6930–6942 *Nucleic Acids Research, 2010, Vol. 38, No. 20* *Published online 28 June 2010*
doi:10.1093/nar/gkq465

When simple sequence comparison fails: the cryptic case of the shared domains of the bacterial replication initiation proteins DnaB and DnaD

Farhat Y. Marston[1], William H. Grainger[2], Wiep Klaas Smits[3], Nicholas H. Hopcroft, Matthew Green[1], Andrea M. Hounslow[1], Alan D. Grossman[3], C. Jeremy Craven[1,*] and Panos Soultanas[2,*]

[1]Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, [2]School of Chemistry, Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK and [3]Department of Biology, Building 68-530, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received March 22, 2010; Revised May 8, 2010; Accepted May 11, 2010

ABSTRACT

DnaD and DnaB are essential DNA-replication-initiation proteins in low-G+C content Gram-positive bacteria. Here we use sensitive Hidden Markov Model-based techniques to show that the DnaB and DnaD proteins share a common structure that is evident across all their structural domains, termed DDBH1 and DDBH2 (DnaD DnaB Homology 1 and 2). Despite strong sequence divergence, many of the DNA-binding and oligomerization properties of these domains have been conserved. Although eluding simple sequence comparisons, the DDBH2 domains share the only strong sequence motif; an extremely highly conserved YxxxIxxxW sequence that contributes to DNA binding. Sequence alignments of DnaD alone fail to identify another key part of the DNA-binding module, since it includes a poorly conserved sequence, a solvent-exposed and somewhat unstable helix and a mobile segment. We show by NMR, *in vitro* mutagenesis and *in vivo* complementation experiments that the DNA-binding module of *Bacillus subtilis* DnaD comprises the YxxxIxxxW motif, the unstable helix and a portion of the mobile region, the latter two being essential for viability. These structural insights lead us to a re-evaluation of the oligomerization and DNA-binding properties of the DnaD and DnaB proteins.

INTRODUCTION

Bacillus subtilis DNA replication initiates at a defined chromosomal origin, *oriC*, where the highly conserved master replication initiator protein DnaA binds to recruit sequentially DnaD, DnaB and the DnaI–DnaC complex (1–6). The hexameric replicative ring helicase DnaC [the homologue of *Escherichia coli* DnaB (beware of the confusing nomenclature of bacterial primosomal proteins: *B. subtilis* DnaB is a distinct protein from the helicases typified by *E. coli* DnaB)] is loaded in an ATP-dependent reaction onto the locally melted AT-rich region by the DnaB–DnaI dual helicase-loading system (7–9). The DnaD, DnaB and DnaI proteins are also components of the replication restart primosome at stalled and collapsed replication forks outside *oriC* (3,10).

Bacillus subtilis DnaD exhibits a complex range of protein oligomerization and DNA recognition behaviors. It binds single stranded (ss) and double stranded (ds) DNA (2), remodels supercoiled DNA by eliminating writhe through increasing negative twist while maintaining an unchanged linking number (11–14). Atomic Force Microscopy indicates that upon binding to DNA, DnaD forms large scaffolds which open up supercoiled DNA (11,12). The isolated DnaD N-terminal domain (DnaD-Nd; DnaD residues 1–128) readily forms aggregates (and crystallises) in the absence of DNA and this has led to the suggestion that the DnaD C-terminal domain (DnaD-Cd; DnaD residues 129–232) masks the aggregation surface of DnaD-Nd: upon binding of DnaD-Cd to DNA this surface is unmasked and oligomerization of DnaD-Nd is triggered (15,16). The

*To whom correspondence should be addressed. Tel: +44 114 222 4323; Fax: +44 114 272 2800; Email: c.j.craven@shef.ac.uk
Correspondence may also be addressed to Panos Soultanas. Tel: +44 115 951 3525; Fax: +44 115 846 8002;
Email: panos.soultanas@nottingham.ac.uk

DNA-remodelling activity of DnaD is the sum of a scaffold-forming activity in its N-terminal domain, DnaD-Nd (15,16), and DNA-binding and DNA-induced oligomerization activities in its C-terminal domain, DnaD-Cd (15).

Bacillus subtilis DnaB also binds ss and dsDNA (2), remodels supercoiled and linear DNA by lateral compaction (12), cooperates as a helicase co-loader with DnaI (the homologue of *E. coli* DnaC) (7,9), acts as an origin attachment protein (17–21), regulates the recruitment of DnaD to the membrane-attached *oriC* (22) and participates in regulation of replication initiation synchrony (23). Trypsin proteolysis resolves *B. subtilis* DnaB into three domains: an N-terminal domain (residues 1–184), a central domain (residues 185–296) and a C-terminal domain (residues 297–432) with an unstructured region at the C-terminal end (residues 434–472) (9). Residues 1–300 (including the N-terminal and central domains) mediate dimerization, tetramerization and ssDNA binding, and residues 365–428 (in the C-terminal domain) mediate higher order oligomerization and ss and dsDNA binding (24). The proteolytically sensitive C-terminus of DnaB seems to regulate higher order oligomerization and association with *oriC in vivo* (24).

Although they do not appear to interact with each other in a yeast two-hybrid assay (22,25,26) the two proteins are needed together to bind to SSB-coated ssDNA, indicating a potential cryptic protein–protein interaction (3). Furthermore, the DnaB-S371P and the DnaD-A166TΔD154Q155 mutant proteins interact with wild-type DnaD and DnaB, respectively, in yeast two-hybrid assays, indicating that a conformational change may be required to expose the cryptic interaction interface (3,22,27). A genetic link between the two proteins has been established. The *dnaB75* (*dnaBS371P*) suppresses the temperature sensitive phenotypes caused by *dnaD23ts* and *dnaB134ts* (3,22,27). It is clear, therefore, that there is a close functional relationship between the two proteins. Work here indicates that this functional relationship is underpinned by a sequence and structural relatedness.

The structure of the N-terminal domain of *B. subtilis* DnaD has been determined by X-ray crystallography (16,28,29), revealing a winged helix-turn-helix fold. Whilst proteins with such a motif commonly, but not exclusively, have a DNA-binding function, this is not the case for the DnaD N-terminal domain, which instead facilitates formation of extensive protein scaffolds (14–16). The DNA-binding properties of DnaD are confined to the C-terminal domain (14,15). An extension at the N-terminus comprising a helix-strand-helix (H1′-S1′-H2′) mediates dimerization and tetramerization via two-stranded β-sheets formed by pairing of the strands in these extensions from two and four molecules, respectively. Helix (H3′) attached at the C-terminal end of the winged helix-turn-helix core appears to be involved in higher order cross-tetramer oligomerization (16).

Two structures of proteins homologous to the C-terminal domain of *B. subtilis* DnaD have been deposited by the Midwest Center for Structural Genomics: a homolog from *Streptococcus mutans* UA 159 (PDB code 2ZC2) and that from *Enterococcus faecalis* (PDB code 2I5U), with pair-wise identities to *B. subtilis* DnaD of 41% and 25%, respectively. Neither of these proteins has been otherwise characterized. The *S. mutans* protein is a fragment of a protein encoded by the gene *SMU 1465C*. Using the Phyre fold recognition server (30), the *SMU 1465C* encoded protein is predicted to contain the N-terminal DnaD domain whereas no such domain is detectable in the full length 2i5u sequence (data not shown but compare predicted secondary structure for DnaD, 2zc2 and 2i5u in Supplementary Figures S1 and S2). The 2zc2 structure, therefore, forms the strongest basis for the construction of models of the full length *B. subtilis* DnaD. The construct in the 2zc2 structure corresponds to residues L129–I201 of *B. subtilis* DnaD, with the final structured residue corresponding to A199 (Figure 1A).

Here, we demonstrate a structural link between DnaD and DnaB, which enables the formulation of a unified understanding of their properties. We show that the two structural domains found in DnaD (which we term DDBH1 and DDBH2 for DnaD-DnaB Homology Domain 1 and 2) are also present in DnaB; sensitive domain detection methods reveal that DnaB has the domain structure (DDBH1)-$(DDBH2)_2$. The similarity is considerably higher for the second DDBH2 domain, the proteins sharing a particularly highly conserved motif YxxxIxxxW, which forms an exposed patch on the surface of the protein. The structured part of the DDBH2 domain contains a more extensive final helix than hitherto assumed. We show that the YxxxIxxxW motif, the final helix and a portion of the unstructured C-terminal tail of DnaD all contribute to DNA binding, and are essential for *B. subtilis* viability.

MATERIALS AND METHODS

Bioinformatics

In addition to resources referenced in the text, we made extensive use of Jalview (31) (http://www.Jalview.org/), ClustalW (32), hmmer3 (33) (http://hmmer.janelia.org/) and cd-hit (34). Other manipulation of data was carried out using in-house python scripts on Linux workstations.

The model of *B. subtilis* DnaD was built from the coordinates of 2zc2 using swissmodel (35) (http://swissmodel.expasy.org/). The additional residues were modelled in using X-PLOR (36) to add random coil residues up to residue F206. Helical dihedral angles were then imposed on residues A199–F206 using an in-house python script, and finally side-chain clashes were removed using X-PLOR.

