtoN2: RKAG, PtoN3: RTKG, PtoN8: RKKG, PtoN9: RKAR, PtoN10: RTKR, PtoN11: QKKG, and PtoN14: RKKR) were promiscuous i.e. binding to multiple different DNA sites (Fig. 3C). We noted that functional PtoN variants always have a lysine at either position 179, or 184, or both. This observation became even clearer after we performed hierarchical clustering of the interaction profile in both the protein and the DNA dimensions (Fig. 3D). A single lysine at either position 179 or 184 is enough to license the DNA-binding capability to PtoN variants (nodes a, b, C, and d on the clustering tree, Fig. 3D), while PtoN1 (QTAG) and PtoN7 (QTAR) that do not possess any lysine at 179/184 are non-functional (node f, Fig. 3D). While a single lysine at 179/184 is enough, their presence at both positions enhanced DNA binding and promiscuity (nodes A, B, C, and D, Fig. 3D). Overall, K179/184 has a “permissive” effect that might permit Q173 and R201 to contact DNA.

Next, we wondered which base of the *NBS* site that Q173 might contact specifically. To find out, we clustered only PtoN variants that share the Q amino acid at position 173 (Fig. 3E). We discovered that those variants preferred DNA sites that possess an Adenine at position 1 (Fig. 3E). We applied the same approach to find the base that residue R201 might contact (Fig. 3F). The emerging trend is that PtoN variants that share an R amino acid at 201 preferred DNA sites with a Cytosine at position 6 (Fig. 4D). Taken together, our results support a model in which each specificity residue has a distinct role: Q173 recognizes Adenine 1, R201 recognizes Cytosine 6, but they can only do so in the presence of a permissive K at either position 179 or 184, or both. In the next section, we confirmed this model using X-ray crystallography data.

Co-crystal structure of the specificity-altered ParB variant with *NBS* reveals the contribution of each specificity residue to the DNA-binding preference

To understand the biophysical mechanism underlying the specificity to *NBS*, we solved the co-crystal structure of a specificity-altered ParB variant (QKKR+K227) mutant with a 20-bp *NBS* DNA duplex to 3.6Å resolution (Fig. S2B). K227 was shown to enhance *NBS* binding but is not part of the core specificity residues (Fig. 3A). Our construct contains the DNA-binding domain and the N-terminal domain of ParB (QKKR+K227) but lacks the C-terminal dimerization domain (Fig. S2B-D and Materials and Methods). While the resolution was modest, several key interactions at the protein-DNA interface were well resolved in the electron density (Fig. S2E-F). By superimposing the structures of the wild-type ParB (DBD)-*parS* complex and the ParB (QKKR+K227)-*NBS* complex, we observed several changes in both the protein and the DNA sites that enabled specific interactions (Fig. 4). First, R173 in wild-

type ParB hydrogen bonds with *parS* Guanine 1, but the shorter side chain of Q173 (in an *NBS*-preferred mutant) is unable to bond with Guanine 1 (Fig. 4A). However, a corresponding base in *NBS* (Adenine 1) positions itself closer to enable hydrogen bonding with this Q173 residue (Fig. 4A); this is possibly due to conformational changes in the *NBS* site that narrows the minor groove width at the Adenine 1:Thymine -1 position (from 7.5 to 3.8Å, Fig. S3). The switch from R173 to Q173 serves to eliminate the ability of ParB to contact *parS* Guanine 1 while simultaneously establishing a new contact with *NBS* Adenine 1. The second notable changes between the two co-crystal structures occurs at position 201 (Fig. 4B). G201 from wild-type ParB has no side chain, hence cannot contact Thymine -6 (Fig. 4B). However, the equivalent residue R201 in an *NBS*-preferred variant readily forms hydrogen bonds with Guanine -6 (Fig. 4B). We also observed DNA unwinding that increased both the minor and the major groove widths at the Cytosine 6:Guanine -6 position of *NBS* (from 7.0 to 9.0 Å, and from 10.2 to 12.5 Å, respectively), possibly to move Guanine -6 outwards to be compatible with the longer side chain of R201 (Fig. S3). Overall, our co-crystal structures are consistent with data from the systematic scanning mutagenesis. Our ParB (QKKR+K227)-*NBS* complex structure, however, did not yield extra information on the specificity residues K179 and K184 because no clear electron density was seen for the side chains of these residues (Fig. 4). It is possible that these residues do not contact specific bases but interact with the phosphate backbone non-specifically at multiple positions, leading to a blurring of the electron-density. The most parsimonious explanation for the permissive capability of K179/184 is that they increase DNA-binding affinity non-specifically to overcome the initial energy barrier and permit specific base contacts from Q173 and R201. If this were true, we predicted that another amino acid with a positively charged side chain could substitute for lysine in its role as a permissive mutation. This turned out to be the case because arginine can readily substitute for lysine at positions 179 and 184 (Fig. 6A). Another hypothesis for the permissive effect of K179/184 is that they induce conformational changes at the recognition helix and helix (res. 200-207), thereby positioning Q173 and R201 favorably to interact with *NBS* Adenine 1 and Guanine -6. However, the two helices were not seen to change their conformations significantly between the two co-crystal structures (Fig. 4), arguing against this hypothesis.

A high throughput bacterial one-hybrid selection reveals multiple combinations of specificity residues that enable *parS* and *NBS* recognition

While the results from our systematic scanning mutagenesis and X-ray crystallography revealed how specificity changed as individual substitutions were introduced sequentially, presumably more variety of amino acids has been sampled by Nature than those presented at the start (RTAG) and end point (QKKR). What are the paths, and are there many, to convert a *parS*-binding protein to an *NBS*-preferred one? Does the order of amino acid substitutions matter? To answer these questions, we explored the entire sequence space at the four specificity residues by generating a combinatorial library of ParB where positions 173, 179, 184, and 201 can be any amino acid (20^4 or 160,000 variants lacking stop codons). We optimized a bacterial one-hybrid (B1H) assay[27] that is based on transcriptional activation of an imidazoleglycerol-phosphate dehydratase encoding gene HIS3 to enable a selection for *parS* or *NBS*-binding variants (Fig. 5A and Fig. S4). ParB variants were fused at their N-termini to the omega subunit of bacterial RNA polymerase. NNS codons (where N = any nucleotide and S = Cytosine or Guanine) were used to randomize the four specificity residues. All ParB variants were also engineered to contain an additional invariable mutation in the N-terminal domain (R104A) that makes ParB unable to spread[16,28] (Fig. 5A). The R104A mutation does not affect the site-specific binding but enables a simpler design of the selection system by converting ParB to a conventional site-specific transcriptional activator (Fig. 5A). If a ParB variant binds to a *parS* or *NBS* site engineered upstream of HIS3, it will recruit RNA polymerase to activate HIS3 expression, thereby enabling a histidine-auxotrophic *E. coli* host to survive on a minimal medium lacking histidine (Fig. 5A and Fig. S4). Deep sequencing of starting libraries revealed that >94% of the predicted variants were represented by at least 10 reads (Fig. S5A-B),

and that libraries prepared on different days were reproducible (R^2>0.90, Fig. S5C).

To assess the ability of each ParB variant to bind to *parS* or *NBS*, we deep-sequenced the relevant region on *parB* variants pre- and post-selection to reveal the underlying sequences and their abundance (Fig. 5A and Fig. S5C). As protein-DNA binding affinity is proportional to the amount of histidine being produced[27], and ultimately to the cell fitness, we quantified the fitness of each variant to rank them (Fig. 5C, and Materials and Methods). We found 1385 and 362 variants that show strong binding to *parS* and *NBS*, respectively (Fig. 5B). We then selected nine variants that either bind *NBS*, *parS*, or both (Fig. 5C) and showed by a pairwise B1H assay and by bio-layer interferometry assay with purified proteins that their functionality is consistent with deep-sequencing data (Fig. S6). To systematically probe the sequence space, we generated a scatter plot of ParB variant fitness when screened for binding to *parS* or *NBS* (Fig. 5C). Of 362 variants that bind *NBS* strongly: 261 are *NBS* specific (i.e. no *parS* binding, magenta box), 19 show strong *NBS* binding but weak-to-medium *parS* binding (pink box), and 82 dual-specificity variants that bind both *parS* and *NBS* (black box) (Fig. 5C). By comparing sequence logos, we observed that *NBS*-specific variants (magenta box) have a high proportion of the Q residue at position 173 but R is allowed; position 201 is dominantly R but polar residues (T and S) are allowed; and positively charged R and K prevail at positions 179 and 184 (Fig. 5C). This sequence logo shares some features with Noc orthologs (dashed magenta box, Fig. 5C). On the other hand, *parS*-specific variants (dark green box) have an invariable R at position 173, same as ParB orthologs (dashed dark green box) (Fig. 5C), but position 201 can be small polar amino acids (C, S, or T, but G is most preferred). Notably, 17 amino acids (except a helix-breaking P or negatively charged D and E) can occupy position 179, and any of the 20 amino acid is tolerable at position 184 (Fig. 5C). Finally, dual-specificity variants (black box) tend to harbor sequence elements from both *parS*- and *NBS*- specific variants (Fig. 5C).

NBS-specific variants predominantly have lysine or arginine at positions 179 and 184

The proportion of *NBS*-specific variants with a K or R amino acid at position 179 is ~58%, higher than a theoretical 10% value if K/R was chosen randomly (Fig. 6A). The same proportion was seen for a K or R at position 184 (Fig. 6A). This proportion increased to ~91% for *NBS*-specific variants with either K or R at either position 179 or 184, and ~19% for those with a K or R at both 179 and 184 (Fig. 6A). The prevalence of positively charged residues supports our model that permissive mutations act by increasing protein-DNA binding affinity non-specifically, most likely via their interactions with the negatively charged phosphate backbone. Also, the observation that positions 179 and 184, which are more than an α-helical turn apart but are equal in their permissive capability, lend a further support to our non-specific DNA-binding model. We noted that K and R are not preferred more than expected from a random chance in *parS*-specific variants (Fig. 6A). Our results suggest that the introduction of permissive substitutions is crucial to acquire a new specificity.

Permissive and specificity-switching mutations were introduced in a defined order to reprogram specificity

We asked if there is an order of substitutions at positions 173, 179, 184, and 201 to create an *NBS*-specific variant. To answer this question, we first reconstructed all possible mutational paths to an *NBS* specificity. We created a force-directed graph that connects functional variants (nodes) together by lines (edges) if they are different by a single amino acid (AA) to visualize the connectivity of functional variants in sequence space (Fig. 6B)[29]. The node size is proportional to its connectivity (number of edges), and node colors represent different classes of functional variants (Fig. 6B). Similarly, we also generated a network graph in which edges represent variants that differ by a single nucleotide (nt) substitution (Fig. S7A-B). Because not all amino acids can be converted to others by a mutation at a single base, a by-nt-substitution network might depict better how long (hard) or short (easy) the mutational paths that *parS*-specific variants might have taken to reprogram their specificity to *NBS*. At first glance, the network is composed of multiple clusters of densely interconnected nodes that share common features in amino acid sequence (Fig. 6B). Furthermore, there are

multiple edges connecting *parS*-preferred variants (dark and light green nodes) to *NBS*-preferred variants (magenta and pink nodes) (Fig. 6B). Supporting this observation, we found that it takes at most four AA (or seven nt) substitutions to convert any *parS*-specific variant to an *NBS*-specific QKKR (Fig. 6C and Fig. S7C). A small number of steps suggested that *NBS*-specific variants can be reached relatively easily from *parS*-specific variants. We focused on *parS*-specific start points RXXG for all analyses below because R173 and G201 are absolutely conserved in all extant ParB orthologs (Fig. 5C). We found all the shortest paths (1,232 in total) that connect *parS*-specific RXXG variants (298 dark green nodes) to an *NBS*-specific QKKR, and quantified the fractions of intermediates in such paths that contain permissive or specificity-switching residues (Fig. 6D). We discovered that permissive substitutions (K or R) at position 179 or 184 happened very early on along the mutational paths (~95% after the first step, Fig. 6D). The fraction of R201 increased more gradually after the introduction of permissive substitutions, and Q173 was introduced last (Fig. 6D). The same order of substitutions was seen when we analyzed a by-nt-substitution network graph (Fig. S7D). In sum, we concluded that the order of amino acid substitutions matters, and that permissive mutations preferentially happened before specificity-switching substitutions.

