JB
Journals.ASM.org

Structure of Bacterial Transcription Factor SpoIIID and Evidence for a Novel Mode of DNA Binding

Bin Chen, Paul Himes, Yu Liu, Yang Zhang, Zhenwei Lu, Aizhuo Liu, Honggao Yan, Lee Kroos

Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA

SpoIIID is evolutionarily conserved in endospore-forming bacteria, and it activates or represses many genes during sporulation of *Bacillus subtilis*. An SpoIIID monomer binds DNA with high affinity and moderate sequence specificity. In addition to a predicted helix-turn-helix motif, SpoIIID has a C-terminal basic region that contributes to DNA binding. The nuclear magnetic resonance (NMR) solution structure of SpoIIID in complex with DNA revealed that SpoIIID does indeed have a helix-turn-helix domain and that it has a novel C-terminal helical extension. Residues in both of these regions interact with DNA, based on the NMR data and on the effects on DNA binding *in vitro* of SpoIIID with single-alanine substitutions. These data, as well as sequence conservation in SpoIIID binding sites, were used for information-driven docking to model the SpoIIID-DNA complex. The modeling resulted in a single cluster of models in which the recognition helix of the helix-turn-helix domain interacts with the major groove of DNA, as expected. Interestingly, the C-terminal extension, which includes two helices connected by a kink, interacts with the adjacent minor groove of DNA in the models. This predicted novel mode of binding is proposed to explain how a monomer of SpoIIID achieves high-affinity DNA binding. Since SpoIIID is conserved only in endospore-forming bacteria, which include important pathogenic *Bacilli* and *Clostridia*, whose ability to sporulate contributes to their environmental persistence, the interaction of the C-terminal extension of SpoIIID with DNA is a potential target for development of sporulation inhibitors.

SpoIIID is a key regulator of transcription during the sporulation process of the bacterium *Bacillus subtilis*. When these rod-shaped cells sense nutrient limitation, they complete DNA replication and synthesize a polar division septum, creating a larger mother cell and a smaller forespore, each of which receives a copy of the chromosome (see Fig. S1 in the supplemental material). The alternative sigma factor, σ^E, becomes active in the mother cell and directs transcription of the *spoIIID* gene and nearly 300 other genes (reviewed in reference 1). Some of the genes under σ^E control code for proteins that cause the mother cell membrane to engulf the forespore and pinch it off as a free protoplast inside the mother cell [reviewed in reference 2] (see Fig. S1). SpoIIID regulates the transcription of more than 100 genes in the mother cell (3). Most of these genes are transcribed by σ^E RNA polymerase initially and then downregulated by SpoIIID as it accumulates. SpoIIID also upregulates transcription of a few genes in the σ^E regulon (3), including directly activating transcription of the gene for the later-acting mother cell sigma factor, σ^K (4, 5) (see Fig. S1). Many genes in the σ^K regulon code for proteins that assemble on the surface of the forespore to produce a multilayered coat that helps the spore withstand environmental stress after it is released by mother cell lysis (reviewed in reference 6). SpoIIID directly represses transcription by σ^K RNA polymerase of at least four genes that code for spore coat proteins, opposing activation of transcription by GerE at these promoters (4, 7–9). GerE up- or downregulates transcription of about 90 genes in the σ^K regulon (3). Both GerE (74 residues) and SpoIIID (93 residues) are small, sequence-specific DNA-binding proteins.

Proteins bind to specific sequences in DNA using a small number of structurally distinct motifs. One of the most prevalent motifs is the helix-turn-helix (HTH), which is found not only in bacterial and eukaryotic transcription factors but also in proteins that participate in DNA repair and RNA metabolism (10). The HTH motif typically consists of a trihelical bundle in which the second and third helices are connected by a sharp turn. The third helix is known as the recognition helix, because it interacts with base pairs in the major groove of DNA. However, this interaction alone does not impart sufficient specificity and affinity. Many HTH proteins overcome this problem by forming homodimers that recognize palindromic sites in DNA (11, 12). A crystal structure of GerE revealed a dimer with an HTH in each monomer, and the recognition helix of each HTH was predicted to interact with inverted repeats matching the degenerate consensus DNA sequence RWWTRGGYNNYY (R means A or G, W means A or T, Y means C or T, and N means A, C, G, or T) (13). SpoIIID has a predicted HTH (14), but a monomer of SpoIIID can bind with high affinity to DNA containing a single match to the degenerate consensus sequence WWRRACARNY (15). Two regions of SpoIIID were shown to be important for DNA binding, the putative recognition helix of the predicted HTH and a basic region near the C terminus. Other HTH proteins, such as homeodomain proteins and winged-helix proteins, use "arms" and "wings," respectively, to make additional contacts with DNA, allowing specific, high-affinity binding by a protein monomer (16, 17).

Here, we report the nuclear magnetic resonance (NMR) solution structure of SpoIIID in complex with DNA. The predicted HTH of SpoIIID is followed by a C-terminal helical extension that

Received 31 December 2013 **Accepted** 26 February 2014

Published ahead of print 28 February 2014

Address correspondence to Honggao Yan, yanh@msu.edu, or Lee Kroos, kroos@msu.edu.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.01486-13.

doi:10.1128/JB.01486-13

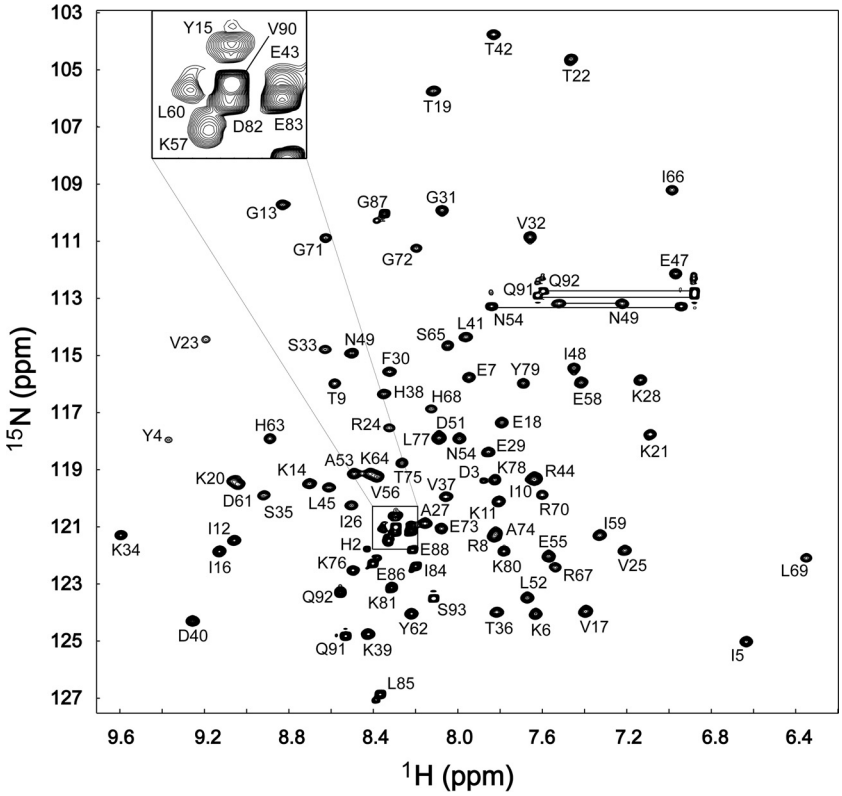

FIG 1 ^{1}H-^{15}N-HSQC spectrum of the SpoIIID-DNA complex. The NMR sample contained ~1.2 mM SpoIIID-DNA complex in a buffer containing 10 mM phosphate, 100 mM KCl, pH 7.0. The spectrum was recorded at 25°C and an ^{1}H frequency of 900 MHz with coherence selection by pulsed-field gradients and sensitivity enhancement. Sequential assignments are indicated with the one-letter amino acid code and residue number. Side chain amides are indicated by horizontal lines. The inset is an expanded view of the boxed region.

is unique for an HTH protein. The NMR data and the effects of substitutions in SpoIIID indicate residues in both the recognition helix of the HTH and in the C-terminal extension that likely interact with DNA. Using an information-driven docking method, we model the SpoIIID-DNA complex. Our results provide strong evidence for a novel mode of DNA binding by SpoIIID.

MATERIALS AND METHODS

NMR sample preparation. SpoIIID was produced in *Escherichia coli*, either singly labeled with ^{15}N or doubly labeled with ^{15}N and ^{13}C, using minimal media with appropriate nitrogen and carbon sources. The protein was purified, and its concentration in the fractions eluted from the heparin column at 1 M NaCl was determined as described previously (15). The pooled fractions were diluted 10-fold to 0.1 M NaCl with 10 mM potassium phosphate buffer, pH 7.0 (buffer 1), mixed at a 1:1.2 molar ratio with a solution of probe 11 DNA, prepared as described previously (15), and incubated on ice for 1 h. Probe 11 forms a 14-bp DNA duplex containing a single copy of the idealized binding site consensus sequence (5′-TAGGACAAGC-3′) (see Fig. S2 in the supplemental material), and analytical ultracentrifugation analyses indicated that an SpoIIID monomer forms a 1:1 complex with probe 11 DNA (15). Complexes were concentrated using Amicon Ultra 4 (5,000 molecular weight cutoff) (Millipore) filtration devices and then centrifuged (16,000 × *g* for 10 min at 4°C). The supernatant (approximately 500 μl) was transferred to a new tube and mixed with 50 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), 0.1% sodium azide, and 10% D_2O. The sample, with a final SpoIIID concentration of 1.2 mM, was placed in an NMR tube and sealed with a septum.