NMR

NMR spectra were acquired at 298 K on a Bruker DRX series 600-MHz spectrometer equipped with triple-resonance cryoprobe. Spectra were processed and analyzed using FELIX (FELIX NMR Inc.). Assignments of backbone resonances were obtained from ^{15}N-HSQC, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB and CBCA(CO)NH spectra. The assignment was completed using FELIX for peak picking and the asstools package as previously described

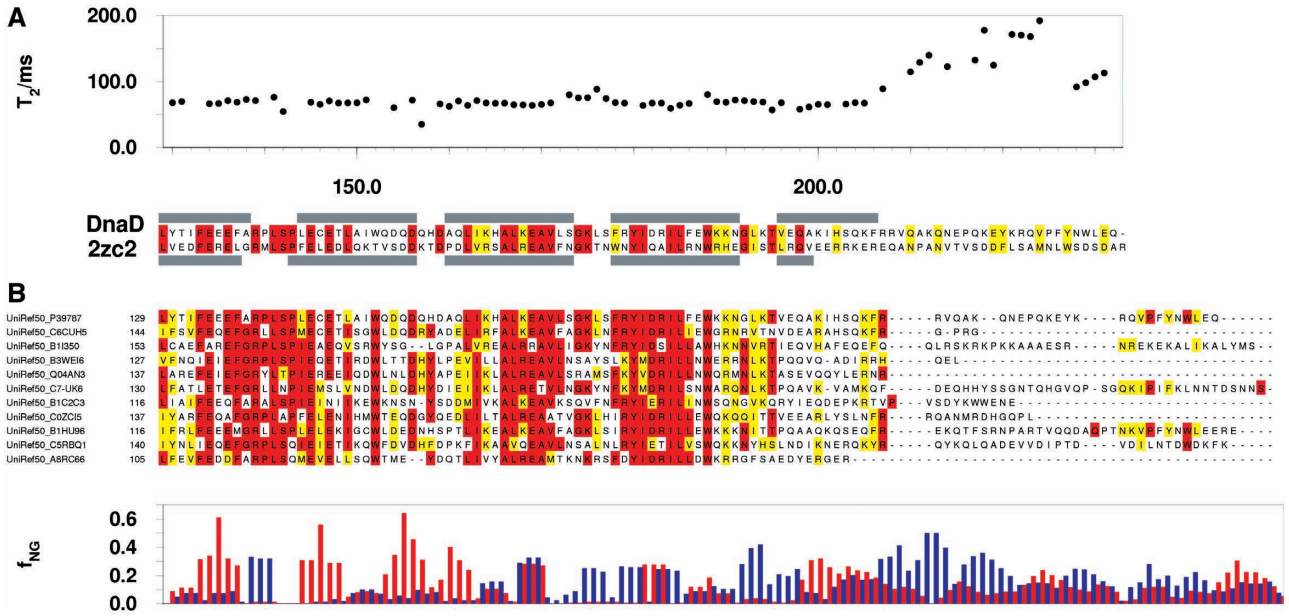

Figure 1. Defining the extent of the C-terminal domain of DnaD. (**A**) NMR ^{15}N relaxation (T_2) data, which defines regions of mobility in solution. The backbone from residue 129 to 206 shows uniform T_2 values, with elevated values beyond residue 206 indicating that the boundary between structure and disorder is approximately at residue 206. The reduced value of T_2 for the four data points shown at the C-terminus suggest some degree of constraint or conformational broadening, but there is no evidence for regular structure in the NOEs, chemical shifts or DNA-binding intensity changes for these residues. An alignment of *B. subtilis* DnaD-Cd with the sequence from *S. mutans*, corresponding to the structure 2zc2 is also shown. The extent of helices in the two structures is shown in grey above and below the sequences. Identical residues are coloured red, and similar residues are colored yellow. (**B**) ClustalW alignment of DnaD sequences. The sequences are sequences 1, 6, 11, 16, etc. drawn from the collection in Supplementary Figure S3. Residues that are identical in at least 50% of sequences (excluding gaps) are colored red. Residues that are similar across at least 50% of sequences are colored yellow. Sequences are marked with their Uniprot identifier. Below the alignment the fraction, f_{NG}, of basic (blue) and acidic (red) residues is shown, expressed as a fraction of the total non-gapped residues at each position in the alignment in Supplementary Figure S3. The profile was smoothed using a three-residue window. Across all parts of this figure the data are aligned vertically by residue number of *B. subtilis* DnaD, except where extensive gapping prevents this beyond residue 215.

(37). For side-chain assignments, CCH-TOCSY and HCCH-TOCSY spectra were employed, and NOEs were assigned in a 100-ms NOESY-($^{15}N/^{13}C$)-HSQC experiment. Samples were prepared in 100 mM sodium phosphate buffer (pH 6.5), 150 mM NaCl, 10 mM DTT and 0.3 mM protein. To provide an optimal sensitivity, 0.3 mM protein was used: higher concentrations led to poorer line shape and sensitivity presumably due to increased transient self-association. DNA titrations were performed using a 2 mM stock of a 10-mer ssDNA (5′-GTTATTGCTC-3′) exchanged into identical buffer conditions to those of the protein solution. Cross-peak intensities were corrected for the volume change during titrations. Amide exchange experiments were initiated by diluting the protein with an equal volume of D_2O buffer. Exchange was detected by comparing intensities to a control spectrum in which the protein was diluted with an equal volume of H_2O buffer. For the measurement of ^{15}N T_2 values, the Bruker pulse program hsqct2etf3gpsi (Avance version 04/01/15) was employed which essentially implements the sequence of Farrow *et al.* (38). CPMG delays of 0, 17, 34, 51, 68, 85, 119, 136 and 189 ms were employed. The correlation time was estimated from the ^{15}N T_2 by simple application of the Lipari–Szabo formula (38) with S^2 set to 0.85 for the rigid backbone, $\tau = 40$ ps and ^{15}N CSA = 160 p.p.m. The 2D ^{15}N-^{1}H correlation spectra were typically acquired for 2.5 h per spectrum (DNA titrations and amide exchange) or 3.5 h per spectrum (T_2).

Mutagenesis and protein purifications

Truncated and full length DnaD-Cd constructs encoding 129-232 (*dnaD-Cd*), 129–177 (*dnaD-Cd177*), 129–196 (*dnaD-Cd196*), 129–206 (*dnaD-Cd206*) and 129–215 (*dnaD-Cd215*) of *B. subtilis DnaD* were generated by PCR using pET22b-DnaD-Cd (*B. subtilis* DnaD residues 129–232) (15) as substrate and cloned as NdeI-HindIII fragments in pET28a and introducing a stop codon immediately prior the HindIII site. Site-directed mutagenesis was carried out with the QuickChange II-E mutagenesis kit (Stratagene). All constructs were verified by DNA sequencing and all oligonucleotides in this study were purchased from Eurogentec.

Full-length DnaD proteins were purified as described before (11,15). DnaD-Cd 129–232, DnaD-Cd 129–215, DnaD-Cd 129–206 and DnaD-Cd 129–196 were prepared by a modified version of these protocols to increase yield for NMR. Proteins were overexpressed in BL21 (DE3) *E. coli* at 37°C in LB or M9 medium with $^{15}NH_4SO_4$ or $^{13}C_6$ D-glucose as the sole carbon and nitrogen sources. Large proportions of the DnaD-Cd proteins were observed in the insoluble fractions, which were solubilized using 5 M guanidine hydrochloride. The proteins were purified using nickel affinity

chromatography, followed by refolding by dialysis. The refolded proteins were passed down the nickel column for a second time before removing the His-tag with thrombin, after which the cleaved tag was removed by a third-nickel affinity step. Finally, the proteins were further purified by gel filtration (Superdex S75). The resulting thrombin cleaved proteins possess the additional residues GSHM at the N-terminus.

Gel shift assays

Agarose gel shift assays were carried out with supercoiled pBSK plasmid as described before (12). Polyacrylamide gel shift assays were carried out with 0.625 nM ssDNA or dsDNA and various concentrations of proteins (DnaD, Y180A, W188A, DnaD-Cd196, DnaD-206 and DnaD-Cd215), as described before (16). The oligonucleotide Y180AF (5′-CTCAGTTTCCGCGCGATTGACCG GATT-3′) was used as ss probe and the complementary Y180AR (5′-AATCCGGTCAATCGCGCGGAAACTG AG-3′) was used to prepare the ds probe. Binding reactions were incubated for 20 min at RT and then resolved through 8% v/v native polyacrylamide gels.

Bacillus subtilis strains

Bacillus subtilis strains used are listed in Table 1. All are isogenic with JH642 and contain the *trpC2* and *pheA1* alleles (39). *dnaD23* is a temperature-sensitive allele that prevents replication initiation at the non-permissive temperature (3,10,40). The transposon insertion Tn*917*ΩHU151 is linked to *dnaD* (22,41,42).

The *dnaD* alleles were fused to the LacI-repressible IPTG-inducible promoter Pspank(hy) and integrated into the chromosome at the non-essential *amyE* or wild-type *dnaD* sites, the *dnaD* open reading frame, including the ribosome-binding site and followed by a TAA-stop codon was amplified by PCR using primers oWKS-178 (5′-CCCAAGCTTAGTAAGAGGTGTGAC GATGAAAAAACAGCAATTTATTGATA-3′) and oWKS-179 (5′-ACATGCATGCTTATTGTTCAAGCC AATTGTAAAAAGG-3′) and chromosomal DNA from a wild-type *B. subtilis* strain (JH642) as template. For the *dnaDY180A* and *dnaDW188A* mutations, PCR was carried out with the same primers but using the pET22b-derived expression plasmids carrying the mutant alleles. For the C-terminally truncated *dnaD*, the coding sequence including ribosome-binding site was amplified using primers oWKS-178 and oWKS-180 (5′-ACATG CATGCTTACACAGTTTTAAGCCCATTTTTCTTCC A-3′), which introduces a TAA stop codon after V196. The rest of the coding sequence of *dnaD* is not present in this construct. PCR products were digested with HindIII and SphI, and cloned into similarly digested pDR111 (43) putting it downstream of the IPTG inducible promoter Pspank(hy). Inserts were sequenced to verify the *dnaD* alleles. Pspank(hy)-*dnaD* constructs were introduced into the *B. subtilis* via double crossover at the non-essential *amyE* locus. Chromosomal DNA from these strain(s) was used to transform a *dnaD23ts* (KPL73) or *oriN oriC*$^-$ strain (MMB208).