Mutational paths that reprogram specificity did not travel across dual-specificity intermediates

We observed that the fraction of variants with C/T/S residues at position 201 did not increase beyond 0% in any step from RXXG variants to QKKR (Fig. 6D and Fig. S7D). Given that dual-specificity variants (black box, Fig. 5C) mostly have T or S amino acid at position 201, it suggests that dual-specificity intermediates might have not been exploited to change specificity. Indeed, no shortest path connecting RXXG and QKKR traversed through any dual-specificity variant (black nodes) (Fig. 6E and Fig. S7E). This proportion is significantly smaller than would be expected by chance (estimated from 1,000 random networks where edges were shuffled randomly, Fig. 6E and Fig. S7E). In contrast, ~51% and ~3% of shortest paths from RXXG variants to QKKR contain light green and pink intermediates, respectively. The proportions of paths with light green or pink intermediates are similar to expected values from random chances (Fig. 6E). The preference for traversing light green nodes, therefore, can be explained by the abundance of such variants in the observed graph (Fig. 6B). Overall, our network analysis revealed that the *parS*-to-*NBS* reprogram did not exploit truly dual-specificity intermediates, and that those with a stricter specificity (light green or pink) were more commonly used.

DISCUSSION

Determinants of specificity and implications for understanding the evolution of protein-DNA interfaces

The *NBS* site differs from the *parS* site by only 2 bases (positions 1 and 6, Fig. 1A) but Noc and ParB recognize and bind them with exquisite specificity. We showed that mutations must have been introduced in a defined order to enable a switch in specificity. Permissive substitutions (K/R at positions 179/184) tend to appear first, presumably to prime *parS*-specific variants for a subsequent introduction of specificity-switching residues (R201 and Q173) which would have otherwise rendered proteins non-functional (Fig. 7). An early introduction of permissive substitutions is likely to be a recurring principle of evolution. For example, a similar prerequisite for permissive mutations were observed in the evolution of influenza resistance to the antiviral drug oseltamivir[30]. Two permissive mutations were first acquired, allowing the virus to tolerate a subsequent occurrence of a H274Y mutation that weakened the binding of oseltamivir to the viral neuraminidase enzyme[30]. These permissive mutations improved the stability of neuraminidase before a structurally destabilizing H274Y substitution was introduced[30,31]. Similarly, a permissive mutation that is far away from the active site of an antibiotic-degrading β-lactamase (TEM1) has little effect on its enzymatic activity by itself but restored stability loss by a subsequent mutation that increased TEM1 activity against cephalosporin antibiotics[32]. In another case, 11 permissive mutations were required to evolve an ancestral steroid hormone receptor from preferring an estrogen response element (ERE) to a new DNA sequence (steroid response

element or SRE)[33]. These 11 mutations were located outside of the DNA-recognition motif but non-specifically increased the affinity for both ERE and SRE, thereby licensing three additional substitutions to alter the specificity to SRE[33]. Additionally, it has been shown that an early introduction of 11 permissive substitutions dramatically increased the number of SRE-binding variants well beyond the historical observed variants[34]. In our case, just a single introduction of a lysine, either at position 179 or 184, was sufficient to permit Q173 and R201 to recognize *NBS* specifically. K179 and/or K184 are within the DNA-recognition helix and increased the affinity to both *parS* and *NBS* non-specifically (Fig. 3C). Taken all together, the results from us and other researchers, each employing a very different model from a wide range of organisms, emphasize the importance of permissive mutations in the molecular evolution of protein-ligand interactions.

Deep mutational scanning in conjunction with network analysis is a powerful approach to reconstruct possible mutational paths that might have been taken to acquire a new function[29,34,35]. Network graph theory was applied to understand the constraints on the evolution of protein-protein interfaces between a histidine kinase and its response regulator partner, between toxin and antitoxin pairs of proteins, and most recently to reveal the alternative evolutionary histories of a steroid hormone receptor[29,34,35]. In our case study, network analysis suggested that mutational paths to a new specificity did not necessarily have to visit dual-specificity intermediates i.e. those that bind *parS* and *NBS* equally strongly (Fig. 6E). Instead, mutational paths to an *NBS*-specific variant tend to be more switch-like, frequently visited dark green nodes (strong *parS* binding, no *NBS* binding) and light green nodes (strong *parS* binding, weak-to-medium *NBS* binding) (Fig. 5C and Fig. 6E). We reason that most black variants, albeit being dual specific, bind both *parS* and *NBS* at a slightly reduced affinity (compared to the wild-type *parS*-specific RTAG or *NBS*-specific QKKR variants, see the scatter plot on Fig. 5C). This might have created an undesirable situation where dual-specificity intermediates neither could compete with the original copy of ParB to bind *parS* nor had high enough affinity themselves to bind *NBS* sites i.e. artificially made non-functional due to competition. A similar principle might also apply to other protein-DNA interactions throughout biology. For example, a reconstructed evolutionary history of a steroid hormone receptor indicated that an ancestral receptor (AncSR1) without permissive mutations must always pass through dual-specificity intermediates to acquire the present-day specificity. On the other hand, the presence of 11 permissive mutations (AncSR1+11P) eliminated the absolute requirement for these dual-specificity intermediates. More dramatically, it has been shown that a single substitution (i.e. a truly switch-like mechanism) was enough to reprogram the specificity of homologous repressor proteins (Arc and Mnt) in bacteriophage P22[36]. Nevertheless, we noted that protein-protein interfaces, particularly in the case of paralogous toxin-antitoxin protein pairs, instead exploited extensively promiscuous intermediates to diversify and evolve. In the case of toxin-antitoxin systems, truly promiscuous intermediates might have been favored because many of them bound to and antagonized cognate and non-cognate toxins equally or even better than wild type[35]. Taken all together, we argue that the topology of the available sequence space and the biology of each system collectively influence the likely paths to evolve a new biological innovation.

Final perspectives

In sum, our work provides a molecular basis for how protein-DNA interaction specificity can change, with a focus on the two proteins ParB and Noc that are widely distributed and important for bacterial chromosome maintenance. We identified a minimal set of four specificity residues at the protein-DNA interface, and dissected at the molecular level the role of individual residues in reprogramming specificity. A small number of specificity residues also enabled a systematic and in-depth analyses of the protein-DNA interface and possible mutational paths that could have changed specificity. Our results indicate that permissive mutations must be introduced well before specificity-switching mutations, and that mutational paths to a new specificity do not necessarily have to visit dual-specificity intermediates. In a broader context, our work might be useful in understanding the evolution of other classes of DNA-binding proteins.

ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is GSE129285. Atomic coordinates for protein crystal structures reported in this paper were deposited in the RCSB Protein Data Bank with the following accession numbers: 6S6H and 6S6P.

SUPPLEMENTARY INFORMATION

Supplementary Information includes Materials and Methods, seven figures, and four tables.

ACKNOWLEDGEMENTS

This study was supported by the Royal Society University Research Fellowship (UF140053) and a BBSRC grant (BB/P018165/1) to T.B.K.L. A.S.B.J's PhD studentship was funded by the Royal Society (RG150448), and N.T.T was funded by the BBSRC grant-in-add (BBS/E/J/000C0683 to the John Innes Centre). We acknowledge Diamond Light Source for access to beamlines I03 and I04 under proposal MX18565 with support from the European Community's Seventh Framework Program (FP7/2007–2013) under Grant Agreement 283570 (BioStruct-X). We also thank Rory Williams for help with early experiments.

REFERENCES

1. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
2. Qian, W. & Zhang, J. Genomic evidence for adaptation by gene duplication. *Genome Res* **24**, 1356–1362 (2014).
3. Conrad, B. & Antonarakis, S. E. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet* **8**, 17–35 (2007).
4. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
5. Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).
6. Livny, J., Yamaichi, Y. & Waldor, M. K. Distribution of Centromere-Like parS Sites in Bacteria: Insights from Comparative Genomics. *J. Bacteriol.* **189**, 8693–8703 (2007).
7. Lin, D. C.-H. & Grossman, A. D. Identification and Characterization of a Bacterial Chromosome Partitioning Site. *Cell* **92**, 675–685 (1998).
8. Lagage, V., Boccard, F. & Vallet-Gely, I. Regional Control of Chromosome Segregation in Pseudomonas aeruginosa. *PLOS Genetics* **12**, e1006428 (2016).
9. Toro, E., Hong, S.-H., McAdams, H. H. & Shapiro, L. Caulobacter requires a dedicated mechanism to initiate chromosome segregation. *PNAS* **105**, 15435–15440 (2008).
10. Fogel, M. A. & Waldor, M. K. A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev.* **20**, 3269–3282 (2006).
11. Ireton, K., Gunther, N. W. & Grossman, A. D. spo0J is required for normal chromosome segregation as well as the initiation of sporulation in Bacillus subtilis. *J. Bacteriol.* **176**, 5320–5329 (1994).
12. Mohl, D. A. & Gober, J. W. Cell cycle-dependent polar localization of chromosome partitioning proteins in Caulobacter crescentus. *Cell* **88**, 675–684 (1997).
13. Gruber, S. & Errington, J. Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in B. subtilis. *Cell* **137**, 685–96 (2009).
14. Wang, X. *et al.* Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. *Genes Dev.* **29**, 1661–1675 (2015).
15. Tran, N. T., Laub, M. T. & Le, T. B. K. SMC Progressively Aligns Chromosomal Arms in Caulobacter crescentus but Is Antagonized by Convergent Transcription. *Cell Rep* **20**, 2057–2071 (2017).
16. Tran, N. T. *et al.* Permissive zones for the centromere-binding protein ParB on the Caulobacter crescentus chromosome. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1192
17. Fisher, G. L. *et al.* The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. *Elife* **6**, (2017).
18. Jakimowicz, D., Chater, K. & Zakrzewska-Czerwi´nska, J. The ParB protein of Streptomyces coelicolor A3(2) recognizes a cluster of parS sequences within the origin-proximal region of the linear chromosome. *Molecular Microbiology* **45**, 1365–1377 (2002).
19. Harms, A., Treuner-Lange, A., Schumacher, D. & Sogaard-Andersen, L. Tracking of chromosome and replisome

