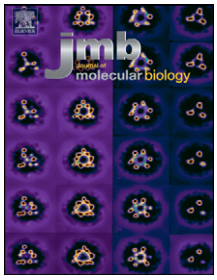

A DNA Mimic: The Structure and Mechanism of Action for the Anti-Repressor Protein AbbA

Ashley T. Tucker [1], Benjamin G. Bobay [1], Allison V. Banse [2], Andrew L. Olson [1], Erik J. Soderblom [3], M. Arthur Moseley [3], Richele J. Thompson [1], Kristen M. Varney [4], Richard Losick [2] and John Cavanagh [1]

1 - Department of Molecular and Structural Biochemistry, *North Carolina State University, Campus Box 7622, 128 Polk Hall, Raleigh, NC 27695, USA*

2 - Department of Molecular and Cellular Biology, *Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA*

3 - Duke Proteomics Core Facility, *Institute for Genome Sciences and Policy, Duke University School of Medicine, Duke University, B02 Levine Sciences Research Center, 450 Research Drive, Durham, NC 27708, USA*

4 - Department of Biochemistry and Molecular Biology, *University of Maryland School of Medicine, 108 North Greene Street, Baltimore, MD 21201, USA*

Correspondence to John Cavanagh: *john_cavanagh@ncsu.edu.*

http://dx.doi.org/10.1016/j.jmb.2014.02.010

Edited by F. Allain

Abstract

Bacteria respond to adverse environmental conditions by switching on the expression of large numbers of genes that enable them to adapt to unfavorable circumstances. In *Bacillus subtilis*, many adaptive genes are under the negative control of the global transition state regulator, the repressor protein AbrB. Stressful conditions lead to the de-repression of genes under AbrB control. Contributing to this de-repression is AbbA, an anti-repressor that binds to and blocks AbrB from binding to DNA. Here, we have determined the NMR structure of the functional AbbA dimer, confirmed that it binds to the N-terminal DNA-binding domain of AbrB, and have provided an initial description for the interaction using computational docking procedures. Interestingly, we show that AbbA has structural and surface characteristics that closely mimic the DNA phosphate backbone, enabling it to readily carry out its physiological function.

Introduction

Bacteria cope with environmental fluctuations and stress by a wide range of adaptive responses [1]. In the bacterium *Bacillus subtilis*, these adaptations include the production of antibiotics and degradative enzymes, killing of siblings by a process known as cannibalism, spore formation, and the formation of complex communities known as biofilms [2,3]. Genes involved in each of these adaptations are subject to regulation by a small number of master regulatory proteins, some of which are dedicated to a particular adaptive response whereas others are global regulators that influence the expression of large numbers of genes involved in diverse adaptations. One such global regulator is the transition state regulator AbrB [4]. This repressor protein helps hold genes involved in responding to stress silent under favorable environmental conditions, when cells are in the exponential phase of growth. However, under stress-inducing conditions, AbrB-mediated repression is relieved, leading to the genome-wide de-repression of large numbers of genes [5].

What is the mechanism of this de-repression? Two pathways contribute to blocking the action of AbrB, both of which involve an upstream regulator known as Spo0A, which is activated by phosphorylation in response to diverse environmental signals [6]. The phosphorylated form of Spo0A, Spo0A~P, is both a repressor and an activator [7]. As a repressor, Spo0A~P binds to and inhibits the transcription of the gene for AbrB, leading to the depletion of AbrB protein by proteolysis [6]. As an activator, Spo0A~P

switches on the gene for an anti-repressor known as AbbA that directly binds to AbrB and blocks its ability to bind to DNA [8].

Just how the binding of AbbA to AbrB inactivates the repressor has been unclear.

In this work, we have determined the high-resolution NMR structure of the functional AbbA dimer, analyzed its interactions with AbrB, and developed the first detailed mechanistic model for how AbbA is able to perform its physiological role based on these data. Unexpectedly, AbbA has surface characteristics that imitate the DNA phosphate backbone and directly interacts with AbrB residues involved in contacting DNA. We conclude that AbbA executes its functional responsibilities by virtue of it being a DNA mimetic.