A deletion mutation in *dnaD* was constructed using the long-flanking homology PCR approach (44). Fragments ~1 kb upstream and downstream of *dnaD* were amplified using primers oWKS-204 (5′-GAAGCTGCGGCAATGG CTTTAG-3′), oWKS-205 (5′-AACGTATTTCCCGCTG TCCTCC-3′), oWKS-206 (5′-TACCGCACAGATGCG TAAGGAGCGTCACACCTCTTACTGTAAAGG-3′) and oWKS-207 (5′-TAATAATGAGATAATGCCGACT GTACTAAAGTGAAAAGGTGACAATCGTG-3′), resulting in two products with regions of homology to a chloramphenicol resistance cassette (*cat*). These products were used as megaprimers in an expand Long Template PCR reaction (Roche) with plasmid pGEMcat (45) as template, followed by a second Expand Long Template PCR using primers oWKS-204 and oWKS-205. Of purified PCR product of the expected size, 30 µl was transformed into strains carrying ectopic copies of *dnaD* or its variants, and double cross-over replacement of the *dnaD* gene with the *cat* cassette in the transformants was verified by PCR and digest.

Table 1. *Bacillus subtilis* strains

Strain	Relevant genotype (reference)
KPL73	*dnaD23*(ts)-Tn917ΩHU151 (*mls*) (22,41,42,)
WKS802	*amyE*::{Pspank(hy)-*dnaD, spc*}
WKS804	*amyE*::{Pspank(hy)-*dnaDY180A, spc*}
WKS806	*amyE*::{Pspank(hy)-*dnaDW188A, spc*}
WKS808	*amyE*::{Pspank(hy)-*dnaD-V196-STOP, spc*}
WKS810	*dnaD23*(ts)-Tn*917*ΩHU151 (*mls*); *amyE*::{Pspank(hy)-*dnaD, spc*}
WKS812	*dnaD23*(ts)-Tn*917*ΩHU151 (*mls*); *amyE*::{Pspank(hy)-*dnaDY180A, spc*}
WKS814	*dnaD23*(ts)-Tn*917*ΩHU151 (*mls*); *amyE*::{Pspank(hy)-*dnaDW188A, spc*})
WKS816	*dnaD23*(ts)-Tn*917*ΩHU151 (*mls*); *amyE*::{Pspank(hy)-*dnaD-V196-STOP, spc*}
MMB208	*spoIIIJ::oriN oriC-S (kan tet)* (51)
WKS849	*spoIIIJ::oriN oriC-S (kan tet); amyE*::{Pspank(hy)-*dnaD, spc*}
WKS865	*amyE*::{Pspank(hy)-*dnaD23, spc*}
WKS886	*dnaD23*(ts)-Tn*917*ΩHU151 (mls); amyE::{Pspank(hy)-*dnaD23, spc*}
WKS876	*amyE*::{Pspank(hy)-*dnaD*, spc} Δ*dnaD* (*cat*)
WKS878	*amyE*::{Pspank(hy)-*dnaDY180A*, spc} Δ*dnaD* (*cat*)
WKS880	*amyE*::{Pspank(hy)-*dnaDW188A*, spc} Δ*dnaD* (*cat*)
WKS883	*amyE*::{Pspank(hy)-*dnaD23*, *spc*} Δ*dnaD* (*cat*)

RESULTS

DnaB contains domains homologous to the DnaD-Nd and DnaD-Cd domains, defining DDBH1 and DDBH2 domains

Pair-wise sequence identity between DnaD and DnaB is insignificant, at only 16% for the full-length proteins in a ClustalW alignment. Evidence for an extensive relationship between DnaB and DnaD was obtained using HHpred (46) (http://toolkit.tuebingen.mpg.de/HHpred). HHpred first performs a PSI-BLAST search based upon the query sequence and forms a Hidden Markov Model (HMM) from the resulting alignment. A HMM is primarily a table of probabilities of different amino acid types being present at each position within a domain (47). The power of HHpred is that it then compares this pattern of

probabilities with HMMs built for alignments based on other sequence families. For example, it is possible to search against HMMs for alignments based on all the sequences with structures in the PDB. This allows rather distant relationships to be detected since it is the conserved features of two families of sequences which are compared rather than features in any one sequence from a particular family. The top hit using full length DnaB as the query was for residues 300–393 to the sequence of the DnaD-like C-terminal domain (PDB code 2zc2, $E = 3.4e\text{-}07$), and a much weaker hit for residues 38–112 to the 3D structure for the DnaD N-terminal domain (PDB code 2v79, $E = 1.1$).

The list of lower scoring HHpred hits for the N-terminal domain search was dominated by helix-turn-helix motifs (in particular winged helix-turn-helix domains), and the hit to the DnaD-Nd structure was only marginally stronger than these. If, however, one looks either side of the region aligned by HHpred to DnaD-Nd, one finds that DnaB and DnaD contain very similar secondary structure elements (Supplementary Figure S1). In particular, a helix-strand-helix motif is predicted at the N-terminus of DnaB. In DnaD, the putatively corresponding H1′-S1′-H2′ elements are key to the formation of both the dimer and the tetramer (16). A dimeric arrangement would also account for the lack of a possible partner for the predicted S1′ strand in a monomeric DnaB.

We term these two regions of homology DDBH1 and DDBH2 domains (DnaD DnaB Homology 1 and 2). The relative positions of these domains in DnaD and DnaB are shown in Figure 2. We note that concurrently with our work the Pfam *DnaD* family was merged into the *DnaB_2* family (48). The evidence for the merge is undocumented. *DnaB_2* (and the former *DnaD*) both encode domains similar in extent to DDBH2.

Another piece of evidence linking the DnaD and DnaB sequences emerged from HHpred analysis of the sequence of residues 146–305 of *B. subtilis* DnaB. This is the sequence between the regions of homology identified above. The top scoring hit was to the 2zc2 structure again, albeit with a poor *E*-value of 2.3. In the absence of other indications this hit would be considered extremely marginal, but in the emerging picture of a structural link between DnaD and DnaB it is very compelling. The fact that DnaD and DnaB contain other strongly matching domains does not influence the scoring of HHpred when a portion of sequence not overlapping these domains is used as the query. Therefore, the central domain of DnaB is a divergent DDBH2 domain.

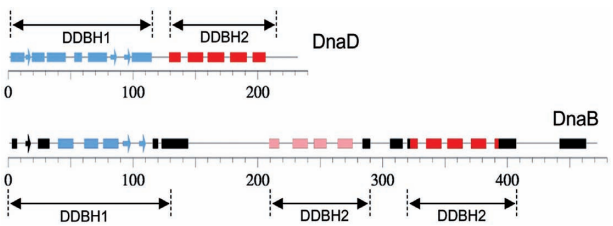

Figure 2. Domain structure of DnaD and DnaB. The figure shows the experimentally determined secondary structure of DnaD and the predicted secondary structure of DnaB (Supplementary Figure S1). The secondary-structure elements are color coded to show the regions for which HHpred detects homology. The estimated extent of the DDBH1 and DDBH2 domains in the two proteins is shown.

The scheme in Figure 2 shows that domains found in DnaD account for the vast majority of the secondary structural elements predicted for *B. subtilis* DnaB, with intervening linkers predominantly predicted as unstructured (Supplementary Figure S1) and is also consistent with the pattern of conservation across the *B. subtilis* DnaB sequences (Supplementary Figure S4). Additionally, there is a predicted helical region (*B. subtilis* DnaB residues 443–463), which we have left unclassified. There are also two conserved regions of predicted order but uncertain secondary structure (*B. subtilis* DnaB residues 155–163 and 425–433) which we have not marked in our domain graphics.

The DnaD/DnaB superfamily

To collect as large a representative set of DnaD and DnaB related proteins as possible we searched the 50% maximum sequence identity Uniprot50 sequence collection using the C-terminal domains (residues 129–232 of DnaD and residues 306–472 of DnaB) as seeds in two separate iterative searches using jackhmmer from the hmmer3 package. The searches yielded 313 and 318 sequences, respectively. The two sets of sequences were strongly overlapping with the combined non-redundant set containing 326 sequences. This overlap is expected if the two query sequences indeed are homologous, and the search procedure is obtaining a sufficiently exhaustive search for homologues. The full-length sequences corresponding to the combined set were then aligned using ClustalW, and using this alignment the pair-wise identities of each sequence to DnaD or DnaB were individually calculated (Supplementary Figure S5). A subset of the sequences clearly has a similarity to DnaD well above the baseline value, and a separate subset has a similarity to DnaB above the baseline value. The lengths of these sequences also show good consistency in these two subsets (Supplementary Figure S5). We define these two subsets as our working set of DnaD and DnaB orthologues. Sequence alignments of these two sets of sequences are shown in Supplementary Figures S3 and S4. A HMM was built (using hmmer3) corresponding to the DDBH2 domain of DnaD in an alignment of the full 326 sequence set of DnaD and DnaB related proteins (Figure 3). Note the prominent YxxxIxxxW motif and the highly conserved threonine (or serine) residue seven residues C-terminal of this motif. The significance of these features is discussed below.

What is the extent of a 'structural' and 'functional' C-terminal domain of DnaD?