dynamics in Myxococcus xanthus reveals a novel chromosome arrangement. *PLoS Genet* **9**, e1003802 (2013).
20. Murray, H., Ferreira, H. & Errington, J. The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. *Molecular Microbiology* **61**, 1352–1361 (2006).
21. Kawalek, A., Bartosik, A. A., Glabski, K. & Jagura-Burdzy, G. Pseudomonas aeruginosa partitioning protein ParB acts as a nucleoid-associated protein binding to multiple copies of a parS-related motif. *Nucleic Acids Res.* **46**, 4592–4606 (2018).
22. Sievers, J., Raether, B., Perego, M. & Errington, J. Characterization of the parB-Like yyaA Gene of Bacillus subtilis. *J. Bacteriol.* **184**, 1102–1111 (2002).
23. Wu, L. J. & Errington, J. Coordination of cell division and chromosome segregation by a nucleoid occlusion protein in Bacillus subtilis. *Cell* **117**, 915–925 (2004).
24. Wu, L. J. *et al.* Noc protein binds to specific DNA sequences to coordinate cell division with chromosome segregation. *EMBO J.* **28**, 1940–1952 (2009).
25. Pang, T., Wang, X., Lim, H. C., Bernhardt, T. G. & Rudner, D. Z. The nucleoid occlusion factor Noc controls DNA replication initiation in Staphylococcus aureus. *PLOS Genetics* **13**, e1006908 (2017).
26. Wu, L. J. & Errington, J. Nucleoid occlusion and bacterial cell division. *Nat. Rev. Microbiol.* **10**, 8–12 (2011).
27. Noyes, M. B. *et al.* A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. *Nucleic Acids Res* **36**, 2547–2560 (2008).
28. Lee, P. S. & Grossman, A. D. The chromosome partitioning proteins Soj (ParA) and Spo0J (ParB) contribute to accurate chromosome partitioning, separation of replicated sister origins, and regulation of replication initiation in Bacillus subtilis. *Mol. Microbiol.* **60**, 853–869 (2006).
29. Podgornaia, A. I. & Laub, M. T. Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. *Science* **347**, 673–677 (2015).
30. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science* **328**, 1272–1275 (2010).
31. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
32. Wang, X., Minasov, G. & Shoichet, B. K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95 (2002).
33. McKeown, A. N. *et al.* Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **159**, 58–68 (2014).
34. Starr, T. N., Picton, L. K. & Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409–413 (2017).
35. Aakre, C. D. *et al.* Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. *Cell* **163**, 594–606 (2015).
36. Raumann, B. E., Knight, K. L. & Sauer, R. T. Dramatic changes in DNA-binding specificity caused by single residue substitutions in an Arc/Mnt hybrid repressor. *Nature Structural & Molecular Biology* **2**, 1115–1122 (1995).

FIG. 1

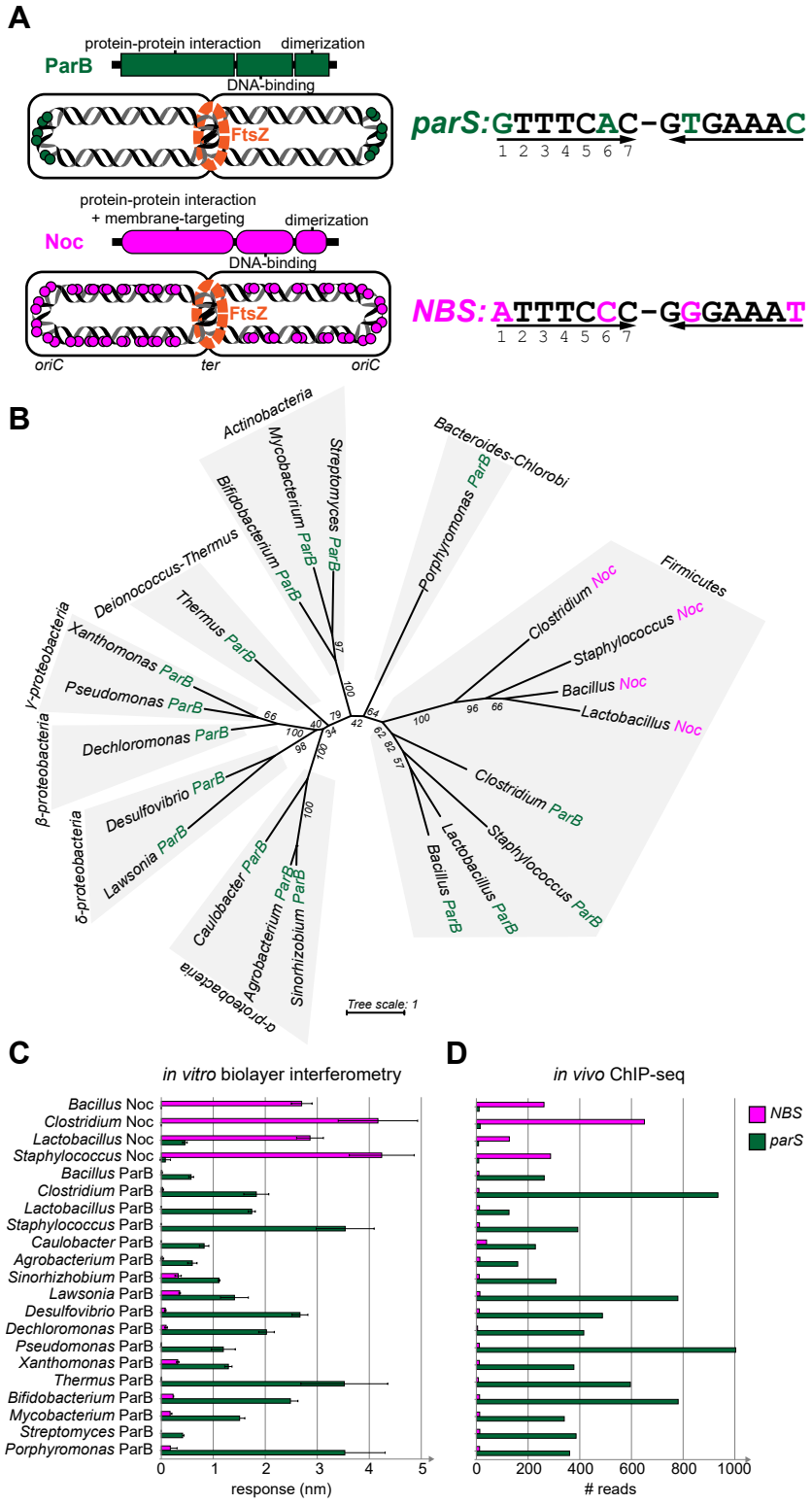


Figure 1. DNA-binding specificity for *parS* and *NBS* is conserved among ParB and Noc orthologs. (A) The domain architecture of ParB (dark green) and Noc (magenta) together with their respective cognate DNA-binding sites, *parS* and *NBS*. Sequence differences between *parS* and *NBS* are highlighted (*parS*: dark green, *NBS*: magenta). The genome-wide distributions of *parS* and *NBS* sites (dark green and magenta circles, respectively) are also shown schematically. **(B)** An unrooted maximum likelihood tree that shows the restrictive distribution of Noc orthologs (magenta branches) to the Firmicutes clade. Bootstrap support values are shown for branches. **(C-D)** *in vitro* binding affinities of ParB-*parS*/Noc-*NBS* correlate to their *in vivo* ChIP-seq enrichment and show the conservation in DNA-binding specificity within each protein family. Bio-layer interferometry was used to measure the binding affinity of ParB/Noc to *parS/NBS in vitro*. The level of protein-DNA interaction was expressed as an averaged response value (unit: nm). Error bars represent standard deviation (SD) from three replicates. For ChIP-seq data, reads in a 100-bp window surrounding the *parS/NBS* site were quantified and used as a proxy for the enrichment of immunoprecipitated *parS* or *NBS* DNA. Results of ChIP-qPCR experiments from three replicates are also shown in Fig. S1B.

FIG. 2

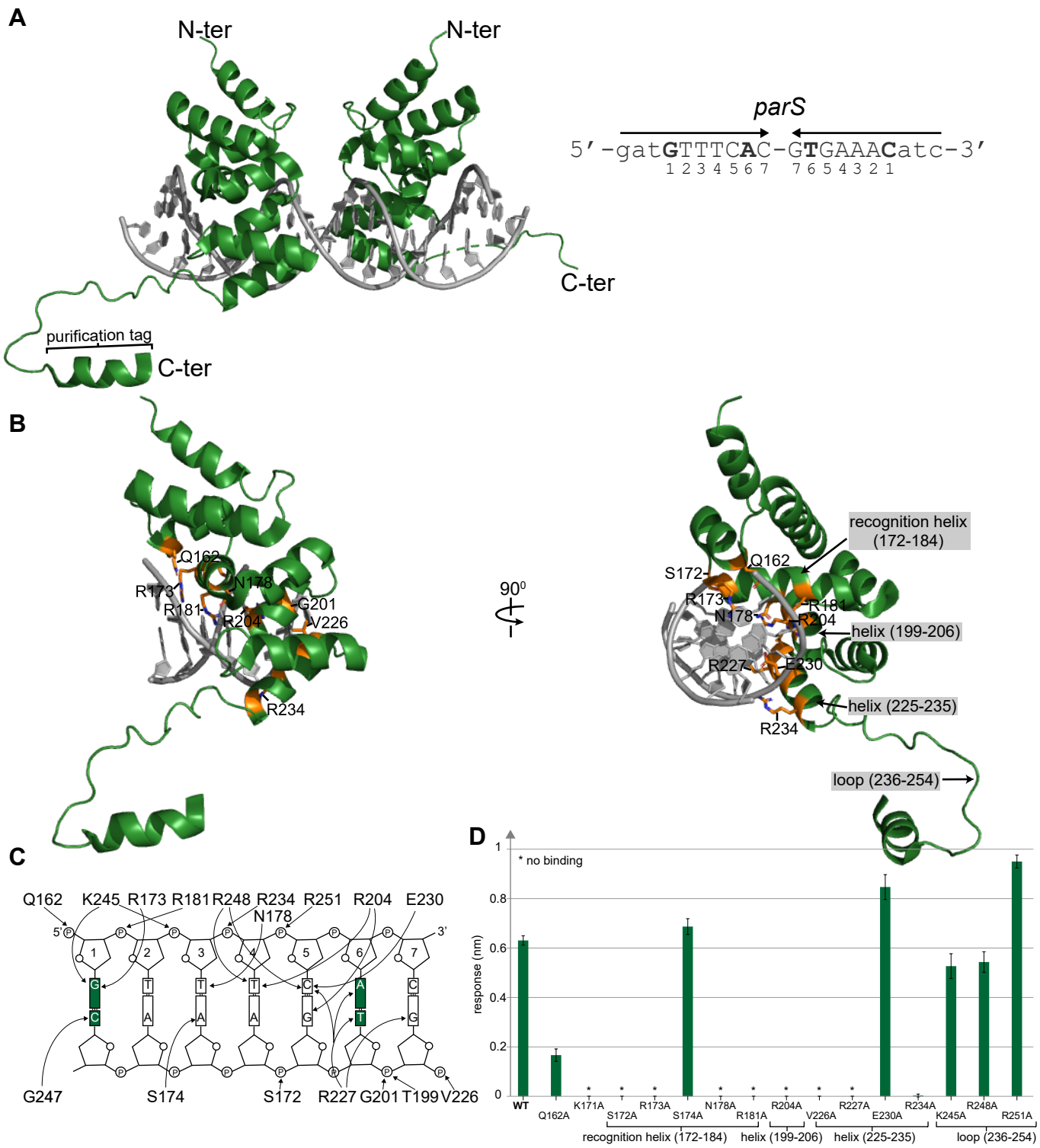


Figure 2. Co-crystal structure of the DNA-binding domain (DBD) of *Caulobacter* ParB with *parS*. **(A)** The 2.4 Å resolution structure of two ParB (DBD) monomers (dark green) in complex with a 20-bp *parS* DNA (grey). The nucleotide sequence of the 20-bp *parS* is shown on the left-hand side; bases (Guanine 1 and Adenine 6) that are different from *NBS* are in bold. The purification tag is also visible in one of the DBD monomers. **(B)** One monomer of ParB (DBD) is shown in complex with a *parS* half-site; residues that contact the DNA are labeled and colored in orange. **(C)** Schematic representation of ParB (DBD)-*parS* interactions. For simplicity, only a *parS* half-site is shown. The two bases at position 1 and 6 that are different between *parS* and *NBS* are highlighted in dark green. **(D)** Alanine scanning mutagenesis and the *in vitro* binding affinities of mutants to *parS* DNA. The level of protein-DNA interaction was expressed as an averaged response value. Error bars represent SD from three replicates.