Results

AbbA amino acid content

The wild-type sequence of AbbA from *B. subtilis* exhibits several interesting features. The sequence is 65 residues in length and, of these residues, 20 are defined as being electrostatic (D, E, K, and R). Of these electrostatic residues, over half (12 out of 20) are negative. This suggests that AbbA likely has significant electrostatic surface character and may use this property for interacting with binding partners. AbbA's sequence was run through a BLAST search on the PDB and no significant similar protein sequences were identified.

AbbA is a dimer

Size-exclusion liquid chromatography (SELC) and native mass spectrometry were used to determine the oligomerization state of AbbA. The expected molecular mass of the AbbA monomer is 7979 Da. The SELC elution profile shows AbbA to be dimeric (Supplementary Fig. 1). The expected molecular mass of AbbA as a dimer is 15,958 Da. The calculated experimental molecular mass from SELC is 17,847 Da. Such differences in expected and experimental molecular masses determined by SELC are well known. SELC separates molecules based on size and shape. If a protein's hydrodynamic radius is not perfectly spherical, the experimentally calculated molecular mass often varies slightly from the expected value. The AbbA dimer was the only oligomeric form of the protein observed in all SELC experiments at all concentrations studied (1 mM to 12.5 μM). These data suggest that AbbA exists solely as a dimer.

In order to confirm its dimeric state, AbbA was also examined using a method independent of shape, native mass spectrometry. Analysis shows that AbbA ion species included both monomers and dimers at a cone voltage of 20 V (Supplementary Fig. 2). Monomeric AbbA shows an *m*/*z* ion series of +10 to +3 (*m*/*z* values, 800.76–2669.80), which, upon transformation, gives a relative molecular mass of 7979.2 Da (expected, 7979 Da). Dimeric AbbA shows an *m*/*z* ion series of +11 to +5 (*m*/*z* values, 1452.9–3203.9), which, upon transformation, gives a relative molecular mass of 15,988.1 Da (expected, 15,958 Da). The differences in expected and relative molecular masses are the result of peak broadening in the spectra caused by using reduced desolvation energy (20 V). This is performed in order to preserve the dimeric species. The monomeric state seen in Supplementary Fig. 2 is an artifact of the dissociation of non-covalent complexes in the source region. Overall, these data suggest that AbbA exists solely as a dimer.

AbbA binds to the N-terminal DNA-binding domain of AbrB

Previous studies have shown that AbbA binds to full-length AbrB [8]. Here, SELC was used to confirm whether AbbA bound to the N-terminal or C-terminal domain of AbrB. The N-terminal DNA-binding domain of AbrB (AbrBN) consists of residues 1–53. The C-terminal multimerization domain of AbrB (AbrBC) consists of residues 54–94. Analysis of size-exclusion chromatography (Supplementary Fig. 3a) shows that AbbA and AbrBN eluted with individual apparent molecular masses of 18.8 kDa and 10.9 kDa, respectively, which is consistent with the formation of homodimers in solution. The AbbA:AbrBN complex eluted at an apparent molecular mass of 23.7 kDa. The expected molecular mass of the complex is 28.1 kDa; however, the differences in experimental and expected molecular masses can be ascribed to the shape and/or hydrodynamic radius of the complex. Supplementary Fig. 3b shows that dimeric AbrBC elutes at an apparent molecular mass of 14.7 kDa. The expected molecular mass of an AbbA:AbrBC complex is 25.8 kDa, which was not seen in the chromatograph of combined AbbA and AbrBC. The combined sample eluted at an apparent molecular mass of 17.4 kDa. Consequently, the SELC studies suggest that AbbA targets the N-terminal DNA-binding domain of AbrB exclusively.