As a preliminary to NMR-based studies of DNA binding, we aimed to define a construct that correctly spanned the functional part of the DnaD C-terminal domain; we show further below that the correct definition of such a domain is not straightforward and previous structural studies have used constructs that are too short. As noted above, the closest available model is the 2zc2 structure, in which there are four major helices (I–IV) and a very short helix V

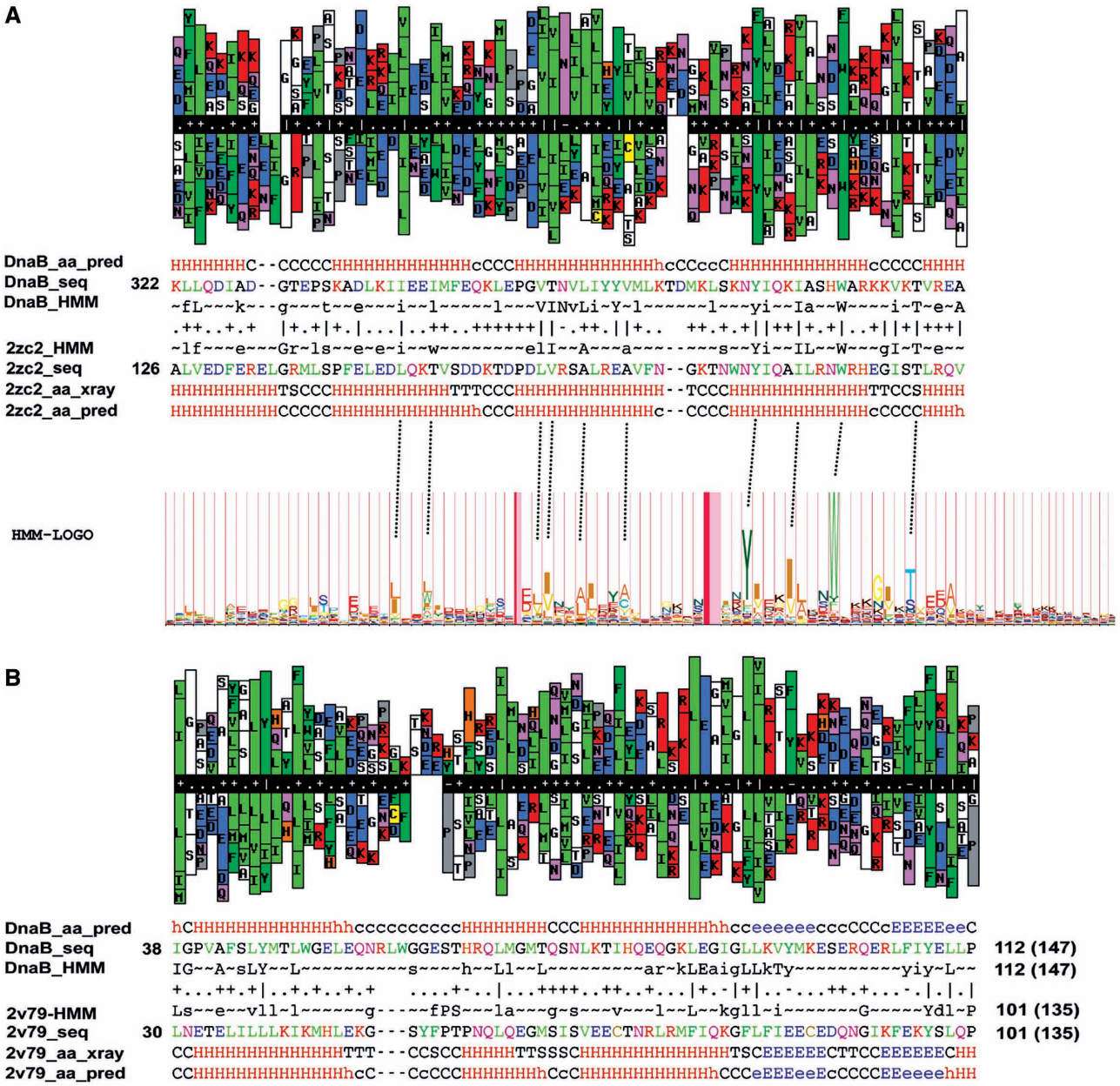

Figure 3. HMM relationships between DnaD and DnaB. (**A**) Top HHpred PDB hit obtained with a query of residues 306–472 of DnaB. The hit is the 2zc2 structure which is closely homologous to the *B. subtilis* DnaD-Cd. The HMM-LOGO for the DDBH2 domain of the combined sequence set of 326 sequences is shown below the HHpred alignment, with the correspondence of the highest probability residue types between the two HMMs shown by dashed lines. (**B**) Top HHpred PDB hit obtained with a query of residues 1–147 of DnaB. The hit is the 2v79 X-ray structure of the *B. subtilis* DnaD-Nd domain. In the HHpred, alignment parts of (A) and (B) the following abbreviations are used: ss_pred, secondary-structure prediction by PSIPRED; seq, sequence; HMM, HHpred representation of HMM where upper and lower case letters denote amino acids with >60 and >40% probability, respectively. Above the alignment are histograms of the DnaB and the PDB HMMs. Bars are proportional to probability of a certain amino acid type at each position, with bars omitted for probabilities <10%. The alignments are images returned by HHpred. The HMM-LOGO was created using the logomat-M server (http://www.sanger.ac.uk/cgi-bin/software/analysis/logomat-m.cgi).

which spans four residues (from L194 to V197, equivalent to V196 to A199 in *B. subtilis* DnaD). Consistent with this, alignment of sequences obtained using PSI-BLAST (49) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with residues 129–232 of *B. subtilis* DnaD as the query yielded an alignment in which good conservation was observed up to approximately *B. subtilis* DnaD residue 197, but was highly divergent after that point, with most sequences continuing for at least 30 residues further (data not shown, but see Figure 1B and Supplementary Figures S3 and S4 for an alignment of a more refined set of sequences based on our analyses below). In contrast, secondary structure prediction programs used by the Phyre server suggested a much longer helix V, extending as far as residue R207 (Supplementary Figure S1), with the remainder of the C-terminal polypeptide disordered. In light of this uncertainty as to the proper extent of a structural and functional C-terminal domain, the first construct that we

studied below by NMR included the whole of the C-terminal tail.

NMR analysis of the structure of the DnaD C-terminal domain

The backbone NMR assignment of *B. subtilis* DnaD residues 129–232 was completed using standard methods, with all non-prolyl residues assigned except residues L144, F226, Y227 and residues from the His-tag remaining after thrombin cleavage. The spectra were relatively poor for a protein of this size, with ^{15}N relaxation data suggesting a correlation time of ~12 ns (Figure 1A). This may be a combination of the effect of the unstructured C-terminal portion, and a degree of transient self association. TALOS analysis (50) of the backbone chemical shift data indicates the presence of five helices: I L129-A138; II L144-D156; III A160-S173; IV F178-N191; V V196-F206 (Figure 1A). ^{15}N T_2 data also confirm that the protein is predominantly disordered from residue 206 onwards (Figure 1A). Therefore, the experimental data clearly support the secondary structure predictions that helix V is considerably more extensive than suggested by available crystal structures 2zc2 and 2i5u, or by sequence alignments.

Analysis of the 3D ^{15}N- and ^{13}C-NOESY HSQC spectra yielded >50 long-range ($|i-j|>4$) NOEs that corroborate the overall fold of the domain based on the 2zc2 structure, however the quality of the spectra was insufficiently high to allow full NMR structure determination. Despite a detailed search of the spectra, no NOEs could be found to properly constrain the final helix V suggesting that this highly hydrophilic helix is not closely packed against the main core fold of the protein. We, therefore, constructed a hybrid homology model based on 2zc2, with the final helix extended as a regular helix to a length indicated by the NMR data (Figure 4A and C). Secondary-structure predictions by Phyre (Supplementary Figure S2) suggest that this helix would be considerably longer in both the 2zc2 and 2i5u full length sequences; it is, therefore, likely that helix V is as long as observed in the *B. subtilis* protein in many, and probably all, DDBH2 domains.

Helices I–IV form a packed structure which corresponds to the well conserved region of the alignments in Figure 1B and Supplementary Figure S3. The structure has a well packed hydrophobic core: residues F133, F137, L141, L149, I163, A166, L167, F178, I181 and L185 correspond (with a two residue shift in numbering) to the core residues in the 2zc2 structure, F131, L135, L139, F147, V161, A164, L165, W176, I179 and L183, respectively. Examples of clearly assignable long-range NOEs that define the core fold were from the aromatic ring of F133 to the methyl groups of L149, I163 and L167, and from the methyl group of A166 to the methyl groups of I181 and L185.

The peripheral nature of helix V is further supported by the observation that deletion of this helix in the DnaD-Cd196 truncated mutant (see below) yielded a readily purifiable protein which is well folded as judged by the relatively small perturbations of its HSQC spectrum compared to the DnaD-Cd232 construct (Supplementary Figure S8D). Furthermore, the strongly up-field shifted signals from the methyl groups of residues I184 and L185 (at −0.60 and −0.85 p.p.m., respectively, in DnaD-Cd) are still observed in 1D spectra of DnaD-Cd196. This shift for I184 arises from close proximity to W188 thus the detailed arrangement of this crucial area of sequence is not perturbed significantly by the deletion of helix V. In contrast further truncation, deleting helix IV, yielded a protein that was insoluble and probably unfolded. Consistent with its exposure, helix V also has markedly more rapid amide hydrogen–deuterium exchange rates than the other helices: exchange is complete for all residues in helix V after 2 h at 298K, pH 6.5, in contrast to the other helices in which observable intensity is still observed at that time (data not shown). The high conservation of *B. subtilis* DnaD T195 (T193 in 2zc2; T389 in *B. subtilis* DnaB) (Supplementary Figures S3 and S4) suggests that its role may be as a helix capping residue for helix V, which may help to fix the orientation of this helix.