FIG. 3

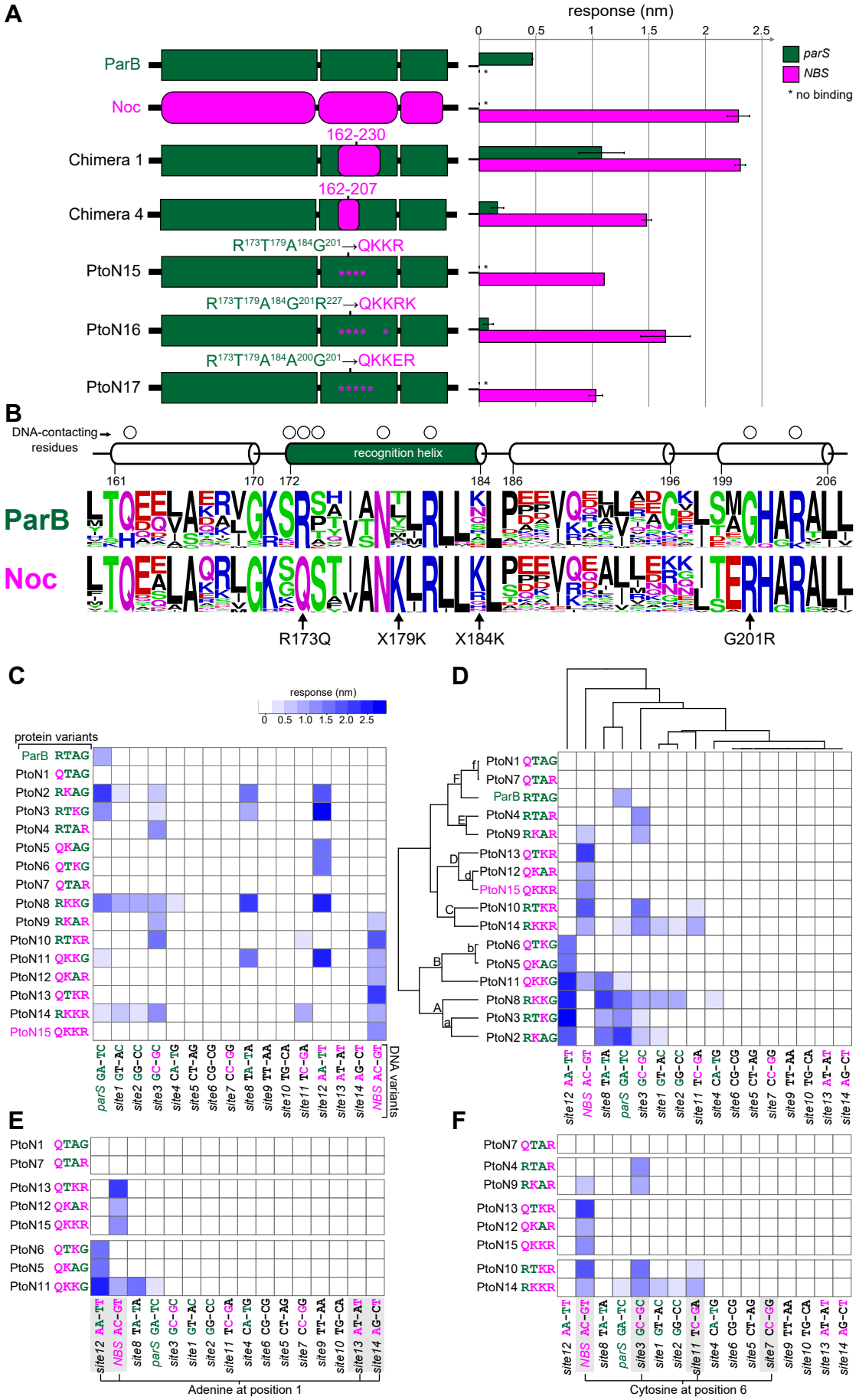


figure legend on the next page

FIG. 3

Figure 3. Four amino acid residues dictate interaction specificity between Noc and *NBS*. (A) Mutations in a subset of residues in the region between res. 162-207 can reprogram interaction specificity. ParB (or segments of amino acids from ParB) and Noc (or equivalent segment in Noc) are shown in dark green and magenta, respectively. The level of protein-DNA interaction was expressed as an averaged response value. Error bars represent SD from three replicates. **(B)** The sequence alignment of ParB (~1800 sequences) and Noc (~400 sequences) orthologs. Amino acids are colored based on their chemical properties (GSTYC: polar; QN: neutral; KRH: basic; DE: acidic; and AVLIPWFM: hydrophobic). The secondary structure of the amino acid region (res. 162-207) is shown above the sequence alignment, together with residues (open circles) that contact DNA in the ParB (DBD)-*parS* structure (Fig. 2). **(C)** Systematic scanning mutagenesis of the protein-DNA interface reveals the contribution of each specificity residue to the DNA-binding preference. Interactions between ParB + 15 PtoN intermediates with 16 DNA sites are represented as a heatmap where each matrix position reflects a response value from our *in vitro* bio-layer interferometry assays. Three replicates for each pairwise interaction were performed, and an averaged response value is presented. Amino acid residues/bases from ParB/*parS* are colored in dark green, and those from Noc/*NBS* in magenta. **(D)** A hierarchical clustering of data in panel **C** in both protein and DNA dimensions. **(E)** A simplified heatmap where only PtoN intermediates with a glutamine (Q) at position 173 are shown. **(F)** A simplified heatmap where only PtoN intermediates with an arginine (R) at position 201 are shown.

FIG. 4

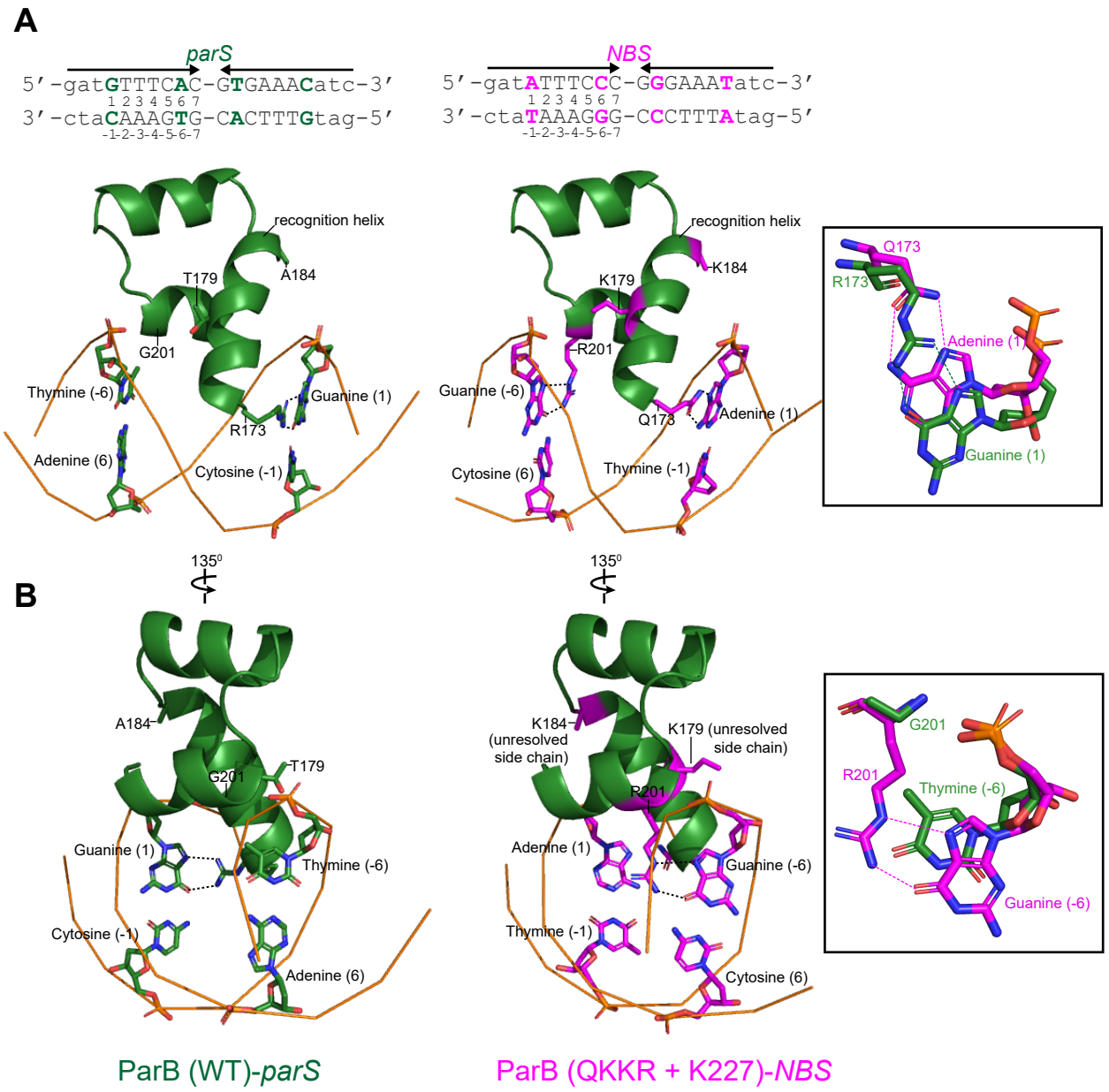


Figure 4. Superimposition of the wild-type ParB-*parS* structure on the ParB (QKKR+K227)-*NBS* structure reveals the contribution of specificity residues to *NBS* binding. To simplify and highlight the roles of specificity residues, only the side chains of specificity residues and their contacting bases are shown. The amino acid region (173-207) and the DNA backbones are shown in cartoon representation. DNA bases are numbered according to their respective positions on *parS/NBS* site. The insets show interactions between either Q/R at position 173 **(A)** or G/R201 **(B)** with their corresponding bases on *parS/NBS*.