NMR data acquisition and analysis. NMR experiments were performed at 25°C on a 900-MHz Bruker Avance spectrometer equipped with a TCI cryoprobe or a 600-MHz Varian Inova spectrometer equipped with a standard triple-resonance probe. Data were processed using the program NMRPipe (18) with chemical shifts referenced to the internal DSS standard and were analyzed using the program NMRView (19). NMR spectra acquired for sequential resonance assignment and structure determinations (20) included two-dimensional (2D) ^{1}H-^{15}N-heteronuclear single quantum coherence (HSQC), ^{1}H-^{13}C-HSQC, 3D HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, HCCH-total correlated spectroscopy (TOCSY), ^{13}C-edited nuclear Overhauser effect spectroscopy (NOESY), ^{15}N-edited NOESY, and ^{13}C-edited NOESY for aromatic regions. The mixing time for all NOESY experiments was set to 70 ms. The 2D ^{1}H-^{15}N-HSQC spectrum of SpoIIID in the complex showed good signal dispersion, indicating that the protein structure is well defined (Fig. 1). Over 97% of the ^{1}H, ^{15}N, and ^{13}C resonances of the protein could be assigned based on the analysis of the spectra mentioned above. A number of 2D and 3D [^{13}C,^{15}N]-filtered NMR spectra, including [^{13}C,^{15}N](ω1)-filtered, [^{13}C](ω2)-edited NOESY; [^{13}C](ω1)-edited, [^{13}C,^{15}N](ω2)-filtered NOESY; [^{13}C,^{15}N](ω1 and ω2)-filtered NOESY; and ^{15}N-edited/^{13}C,^{15}N-filtered and ^{13}C-edited/^{13}C,^{15}N-filtered NOESY, were collected to try to obtain information about the DNA structure and intermolecular NOEs (21, 22). For the most part, our efforts were unsuccessful due to the insensitivity of the filtered experiments in this biomolecular system. However, a few intermolecular NOEs were obtained from the 3D ^{13}C,^{15}N-filtered (F1), ^{15}N-edited (F3) NOESY spectrum that were not obtained with the 3D ^{15}N-edited NOESY spectrum (see Table S1 in the supplemental material), which provided us useful information to generate a model of

the protein-DNA complex (23). Backbone [^{1}H]-^{15}N heteronuclear NOEs were measured and analyzed as described previously (24).

Calculation of the SpoIIID structure. The solution structure of SpoIIID complexed with DNA was calculated using a torsion angle dynamic-simulated annealing protocol with the program CYANA 2.1 (25). The structure calculation was performed for all residues of SpoIIID, although the N- and C-terminal regions were unstructured. The NOE distance restraints were obtained from 3D ^{13}C-edited and ^{15}N-edited NOESY spectra and categorized on the basis of the NOE peak intensities. Dihedral angle values were derived from TALOS (26). Only residues with all 10 predictions lying in the same region of the Ramachandran plot were used. The isomerization state of all proline residues was determined as *trans* (27). Backbone hydrogen bond restraints were applied for residues that showed helical ^{13}C chemical shifts and regular helical secondary structure NOEs (27). The initial structure calculation was carried out using both distance restraints and dihedral angle restraints. A total of 100 conformers was generated, the first 20 of which, with lowest target function, were used for structural analysis.

Binding of SpoIIID to DNA. Plasmids and strains used to express wild-type and single-Ala-substituted forms of SpoIIID are described in Table S2 in the supplemental material. Mutations were introduced into *spoIIID* using the QuikChange site-directed mutagenesis kit (Stratagene) and mutagenic oligonucleotides that are listed in Table S3. All mutant *spoIIID* genes were sequenced at the Michigan State University Research Technology Support Facility to confirm that no undesired mutations were present. Overexpression of SpoIIID in *E. coli* BL21(DE3) (Novagen) and purification for electrophoretic mobility shift assays (EMSAs) were described previously (15). The concentrations of wild-type and Ala-substituted SpoIIID were determined as described previously (15), except the concentrations of SpoIIID R44A and SpoIIID K76A, which were too low for reliable absorbance measurements, were estimated from Coomassie blue staining (see Fig. S3). EMSAs were performed as described previously (15), except a lower concentration (0.1 nM) of labeled probe DNA was used, so the apparent K_d (dissociation constant) could be determined by plotting the linear range of the log of the ratio of bound to free probe DNA versus the log [SpoIIID] and observing the [SpoIIID] at which the line intersected the *x* axis (i.e., [bound DNA] = [free DNA]).

Structural modeling of the protein-DNA complex. Models were calculated using the information-driven biomolecular docking program HADDOCK v2 (28–30). Intermolecular NOEs (see Table S1 in the supplemental material), mutational data (i.e., effects on DNA binding of Ala substitutions in SpoIIID and of base pair changes in DNA presented here), and conserved sequences in natural SpoIIID binding sites (3, 4, 9, 31) were translated into ambiguous interaction restraints (AIRs) to drive the docking process (see Table S4). Thirty-one AIRs with a 2-Å distance definition were used in the HADDOCK docking. The 20 conformers of SpoIIID in the DNA-bound conformation with the unstructured C-terminal region of SpoIIID (residues 82 to 93) removed were used for the docking. Residues showing intermolecular NOEs and/or effects on DNA binding when changed to Ala were selected as active residues (see Table S4). Based on the intermolecular NOEs on α-helical regions of SpoIIID, semiflexible regions were defined as residues 34 to 44, 57 to 65, and 67 to 81. Because no structural information of the DNA portion was available, a model structure of a standard B-form 14-mer DNA duplex was generated using the 3D-DART server (32). DNA base pair restraints were defined as described previously (30). Base pairs that, when mutated, affected DNA binding and/or base pairs conserved in natural SpoIIID binding sites were selected as active base pairs (see Table S4). Additional restraints to maintain base pair planarity and Watson-Crick bonds were introduced for the DNA. A total of 1,000 structures were calculated in the rigid-body docking stage. The 200 structures with the lowest intermolecular energy were selected for semiflexible simulated annealing with the AIRs as intermolecular restraints, all NMR experimental restraints used earlier for the protein structure determination, and Waston-Crick bonds and planarity restraints as intramolecular restraints for the DNA. During this stage, DNA was considered fully flexible and the protein side chains of the three semiflexible regions were allowed to move (30). Further refinement of the 200 structures was performed in an explicit solvent with all restraints mentioned above. Finally, the 200 refined structures were clustered using HADDOCK2.1 package scripts with a backbone root mean square deviation (RMSD) cutoff of 4.5 Å for the protein and DNA. This generated just one cluster of 200 structures that was analyzed further. The top 10 structures with the lowest HADDOCK score were selected to represent the protein-DNA complex.

Protein structure accession number. The structure of SpoIIID complexed with DNA was deposited in the Protein Data Bank under accession number 2L0k.

RESULTS

Structure of SpoIIID in complex with DNA. The three-dimensional structure of SpoIIID bound to DNA was determined using restraints, including 2,194 NOE distance and 132 torsion angle restraints (Table 1), derived from heteronuclear multidimensional NMR spectroscopy (Fig. 1). The DNA structure could not be determined (see below), but it consisted of a 14-bp duplex containing a single copy of the 10-bp idealized binding site consensus sequence (5′-TAGGACAAGC-3′) (see Fig. S2 in the supplemental material), and previous work showed that an SpoIIID monomer forms a 1:1 complex with this DNA molecule (15). Figure 2A shows an ensemble of 20 conformers of SpoIIID with the lowest target functions, representing its three-dimensional NMR structure in complex with DNA. The structure is well defined for the structured region (residues 4 to 81), with an RMSD of 0.43 Å for the protein backbone and 0.86 Å for all heavy atoms (Table 1). No distance or torsion angle restraints are violated by more than 0.3 Å or 5°, respectively. Ramachandran plot analysis of the structures with PROCHECK-NMR (33) showed that of the nonglycine and nonproline residues, 87.5% and 12.5% are in the most favorable and additionally allowed regions, respectively. The N-terminal tripeptide and the C-terminal residues 82 to 93 are disordered, as indicated by negative backbone [^{1}H]-^{15}N heteronuclear NOEs (Fig. 2B), the lack of proton resonances for residues His2 and Asp3, and the small number of medium-range NOEs and lack of long-range NOEs between C-terminal residues 82 to 93 and the rest of the protein (see Fig. S4).