Across the extended family of putative DnaD-Cd orthologues represented in Supplementary Figure S3 (and the subset shown in Figure 1B), the YxxxIxxxW motif (and key residues involved in its packing into the overall structure) is extremely highly conserved. The six most highly conserved hydrophobic residues in the alignment are Y180 (100%), L185 (98%), W188 (96%), I184 (94%), A166 (92%) and A170 (92%). The mean absolute conservation of the hydrophobic residues in this alignment in the structured part of the C-terminal domain is 62% (SD = 26%). The structural model shows that residues Y180, I184 and W188 lie on an exposed face of helix IV, whereas L185, A166 and A170 appear to be more packed into the core of the protein, anchoring this helix (Figure 4A).

DNA binding monitored by NMR

To investigate how DnaD-Cd binds to DNA, we titrated a 10-mer ssDNA into $^{15}N/^{13}C$ labeled DnaD-Cd and collected 1D proton, 2D ^{15}N-HSQC and 2D ^{13}C-HSQC spectra. Given the observation from gel shifts that very large complexes were formed between DnaD-Cd and DNA (15), it was anticipated that on adding DNA to DnaD-Cd the effect on the NMR spectrum might be to simply reduce the intensity of peaks across the spectrum, as increasing amounts of protein were sequestered into DNA:DnaD-Cd complexes. Surprisingly, the effects on peak intensities observed were far from uniform as shown in Figure 4B and Supplementary Figure S6. (We return to the reconciliation of the NMR and gel shift data in the discussion.) For some residues (e.g. E134) the crosspeaks decreased in intensity, but reached a plateau at ~30% of the original intensity; for some residues this drop in intensity was much more severe (e.g. I184); and for some residues, the crosspeaks were effectively unchanged (e.g. W229). We interpret the decay to ~30% intensity to be due to an increase in the overall tumbling time of the protein–DNA complex. The drop to 30% intensity is only compatible with the formation of rather small DNA:DnaD-Cd complexes, certainly no larger than

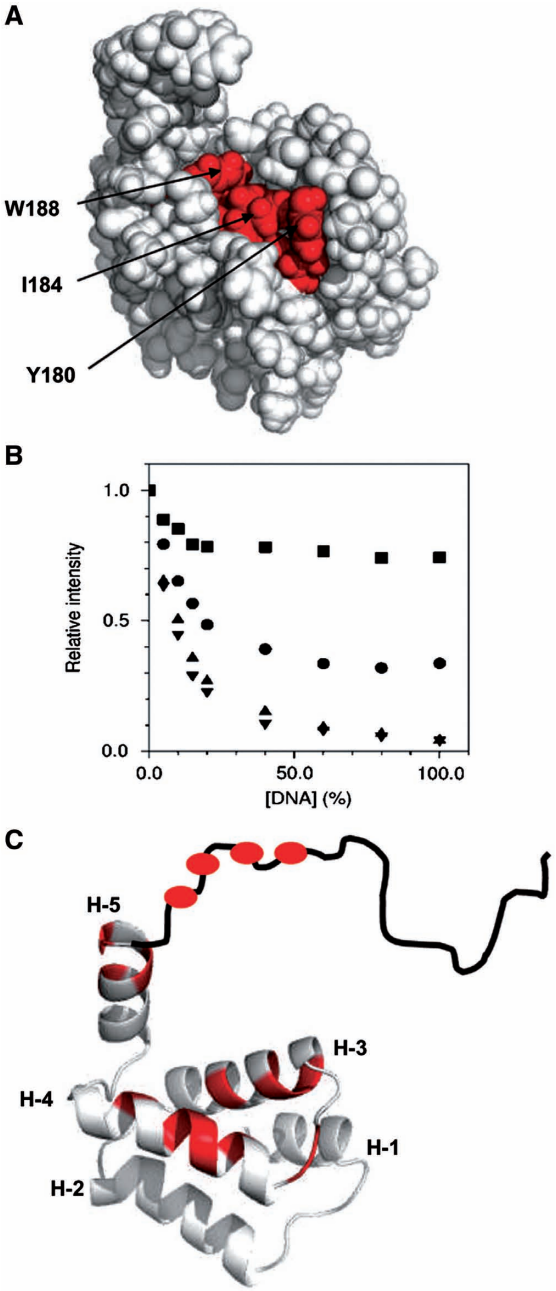

Figure 4. The conserved YxxxIxxxW motif and DNA-binding data mapped onto a structural model of DnaD-Cd. (**A**) A space filling representation of the DnaD-Cd model, with the six most highly conserved residues in DnaD-Cd (A166, A170, Y180, I184, L185 and W188) colored red. Three of these residues (W188, Y180 and I184) form an exposed hydrophobic patch, whereas A166, A170 and L185 appear to anchor this segment to the main hydrophobic core of the protein. (**B**) Examples of three types of behavior observed on titration of 10 mer DNA into DnaD-Cd. HSQC crosspeaks for the majority of residues (such as E134, circles) show a moderate intensity reduction to ~30% of initial value. Some residues show a marked intensity reduction (residue I184, triangles; residue K205, inverted triangles). Some residues (residue W229, squares) show only a very small reduction in intensity. DNA concentration is expressed as fraction of protein concentration. Data for all residues are presented in Supplementary Figure S6. (**C**) Localization of DNA binding by perturbation of ^{15}N HSQC crosspeaks. The structure is shown in cartoon representation in the same orientation as (A), with the C-terminal end of helices labeled. The mobile portion is depicted in sketch form. Residues for which the crosspeak intensity is reduced by >90% after addition of 1 : 1 10-mer DNA to DnaD-Cd215 are colored red. These are residues 168, 170, 171, 176, 177, 181, 183, 184, 185, 188, 201, 203, 205, 207, 208, 210 and 211. This figure is based on data for DnaD-Cd215 sample since it gave superior HSQC spectra (presumably due to superior tumbling properties). A closely similar subset of residues is selected with an 80% cutoff in the titration with DnaD-Cd. The slightly greater drop in intensity in the DnaD-Cd215 titration probably arises from the increased sharpness of the lines for the apo-protein in this construct.

perhaps a total of 2–3 protein and/or DNA chains. The residues that show negligible decrease in intensity are the most C-terminal of the mobile residues in the apo-protein, and presumably retain sufficient mobility in the complex to be unaffected by the increase in the overall tumbling time.

The most information comes from the residues that show much greater cross-peak attenuation than caused by the increase in tumbling time. These residues must be affected by conformational exchange in the complex, and thus reflect on the DNA-binding region. The greatest such changes in intensity are represented on the 3D structure of DnaD in Figure 4C. This identifies the region of the protein that binds DNA to include residues in helices III and IV (which contains the highly conserved YxxxIxxxW motif), residues in helix V and, most significantly, a large portion of the mobile region (up to residue 215).

^{13}C-HSQC spectra showed effects in line with these interpretations of the ^{15}N-HSQC spectra (Supplementary Figure S7), but did not contribute significantly greater detail. Only very small chemical shift changes (<0.05 p.p.m. for a weighted combined ^{1}H and ^{15}N amide shift) were observed in the ^{15}N HSQC spectra (Supplementary Figure S8), even for residues showing substantial intensity changes.

A conserved YxxxIxxxW in helix IV affects DNA binding

To test the involvement of the highly conserved YxxxIxxxW motif in DNA binding, residues Y180 and W188 were mutated to alanines in full-length DnaD and the DNA-binding properties of the mutant proteins were examined. Native DnaD binds to supercoiled pBSK and forms scaffold complexes that can be detected as shifted bands in agarose gels (12,15). The Y180A and W188A mutant proteins were able to shift pBSK in gel shift assays but at higher concentrations relative to native DnaD indicating a DNA-binding defect (Figure 5A). Direct verification was obtained with comparative gel shift assays showing that Y180A and W188A exhibited defective DNA-binding activities with ds and ssDNA (Figure 5B). This defect was more pronounced with the Y180A protein. We conclude that Y180 and W188 are directly involved in DNA binding.