FIG. 5

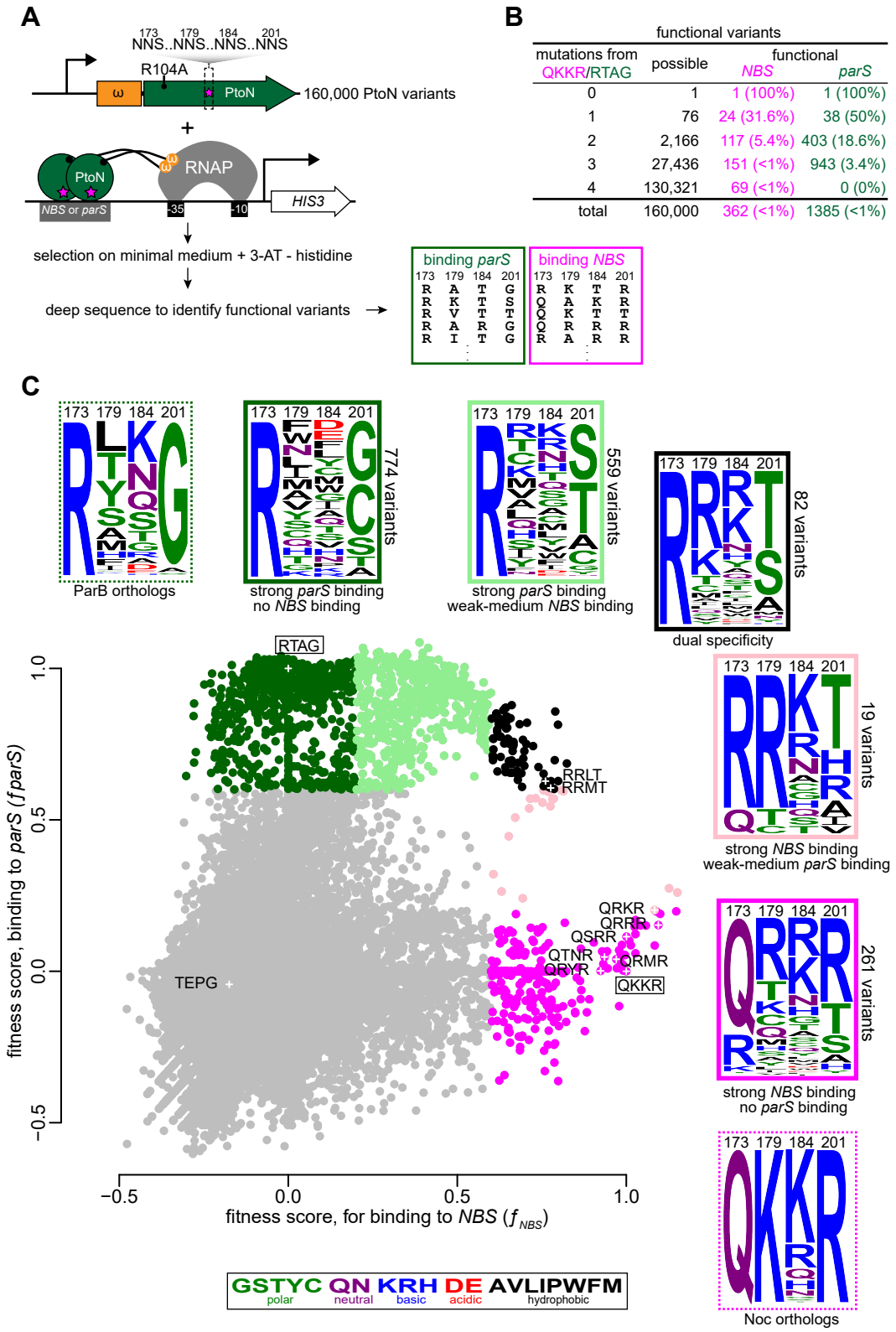


Figure 5. High-throughput mapping of the fitness of protein-DNA interface mutants. (A) The principle and design of the deep mutational scanning experiment which was based on bacterial one-hybrid assay and high-throughput sequencing. **(B)** Summary of functional *NBS*-binding and *parS*-binding variants. **(C)** Fitness scores of variants as assessed by their ability to bind *NBS* (x-axis) or *parS* (y-axis). Dark green: strong *parS* binding, no *NBS* binding (fitness score: f_{parS}≥0.6, f_{NBS}≤0.2); light green: strong *parS* binding, weak-to-medium *NBS* binding (f_{parS}≥0.6, 0.2≤f_{NBS}≤0.6); magenta: strong *NBS* binding, no *parS* binding (f_{NBS}≥0.6, f_{parS}≤0.2); pink: strong *NBS* binding, weak-to-medium *parS* binding (f_{NBS}≥0.6, 0.2≤f_{parS}≤0.6); black: dual specificity (f_{NBS}≥0.6, f_{parS}≥0.6). Frequency logos of each class of variants are shown together with ones for ParB/Noc orthologs. Amino acids are colored according to their chemical properties. The positions of WT ParB (RTAG), Noc (QKKR), and nine selected variants for an independent validation are also shown and labeled on the scatterplot.

FIG. 6

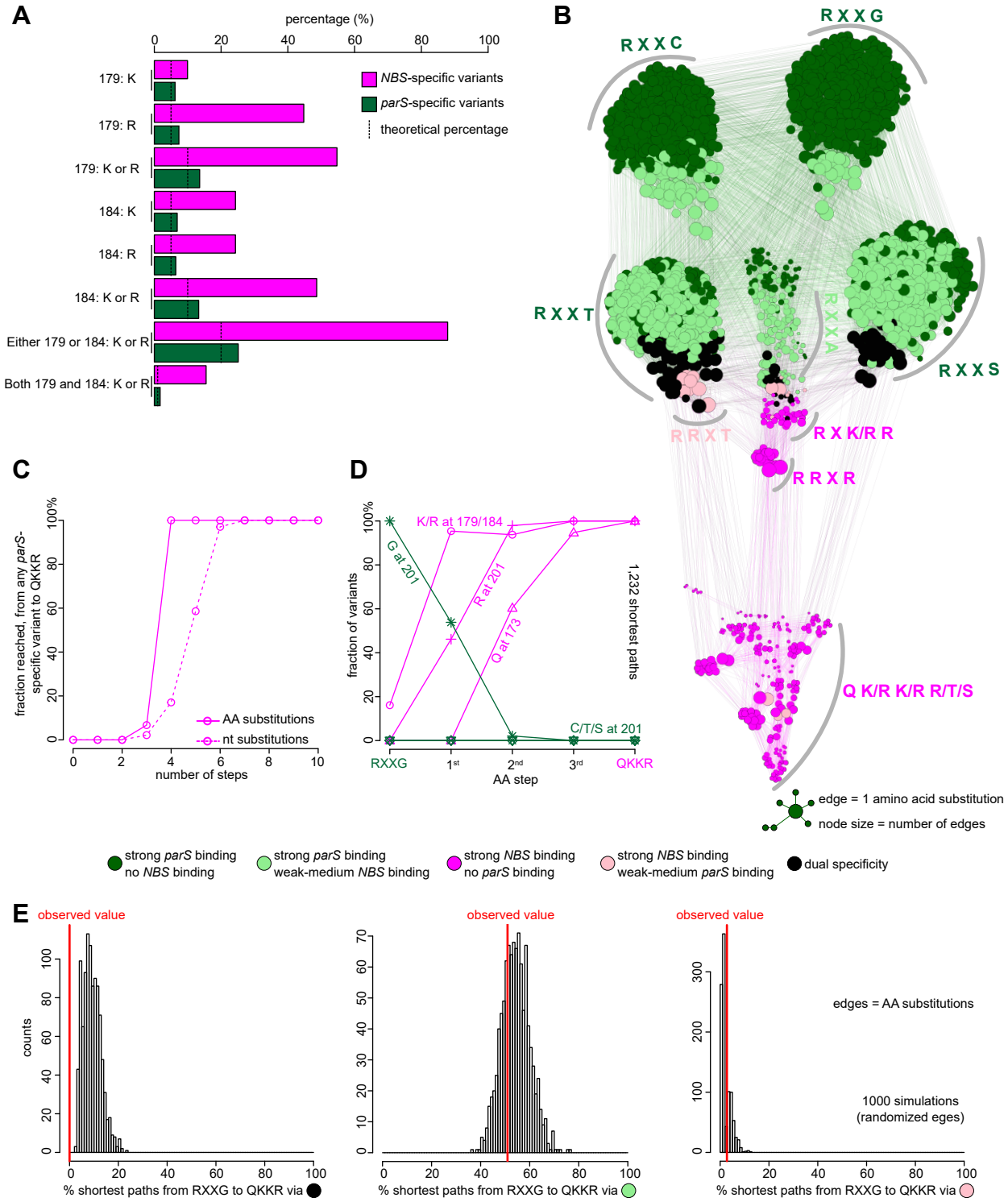


Figure 6. Deep mutational scanning experiments reveals the common properties of the mutational paths to a new DNA-binding specificity. (A) Fractions of arginine or lysine residues at either position 179, 184, or both, in *parS*-specific (dark green) and *NBS*-specific (magenta) variants. The dotted lines indicate the expected percentage if arginine/lysine was chosen randomly from 20 amino acids. **(B)** A force-directed network graph connecting strong *parS*-binding variants to strong *NBS*-binding variants. Nodes represent individual variants, and edges represent single amino acid (AA) substitutions. Node sizes are proportional to their corresponding numbers of edges. Node colors correspond to different classes of variants. **(C)** Cumulative fraction of highly *parS*-specific variants that reached an *NBS*-specific QKKR variant in a given number of amino acid (solid line) or nucleotide (dotted line) substitutions (see also Fig. S7A). **(D)** Fraction of intermediates on all shortest paths from highly *parS*-specific RXXG variants to the *NBS*-preferred QKKR that have permissive amino acids (K/R) at either position 179/184 or both, or have R at position 201, or Q at position 173, or C/T/S at position 201 after a given number of AA steps (see also Fig. S7D). **(E)** Percentage of shortest paths that traversed black, light green, or pink variants to reach QKKR from any of the highly *parS*-specific RXXG variants (red lines). The result was compared to ones from 1,000 simulations where the edges were shuffled randomly while keeping the total number of nodes, edges, and graph density constant.

FIG. 7

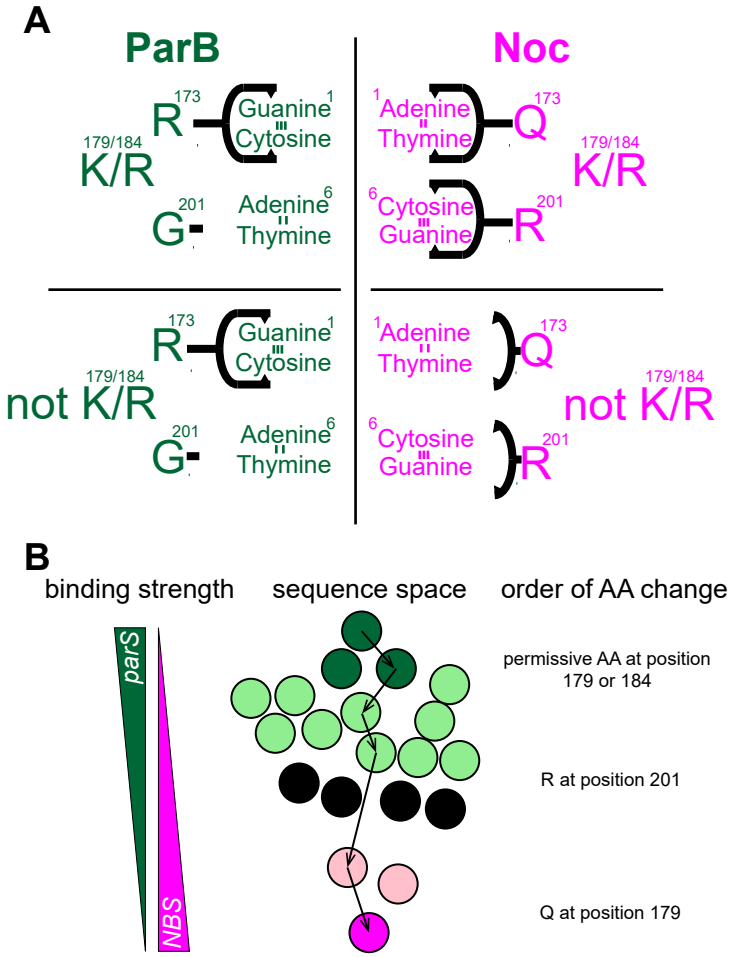


Figure 7. A model for the evolution of *NBS*-binding specificity. (A) Contributions of each specificity residue to enable a switch in binding specificity from *parS* to *NBS*. An R173Q substitution enabled interactions with Adenine 1:Thymine -1 (of *NBS*). A G201R substitution enabled interactions with Cytosine 6: Guanine -6 (of *NBS*). Q173 and R201 could only do so in the presence of permissive residues K at either 179, 184, or both. Without K179/184, Q173 and R201 were poised to interact with specific bases but could not, possibly because of insufficient DNA affinity. **(B)** Analysis of mutational paths that traversed the network of functional variants showed that the order of introducing specificity-switching substitutions matters, and that the shortest paths to *NBS*-specific variants do not necessarily involve a dual-specificity nodes to evolve a new DNA-binding preference.