The structured region of SpoIIID (residues 4 to 81) consists of five helices (Fig. 2C). The first four helices form a rigid and compact architecture in the order α1 (residues 4 to 19), the HTH motif (α2-turn-α3; residues 23 to 48), and α4 (residues 51 to 65). The HTH motif is connected to the C-terminal end of α1 by a turn involving Lys20, Lys21, and Thr22, and there are extensive side chain interactions between α1 and the HTH motif, including a salt bridge between the side chains of Arg8 in α1 and Asp40 in α3 that likely explains the instability of D40K-substituted SpoIIID (15). α4 folds back onto this structure via a sharp turn centered at Asn49 (with a dihedral angle, ϕ, of −149°) that positions the main chain of α4 almost antiparallel to α3 at an ~30° angle, contributing to the formation of the hydrophobic core of the protein (Fig. 2D). The C-terminal end of α4 protrudes from the core but is associated with it via interactions between the side chain of His63 in α4 and the side chains of Thr22 in the first turn and Val23 in α2. All four α-helices (α1 to α4) are amphipathic, with their hydrophobic residues oriented toward the center of the bundle. The side chains of Ile12, Ile16, Ile26, Val37, Leu41, Leu45, Leu52, and Val56 are deeply buried in the hydrophobic core, which is so compact that a water molecule cannot fit inside. The side chain of Ile12

TABLE 1 Structural statistics for the SpoIIID ensemble

Parameter	Value for SpoIIID
NMR distance and dihedral restraints	2,418
NOE distance restraints	2,194
Intraresidue (|i-j| = 0)	628
Sequential (|i-j| = 1)	633
Medium range (|i-j|≤4)	621
Long range (|i-j|>4)	312
Hydrogen bonds	92
Dihedral angle restraints[a]	132
ϕ	66
ψ	66
Structure statistics	
Residual CYANA target function (Å)	1.42 ± 0.0052
Violations from exptl restraints from the 20 structures	
No. of distance restraint violations of >0.30 Å	0
No. of dihedral angle restraint violations of >5.0°	0
No. of van der Waals violations of >0.30 Å	0
Maximum dihedral angle restraint violation (°)	4.39
Maximum distance constraint violation (Å)	0.28
Maximum van der Waals violations (Å)	0.27
Deviations from idealized geometry[b]	
Bond lengths (Å)	0.005
Bond angles (°)	0.7
Coordinate precision[c] (Å)	
Protein backbone	0.43 ± 0.09
Protein heavy atoms	0.86 ± 0.12
Ramachandran statistics[d] (%)	
Most favored regions	87.5
Additionally allowed regions	12.5
Generously allowed regions	0.0
Disallowed regions	0.0

[a] The dihedral angles were predicted by using the program TALOS (30).
[b] The data were generated from the ADIT Validation Server at the RCSB Protein Data Bank.
[c] The coordinate precision is defined as the average RMSD between the 20 SpoIIID structures and the mean coordinates. The reported values are for residues Tyr4-Lys81, and the backbone refers to the N, Cα, and CO atoms.
[d] PROCHECK statistics calculated over the ensemble of 20 structures.

in α1 inserts into the hydrophobic pocket between α2 and α3, with one of its Cγ1 protons packed above the aromatic ring of Phe30 in α2 (so close to the ring that one of the Cγ1 protons is shielded and its chemical shift is shifted to −0.485 ppm because of the ring current effect, compared with 1.42 ppm for the other Cγ1 proton of Ile12). The architecture of the turn in the HTH is maintained by hydrophobic interactions between the side chains of Val32 in the turn and Ala27 in α2. α3 is capped by the hydroxyl groups of Ser35 and Thr36 at its N-terminal end, which, together with that of Ser33 in the turn of the HTH, form a cluster of hydroxyl groups in a triangle arrangement. The final α-helix, α5 (residues 67 to 81), is connected to α4 by a kink at Ile66 (with a dihedral angle, ϕ, of −116°). It extends away from the central structured region, and its C-terminal end displays some mobility, as indicated by smaller backbone [^{1}H]-^{15}N heteronuclear NOEs (Fig. 2B). However, α5 plays a critical role in DNA binding (see below).

Structural comparison to other proteins. Structural alignments with other proteins revealed that SpoIIID bound to DNA is unique among proteins containing HTH domains. All HTH domains contain a prototypic core structure composed of three helices arranged in a triangular fashion, and different families of HTH domains are distinguished by various extensions of the prototypic core structure (10). The prototypic core structure of SpoIIID consists of the first three helices, which is best aligned with the σ_4 domain of the primary σ factor from *Thermus aquaticus* (Protein Data Bank [PDB] entry 1KU3) (34) with an RMSD of only 1.44 Å over 44 aligned residues (C_α atoms). The HTH motif in the σ_4 domain recognizes the −35 element of bacterial promoters.

When the entire structured region of SpoIIID was used for a structural similarity search of the Protein Data Bank using the DaliLite server (35), however, no protein was found with an extension similar to helices 4 and 5 of SpoIIID. The top three matches were the N-terminal HTH domains of a probable transcriptional regulator (PA0477) from *Pseudomonas aeruginosa* (PDB entry 2ESN, chain B; 300 residues) with a Z-score of 5.8, a transcription regulator (TM1602) from *Thermotoga maritima* (PDB entry 1J5Y; 172 residues) with a Z-score of 5.5, and the manganese transport regulatory protein, MntR, from *E. coli* (PDB entry 2H09; 155 residues) (36) with a Z-score of 5.3. These three HTH domains are all winged HTH domains with a two-stranded β-sheet extension to the C terminus of the prototypic core structure. In contrast, the extension to the C terminus of the prototypic core structure of SpoIIID features two helices. The three HTH domains were selected as top matches with the best Z-scores by the DaliLite server, presumably because the small β-sheets of these HTH domains have an orientation similar to that of helix 4 of SpoIIID, as illustrated for PA0477 (Fig. 3A), which has an RMSD of 3.1 Å over 57 aligned residues (C_α atoms) with 14% sequence identity to SpoIIID. Conversely, presumably because the helical extensions have rather different orientations than helices 4 and 5 of SpoIIID, none of the HTH domains with helical extensions was selected among the top matches by the DaliLite server. This is supported by the result of a structural similarity search using the FATCAT server (37), which performs flexible structural alignment by allowing twists. The top match of this structural similarity search was the σ_4 domain of the flagellar σ factor σ^{28} from *Aquifex aeolicus* (PDB entry 1RP3, chain A; 239 residues) (38). The σ_4 domain has large helical extensions to both the N and C termini of the prototypic HTH core structure. With rigid structure alignment, the two structures could be aligned only for the three helices of the core structure, with an RMSD of 1.5 Å for 44 aligned residues (C_α atoms) (Fig. 3B). With flexible structure alignment with one twist, resulting in a break between residues 136 and 137, the two structures could be aligned with an RMSD of 1.97 Å for 68 residues (Fig. 3C). Taken together, the structural similarity searches showed that SpoIIID has a novel helical extension to the prototypic HTH core structure and represents a new family of HTH domain-containing proteins.

Interaction of SpoIIID with DNA. We next tried to determine the structure of the DNA in the complex and investigate the binding interface between SpoIIID and the DNA by recording a number of 2D and 3D ^{13}C/^{15}N-filtered NMR spectra on a 900-MHz NMR spectrometer. We were not able to assign the resonances

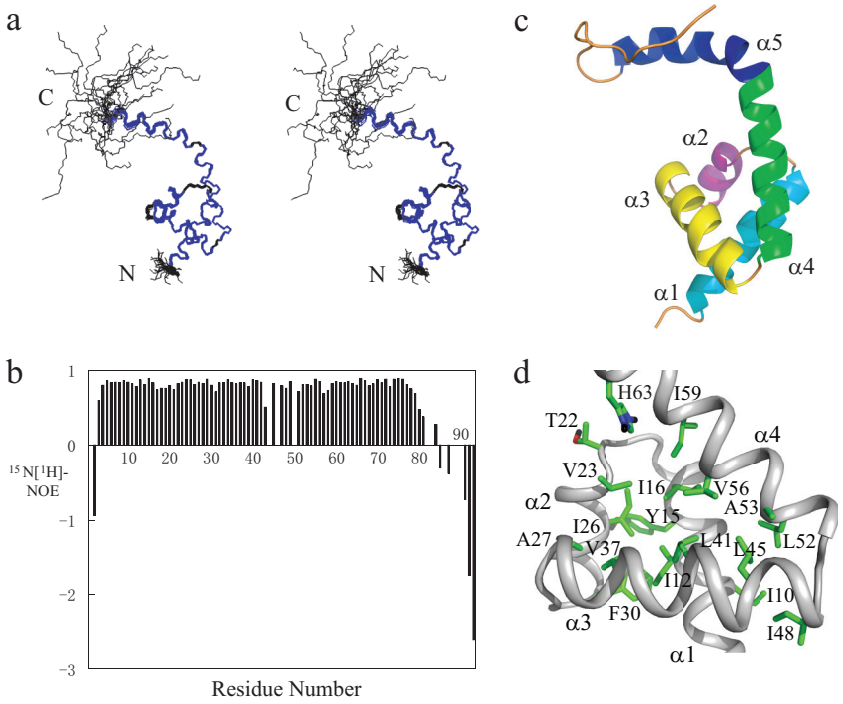

FIG 2 Three-dimensional protein structure and [^{1}H]-^{15}N heteronuclear NOEs of the SpoIIID-DNA complex. (A) Stereo view of the superimposed backbone traces of the 20 NMR-derived lowest-energy structures with disordered and loop regions colored black and well-defined regions colored blue. (B) [^{1}H]-^{15}N heteronuclear NOEs plotted against the amino acid sequence. (C) Ribbon representation of the average structure derived from the 20 lowest-energy structures. (D) Hydrophobic contacts between helices.

from the DNA, as the isotopically filtered NMR experiments performed on this biomolecular system were not sensitive enough to yield data of sufficient quality for such an analysis. Because the NMR signals of the DNA could not be assigned to its constituent atoms, structure determination of the bound DNA was not possible. However, the 3D ^{15}N-edited/^{13}C,^{15}N-filtered NOESY spectrum allowed us to identify intermolecular NOEs between the amide protons of the protein and DNA bases or riboses (Fig. 4A