AbrBN residues R8, R15, and R23 are critical for binding AbbA

Previous mutagenic and structural analyses identified four conserved arginine residues (R8, R15, R23, and R24) as critical for AbrB's ability to bind DNA [9,10]. To investigate if these residues were also important for AbbA binding, we purified wild-type His_6-AbrB, His_6-AbrB^{R8A}, His_6-AbrB^{R15A}, His_6-AbrB^{R23A}, and His_6-AbrB^{R24A}. Each of these proteins were incubated with purified AbbA, applied to a Ni-NTA agarose affinity column, and eluted with

imidazole. AbbA was retained on the column when either wild-type His_6-AbrB or His_6-$AbrB^{R24A}$ was used, but AbbA was not retained on the column with $AbrB^{R8A}$, $AbrB^{R15A}$, or $AbrB^{R23A}$ (Fig. 1a). Thus, it appears that three of the four conserved arginine residues in AbrB critical for DNA binding (R8, R15, and R23) are required for binding to AbbA as well [4].

AbbA binds to AbrB with a K_d that is competitive with DNA

In order to determine if the AbbA:AbrB interaction is competitive with AbrB:DNA interactions, we used isothermal titration calorimetry (ITC) to obtain the dissociation constant (K_d) for the binding between AbrB and AbbA. K_d values for a variety of AbrB:DNA complexes have been measured previously [11,12]. The K_d values for AbrB binding to six different target promoters have been measured in the range 6–43 nM (average K_d of 25 nM; average length of 361 bp). Figure 1b shows representative ITC data for the full-length AbbA:AbrB binding event. The K_d for the interaction was found to be in the range 11–21 nM (average K_d of 16 nM) with N values ranging from 0.868 to 0.969 (average N value of 0.9185). The stoichiometry of the interaction is addressed in greater detail below with analytical ultracentrifugation (AUC) experiments. This K_d range suggests that the AbbA:AbrB interaction is extremely competitive with AbrB:DNA interactions.

NMR structural characterization of AbbA

Chemical shift assignments for AbbA were obtained for 92% of the backbone resonance (HN, N, C^{α}, and C′) and 94% of the total atoms (including side chains); missing assignments were at the N-terminus (residues 1–3) and loop 1 between helices 1 and 2 (residues 26–32). The ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum of AbbA shows good spectral dispersion and all NH resonances were assigned with the exception of 1–3, 26, 29, and 30 (Supplementary Fig. 4). Statistics for the 10 lowest energy structures of AbbA are shown in Table 1. These structures were determined with a total of 1846 nuclear Overhauser enhancements (NOEs) of which 336 were intraresidue, 578 were sequential, 426 were medium range, and 406 were long range. We determined 98 intermolecular NOEs using ^{13}C-NOE spectroscopy (NOESY)-HSQC isotope-filtered experiment on a sample composed of a mixture of $^{13}C/^{15}N$ labeled AbbA and $^{12}C/^{14}N$ AbbA. The final 10 structures had zero NOE, hydrogen bond, and dihedral violations per structure. The structural ensemble showed an average C^{α} root-mean-square deviation (r.m.s.d.) of 0.428 ± 0.075 Å for secondary structure (backbone atoms) and of 0.861 ± 0.106 Å for secondary structure (heavy atoms; i.e., all non hydrogen atoms). PSVS (*P*rotein *S*tructure *V*alidation *S*erver), MolProbity, WHATCHECK, and PROCHECK were used to analyze the ensemble that was subsequently determined to be of good quality (Table 1) [13–16]. Ramachandran analysis confirmed that 99.5% of residues lie within the generously allowed or better conformational space. An overlay of the 10 lowest energy structures is shown in Fig. 2. The monomeric subunit is composed of three α-helices between residues 17–25 (α1), 31–45 (α2), and 51–67 (α3) with two loop regions between residues 26–30 (L1) and 46–50 (L2) and relatively unstructured first 15 residues on the N-terminus.