FIG. S1

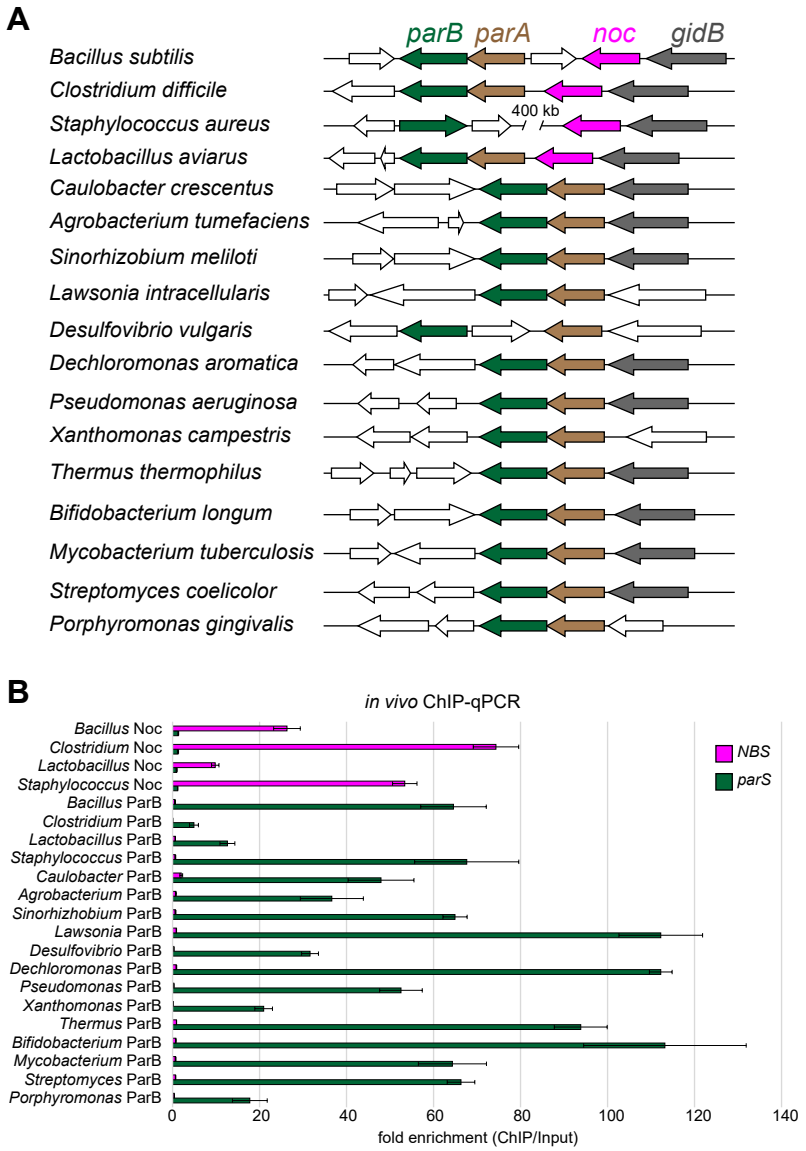


Figure S1. DNA-binding specificity for *parS* and *NBS* is conserved among ParB and Noc orthologs. (A) Genomic context of ParB- and Noc-encoding genes in various bacterial species. *parB*, *parA*, *noc*, and the highly conserved *gidB* gene, are colored in dark green, brown, magenta, and grey, respectively. Genes at the border of the *parB-parA-noc* cluster (open arrows) vary between bacterial species. **(B)** The *in vivo* binding preferences of ParB-*parS*/Noc-*NBS.* Results of ChIP-qPCR experiments from three independent replicates are also shown in Fig. S1B. Error bars represent standard deviation (SD) from three replicates.

FIG. S2

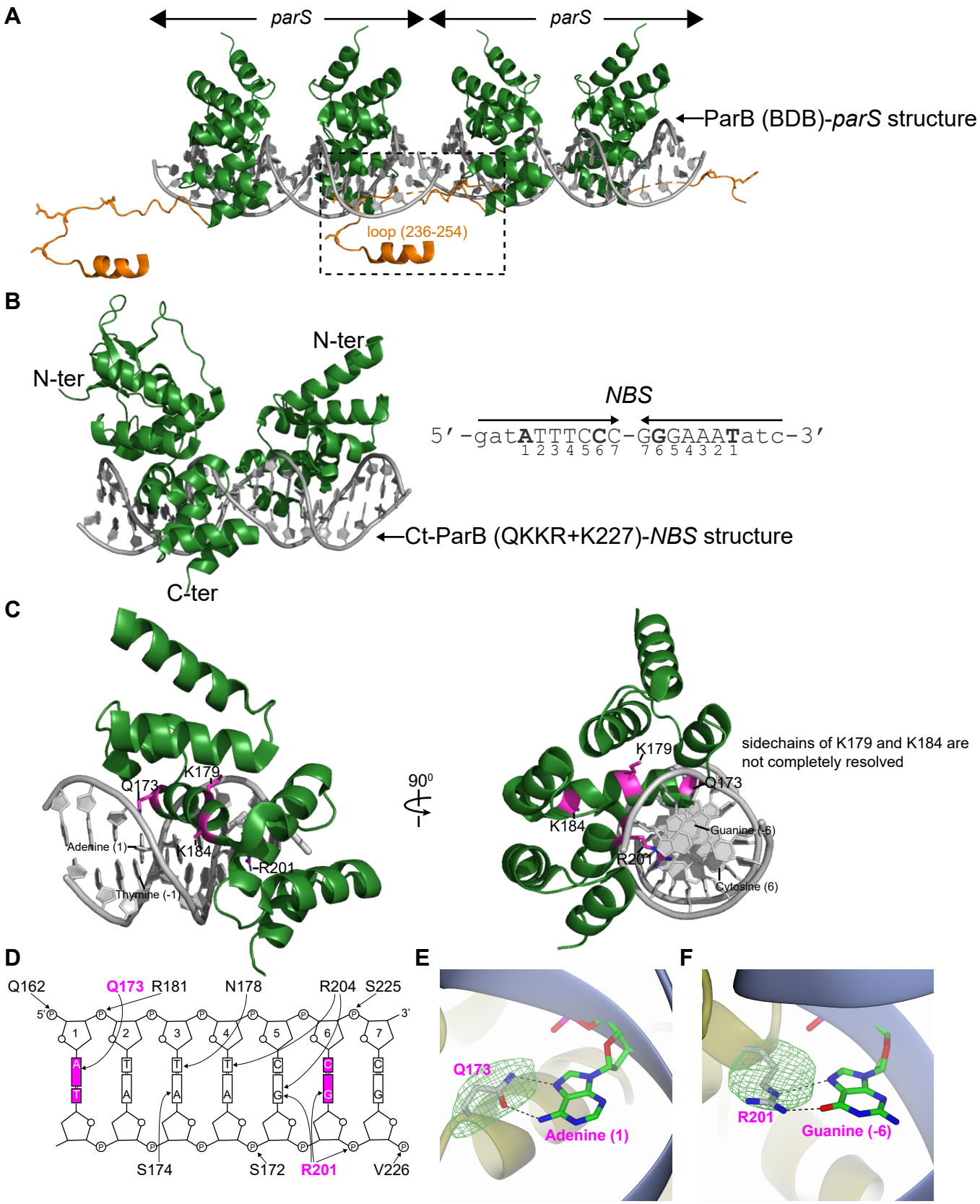


figure legend on the next page

FIG. S2

Figure S2. Co-crystal structures of the ParB (DBD)-*parS* complex and of the *NBS*-preferred *Caulobacter* Ct-ParB variant with *NBS*. (A) Interactions between the loop (res. 236-254) (orange) and the minor groove of *parS* DNA from an adjacent ParB (DBD)-*parS* complex in the crystal lattice (dashed box). DBD stands for DNA-binding domain. **(B)** The 3.6 Å structure of two C-terminally truncated Ct-ParB (QKKR + K227) monomers (dark green) in complex with a 20-bp *NBS* DNA (grey). Each Ct-ParB (QKKR + K227) monomer contains the N-terminal domain and the central DNA-binding domain, but lacks the C-terminal dimerization domain. The nucleotide sequence of the 20-bp *NBS* site is shown on the left-hand side; bases (Adenine 1 and Cytosine 6) that are different from *parS* are in bold. **(C)** One monomer of ParB (QKKR + K227) (DBD) is shown in complex with an *NBS* half-site; only the specificity residues are labeled and colored in magenta. **(D)** Schematic representation of ParB (QKKR + K227)-*NBS* interactions. For simplicity, only half of *NBS* is shown. The two bases at position 1 and 6 that are different between *parS* and *NBS* are highlighted in magenta. (**E-F**) Omit electron density for mutated residues in the Ct-ParB (QKKR+K227)-*NBS* complex. For two of the substitutions, namely R173Q and G201R, it was possible to build in the newly introduced side-chains with confidence. Shown are 3.6 Å resolution omit mF_{obs}-DF_{calc} difference electron density maps for the side-chains calculated using phases from the final model without the side-chains present and re-refining (shown at 3.0σ in green mesh).