To examine whether a DnaD polypeptide up to and including helix IV, but lacking helix V and the unstructured C-terminal region, was sufficient for DNA binding we truncated DnaD-Cd at position V196 (DnaD-Cd196). DnaD-Cd binds DNA but is unable to form scaffold structures detectable by plasmid shifts and its DNA-binding

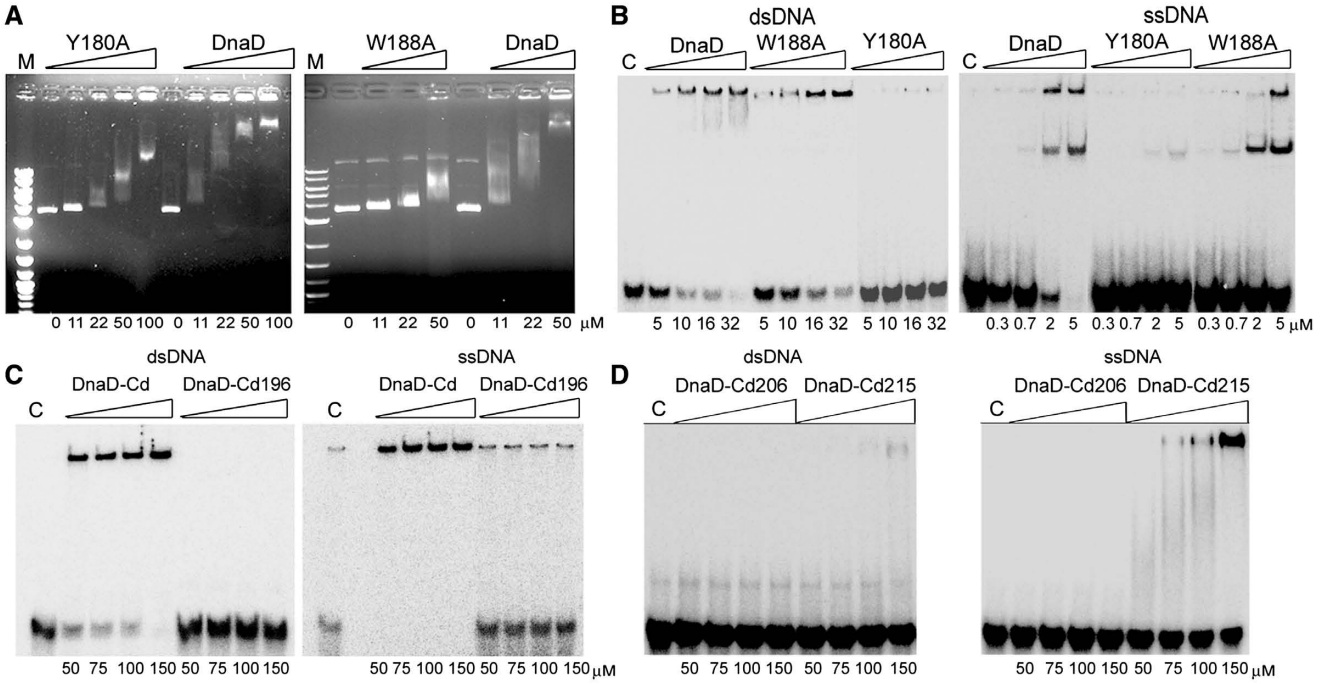

Figure 5. DNA binding of DnaD-Cd and truncation mutants. (**A**) Y180A and W188A have reduced DNA-binding activity. Agarose gel shift assays using DnaD, Y180A and W188A complexes and supercoiled pBSK plasmid. Comparative experiments were carried out with increasing concentrations of proteins, as indicated. Both mutant proteins were defective in DNA binding compared to wild-type DnaD. (**B**) Direct comparison of the DNA-binding activities of full length DnaD and the mutant Y180A, W188A proteins. Gel shift assays were carried out in 100 mM NaCl, 50 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 2.2% v/v glycerol with 0.625 nM ssDNA or 0.125 nM dsDNA probes at increasing concentrations of proteins, as indicated. (**C**) DnaD-Cd196 is deficient in DNA binding. DNA-binding reactions were carried out with increasing concentrations of DnaD and DnaD-Cd196 proteins as indicated in 30 mM NaCl, 50 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT and 0.625 nM ss or dsDNA probes, as indicated. (**D**) DnaD-Cd215 binds DNA but DnaD-Cd206 does not. DNA-binding reactions were carried out with increasing concentrations of DnaD-Cd206 and DnaD-Cd215 proteins as in described in panel (C).

activity can therefore only be detected by classical gel shift assays (15). DnaD-Cd196 did not bind DNA in gel shift assays (Figure 5C) and NMR DNA titration experiments (Supplementary Figure S8D), suggesting that helix IV by itself is not sufficient for DNA binding, despite the fact that it is a part of the DNA-binding module.

Helix V and an unstructured C-terminal region containing residues 206–215 are involved in DNA binding

Results of the NMR DNA titrations suggested that helix V and part of the unstructured C-terminus up to residue 215 also participate in DNA binding. We constructed two more truncated versions of DnaD-Cd, the DnaD-Cd206 and DnaD-Cd215 polypeptides. Gel shift assays demonstrated that DnaD-Cd206 did not bind DNA (Figure 5D), whereas DnaD-Cd215 did (though considerably weaker than full length DnaD-Cd (compare Figure 5B and C). NMR DNA titration experiments confirmed that DnaD-Cd206 did not bind DNA but DnaD-Cd215 did (Supplementary Figures S6 and S8). The combined data show that helix V and part of the unstructured C-terminus (residues 206–215) are essential for DNA binding and together with YxxxIxxxW motif in helix IV constitute the complete DNA–binding module of DnaD.

Helix V and the C-terminal unstructured region of DnaD are essential for *B. subtilis* viability

Bacillus subtilis DnaD is essential for viability. To determine if the mutant DnaD proteins with altered DNA-binding activity *in vitro* are functional *in vivo* and able to support cell growth, we fused the mutant alleles (*dnaDY180A*, *dnaDW188A* and *dnaDV196-stop*) to the IPTG-inducible promoter Pspank(hy), introduced these alleles into an ectopic site on the chromosome (*amyE*), and tested for the ability to complement loss of function mutations in *dnaD*. As controls, we fused the wild-type *dnaD* and the temperature sensitive *dnaD23ts* mutation (22,41,42) to Pspank(hy). Using western blotting with polyclonal anti-DnaD antibodies, we verified that induction by IPTG led to overproduction of protein, demonstrating that these proteins are not inherently unstable *in vivo* (data not shown).

We assessed whether ectopic copies of wild-type or mutant *dnaD* could rescue the growth defect of a *dnaD23ts* strain at the non-permissive temperature. Since Pspank(hy) is not fully repressed, these experiments were carried out in the absence of IPTG and relied on the basal, leaky, expression of the *dnaD* alleles from Pspank(hy). We found that wild-type *dnaD*, *dnaDY180A* and *dnaDW188A*, but not *dnaD-V196-STOP*, were capable of restoring growth at the non-permissive temperature as judged by growth on LB agar (Figure 6A). The lack of

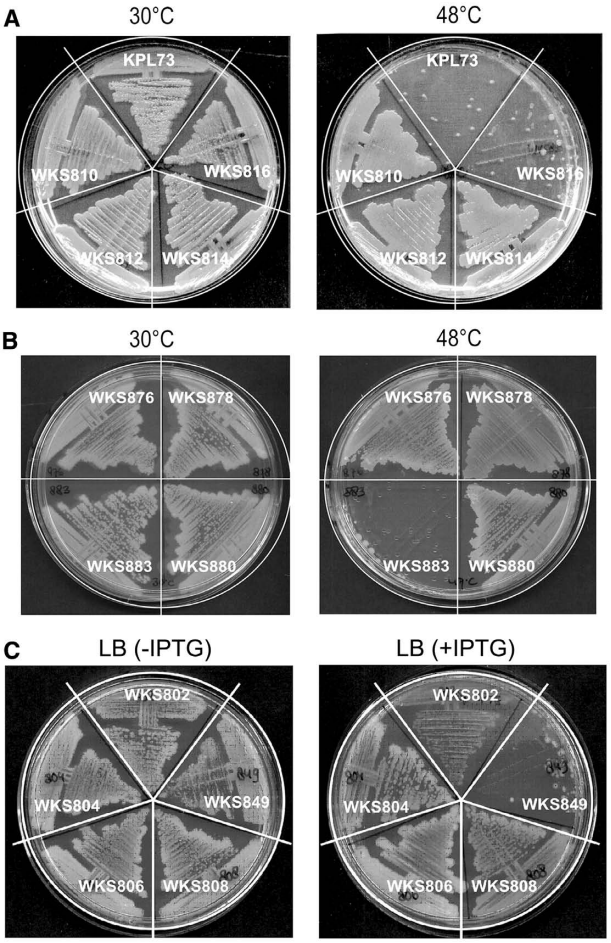

Figure 6. Biological significance of DnaD binding to DNA. (**A**) Truncated DnaD-V196-STOP does not rescue growth of *dnaD23ts* at the non-permissive temperature. The *dnaD23ts* strains carrying an ectopic inducible wild-type *dnaD* (WKS810), *dnaDY180A* (WKS812), *dnaDW188A* (WKS816) or *dnaD-V196-STOP* (WKS818) were grown on LB agar. A *dnaD23ts* strain without inducible *dnaD* (KPL73) is shown as control. Plates were incubated overnight at 30°C or 48°C as indicated. (**B**) DnaD-Y180A and DnaD-W188A sustain growth. Strains carrying an inducible wild-type DnaD (WKS876), Y180A (WKS878), W188A (WKS880) and DnaD23 (WKS883) can support growth at 30°C when the native copy of *dnaD* is knocked out. LB agar plates were incubated over night at the indicated temperature. As a control, 48°C is shown for WKS883. (**C**) Overproduction of wild-type DnaD causes growth inhibition. Strains were grown on LB agar plates, without or with 1-mM IPTG over night at 37°C. WKS802-808 contain an ectopic inducible *dnaD* (or mutant thereof) in an otherwise wild-type background, whereas WKS849 contains a heterologous origin and is inactivated for *oriC*. Information on all strains is provided in Table1.

complementation by *dnaD-V196-STOP* indicates that an essential function of DnaD, presumably DNA-binding activity, is located in the C-terminal region, after V196. We also verified that an ectopic inducible copy of *dnaD23* did not complement a *dnaD23ts* mutation (data not shown), consistent with previous findings (3).