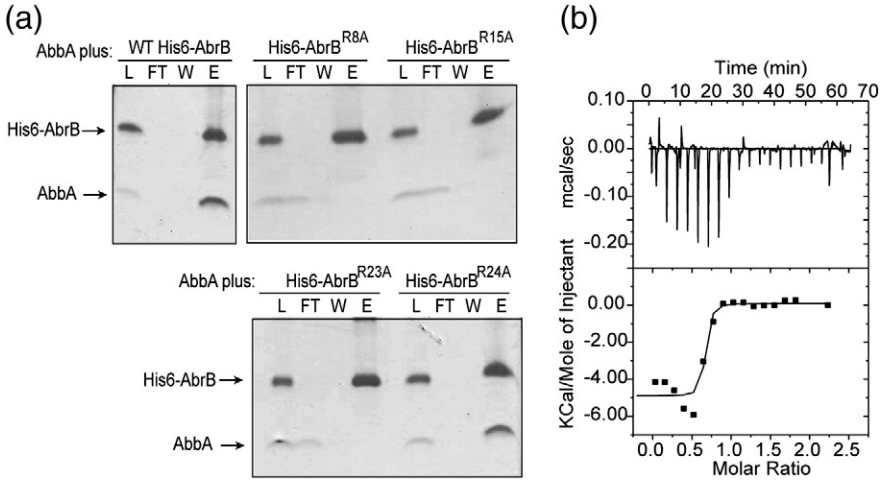

Fig. 1. Binding specificity and affinity of AbbA and AbrB. (a) Purified AbbA was incubated with purified wild-type His_6-AbrB, His_6-$AbrB^{R8A}$, His_6-$AbrB^{R15A}$, His_6-$AbrB^{R23A}$, or His_6- $AbrB^{R24A}$ and applied to a Ni-NTA agarose column. The presence of AbbA, wild-type His_6-AbrB, His_6-$AbrB^{R8A}$, His_6-$AbrB^{R15A}$, His_6-$AbrB^{R23A}$, and His_6-$AbrB^{R24A}$ was monitored in the load (L), flow-through (FT), wash (W), and eluate (E) by Coomassie staining. Elution fractions were concentrated 5-fold relative to the load. (b) ITC binding isotherms of the AbbA:AbrB interaction.

Table 1. NMR and refinement statistics for AbbA.

NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1846
Intraresidue	336
Interresidue	1510
Sequential ($\|i-j\|=1$)	578
Medium range ($\|i-j\|<4$)	426
Long range ($\|i-j\|>5$)	406
Intermolecular	98
Hydrogen bonds	96
Dihedral angles (φ,ψ)	142
$^1D_{NH}$ RDCs	74
Correlation coefficient (%)	0.937 ± 0.002
$Q_{normalized}$ (%)	10.992 ± 0.217
Structure statistics	
Violations (mean and S.D.)	
Distance constraints (Å)	0.08 ± 0.03
Dihedral angle constraints (º)[a]	N/A
Maximum dihedral angle violation (º)[a]	N/A
Maximum distance constraint violation (Å)	0.179
Deviations from idealized geometry	
Bond lengths (Å)	0.014 ± 0.0008
Bond angles (º)	2.30 ± 0.009
Improper *Z*-score[b]	1.001
Average pairwise r.m.s.d.[c] (Å)	
Heavy	0.861 ± 0.106
Backbone	0.428 ± 0.075

[a] No dihedral angle violations.
[b] As calculated by WHATCHECK [16].
[c] Pairwise r.m.s.d. was calculated among 10 refined structures (over residues 15–65 from each monomer).

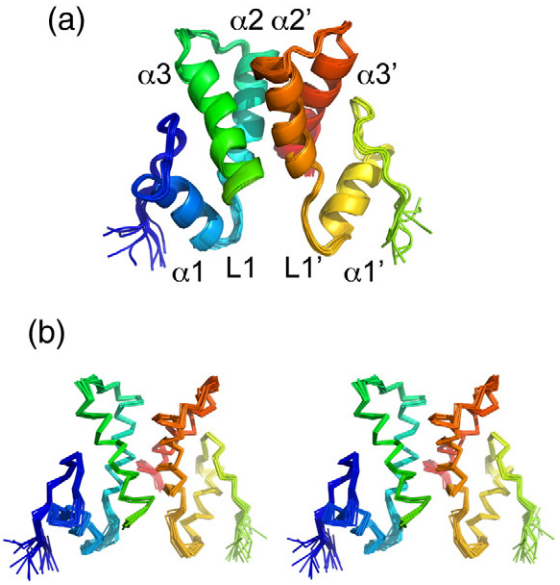

Fig. 2. The NMR solution structure of AbbA. (a) The lowest 10 energy structures. One monomer is colored blue to green whereas the other monomer is colored yellow to red. (b) A stereogram of the lowest 10 energy structures of unbound AbbA. Pertinent helices and loops are identified.