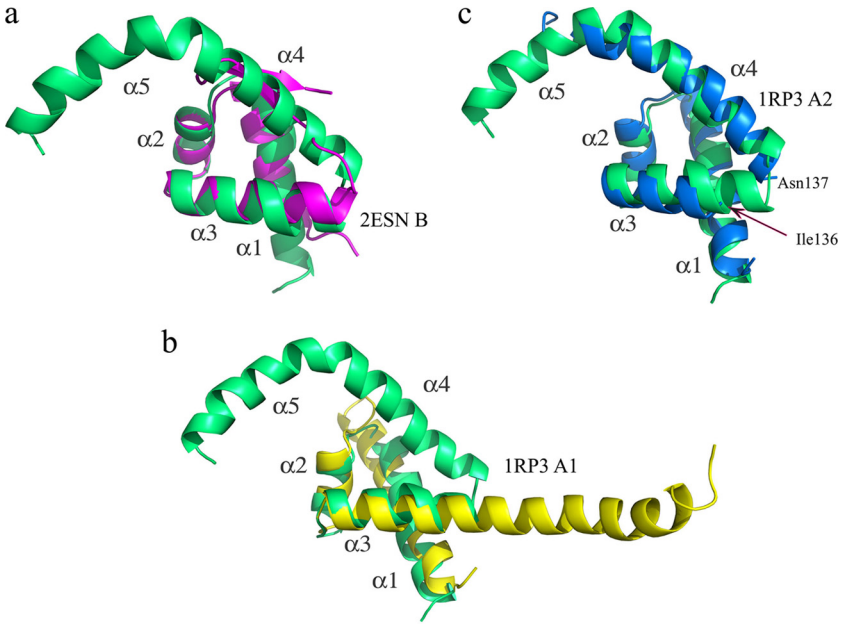

FIG 3 Structural comparison to other proteins. SpoIIID (green) is superimposed with the winged HTH domain of a putative transcriptional regulator, PA0477 from *P. aeruginosa* (PDB entry 2ESN, chain B) (A), and the σ_4 domain of the flagellar σ factor σ^{28} from *Aquifex aeolicus* (PDB entry 1RP3, chain A) without any twist (B) or with a twist resulting in a break at residue 136 (C).

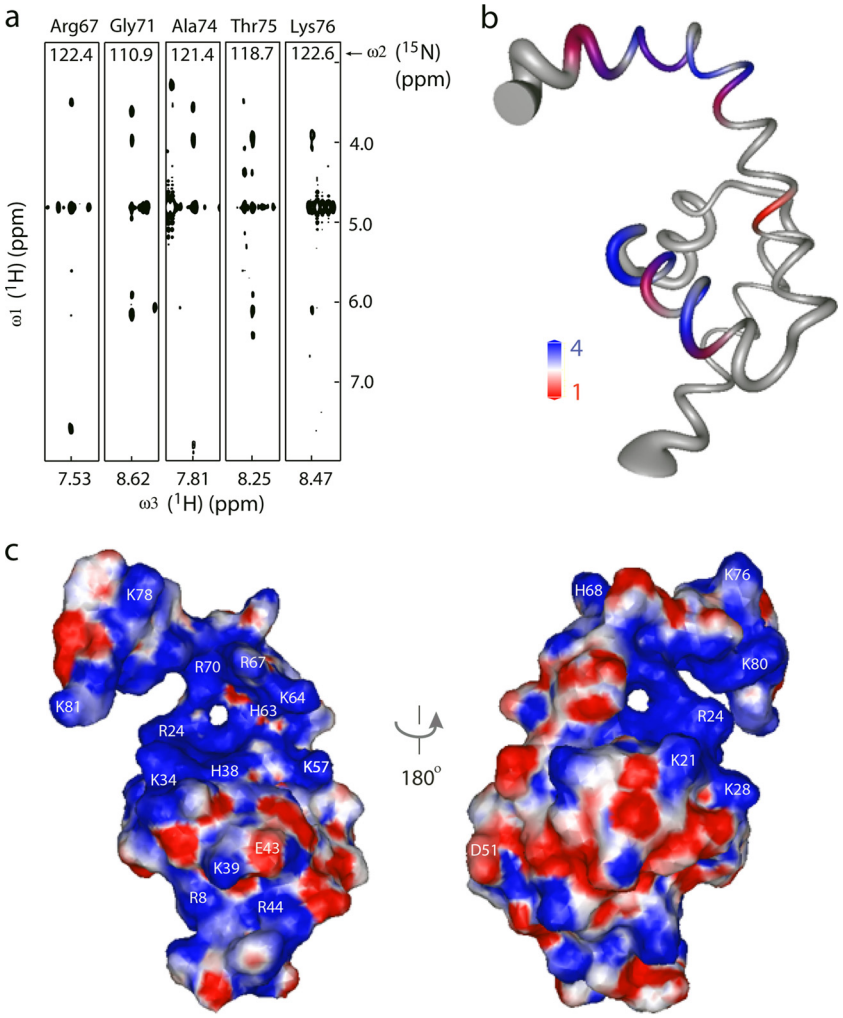

FIG 4 Molecular determinants of the SpoIIID-DNA interface. (A) Selected strips from a 3D ^{15}N-edited, [^{13}C,^{15}N]-filtered NOESY spectrum for several residues in the C-terminal α-helix (α5) of *spoIIID*, indicative of intermolecular NOEs with DNA. (B) SpoIIID (residues 1 to 81) in the same orientation as that of Fig. 2C with residues color coded according to the number (1 to 4) of intermolecular NOEs. (C) Surface electrostatic potential representation of SpoIIID (residues 1 to 81). Positive, blue; negative, red; neutral, gray. The orientation of the left panel, which we designate the front, is the same as that shown in Fig. 2C.

and B; also see Table S1 in the supplemental material). As anticipated, residues Ser35, Thr36, Glu43, and Arg44 in the putative recognition helix (α3) of the HTH motif exhibited intermolecular NOEs. Strikingly, Lys64 in α4 and Arg67, Gly71, Gly72, Ala74, Thr75, and Lys76 in α5 also displayed intermolecular NOEs, indicating that these regions interact with DNA.

Electrostatic surface potential representations show that DNA-bound SpoIIID (residues 1 to 81) has two positively charged patches over its front (Fig. 4C), an extensive tract between the HTH motif and helices α4 and α5 involving Arg24, Lys34, His38, Lys57, His63, Lys64, Arg67, Arg70, Lys78, and Lys81 that form a positively charged groove, and a smaller patch on the lower part of the front between α1 and α3 involving Arg8, Lys39, and Arg44. The back of the structure shows the presence of evenly distributed negative and positive charges (Fig. 4C).

The charge distribution on the front of SpoIIID and the intermolecular NOE data are consistent with the suggestion from previous analysis of alterations to SpoIIID that two regions are important for DNA binding, the putative recognition helix (α3) of the HTH motif and a basic region (α5) near the C terminus (15). However, the previous mutational analysis employed charge reversal substitutions in SpoIIID, which are more likely to affect its structure and interaction with DNA than Ala substitutions. Therefore, we expressed in *E. coli* and purified wild-type SpoIIID and 9 altered forms of SpoIIID, each with a single-Ala substitution in the putative recognition helix of the HTH or in a basic (Lys) residue near the C terminus. All of the Ala-substituted proteins bound probe 10, containing a single copy of the idealized binding site consensus sequence (Fig. 5A) (15), with affinity similar to that of wild-type SpoIIID in EMSAs, except E43A, which had an approximately 2-fold lower affinity (Table 2).

Since binding to probe 10 failed to reveal differences in affinity among most of the SpoIIID proteins, we measured binding to probes 15 to 17, each differing from probe 10 by a single base pair in the highly conserved ACA sequence in the center of the binding site consensus (Fig. 5A and B). Wild-type SpoIIID bound probes 15 and 16 with reduced affinity (Table 2), indicating the importance of the ADE21/THY8 and CYT22/GUA7 base pairs for DNA

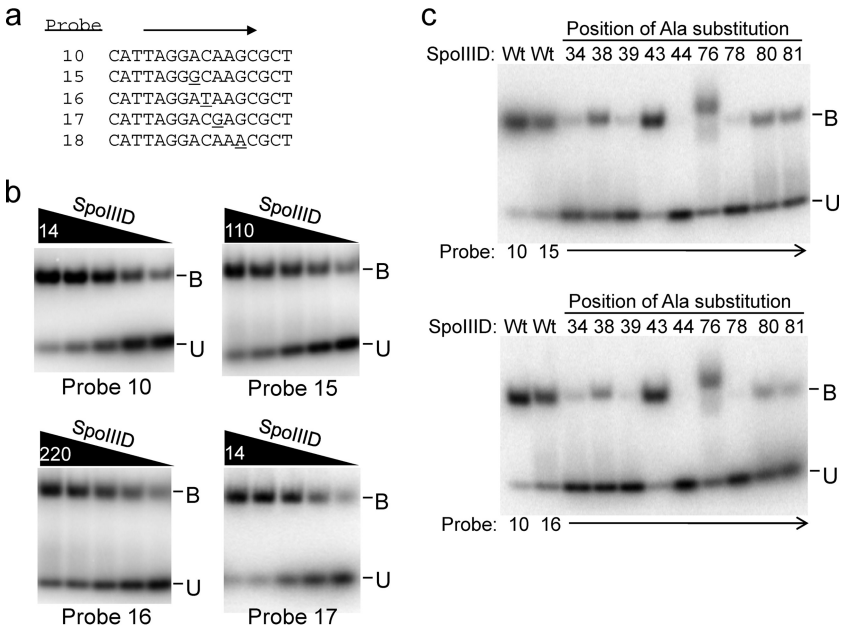

FIG 5 Binding of SpoIIID to DNA. (A) Sequences of DNA probes. Only one strand of each probe is shown. The arrow denotes the idealized binding site consensus sequence in probe 10, and underlined bases in the other probes indicate differences from probe 10. (B) EMSAs of wild-type SpoIIID binding to different DNA probes. A 2-fold dilution series of SpoIIID beginning at the concentration (nM) above the leftmost lane in each panel is shown for different probes (indicated below each panel) (0.1 nM). Bound (B) and unbound (U) probe are indicated. (C) EMSAs of wild-type and altered SpoIIID binding to DNA. Wild-type (Wt) or single-Ala-substituted forms of SpoIIID (840 nM) were tested with probe 10 (both panels), probe 15 (top panel), or probe 16 (bottom panel) (0.1 nM).

binding (see Fig. S2 in the supplemental material for numbering of bases). Strikingly, 7 Ala-substituted SpoIIID proteins bound weakly or not at all to probes 15 and 16 (Fig. 5C and Table 2). Given the high-affinity binding of these proteins to probe 10 (presumably indicative of proper folding), their impaired binding to probes 15 and 16 strongly suggests that Lys34, His38, Lys39, Arg44, Lys78, Lys80, and Lys81 of SpoIIID are important for binding to DNA (at least to sites with 1 or more mismatches to the idealized binding site consensus sequence). Lys76 also contributes to binding to probes 15 and 16, since K76A had about 4-fold lower affinity than wild-type SpoIIID (Table 2). K76A produced two complexes with slightly different mobilities from that of the complex produced by wild-type SpoIIID or the other Ala-substituted proteins (Fig. 5C). K76A might exhibit two modes of binding to DNA that differ from that of wild-type SpoIIID (see Discussion). Interestingly, despite its 2-fold lower affinity for probe 10, E43A bound probes 15 and 16 with affinity similar to that of wild-type SpoIIID (Table 2).