FIG. S3

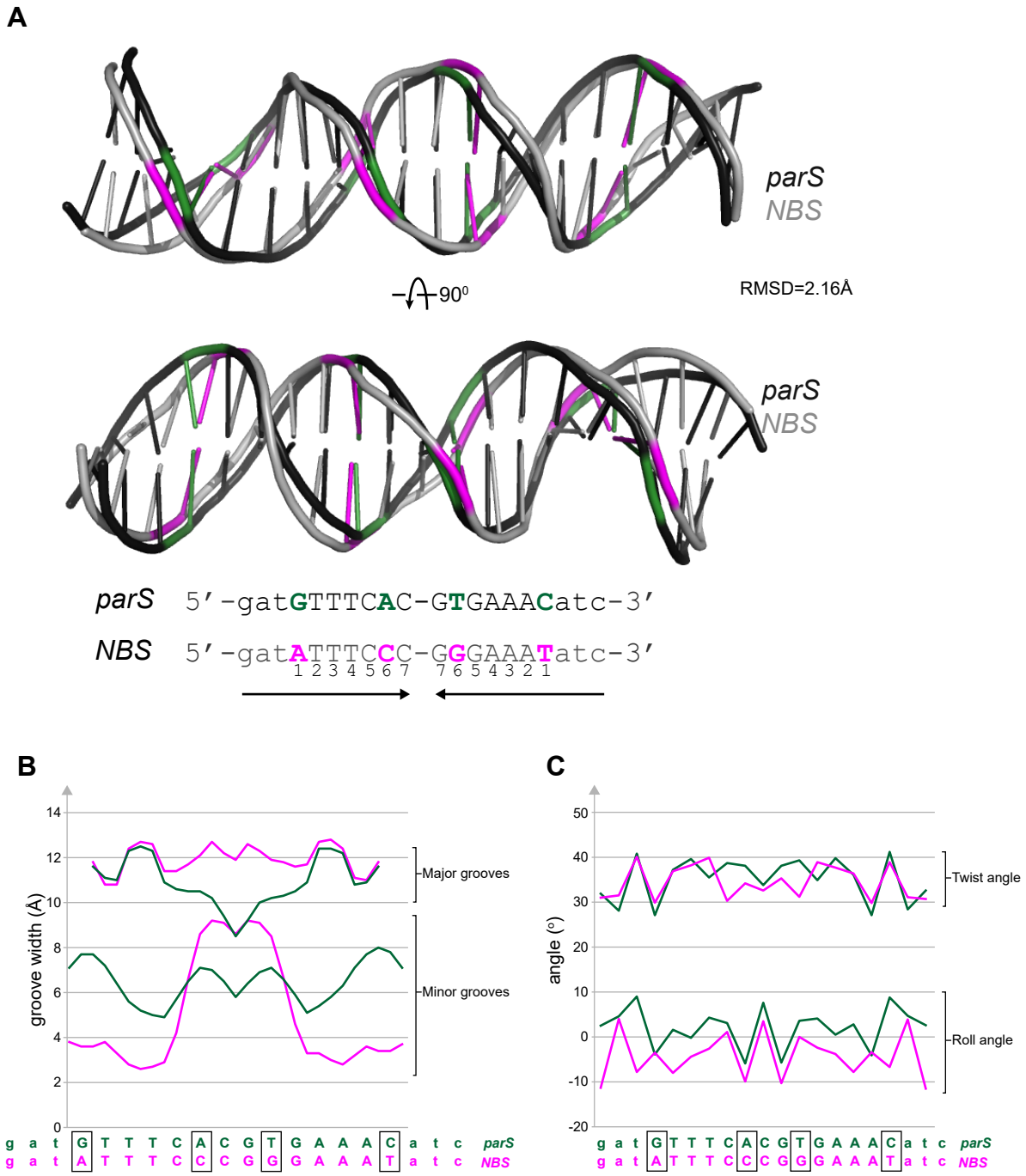


Figure S3. Conformational changes at *parS* and *NBS* DNA within the two co-crystal structures. (A) A superimposition of *parS* and *NBS* DNA structures, root-mean-square deviation (RMSD) value is also shown. Bases that differ between *parS* (dark green) and *NBS* (magenta) are highlighted. **(B)** The major and minor groove widths of the bound DNA (*parS*: dark green, *NBS*: magenta). **(C)** The roll and twist angles for each base pair step of the bound DNA (*parS*: dark green, *NBS*: magenta).

FIG. S4

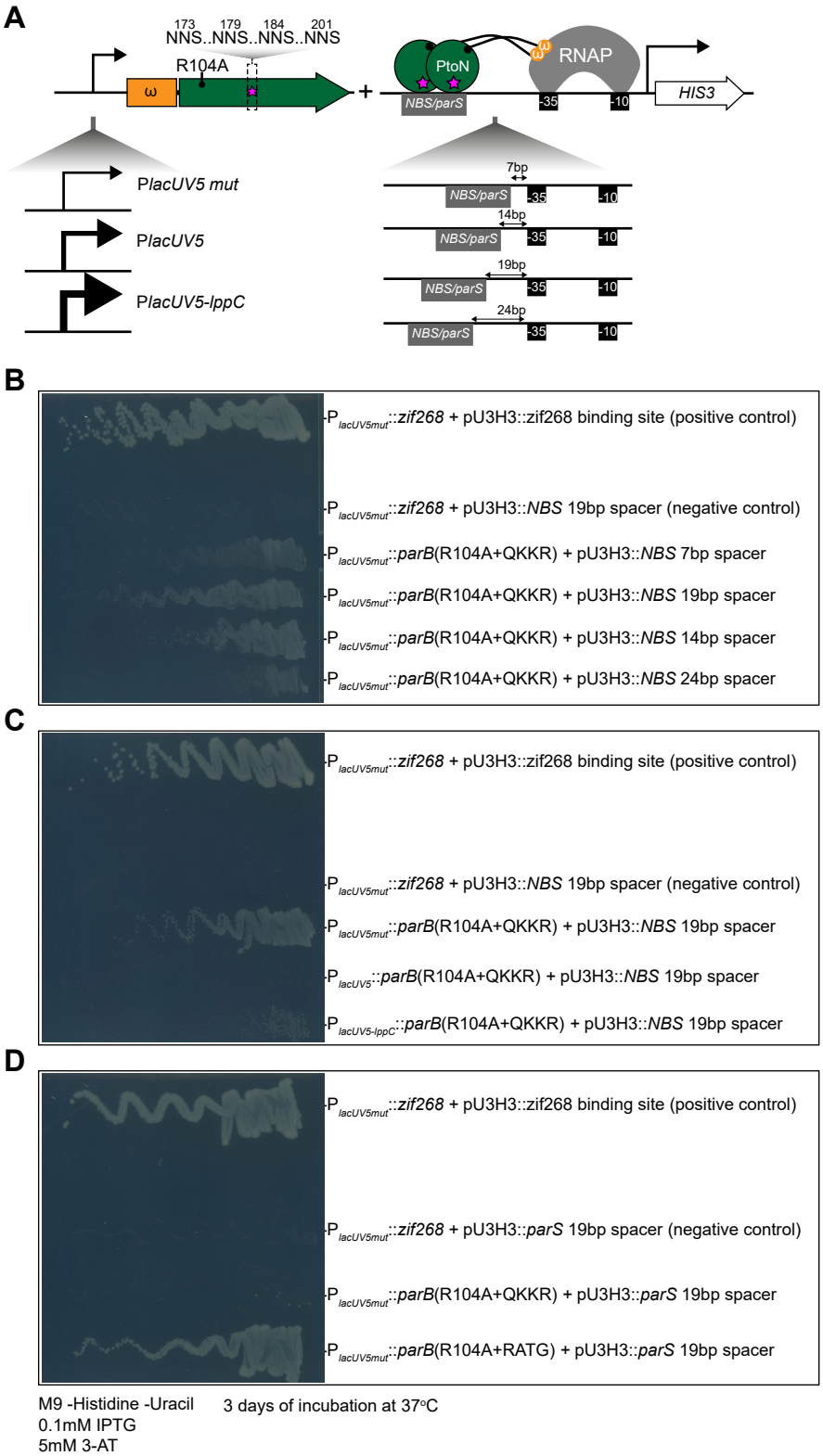
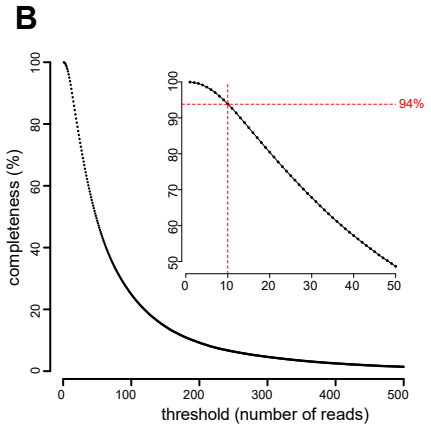


Figure S4. Optimization of bacterial one-hybrid (B1H) assay to select for variants that bind *parS* or *NBS*. (A) The strength of the promoter that drives the expression of *parB* variants and the distances between the *parS* or *NBS* binding site to the core -10 -35 promoter were optimized. **(B)** A 19-bp gap between *NBS/parS* and the core promoter is optimal, based on the streak test for cell growth in a minimal medium lacking histidine. **(C)** A weak promoter ($P_{lacUV5mut}$) is optimal, based on the streak test for cell growth in a minimal medium lacking histidine. **(B and D)** The presence of *parS* or *NBS* upstream of HIS3, in the absence of ParB variants, did not auto-activate its expression.

FIG. S5

A

	number of quality-filter reads
Pre-selection, replicate 1	6,954,115
Pre-selection, replicate 2	8,623,273
Pre-selection, replicate 3	7,201,395
Total	22,778,783
Post-selection NBS, replicate 1	6,294,658
Post-selection NBS, replicate 2	7,487,037
Post-selection NBS, replicate 3	6,689,637
Total	20,471,332
Post-selection parS, replicate 1	5,380,569
Post-selection parS, replicate 2	8,572,521
Post-selection parS, replicate 3	7,269,057
Total	21,222,147