From this point on, residues in the second monomeric subunit of the AbbA homodimer will be identified by a prime superscript (′).

The dimer quaternary structure has extensive interactions between the individual AbbA monomers. The second and third helices on each monomeric subunit comprise the majority of the intermolecular interface. Dimer interface interactions are defined by a range of NOEs among E33-F64′; L34-F64′; S36-Y61′ and F64′; S37-Y61′, F64′, and R65′; I39-I39′, L57′, V58′, and Y61′; N40-W61′ and R65′; T43-Y54′; and T44-V58′ and Y61′ (Supplementary Table 1). Upon dimerization, a surface area of 770.3 Å^2 from each monomeric subunit becomes buried through notable hydrophobic interactions between helices α2, α3, α2′, and α3′. The helices lie at an angle ~50° and ~70° to each other across the interface (comparing α2 to α2′ and α3 to α3′, respectively). A smaller amount of hydrophobic interactions are also found between α1 and α2 (α1′ and α2′) (Supplementary Fig. 5a). The AbbA sequence has 12 negatively charged residues (D and E) per monomer that are primarily distributed in the first 30 residues and in the last 10 residues, indicating a functionally charged domain (Supplementary Fig. 5b).

To both query the uniqueness of the dimeric AbbA structure and assist in the identification/characterization of the dimerization interface, we ran the monomeric structure of AbbA through the DALI server. Initial structure calculations for AbbA focused on a single monomeric unit and composed of dihedral constraints, J-coupling constants, and carbon (120 ms mixing time) and nitrogen NOESY experiments (120 and 150 ms mixing times). This structure was then run through the DALI Web server program to search the PDB for "like" structures [17]. Through DALI, we obtained one significant structural similarity hit with similar amino acid length, 2JS1 (solution NMR structure of the homodimer protein YVFG from *B. subtilis*). 2JS1 is a dimeric protein of 80 residues in length, described as an "uncharacterized protein" in UniProt, and has a 52% similar or better residue alignment with AbbA (14% identity, 28% strongly similar, 10% weakly similar, and 48% dissimilar). These results were used along with the ^{13}C-NOESY-HSQC isotope-filtered experiment NOEs to drive the characterization of the unknown structure and dimerization interface of AbbA in solution.

Chemical crosslinking and mass spectrometry of the AbbA:AbrBN complex

Chemical crosslinking studies between pairs of primary amines (lysines) and carboxyl (glutamates and aspartates) to primary amines were used to help characterize the interface between AbbA and AbrBN. Chemical crosslinkers of varying lengths have been used previously to map three-dimensional (3D)

structures of proteins and protein complexes [18–21]. Specifically, three chemical crosslinkers were employed: (i) bis(sulfosuccinimidyl)suberate (BS3), with a spacer length of 11.4 Å; (ii) disuccinimidyl glutarate (DSG), with a spacer length of 7.7 Å; and (iii) 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), which is a zero-length spacer. Supplementary Fig. 6 shows the crosslinked complexes analyzed via SDS-PAGE. Multiple bands corresponding to crosslinked complexes were present. Bands with molecular masses above 28 kDa were considered aggregation artifacts since little/no crosslinked species were detected when analyzed (data not shown). The bands below 14 kDa are non-crosslinked monomers and uncomplexed crosslinked dimers. The target bands corresponding to an AbbA homodimer bound to an AbrBN homodimer in the 20- to 28-kDa range were analyzed via mass spectrometry.