Although wild-type SpoIIID bound probe 17 with affinity similar to that of probe 10, only 2 Ala-substituted SpoIIID proteins exhibited this behavior, E43A and K81A (Table 2). The other proteins had lower affinity for probe 17 than for probe 10, with H38A, K76A, and K80A being about 2-fold lower, K39A and K78A being 3- to 4-fold lower, and K34A and R44A being 19- and 9-fold lower, respectively. These results indicate that the ADE23/THY6 base pair is somewhat important for DNA binding (at least by Ala-substituted SpoIIIDs) and suggest that Lys34 and Arg44 contribute the most to binding to probe 17, followed by Lys39 and Lys78 and finally His38, Lys76, and Lys80.

To search for a second probe capable of distinguishing the relative contributions of SpoIIID residues to DNA binding, we screened probes 18 to 22 for binding to wild-type SpoIIID, K39A, and R44A (see Fig. S5 in the supplemental material). Wild-type SpoIIID bound probe 18 with affinity similar to that of probe 10, but K39A and R44A bound weakly to probe 18. Therefore, we measured the affinity of all 9 Ala-substituted SpoIIID proteins for probe 18 (Table 2). All of the proteins except E43A exhibited lower affinity for probe 18 than for probe 17, indicating that the GUA25/CYT4 base pair is more important for binding of Ala-substituted SpoIIIDs than the ADE23/THY6 base pair. Lys34 and Arg44 were crucial for binding to probe 18, the same two residues that appeared to contribute most to probe 17 binding. In addition, Lys39 was crucial for binding to probe 18, followed by Lys78, His38,

TABLE 2 Binding of wild-type and altered SpoIIID to DNA

SpoIIID	Binding with probe no.: 10	15	16	17	18
Wild type	2.4 ± 0.2[a]	28 ± 0.8	69 ± 8	3.0 ± 0.5	2.7 ± 0.2
Altered					
K34A	2.2 ± 0.7	>7,700[b]	>7,700	42 ± 5	>840
H38A	2.6 ± 0.6	>2,500	>2,500	5.7 ± 0.2	38 ± 7
K39A	1.8 ± 0.1	>12,000	>12,000	8 ± 1	>840
E43A	5.4 ± 0.8	32 ± 0.6	54 ± 5	5.3 ± 0.1	5.9 ± 0.1
R44A	2.7 ± 0.3	>840	>840	25 ± 4	>840
K76A	3.2 ± 0.6	140 ± 30	300 ± 50	5.3 ± 0.7	11 ± 2
K78A	2.5 ± 0.8	>11,000	>11,000	7 ± 1	120 ± 20
K80A	1.8 ± 0.3	>14,000	>14,000	3.5 ± 0.4	4.8 ± 0.2
K81A	2.0 ± 0.9	>12,000	>12,000	2.6 ± 0.4	3.9 ± 0.3

[a] Apparent K_d (nM) measured by EMSAs. Averages and standard deviations from at least 3 determinations are shown. EMSAs were performed as described previously (15). The binding reaction buffer contained 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM EDTA, 5% glycerol, and 0.1 mM double-stranded poly(dI-dC).

[b] The number is the highest concentration of altered SpoIIID (nM) that was tested, and less than half of the probe was bound at that concentration.

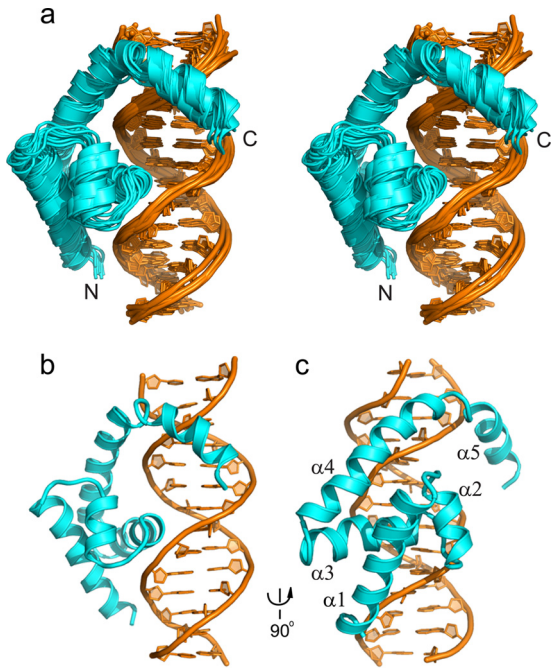

FIG 6 Models of the protein-DNA complex generated using HADDOCK. (A) The ensemble of the top 10 models in stereo view. (B and C) The best model in cartoon representation viewed from different angles.

Lys76, Lys80, and Lys81, in order of decreasing apparent contribution to probe 18 binding, in excellent agreement with their apparent contributions to probe 17 binding.

Our DNA-binding data indicate that the ADE21/THY8 and CYT22/GUA7 base pairs near the center of the consensus sequence are the most important for SpoIIID binding, followed by GUA25/CYT4 and then ADE23/THY6. Our data also suggest that Lys34 and Arg44 of SpoIIID contribute most to its affinity for DNA, followed by Lys39 and Lys78 and then His38. The other residues tested contribute less to DNA-binding affinity, and their relative contributions vary for different mutant DNA probes. These mutational data, together with our other data and information about conserved sequences in natural SpoIIID binding sites, were used to derive AIRs (see Table S4 in the supplemental material) for information-driven docking of SpoIIID to its idealized binding site consensus sequence.

Structural modeling of the protein-DNA complex. Although as mentioned above the intermolecular NOE-derived protons from the DNA molecule could not be unambiguously identified, the intermolecular NOEs involving residues of SpoIIID (Fig. 4A and B; also see Table S1 in the supplemental material) combined with our mutational data (Fig. 5 and Table 2) and conserved sequences in natural SpoIIID binding sites (3, 4, 9, 31) could be translated into 31 AIRs (see Table S4) to facilitate modeling of the protein-DNA complex (39) using the information-driven docking program HADDOCK (28–30). After rigid body docking, semiflexible simulated annealing, and explicit water refinement, 200 models were clustered using a pairwise backbone RMSD of 4.5 Å as a cutoff. Importantly, this resulted in only one cluster, with an average HADDOCK score of −717 ± 115 kcal/mol. The finding that all 200 models produce a single cluster using a 4.5-Å pairwise RMSD cutoff indicates that the information in the 31 AIRs sufficiently restrains the models to one orientation of SpoIIID with respect to the DNA, and that the location of SpoIIID with respect to the DNA (and the consensus sequence) is fairly well defined. In all of the models, the DNA-interacting surface of SpoIIID is composed primarily of residues from two regions: (i) the recognition helix α3, which inserts into the major groove of the DNA near the consensus sequence, and (ii) helices α4 and α5, which interact with the adjacent minor groove of the DNA. The latter interaction of the unique C-terminal helical extension of SpoIIID with the minor groove of DNA can explain how a monomer of SpoIIID achieves high-affinity DNA binding, and the modeling provides strong evidence for this novel mode of DNA binding by an HTH protein. The models also reveal a third region in helix α1 that interacts with DNA, although less extensively.