The complementation of *dnaD23ts* by *dnaDY180A* and *dnaDW188A* could be due to true complementation or to the formation of mixed multimers with DnaD23 protein. To test if the *dnaDY180A* and *dnaDW188A* alleles truly function to support cell growth, we introduced a deletion–insertion mutation of *dnaD* into strains expressing each of the different *dnaD* alleles from Pspank(hy). We found that the native *dnaD* could be deleted in strains expressing wild-type *dnaD*, *dnaDY180A* and *dnaDW188A* (Figure 6B), but not *dnaD-V196-STOP*. As expected, knocking out the native *dnaD* from the strain carrying the inducible *dnaD23ts* rendered this strain temperature sensitive (Figure 6B). These results demonstrate that the Y180A and W188A missense mutants, that have reduced DNA-binding activity *in vitro* (Figure 5A and B), can support growth whereas the C-terminally truncated DnaD, that has little or no DNA-binding activity, cannot. In line with these observations, we were not able to obtain transformants with a plasmid that would result in a truncation of the *dnaD* gene at the native locus (corresponding to DnaD-V196-STOP), whereas it was readily introduced in a merodiploid strain harboring an ectopic copy of wild-type *dnaD* (data not shown).

Even though the missense mutants DnaDY180A and DnaDW188A were able to complement loss of function mutations in *dnaD*, they had altered function *in vivo*. In the course of analyzing the phenotypes of the Pspank(hy)-*dnaD* fusions, we found that overexpression of wild-type DnaD upon the addition of 1 mM IPTG caused a growth defect as judged by growth on LB agar plates (Figure 6C, strain WKS802). This inhibitory effect was not due to DnaD acting on the chromosomal origin of replication, *oriC*. In a strain containing the heterologous origin of replication *oriN* and a deletion in *oriC* (51), overexpression of DnaD (with 1 mM IPTG) still caused a severe defect in cell growth (Figure 6, strain WKS849). In contrast, overexpression of DnaDY180A, DnaDW188A and DnaD-V196-STOP in otherwise wild-type cells did not cause an obvious defect in growth as judged on LB agar plates (Figure 6C). Since these mutant proteins have altered DNA-binding activity *in vitro*, we suggest that the inability to inhibit growth when overexpressed is due to decreased DNA binding *in vivo*.

We also compared other growth characteristics of strains expressing either *dnaDY180A* or *dnaDW188A* with those of strains expressing wild-type DnaD. These experiments were carried out in cells deleted for the endogenous wild-type *dnaD* (*ΔdnaD*) and expressing each of the *dnaD* alleles from Pspank(hy). We tested for growth on agar plates made with defined minimal medium and succinate as carbon source. The mutants were able to grow under these conditions indicating that they support viability under rapid and slow growth conditions. The mutants also grew and sporulated on agar plates with Difco sporulation medium (data not shown). Finally, we measured doubling times in liquid LB medium and minimal succinate medium. The wild-type and *dnaDW188A* mutant had similar doubling times, but the *dnaDY180A* mutant had a slightly longer doubling time indicating a small defect in cell growth (Table 2). Taken together, these data indicate that the DnaDY180A and DnaDW188A, with a reduced ability to bind to DNA, likely have reduced activity *in vivo*, but nonetheless support cell growth, viability and sporulation.

Table 2. Growth rate of *dnaD* mutants

Strain	Doubling time (minutes ± SE)[a]	
	LB	Succinate
WKS876 (wt)	32.0 ± 0.5	64.6 ± 3.5
WKS878 (Y180A)	35.5 ± 0.7	75.21 ± 2.3
WKS880 (W188A)	30.8 ± 0.3	65.72 ± 2.0

[a]Growth rates of cells deleted for the native *dnaD* (*ΔdnaD*) and harboring an ectopic copy of wild-type *dnaD* (WKS876), *dnaDY180A* (WKS878) or *dnaDW188A* (WKS880), each under control of Pspank(hy), were grown in liquid LB or defined minimal medium with succinate as carbon source at 37°C in the absence of IPTG. The culture doubling time is indicated ± the standard error of the mean ($n = 3$).

DISCUSSION

The structural similarities between DnaD and DnaB correlate with many functional properties of the two proteins. A common property of the DDBH1 domains of both DnaD and DnaB is that they form dimers which associate into tetramers (15,24). In the DDBH1 domain of *B. subtilis* DnaD, dimer–tetramer association is mediated by a helix-strand-helix (H1′-S1′-H2′) element flanking the N-terminus of the winged helix-turn-helix core while a helix (H3′) flanking the C-terminus of the helix-turn-helix core is involved in higher order cross-tetramer oligomerization (16). As both structural elements are preserved in DnaB, the same oligomerization properties are likely to be also conserved in the DDBH1 domain of DnaB.

The DDBH2 domains of DnaD and DnaB are monomeric but oligomerize upon binding to DNA (15,24). Our data show that the C-terminal part of the *B. subtilis* DnaD DDBH2 domain and part of its disordered C-terminal extension are involved in DNA binding. The equivalent region of *B. subtilis* DnaB, encompassing part of the C-terminal region of its second DDBH2 (residues 365–400) and its C-terminal extension (residues 400–428), has also been shown experimentally to bind DNA and to mediate DNA–induced higher order oligomerization (24). DNA-binding and DNA-induced oligomerization appear to be functionally and structurally conserved in the DDBH2 domain in both proteins. Furthermore, in alignments of DnaD sequences the segment corresponding to helices I and II is predominantly acidic, whilst the segment comprising the DNA-binding module (helices IV, V, and mobile residues to approximately *B. subtilis* DnaD residue 215) is predominantly basic (Figure 1B), consistent with the latter having a DNA-binding function. This acidic/basic charge distribution is closely mimicked in DnaB (data not shown).

DNA-binding winged helix-turn-helix motifs bind exclusively to dsDNA (52). However, a polypeptide fragment of *B. subtilis* DnaB (residues 1–300), encompassing the (DDBH1)-(DDBH2) fragment, binds ssDNA and not dsDNA (24) and a polypeptide fragment of *B. subtilis* DnaD (residues 1–128), encompassing the DDBH1 domain, does not bind DNA (14,15). It therefore appears that the function of the DDBH1 domains in both DnaD and DnaB is not to bind DNA but to mediate dimerization–tetramerization. From the above we predict that the ssDNA-binding activity of the *B. subtilis* DnaB fragment encompassing (DDBH1)-(DDBH2) resides on the DDBH2 domain. Hence, the middle DDBH2 domain of *B. subtilis* DnaB has likely evolved to bind exclusively ssDNA and lost the ability to bind dsDNA.

DDBH2 motifs are found in many bacterial and phage proteins of unknown functions (15). In several of these proteins two tandem DDBH2 domains are present, equivalent to the tandem DDBH2 domains found in *B. subtilis* DnaB, while in other proteins DDBH2 domains are fused in putative phage initiation factors. In the orf4 protein from phage 7201 of *S. thermophilus* the DDBH2 domain is fused to a RepA-like domain associated with initiation of DNA replication (15). Similar modular architectures have been observed in a putative replication initiation protein from phage A118 of *Listeria monocytogenes* and in Protein 20 from *Lactobacillus plantarum* (22). Therefore, the modular character of the DDBH1 and DDBH2 motifs and their structural/functional properties appear to be conserved across many bacterial and phage proteins beyond the DnaD and DnaB primosomal proteins.

Bacillus subtilis DnaD-Cd oligomerizes upon binding to DNA (15). We, therefore, expected that titration of DNA into DnaD-Cd would lead to a progressive, complete and uniform decrease of signal intensity in NMR experiments, and would thus lead to little useful information. However, the attenuation of signals was far from uniform, and did not lead to complete loss of signal intensity. The observation of conformational broadening on binding oligonucleotides is in contrast quite common, and generally ascribed to the possibility of a variety of binding registers. Calculations show that the observation of very small shift changes and differential broadening in a titration is compatible with exchange (substantially on the slow side of intermediate exchange) between the free protein and a bound form in which peaks corresponding to certain residues are severely broadened (data not shown). The observation of large complexes in the gel shift experiments, and the rather small complexes suggested in the NMR experiments, may be a consequence of the different relative concentrations of DNA and protein in these two experiments if the complexes contain a superstoichiometric ratio of protein to DNA. Detailed analysis of these complexes will be needed with alternative methods to resolve this issue.

In summary, the bacterial replication initiation proteins DnaD and DnaB share similar domain organizations. Elucidation of their domain structures greatly helps to unify and explain their similar properties. The DDBH1 domain contains a conserved N-terminal helix-strand-helix element that forms β-sheets in dimers–tetramers, and a C-terminal helix that mediates cross-tetramer oligomerization. The DDBH2 domain has a C-terminal DNA-binding module comprised of a conserved YxxxIxxxW motif (in the penultimate helix), the last helix and part of the unstructured region at the C-terminus. Thus the DDBH2 domain (which is the most strongly conserved domain between the two

proteins) displays the curious contrast of a very well conserved motif, built on a well conserved protein core structure, coupled to a loosely associated helix and an unstructured region, both of the latter being ill conserved. Together these elements achieve DNA binding and high order oligomerization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Biotechnology and Biological Sciences Research Council (BBSRC) (grant BB/E006450/1 to P.S.); BBSRC Ph.D. studentship (to F.Y.M.); National Institute of Health (grant GM41934 to A.D.G.); Rubicon fellowship from the Netherlands Organization for Scientific Research (to W.K.S.). Funding for open access charge: BBSRC.

Conflict of interest statement. None declared.