Both AbbA and AbrBN were present in all the bands analyzed. A list of the through-space crosslinks between AbbA and AbrBN are listed in Supplementary Table 2. In the BS3 band, one intrapeptide crosslink and no through-space crosslinks were detected. In the DSG band, one through-space crosslink between AbbA and AbrBN was detected. Ordinarily, amine-to-amine crosslinking is the most commonly used crosslinking chemistry; however, this method proved inadequate in this study due to the lack of appropriately positioned lysine residues in each protein. While AbrBN contains eight lysines and/or free amines dispersed throughout the protein (K2, K9, K31, K42, K46, K47, K49, and M1), AbbA contains only five lysines and/or free amines (K19, K47, K55, K56, and G1) with each presumably distal to the interaction site. On the other hand, the EDC crosslinker (carboxyl-to-amine) provided a wealth of information due to AbbA's favorable acidic sequence characteristics (12 acidic residues per monomer clustered around the presumed interaction site) and AbrBN's dispersion of lysine residues (Supplementary Fig. 5b). In the EDC band at ~28 kDa, there were eight through-space crosslinks between AbbA and AbrBN detected. In the ~20-kDa band, there were seven through-space crosslinks between AbbA and AbrBN detected. The two EDC bands had five crosslinks in common, which is an indication that these bands contained very similar species with respect to conformation/structure (Supplementary Table 2). The majority of the crosslinks suggest close contacts between residue E11, E15, and G1 and the C-terminal end of α3 (D62 and E67) of AbbA and the DNA-binding interface composed of β1, L1, and L2′ of AbrBN (K9, E12, and E30, respectively) [4]. There were two crosslinks to K55 of AbbA from D11 and E30 of AbrBN that are not within the proposed binding region. These crosslinks are likely artifacts due to the high protein concentrations used in the experiments.

Sedimentation equilibrium of the AbbA:AbrBN complex

Sedimentation equilibrium AUC was used to confirm the stoichiometry of the AbbA and AbrBN complex (Supplementary Fig. 7). The resulting mass-averaged molecular mass values at equilibrium were 18,024 ± 355 Da for AbbA, 11,644 ± 305 Da for AbrBN, and 27,607 ± 280 Da for the AbbA:AbrBN complex. The individual molecular masses for AbbA and AbrBN indicate values close to their expected homodimeric molecular masses of 15,958 Da and 12,200 Da, respectively. Based on the molecular mass results from the AbbA:AbrBN complex, there is a 2:2 stoichiometric ratio (i.e., one AbbA homodimer binds to one AbrBN homodimer).

Modeling the interaction between AbbA and AbrBN

The HADDOCK molecular docking approach was employed to develop a structural model of the interaction between AbbA and AbrBN (PDB codes: 2LZF and 1Z0R) [22]. Restraints for the docking came from the mutational and crosslinking mass spectrometry data discussed above and surface accessibility information [23]. The molecular docking between AbbA and AbrBN resulted in five clusters of docked structures based on a maximum C^{α} r.m.s.d. of 5.5 Å and a minimum of four structures per cluster. By far, cluster 1 resulted in the most structures (157 out of 200 or 78.5% of the docked solutions; Supplementary Figs. 8 and 9 and Supplementary Tables 3 and 4). The C^{α} r.m.s.d. between unbound AbbA and the lowest energy structure from cluster 1 of bound AbbA is approximately 5.5 Å. The C^{α} r.m.s.d. between unbound AbrBN and the lowest energy structure from cluster 1 of bound AbrBN is approximately 6.3 Å. Notably, this 6.3 Å r.m.s.d. is greater than that observed when unbound AbrBN binds to DNA (2.8 Å). This suggests that AbrBN undergoes a greater conformational change when it binds to AbbA than when it binds to DNA [4].

The interface between AbbA and AbrBN reveals an extended network of intermolecular hydrogen bonds that can stabilize the complex. The overall model for the AbbA:AbrBN complex, including putative intermolecular hydrogen bonds, is shown in Fig. 3a. The highly negative charged region of AbbA (Supplementary Fig. 5b) is seen to interact with the DNA binding face of AbrBN. The interface is composed of interactions between residues 1, 12, 20, 23–33, and 66–68 (AbbA) and residues 9–15, 20–21, 22–25, and 30–31 (AbrBN). The interaction interface is composed of 18 (36 total for the complex) unique potential hydrogen bonds and 16 (32 total for the complex) unique salt bridges. This leads to an interface of ~1200 $Å^2$ buried surface area.