An ensemble of the top 10 SpoIIID-DNA models of the cluster is displayed in Fig. 6A, and it has a pairwise RMSD of 1.05 ± 0.33 Å over all of the backbone atoms and an average HADDOCK score of −933 ± 27 kcal/mol (Table 3). Ramachandran plot analysis of the top 10 models indicated that 93.4% of the protein residues are in the most-favored regions. While the top 10 models are in excellent agreement (i.e., the orientation and location of SpoIIID with respect to the DNA is well defined), there remains uncertainty in the details of the interaction surface between SpoIIID and DNA (i.e., specific hydrogen bond and van der Waals interactions). Figure S6A in the supplemental material illustrates this point for hydrogen bond interactions. Despite this uncertainty, it is worth noting that in the majority of the top 10 models, as indicated by the black, red, and blue lines in Fig. S6A, there are 16 instances of an SpoIIID residue predicted to form at least one hydrogen bond with a sugar (3 instances) or phosphate (13 instances) of the DNA

TABLE 3 Structural statistics for the top 10 SpoIIID-DNA complex models

Parameter	Value for SpoIIID-DNA
Energy statistics	
van der Waals energy (kcal/mol^{-1})	−704.6 ± 18.4
Electrostatic energy (kcal/mol^{-1})	−5,402.0 ± 97.9
Desolvation energy (kcal/mol^{-1})	21.4 ± 2.9
AIR-energy (kcal/mol^{-1})	19.6 ± 2.1
AIR-violations (Å)	1.1 ± 0.3
AIR RMS (Å)	0.2 ± 0.01
HADDOCK score (kcal/mol^{-1})	−933.1 ± 26.5
Structural statistics	
Deviation from ideal geometry	
Bond (Å)	0.0322 ± 0.00005
Angle (°)	2.499 ± 0.0016
Impropers (°)	0.378 ± 0.0011
Average RMS difference backbone (Å)	
Interface all	1.00 ± 0.37
All	1.05 ± 0.33
Interface protein	1.17 ± 0.44
Interface DNA	0.88 ± 0.25
Buried surface area (Å^2)	2,234 ± 94
Ramachandran plot (%)	
Residues in most favored regions	93.4
Residues in additionally allowed regions	6.48
Residues in generously allowed regions	0.12
Residues in disallowed regions	0

backbone and only 4 such interactions with a DNA base. Hence, the models predict that SpoIIID forms many hydrogen bonds with the sugar-phosphate backbone of DNA and fewer with the DNA bases, consistent with the high affinity and moderate sequence specificity of DNA binding observed for SpoIIID in previous studies (3, 4, 9, 31). Many residues in the positively charged groove (His38, Lys57, His63, Arg70, Lys78, and Lys81) and one in the positively charged patch (Arg8) on the front of SpoIIID (Fig. 4C) are predicted to interact with a phosphate of the DNA backbone in the majority of the top 10 models (see Fig. 6SA). In contrast, only 4 residues (Ser35, Lys39, Arg67, and His68) are predicted to interact with a DNA base in the majority of the top 10 models, and one of these (between His68 and C28) is to a base outside the consensus sequence, so how SpoIIID achieves sequence-specific DNA binding is not well predicted by the models.

The best SpoIIID-DNA model, i.e., the model with the lowest HADDOCK score, is shown in Fig. 6B and C. Figure S6B in the supplemental material illustrates the hydrogen bond interactions between SpoIIID residues and DNA in the best model. Five residues are predicted to make more than one hydrogen bond with a phosphate of the DNA backbone. In all, 12 residues of SpoIIID are predicted to form 21 hydrogen bonds with a sugar (5 bonds) or phosphate (16 bonds) of the DNA backbone, and 3 residues of SpoIIID (Ser35, Thr36, and Lys39) in the recognition helix of the HTH are predicted to form 5 hydrogen bonds with a DNA base of the consensus sequence.

Taken together, our modeling provides strong evidence for a novel mode of DNA binding by SpoIIID in which its unique C-terminal extension interacts with the minor groove. This interaction and that of the HTH recognition helix involves many hydrogen bonds to the sugar-phosphate backbone of DNA and far fewer to DNA bases according to the top 10 models, providing a possible explanation for the general DNA-binding characteristics of SpoIIID (i.e., high-affinity binding as a monomer to sequences matching a degenerate consensus, indicative of moderate sequence specificity). The models also provide possible explanations for many of the observed effects of substitutions in SpoIIID on transcription *in vivo* (15) and on binding to DNA *in vitro* (Table 2); however, the models do not explain all of the experimental observations, and the models make some predictions that remain to be tested (see Discussion).

DISCUSSION

The structure of DNA-bound SpoIIID revealed a new type of HTH-containing protein with a novel C-terminal extension comprised of two helices connected by a kink. Intermolecular NOEs and the effects of Ala substitutions in SpoIIID indicate that both the HTH and the C-terminal extension interact with DNA. The interaction of the unique C-terminal extension of SpoIIID with DNA presumably explains how a monomer of this HTH protein achieves high-affinity binding. Information-driven modeling of the SpoIIID-DNA complex resulted in a single cluster of models in which the recognition helix of the HTH interacts with the major groove of DNA and the C-terminal extension interacts with the adjacent minor groove. The modeling provides strong evidence for a novel mode of DNA binding by SpoIIID.

The NMR solution structure of SpoIIID in complex with DNA is well defined for residues 4 to 81 of SpoIIID. The C-terminal residues 82 to 93 are disordered but are not needed for DNA binding *in vitro* or for SpoIIID-dependent transcription *in vivo* (15). The disordered region might have interfered with previous efforts to crystallize SpoIIID in complex with DNA (P. Himes, J. Geiger, and L. Kroos, unpublished data). This effort should be revisited with truncated SpoIIID lacking the disordered region. We tried to determine the structure of SpoIIID in the absence of DNA, but SpoIIID was unstable in solution without DNA, making it impossible to collect a set of NMR data suitable for structure determination of the apo protein.

A new type of HTH-containing protein. SpoIIID represents a new family of HTH domain-containing proteins due to its C-terminal extension from residues 51 to 81, which features two helices connected by a kink at residue 66. The novel C-terminal extension of *B. subtilis* SpoIIID is likely conserved among SpoIIID orthologs, which exhibit 39% identity and 79% similarity to residues 52 to 79 of the *B. subtilis* protein (see Fig. S7 in the supplemental material). The orthologs are found only in endospore-forming bacteria, so they likely play a similar role in governing gene expression critical for sporulation, although this largely remains to be tested. Recently, SpoIIID of *C. difficile* was shown to play an important role in sporulation of this emergent pathogen, upregulating transcription of *sigK* (40), as in *B. subtilis* (4, 5). Understanding how SpoIIID binds to DNA and activates transcription of genes crucial for sporulation could reveal potential targets for development of sporulation inhibitors. Sporulation contributes to the environmental persistence and transmission of pathogenic *Bacilli* and *Clostridia* (41–43). Sporulation likely also contributes to persistence in the host upon antibiotic treatment (44). Spores have been shown to germinate and resporulate in the mouse gastrointestinal tract (45, 46), so the ability to inhibit sporulation may increase the efficacy of therapeutics (47, 48).

High-affinity DNA binding as a monomer. Our work provides new insight into how an SpoIIID monomer binds DNA with high affinity. Previous work implicated the putative recognition helix of the HTH and a C-terminal basic region in DNA binding (15). Charge reversal substitutions in the putative recognition helix of SpoIIID greatly impaired or eliminated DNA binding *in vitro*, as did a C-terminal truncation ending at residue 75, but not one ending at residue 81. To extend this analysis, we chose 5 residues for which charge reversal substitutions had been tested (Lys34, His38, Lys39, Glu43, and Arg44) and 4 basic residues in the C-terminal region (Lys76, Lys78, Lys80, and Lys81) to test the effects of single-Ala substitutions on DNA binding. Surprisingly, none of the single-Ala substitutions impaired binding of SpoIIID to DNA containing the idealized consensus sequence (Table 2, probe 10). Apparently, loss of contacts due to single-Ala substitutions did not impair binding to this sequence sufficiently to be detected by EMSAs, whereas the charge reversal substitutions studied previously had introduced unfavorable interactions (15). Two mutations in the highly conserved ACA sequence of the consensus reduced binding of wild-type SpoIIID about 10- to 30-fold and greatly impaired or eliminated binding of the Ala-substituted proteins (Table 2, probes 15 and 16, and Fig. 5). Other mutations in the consensus sequence did not affect binding of wild-type SpoIIID and impaired binding of only some Ala-substituted proteins (Table 2, probes 17 and 18; also see Fig. S5 in the supplemental material). Of the sequences mutated, the highly conserved AC sequence of the consensus is the most important for SpoIIID binding, and of the residues in SpoIIID tested, all except Glu43 are important for DNA binding, with Lys34, Lys39, Arg44, and Lys78 being the most important. These results extend previous work

(15) by showing that Ala substitutions in the putative recognition helix of the HTH or in the C-terminal basic region can impair DNA binding, supporting the model that these two regions allow a monomer of SpoIIID to bind DNA with high affinity. Further, our results highlight the importance of the AC sequence that is highly conserved in SpoIIID binding sites.

Our NMR data provide additional insight into how an SpoIIID monomer binds DNA with high affinity. Analysis of intermolecular NOEs revealed that amide protons of several SpoIIID residues in the region spanning from Lys64 to Thr75 of the novel C-terminal extension likely interact with DNA bases or riboses (Fig. 4A and B; also see Table S1 in the supplemental material). Hence, the C-terminal extension appears to interact with DNA extensively, not just via the basic region from Lys76 to Lys81.

A novel mode of DNA binding. Using the structure of DNA-bound SpoIIID and all of the available information about the protein-DNA interaction, there were enough restraints for the docking program HADDOCK to produce a single cluster of SpoIIID-DNA complex models with a pairwise RMSD cutoff of 4.5 Å. The general agreement of all of the models provides strong evidence for a novel mode of DNA binding in which the recognition helix of the HTH interacts with the major groove of DNA near the consensus sequence and the C-terminal extension interacts with the adjacent minor groove. The novel aspect of the predicted binding mode is the interaction of the C-terminal extension with the minor groove of DNA. A sharp turn after recognition helix α3 is centered at Asn49 and allows helix α4 to interact extensively with a sugar-phosphate backbone of the adjacent minor groove (Fig. 6C). The kink at Ile66 between α4 and α5 allows α5 to make many additional predicted interactions, primarily with the same sugar-phosphate backbone as α4 but ultimately reaching across the major groove to interact with the other backbone in the vicinity of the HTH turn (i.e., the turn between α2 and α3) (Fig. 6B and C).