REFERENCES

1. Bruand,C., Ehrlich,S.D. and Janniere,L. (1995) Primosome assembly site in *Bacillus subtilis*. *EMBO J.*, **14**, 2642–2650.
2. Marsin,S., McGovern,S., Ehrlic,H.S.D., Bruand,C. and Polard,P. (2001) Early steps of *Bacillus subtilis* primosome assembly. *J. Biol. Chem.*, **276**, 45818–45825.
3. Bruand,C., Velten,M., McGovern,S., Marsin,S., Serena,C., Ehrlich,S.D. and Polard,P. (2005) Functional interplay between the *Bacillus subtilis* DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Mol. Microbiol.*, **55**, 1138–1150.
4. Mott,M.L. and Berger,M. (2007) DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.*, **5**, 343–354.
5. Zakrzewska-Czerwinska,J., Jakimowicz,D., Zawilak-Pawlik,A. and Messer,W. (2007) Regulation of the initiation of chromosomal replication in bacteria. *FEMS Microbiol. Rev.*, **31**, 378–387.
6. Smits,W.K., Goranov,A.I. and Grossman,A.D. (2010) Ordered association of helicase loader proteins with the *Bacillus subtilis* origin of replication *in vivo*. *Mol. Microbiol.*, **75**, 452–461.
7. Velten,M., McGovern,S., Marsin,S., Ehrlich,S.D., Noirot,P. and Polard,P. (2003) A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol. Cell*, **11**, 1009–1020.
8. Ioannou,C., Schaeffer,P.M., Dixon,N.E. and Soultanas,P. (2006) Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in *Bacillus subtilis*. *Nucleic Acids Res.*, **34**, 5247–5258.
9. Nunez-Ramirez,R., Velten,M., Rivas,G., Polard,P., Carazo,J.M. and Donate,L.E. (2007) Loading a ring: Structure of the *Bacillus subtilis* DnaB protein, a co-loader of the replicative helicase. *J. Mol. Biol.*, **367**, 764–769.
10. Bruand,C., Farache,M., McGovern,S., Ehrlich,S.D. and Polard,P. (2001) DnaB, DnaD and DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Mol. Microbiol.*, **42**, 245–255.
11. Turner,I.J., Scott,D.J., Allen,S., Roberts,C.J. and Soultanas,P. (2004) The *Bacillus subtilis* DnaD protein: a putative link between DNA remodeling and initiation of DNA replication. *FEBS Lett.*, **577**, 460–464.
12. Zhang,W., Carneiro,M.J.V.M., Turner,I.J., Allen,S., Roberts,C.J. and Soultanas,P. (2005) The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodeling activities. *J. Mol. Biol.*, **351**, 66–75.
13. Zhang,W., Allen,S., Roberts,C.J. and Soultanas,P. (2006) The *Bacillus subtilis* primosomal protein DnaD untwists supercoiled DNA. *J. Bacteriol.*, **188**, 5487–5493.
14. Zhang,W., Machón,C., Orta,A., Phillips,N., Roberts,C.J., Allen,S. and Soultanas,P. (2008) Single-molecule atomic force spectroscopy reveals that DnaD forms scaffolds and enhances duplex melting. *J. Mol. Biol.*, **377**, 706–714.
15. Carneiro,M.J., Zhang,W., Ioannou,C., Scott,D.J., Allen,S., Roberts,C.J. and Soultanas,P. (2006) The DNA-remodelling activity of DnaD is the sum of oligomerization and DNA-binding activities on separate domains. *Mol. Microbiol.*, **60**, 917–924.
16. Schneider,S., Zhang,W., Soultanas,P. and Paoli,M. (2008) Structure of the N-terminal oligomerization domain of DnaD reveals a unique tetramerization motif and provides insights into scaffold formation. *J. Mol. Biol.*, **376**, 1237–1250.
17. Winston,S. and Sueoka,N. (1980) DNA membrane association is necessary for initiation of chromosomal and plasmid replication in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*, **77**, 2834–2838.
18. Hoshino,T., McKenzie,T., Schmidt,S., Tanaka,T. and Sueoka,N. (1987) Nucleotide sequence of *Bacillus subtilis dnaB*: a gene essential for DNA replication initiation and membrane attachment. *Proc. Natl Acad. Sci. USA*, **84**, 653–657.
19. Sueoka,N. (1998) Cell membrane and chromosome replication in *Bacillus subtilis*. *Prog. Nucleic Acids Res.*, **59**, 34–53.
20. Watabe,K. and Forough,R. (1987) Effects of temperature sensitive variants of the *Bacilus subtilis dnaB* gene on the replication of low-copy number plasmid. *J. Bacteriol.*, **169**, 4141–4146.
21. Sato,Y., McCollum,M., McKenzie,T., Laffan,J., Zuberi,A. and Sueoka,N. (1991) *In vitro* type II binding of chromosomal DNA to membrane in *Bacillus subtilis*. *J. Bacteriol.*, **173**, 7732–7735.
22. Rokop,M.E., Auchtung,J.M. and Grossman,A.D. (2004) Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol. Microbiol.*, **52**, 1757–1767.
23. Li,Y., Kurokawa,K., Reutimann,L., Mizumura,H., Matsuo,M. and Sekimizu,K. (2007) DnaB and DnaI temperature-sensitive mutants of *Staphylococcus aureus*: evidence for involvement of DnaB and DnaI in synchrony regulation of chromosome replication. *Microbiol.*, **153**, 3370–3379.
24. Grainger,W.H., Machón,C., Scott,D.J. and Soultanas,P. (2010) DnaB proteolysis *in vivo* regulates oligomerization and its localization at *oriC* in *Bacillus subtilis*. *Nucleic Acids Res.*, **38**, 2851–2864.
25. Ishigo-Oka,D., Ogasawara,N. and Moriya,S. (2001) DnaD protein of *Bacillus subtilis* interacts with DnaA, the initiator protein of replication. *J. Bacteriol.*, **183**, 2148–2150.
26. Noirot-Gros,M.F., Dervyn,E., Wu,L.J., Mervelet,P., Errington,J., Ehrlich,S.D. and Noirot,P. (2002) An expanded view of bacterial DNA replication. *Proc. Natl Acad. Sci. USA*, **99**, 8342–8347.
27. Rokop,M.E. and Grossman,A.D. (2009) Intragenic and extragenic suppressors of temperature sensitive mutations in the replication initiation genes *dnaD* and *dnaB* of *Bacillus subtilis*. *PLoS ONE*, **4**, e6774.
28. Schneider,S., Carneiro,M.J., Ioannou,C., Soultanas,P. and Paoli,M. (2007) Crystallization and X-ray diffraction analysis of the DNA-remodelling protein DnaD from Bacillus subtilis. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **63**, 110–113.
29. Huang,C.Y., Chang,Y.W. and Chen,W.T. (2008) Crystal structure of the N-terminal domain of *Geobacillus kaustophillus* HTA426 DnaD protein. *Biochem. Biophys. Res. Comm.*, **375**, 220–224.
30. Kelley,L.A. and Sternberg,M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
31. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
32. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) ClustalW and ClustalX version 2. *Bioinformatics*, **23**, 2947–2948.
33. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

34. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
35. Arnold,K., Bordoli,L., Kopp,J. and Schwede,T. (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling. *Bioinformatics*, **22**, 195–201.
36. Brünger,A.T. (1992) *X-PLOR Version 3.1*. Yale University Press, New Haven, Connecticut.
37. Reed,M.A., Hounslow,A.M., Sze,K.H., Barsukov,I.G., Hosszu,L.L., Clarke,A.R., Craven,C.J. and Waltho,J.P. (2003) Effects of domain dissection on the folding and stability of the 43 kDa protein PGK probed by NMR. *J. Mol. Biol.*, **330**, 1189–11201.
38. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochem.*, **33**, 5984–6003.
39. Perego,M., Spiegelman,G.B. and Hoch,J.A. (1988) Structure of the gene for the transition state regulator, abrB: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis*. *Mol. Microbiol.*, **2**, 689–699.
40. Karamata,D. and Gross,J.D. (1970) Isolation and genetic analysis of temperature-sensitive mutants of *B. subtilis* defective in DNA synthesis. *Mol. Gen. Genet.*, **108**, 277–287.
41. Goranov,A.I., Katz,L., Breier,A.M., Burge,C.B. and Grossman,A.D. (2005) A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proc. Natl Acad. Sci. USA*, **102**, 12932–12937.
42. Lemon,K.P., Kurtser,I., Wu,J. and Grossman,A.D. (2000) Control of initiation of sporulation by replication initiation genes in *Bacillus subtilis*. *J. Bacteriol.*, **182**, 2989–2991.
43. Britton,R.A., Eichenberger,P., Gonzalez-Pastor,J.E., Fawcett,P., Monson,R., Losick,R. and Grossman,A.D. (2002) Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. *J. Bacteriol.*, **184**, 4881–4890.
44. Wach,A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae Yeast*, **12**, 259–265.
45. Youngman,P., Poth,H., Green,B., York,K., Olmedo,G. and Smith,K. (1989) Methods for genetic manipulation, cloning, and functional analysis of sporulation genes in *Bacillus subtilis*. In Smith,I., Slepecky,R. and Setlow,P. (eds), *Regulation of Procaryotic Development*. American Society for Microbiology, Washington, DC, pp. 65–87.
46. Söding,J., Biegert,A. and Lupas,A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
47. Schuster-Böckler,B., Schultz,J. and Rahmann,S. (2004) HMM Logos for visualization of protein families. *BMC Bioinformatics*, **5**, 7.
48. Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,J.S., Hotz,H.R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L. *et al*. (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
49. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
50. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
51. Berkmen,M.B. and Grossman,A.D. (2007) Subcellular positioning of the origin region of the *Bacillus subtilis* chromosome is independent of sequences within oriC, the site of replication initiation, and the replication initiator DnaA. *Mol. Microbiol.*, **63**, 150–165.
52. Gajiwala,K.S. and Burley,S.K. (2000) Winged helix proteins. *Curr. Opin. Struct. Biol.*, **10**, 110–116.