Upon complex formation, there are minimal conformational perturbations in β3, β4, β3′, and β4′ of AbrBN and α2, α3, α2′, and α3′ of AbbA. However, AbrBN's DNA binding region (residues 6–32 and 29′–33′ and residues 6′–32′ and 29–33) appears to undergo significant changes, similar to those we previously determined in the AbrBN:*abrb8* modeled complex [4]. Notable conformational changes are also observed in AbbA within its first 30 residues (Fig. 3b, blue/broad regions indicate larger conformational changes upon complex formation compared to red/narrow regions). tCONCOORD [24] conformational space analysis (see Supplementary Methods) reveals that unbound AbbA has the propensity to sample a larger conformational space within its first 30 residues compared to the rest of the protein (Fig. 3c).

Principal components analysis (PCA) was also performed on the NMR structure ensembles of AbbA (see Supplementary Methods). PCA is a standard tool in the field of multivariate analysis for extracting from a set of interrelated variables a much smaller set that retains most of the variation contained in the full set. By setting up a correlation matrix whose elements are the ensemble average of the pairwise products of displacements from their average position of landmarks (such as the C^{α} positions in a protein), PCA can be very helpful in identifying, from an NMR structural ensemble, correlations in conformational rearrangements within a protein [25]. A structural PCA plot for the first principal component of the NMR structure of AbbA (Fig. 3d) reveals the following: on average, within the NMR structure ensemble, (i) residues 1–30 (α1 and L1) move in an anti-correlated (opposite) fashion with residues 1′–30′ (α1′ and L1′); (ii) residues 66–68 move in an anti-correlated fashion with residues 66′–68′; and (iii) the dimerization interface (α2, α3, α2′, and α3′) move in an anti-correlated fashion also. Most important here is the tendency for a general concerted motion involving α1/L1 and α1′/L1′. This concerted, anti-correlated event can be viewed as α1/L1 moving *away* from α1′/L1′ to accommodate AbrBN binding. These concerted motions allow the domains to appropriately reorient during complex formation with AbrB, enabling critical negatively