The predicted minor groove binding by SpoIIID helices α4 and α5 appears to be very different from that by homeodomain proteins or winged-helix proteins that use arms or wings, respectively, to make additional contacts with DNA (16, 17), and also quite different from minor groove binding by the "hinge" helix of PurR and other LacI family members. These dimeric or tetrameric proteins recognize palindromic DNA sequences with HTH motifs that contact major grooves and with symmetric hinge helices, each containing a residue (Leu54 in PurR) that intercalates between base pairs, kinks the DNA, and opens the minor groove for additional interactions with residues of the hinge helices, including one base-specific hydrogen bond in some cases (Lys55 in PurR) (49). Our models predict that SpoIIID α4 and α5 interact much more extensively with an unkinked, unopened minor groove; however, we cannot rule out the possibility that SpoIIID distorts the DNA, as proposed previously based on DNase I hypersensitivity adjacent to some sites in DNA protected by SpoIIID *in vitro* in footprinting experiments (3, 4, 9, 31).

Three observations suggest that the C-terminal basic region from Lys76 to Lys81 in helix α5 interacts flexibly with DNA. First, from residue 76 to residue 81, this region becomes progressively more flexible based on the backbone [^{1}H]-^{15}N heteronuclear NOEs (Fig. 2B). Second, SpoIIID K76A produced two complexes with slightly different mobilities than that produced by wild-type SpoIIID (Fig. 5C), perhaps reflecting different interactions of the basic region (lacking Lys76) with DNA. Third, K76A affected binding to probe 18 more than did K80A or K81A, but the opposite was observed for binding to probes 15 and 16 (Table 2), as if changes in the basic region influence interactions elsewhere in the SpoIIID-DNA complex. Perhaps analogously, the N-terminal arm of homeodomain proteins can exhibit different modes of minor groove binding that influence how the recognition helix of the HTH interacts in the major groove (50). Among SpoIIID orthologs, the C-terminal basic region (corresponding to residues 76 to 81 of *B. subtilis*) is conserved in *Bacilli*, so a flexible minor groove interaction might be conserved, but in *Clostridia* and other distant relatives, only the motif (K, R, Q)XKY (corresponding to residues 76 to 79 of *B. subtilis*) is conserved (see Fig. S7 in the supplemental material).

Predictions and explanations based on SpoIIID-DNA models and the SpoIIID structure. A more detailed analysis of the top 10 SpoIIID-DNA models indicated that despite excellent overall agreement among the models (Table 3), there remains considerable uncertainty in the interaction surface between SpoIIID and DNA at the level of predicted hydrogen bond (see Fig. S6A in the supplemental material) or van der Waals interactions. This is due in part to uncertainty in the position of side chains of surface residues in the SpoIIID structure despite a well-defined backbone. Nevertheless, focusing just on predicted hydrogen bond interactions in the majority of the top 10 models implicates His2, Arg8, Ser33, Thr36, His38, Thr42, Lys57, His63, Arg67, Arg70, Thr75, Lys78, and Lys81 of SpoIIID in hydrogen bonding to the sugar-phosphate backbone of DNA and Ser35, Lys39, Arg67, and His68 of SpoIIID in hydrogen bonding to bases of DNA (see Fig. S6A). In the best model, Arg8, Ser33, Arg70, Lys78, and Lys81 are predicted to make more than one hydrogen bond with a phosphate of the DNA backbone (see Fig. S6B). The top models predict extensive hydrogen bonding between SpoIIID and the sugar-phosphate backbone of DNA and much less hydrogen bonding between SpoIIID and bases of DNA. This may explain how SpoIIID achieves high-affinity DNA binding (apparent K_d of $<$10 nM) (15) (Table 2) but with moderate sequence specificity (i.e., its binding site consensus sequence of WWRRACARNY is quite degenerate). It may also explain the observation that SpoIIID binds to the coding region of some genes (3, 4), including some it does not regulate (3).

Our structure of SpoIIID and modeling of the SpoIIID-DNA complex provide possible explanations for the effects of substitutions in SpoIIID on transcription *in vivo* (15) and on binding to DNA *in vitro* (Table 2). Substitutions in SpoIIID that reduced expression of an SpoIIID-dependent reporter more than 3-fold include H2E, R8E, V23K, R24E, I26E, V32E, S33R, K34E, S35E, T36E, V37E, H38E, K39E, D40K, R44E, D51K, H63E, K64E, H68E, K76E, K78E, and D82K (15). Many of these residues play important roles in forming the structure of SpoIIID: Arg8 and Asp40 form a salt bridge, Val23 and His63 side chains interact (Fig. 2D shows hydrophobic contacts), Ile26 and Val37 side chains help form the hydrophobic core, Val32 and Ala27 side chains interact, and the Ser33, Ser35, and Thr36 hydroxyl groups form a cluster. Some of these residues (Arg8, Ser33, Ser35, Thr36, and His63) are also predicted to form a hydrogen bond with DNA by the majority of the top 10 SpoIIID-DNA models, as are several other residues of SpoIIID in which substitutions impaired reporter expression (His2, His38, Lys39, His68, and Lys78) (see Fig. S6A in the supplemental material). Loss of hydrogen bond interactions with DNA might also explain why Ala substitutions for

His38, Lys39, or Lys78 in SpoIIID impaired binding *in vitro* to DNA probes differing by 1 bp from the idealized consensus sequence (Table 2).

The top models of the SpoIIID-DNA complex make some predictions that remain to be tested. GUA20 is predicted to form a hydrogen bond with Lys39 in the majority of the top 10 models (see Fig. S6A in the supplemental material). A DNA probe differing by 1 bp from probe 10 at this position should be tested for binding of wild-type and Ala-substituted forms of SpoIIID. Ser35 and Arg67 are predicted to form hydrogen bonds with ADE21 and CYT4, respectively, in the majority of the top 10 models (see Fig. S6A). S35A and R67A forms of SpoIIID should be purified, and binding to probes 15 (with GUA replacing ADE21) and 18 (with THY replacing CYT4), respectively, should be compared with binding to other DNA probes. The finding that the mutation in probe 18 had a greater effect on binding of Ala-substituted SpoIIIDs than the mutation in probe 17 (Table 2) is consistent with the prediction that Arg67 forms a hydrogen bond with CYT4, while the THY5/ADE24 base pair (mutated to CYT5/GUA24 in probe 17; Fig. 5A) is not predicted to form a hydrogen bond with SpoIIID (see Fig. S6A).

Transcriptional regulation by SpoIIID. SpoIIID up- or downregulates transcription of more than 100 genes in the mother cell during sporulation (3), but only 20 SpoIIID binding sites have been mapped by DNase I footprinting (3, 4, 9, 31). Based on the positions of the binding sites mapped so far, it appears that SpoIIID represses transcription by interfering with binding of σ^E- or σ^K-RNA polymerase or binding of the GerE activator protein. SpoIIID likely activates transcription by contacting σ^E- and σ^K-RNA polymerase (4). Asp51 and the C-terminal basic region of SpoIIID have been proposed as potential contact points with σ^E- and σ^K-RNA polymerase, since D51K and D82K substitutions nearly eliminated expression of an SpoIIID-dependent reporter (15). Asp51 is the first residue of helix α4 and is predicted to be highly exposed on the surface of SpoIIID bound to DNA (Fig. 4C), so it remains a good candidate for contacting σ^E- and σ^K-RNA polymerase. Asp82 is not a candidate for contacting σ^E- and σ^K-RNA polymerase, since truncation of SpoIIID at Lys81 did not prevent reporter expression; thus, it was speculated that the D82K substitution prevents other residues in the C-terminal basic region from contacting σ^E- and σ^K-RNA polymerase (15). In light of our findings that single-Ala substitutions for Lys76, Lys78, Lys80, or Lys81 of SpoIIID impaired DNA binding *in vitro* (Table 2, probes 15 and 16) and our observations that suggest the C-terminal basic region of SpoIIID interacts flexibly with DNA, it seems more likely that the D82K substitution in SpoIIID perturbs the interaction of its C-terminal basic region with DNA.

In terms of a target for development of sporulation inhibitors, the interaction of the novel C-terminal extension of SpoIIID with DNA currently looks most promising. SpoIIID orthologs not only in *Bacilli* but also in *Clostridia* and other distant relatives exhibit high similarity to *B. subtilis* SpoIIID residues 52 to 79 (see Fig. S7 in the supplemental material). In contrast, even if Asp51 of *B. subtilis* SpoIIID does contact σ^E- and σ^K-RNA polymerase, the interaction may not be broadly conserved, since Asp or Glu is not typically found at the corresponding position of SpoIIID orthologs in *Clostridia* and other distant relatives, although SpoIIID orthologs in *Bacilli* have Asp or Glu at the corresponding position (see Fig. S7).

ACKNOWLEDGMENTS

We thank T. Kwaku Dayie (University of Maryland) for help with the structural modeling.

This work was supported by National Institutes of Health grants GM43585 (to L.K.) and GM58221 (to H.Y.) and by Michigan State University AgBioResearch. This study made use of NMR spectrometers funded in part by NSF grant BIR9512253 and the Michigan Economic Development Corporation.