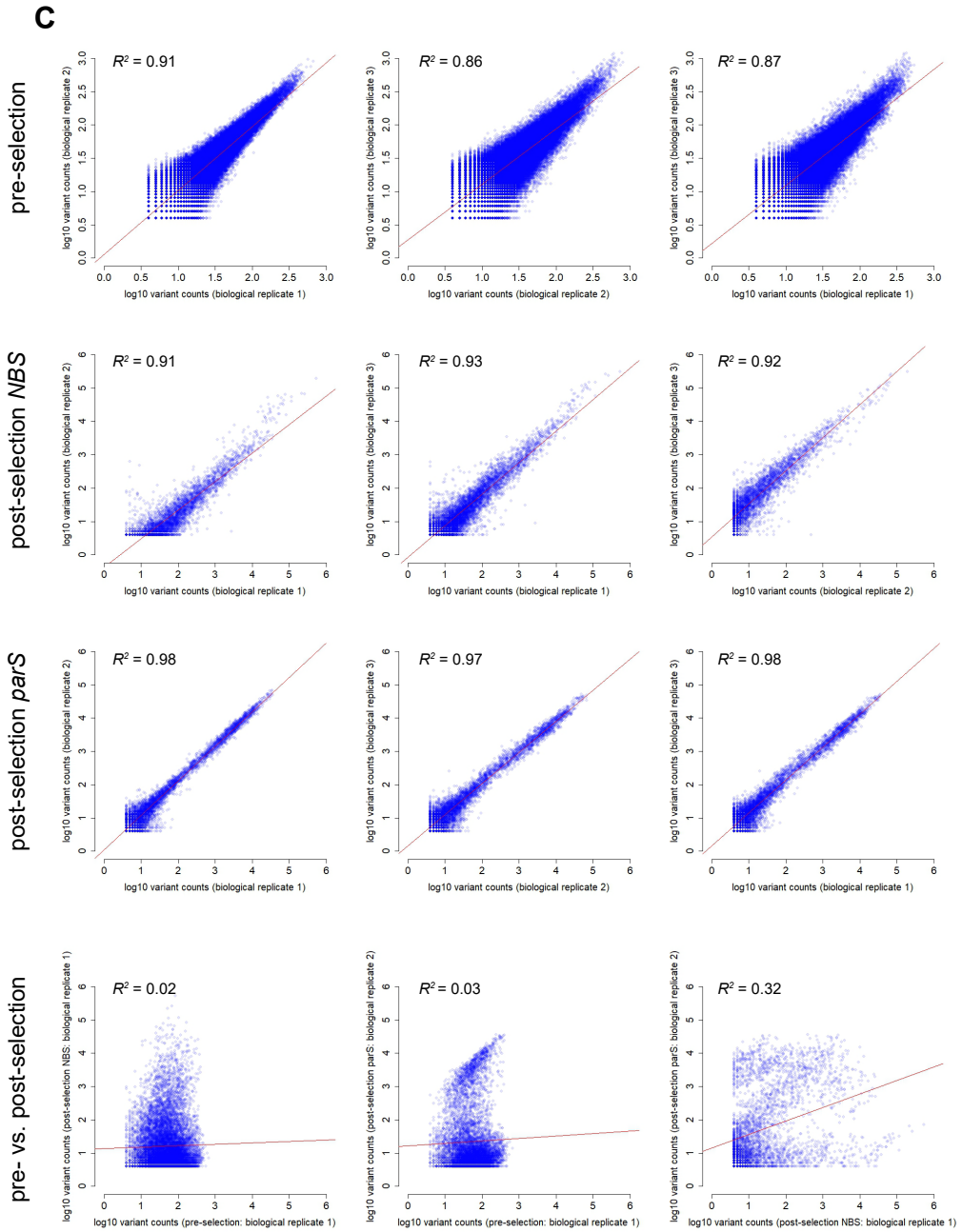


Figure S5. Statistics of deep-sequencing reads and completeness of starting libraries. (A) Number of quality-filtered reads for each biological replicate, for pre- and post-selection libraries. **(B)** The completeness of pre-selection libraries at different thresholds. A completeness of 100% means that all 160,000 variants lacking stop codons were present in the pre-selection library. In the starting library, greater than 94% of the predicted variants were represented by at least 10 reads. **(C)** Reproducibility of biological replicates: pre- vs. pre-selection replicates and pre- vs. post-selection replicates. Pearson's correlation coefficients (R^2) are also shown. Red lines show least squares best fits.

FIG. S6

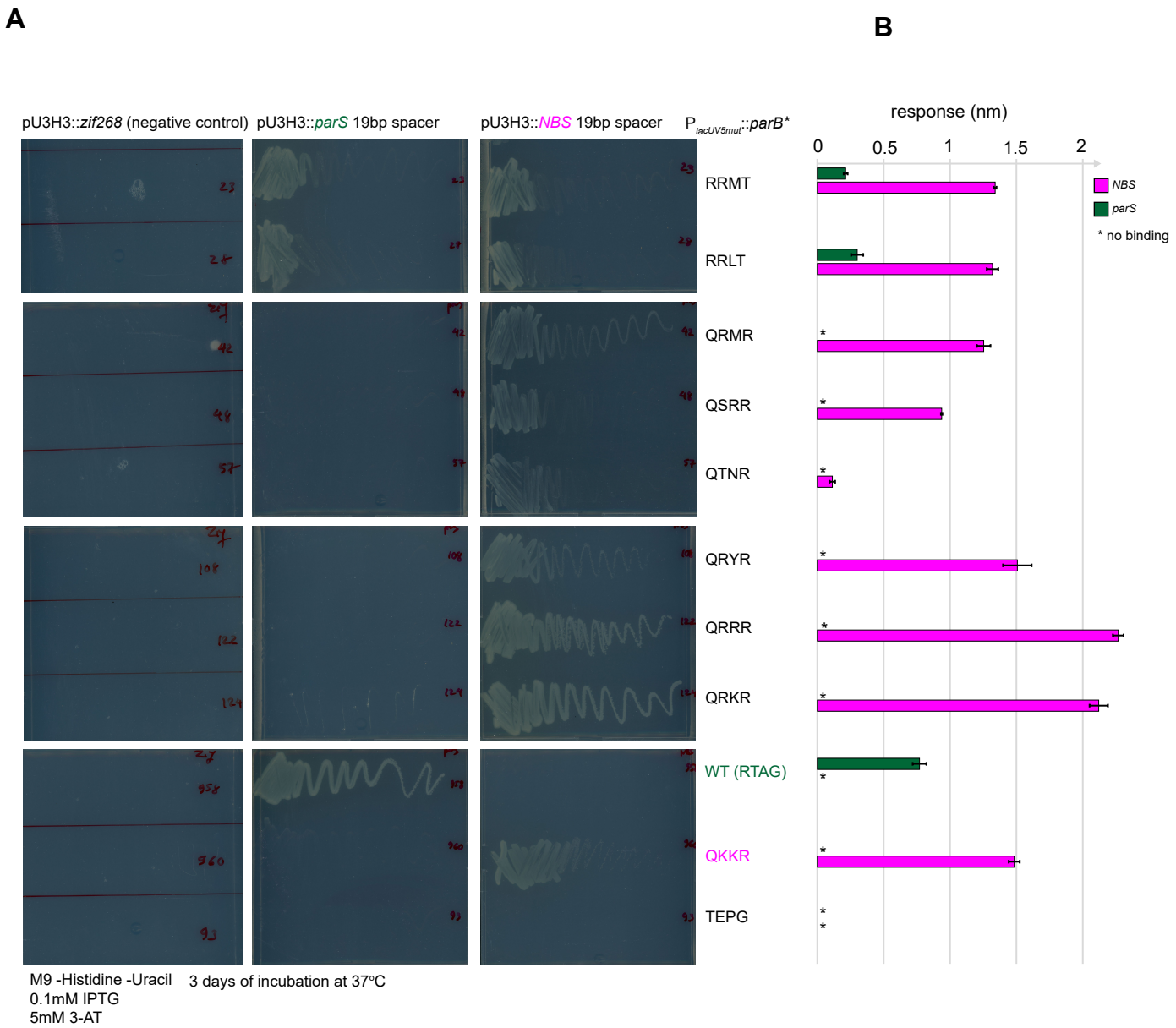


Figure S6. Validation of selected variants from the deep mutational scanning experiments. (A) Validation by pairwise bacterial one-hybrid assays. The ability of nine selected variants to grow on a minimal medium lacking histidine (but supplemented with 5mM 3-AT to increase the stringency) was assessed by a streak test. Plasmid harboring a binding site of an eukaryotic transcription factor (*zif268*) served as a negative control. **(B)** Validation by bio-layer interferometry assays. Selected variants were expressed, purified, and subsequently analyzed in an *in vitro* bio-layer interferometry assay that directly assess their binding to a *parS*-containing or *NBS*-containing duplex DNA. Three replicates for each pairwise interaction were performed, and an averaged response value is presented.

FIG. S7

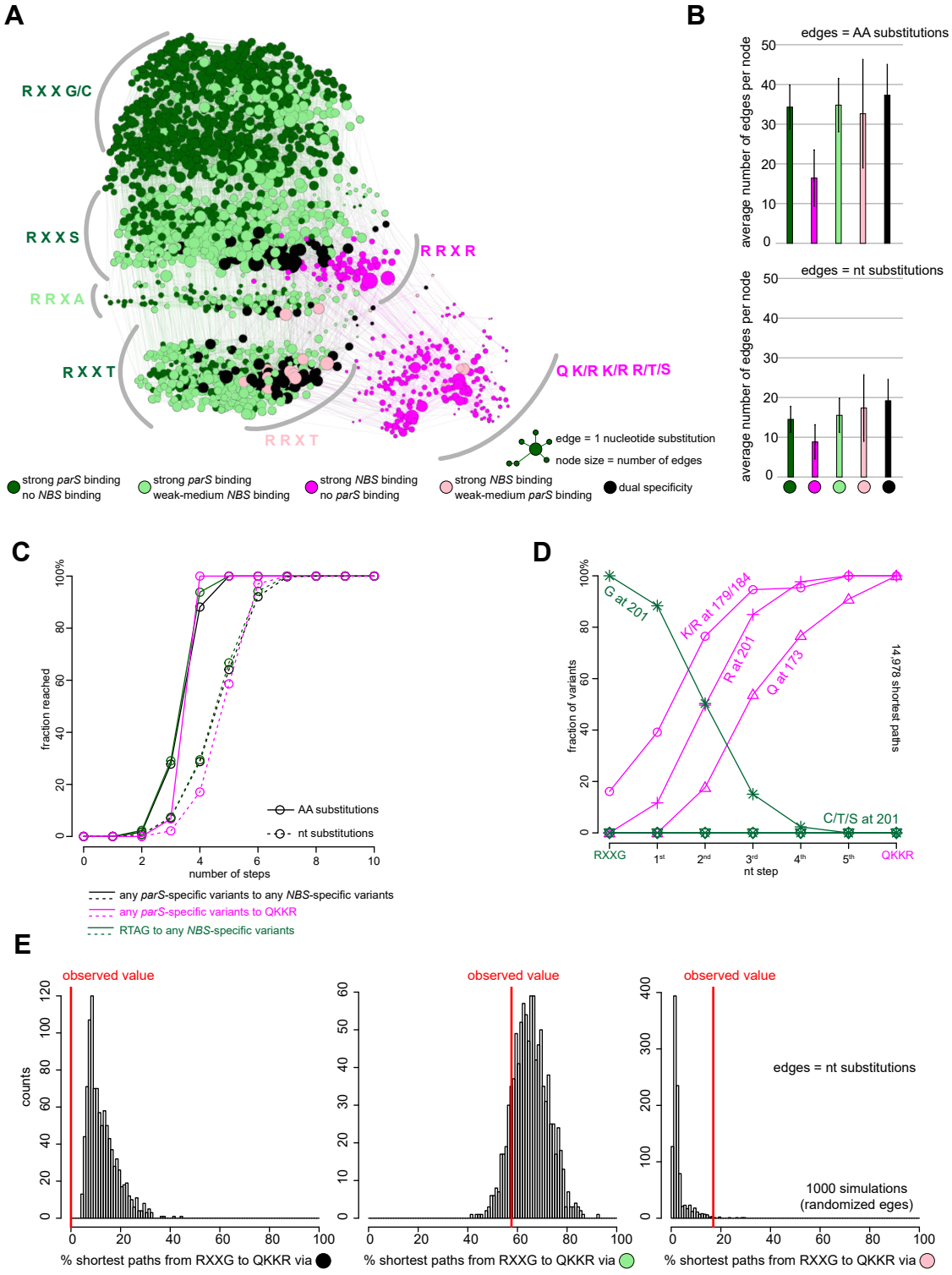


Figure S7. Deep mutational scanning experiments reveal the common properties of mutational paths. (A) A force-directed network graph connecting strong *parS*-binding variants to strong *NBS*-binding variants. Nodes represent individual variants, and edges represent single nucleotide (nt) substitutions. Node sizes are proportional to their corresponding numbers of edges. Node colors correspond to different classes of variants. **(B)** Average number of edges per node. **(C)** Cumulative fraction of variants that reached their destinations in a given number of amino acid (solid line) or nucleotide (dotted line) substitutions. Black lines: from any *parS*-specific variants to any *NBS*-specific variants. Magenta lines: from any *parS*-specific variants to QKKR. Dark green lines: from RTAG to any *NBS*-specific variants. **(D)** Fraction of intermediates on all shortest paths from highly *parS*-specific RXXG variants to the *NBS*-preferred QKKR that have permissive amino acids (K/R) at either position 179/184 or both, or have R at position 201, or Q at position 173, or C/T/S at position 201 after a given number of nt steps. **(E)** Percentage of shortest paths that traversed black, light green, or pink variants to reach QKKR from any of the highly *parS*-specific RXXG variants (red lines). The result was compared to ones from 1,000 simulations where the by-nt-substitution edges were shuffled randomly while keeping the total number of nodes, edges, and graph density constant.