REFERENCES

1. **Kroos L.** 2007. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu. Rev. Genet. **41:**13–39. http://dx.doi.org/10.1146/annurev.genet.41.110306.130400.
2. **Hilbert DW, Piggot PJ.** 2004. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. Microbiol. Mol. Biol. Rev. **68:**234–262. http://dx.doi.org/10.1128/MMBR.68.2.234-262.2004.
3. **Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS Biol. **2:**e328. http://dx.doi.org/10.1371/journal.pbio.0020328.
4. **Halberg R, Kroos L.** 1994. Sporulation regulatory protein SpoIIID from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. J. Mol. Biol. **243:**425–436. http://dx.doi.org/10.1006/jmbi.1994.1670.
5. **Kroos L, Kunkel B, Losick R.** 1989. Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science **243:**526–529. http://dx.doi.org/10.1126/science.2492118.
6. **McKenney PT, Driks A, Eichenberger P.** 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nat. Rev. Microbiol. **11:**33–44. http://dx.doi.org/10.1038/nrmicro2921.
7. **Zhang J, Ichikawa H, Halberg R, Kroos L, Aronson AI.** 1994. Regulation of the transcription of a cluster of *Bacillus subtilis* spore coat genes. J. Mol. Biol. **240:**405–415. http://dx.doi.org/10.1006/jmbi.1994.1456.
8. **Zheng L, Halberg R, Roels S, Ichikawa H, Kroos L, Losick R.** 1992. Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes. J. Mol. Biol. **226:**1037–1050. http://dx.doi.org/10.1016/0022-2836(92)91051-P.
9. **Ichikawa H, Kroos L.** 2000. Combined action of two transcription factors regulates genes encoding spore coat proteins of *Bacillus subtilis*. J. Biol. Chem. **275:**13849–13855. http://dx.doi.org/10.1074/jbc.275.18.13849.
10. **Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM.** 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol. Rev. **29:**231–262. http://dx.doi.org/10.1016/j.femsre.2004.12.008.
11. **Pabo CO, Sauer RT.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095. http://dx.doi.org/10.1146/annurev.bi.61.070192.005201.
12. **Huffman JL, Brennan RG.** 2002. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr. Opin. Struct. Biol. **12:**98–106. http://dx.doi.org/10.1016/S0959-440X(02)00295-6.
13. **Ducros VM, Lewis RJ, Verma CS, Dodson EJ, Leonard G, Turkenburg JP, Murshudov GN, Wilkinson AJ, Brannigan JA.** 2001. Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in *Bacillus subtilis*. J. Mol. Biol. **306:**759–771. http://dx.doi.org/10.1006/jmbi.2001.4443.
14. **Kunkel B, Kroos L, Poth H, Youngman P, Losick R.** 1989. Temporal and spatial control of the mother-cell regulatory gene *spoIIID* of *Bacillus subtilis*. Genes Dev. **3:**1735–1744. http://dx.doi.org/10.1101/gad.3.11.1735.
15. **Himes P, McBryant S, Kroos L.** 2010. Two regions of *Bacillus subtilis* transcription factor SpoIIID allow a monomer to bind DNA. J. Bacteriol. **192:**1596–1606. http://dx.doi.org/10.1128/JB.01506-09.
16. **Tullius T.** 1995. Homeodomains: together again for the first time. Structure **3:**1143–1145. http://dx.doi.org/10.1016/S0969-2126(01)00250-7.
17. **Brennan RG.** 1993. The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell **74:**773–776. http://dx.doi.org/10.1016/0092-8674(93)90456-Z.
18. **Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A.** 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6:**277–293.

19. **Johnson BA, Blevins RA.** 1994. NMRView: a computer program for the visualization and analysis of NMR data. J. Biomol. NMR **4:**603–614. http://dx.doi.org/10.1007/BF00404272.
20. **Cavanagh J, Fairbrother W, Palmer AG, Rance M, Skleton NJ.** 2006. Protein NMR spectroscopy–principles and practice, 2nd ed. Elsevier Academic Press, Burlington, MA.
21. **Lee W, Revington MJ, Arrowsmith C, Kay LE.** 1994. A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. FEBS Lett. **350:**87–90. http://dx.doi.org/10.1016/0014-5793(94)00740-3.
22. **Ogura K, Terasawa H, Inagaki F.** 1996. An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wideband inversion shaped pulse. J. Biomol. NMR **8:**492–498.
23. **Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE.** 1997. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. J. Am. Chem. Soc. **119:**6711–6721. http://dx.doi.org/10.1021/ja970224q.
24. **Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE.** 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry **33:**5984–6003. http://dx.doi.org/10.1021/bi00185a040.
25. **Guntert P.** 2004. Automated NMR structure calculation with CYANA. Methods Mol. Biol. **278:**353–378. http://dx.doi.org/10.1385/1-59259-809-9:353.
26. **Cornilescu G, Delaglio F, Bax A.** 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR **13:**289–302. http://dx.doi.org/10.1023/A:1008392405740.
27. **Wüthrich K.** 1986. NMR of proteins and nucleic acid. Wiley, New York, NY.
28. **Dominguez C, Boelens R, Bonvin AM.** 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. **125:**1731–1737. http://dx.doi.org/10.1021/ja026939x.
29. **van Dijk M, Bonvin AM.** 2010. Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. Nucleic Acids Res. **38:**5634–5647. http://dx.doi.org/10.1093/nar/gkq222.
30. **van Dijk M, van Dijk AD, Hsu V, Boelens R, Bonvin AM.** 2006. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Res. **34:**3317–3325. http://dx.doi.org/10.1093/nar/gkl412.
31. **Zhang B, Daniel R, Errington J, Kroos L.** 1997. *Bacillus subtilis* SpoIIID protein binds to two sites in the *spoVD* promoter and represses transcription by σ^E RNA polymerase. J. Bacteriol. **179:**972–975.
32. **van Dijk M, Bonvin AM.** 2009. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. **37:**W235–W239. http://dx.doi.org/10.1093/nar/gkp287.
33. **Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM.** 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR **8:**477–486.
34. **Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA.** 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol. Cell **9:**527–539. http://dx.doi.org/10.1016/S1097-2765(02)00470-7.
35. **Holm L, Kaariainen S, Rosenstrom P, Schenkel A.** 2008. Searching protein structure databases with DaliLite v. 3. Bioinformatics **24:**2780–2781. http://dx.doi.org/10.1093/bioinformatics/btn507.
36. **Tanaka T, Shinkai A, Bessho Y, Kumarevel T, Yokoyama S.** 2009. Crystal structure of the manganese transport regulatory protein from *Escherichia coli*. Proteins **77:**741–746. http://dx.doi.org/10.1002/prot.22541.
37. **Ye YZ, Godzik A.** 2004. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. **32:**W582–W585. http://dx.doi.org/10.1093/nar/gkh430.
38. **Sorenson MK, Ray SS, Darst SA.** 2004. Crystal structure of the flagellar σ/anti-σ complex σ^{28}/FlgM reveals an intact σ factor in an inactive conformation. Mol. Cell **14:**127–138. http://dx.doi.org/10.1016/S1097-2765(04)00150-9.
39. **Kobayashi M, Ab E, Bonvin AM, Siegal G.** 2010. Structure of the DNA-bound BRCA1 C-terminal region from human replication factor C p140 and model of the protein-DNA complex. J. Biol. Chem. **285:**10087–10097. http://dx.doi.org/10.1074/jbc.M109.054106.
40. **Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I.** 2013. Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. PLoS Genet. **9:**e1003756. http://dx.doi.org/10.1371/journal.pgen.1003756.
41. **Wilcox MH.** 2003. Gastrointestinal disorders and the critically ill. Clostridium difficile infection and pseudomembranous colitis. Best Pract. Res. Clin. Gastroenterol. **17:**475–493. http://dx.doi.org/10.1016/S1521-6918(03)00017-9.
42. **Setlow P.** 2006. Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. J. Appl. Microbiol. **101:**514–525. http://dx.doi.org/10.1111/j.1365-2672.2005.02736.x.
43. **Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, Wren BW, Fairweather NF, Dougan G, Lawley TD.** 2012. The *Clostridium difficile spo0A* gene is a persistence and transmission factor. Infect. Immun. **80:**2704–2711. http://dx.doi.org/10.1128/IAI.00147-12.
44. **Bartlett JG.** 2007. *Clostridium difficile*: old and new observations. J. Clin. Gastroenterol. **41:**S24–S29. http://dx.doi.org/10.1097/MCG.0b013e31803d16ec.
45. **Hoa TT, Duc LH, Isticato R, Baccigalupi L, Ricca E, Van PH, Cutting SM.** 2001. Fate and dissemination of *Bacillus subtilis* spores in a murine model. Appl. Environ. Microbiol. **67:**3819–3823. http://dx.doi.org/10.1128/AEM.67.9.3819-3823.2001.
46. **Tam NK, Uyen NQ, Hong HA, Duc le H, Hoa TT, Serra CR, Henriques AO, Cutting SM.** 2006. The intestinal life cycle of *Bacillus subtilis* and close relatives. J. Bacteriol. **188:**2692–2700. http://dx.doi.org/10.1128/JB.188.7.2692-2700.2006.
47. **Ochsner UA, Bell SJ, O'Leary AL, Hoang T, Stone KC, Young CL, Critchley IA, Janjic N.** 2009. Inhibitory effect of REP3123 on toxin and spore formation in *Clostridium difficile*, and in vivo efficacy in a hamster gastrointestinal infection model. J. Antimicrob. Chemother. **63:**964–971. http://dx.doi.org/10.1093/jac/dkp042.
48. **Mathur T, Kumar M, Barman TK, Kumar GR, Kalia V, Singhal S, Raj VS, Upadhyay DJ, Das B, Bhatnagar PK.** 2011. Activity of RBx 11760, a novel biaryl oxazolidinone, against *Clostridium difficile*. J. Antimicrob. Chemother. **66:**1087–1095. http://dx.doi.org/10.1093/jac/dkr033.
49. **Schumacher MA, Choi KY, Zalkin H, Brennan RG.** 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science **266:**763–770. http://dx.doi.org/10.1126/science.7973627.
50. **Frazee RW, Taylor JA, Tullius TD.** 2002. Interchange of DNA-binding modes in the deformed and ultrabithorax homeodomains: a structural role for the N-terminal arm. J. Mol. Biol. **323:**665–683. http://dx.doi.org/10.1016/S0022-2836(02)00996-8.