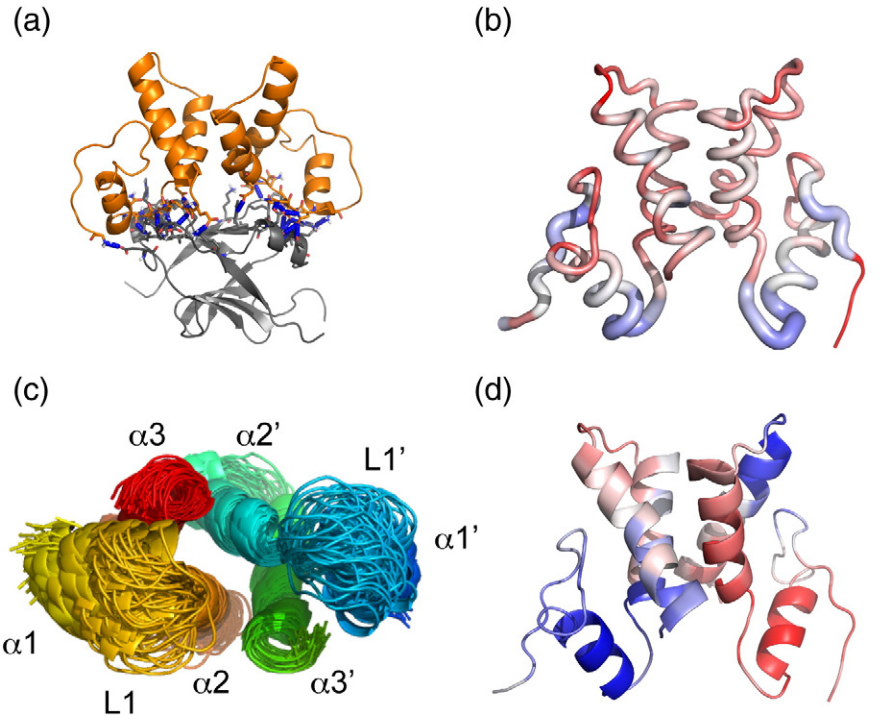

Fig. 3. AbbA's interactions and motions upon binding AbrB. For (a), AbbA is colored orange, AbrBN is colored gray, and hydrogen bonds are noted by blue broken lines. (a) Lowest energy model from cluster 1 from the docking protocol between AbbA and AbrBN by HADDOCK. Hydrogen bonds between AbbA and AbrBN shown as blue broken lines with those residues involved in the interaction shown as sticks. (b) A representation of the r.m.s.d. between the unbound and bound AbbA. The r.m.s.d. is noted in two separate manners, the r.m.s.d. goes from low to high based on color (red to white to blue) and thickness of cartoon. (c) The conformational space sampling unbound AbbA can undergo. The structure is rotated 90° about the *x*-axis from (d) to highlight the inherent flexibility of the first helix and the loop between helix 1 and helix 2 (yellow to orange; blue to cyan). (d) PCA plots of the maximum likelihood correlation matrix for the unbound structure of ensembles of Abba, first principle component. Regions colored similarly (red or blue) are regions with potential self-correlated movement, whereas regions colored differently (red *versus* blue) represent potential anti-correlated movement.

charged residues on AbbA (the majority of which are found within the first 30 residues and C-terminus) to maximize their interaction with residues important to AbrB's DNA recognition and binding (R8, R15, and R23). We also note that the AbbA dimerization interface (composed of α2, α3, α2′, and α3′; Fig. 3b) shows a relatively low propensity for motion. This allows AbbA to readily retain its homodimeric state upon binding AbrB and provides a fulcrum point for the larger conformational shifts within its first 30 residues.

These structural/modeling results can be summarized as follows: the first 30 residues of AbbA, which have the greatest propensity for motion (Fig. 3c), show the largest conformational change between the unbound and the bound states (Fig. 3b), with the α1/L1 and α1′/L1′ regions moving away from one another (Fig. 3d) to accommodate formation of the AbbA:AbrBN complex (Fig. 3a).

Discussion

Previously, we solved the structure of the DNA-binding N-terminal domain of AbrB and provided a model for its DNA-binding proclivities [4]. However, there was no detailed information on the interaction between AbrB and AbbA or why AbbA would effectively compete with DNA for binding.

Generally, the most straightforward manner to prevent a protein binding to DNA is to make the protein's DNA-binding site inaccessible to the DNA. This entails the direct binding of an “inhibitor” protein at the DNA-binding site on the original protein. Such a mechanism necessitates a protein–protein binding affinity that can readily contest DNA binding. Since imitation is the sincerest form of flattery, it is no surprise that some proteins have evolved structural and surface traits that closely mimic important DNA characteristics in order to assure competitive binding. In this case, rather than two proteins competing for the same section of DNA in order to control transcriptional activities, there is a competitive interaction between the inhibitor protein and the DNA binding sequence of the DNA-binding protein.

Though a few protein DNA mimics are known [26–33], they are still relatively uncommon, with perhaps less than 20 having been discovered at this time. While those that have been discovered appear to exhibit quite diverse cellular functions, they generally share a common structural feature: the use of appropriately spaced glutamic acid and aspartic acid amino acids to imitate DNA backbone phosphates. Many, but not all, also have highly hydrophobic cores to stabilize the protein fold, which has to cope with both a preponderance of charged residues that mimic the phosphate backbone and the small internal volume imposed by the restrictions placed by the size of DNA. For example, the dimeric Ocr (*o*vercome *c*lassical *r*estriction) protein is expressed by bacteriophage T7 upon infection of *Escherichia coli*. Ocr rapidly reduces Type I DNA restriction systems within the host cell and enables successful bacterial infection of the bacterium [26]. Each Ocr monomer forms a bundle of three long α-helices with an additional shorter α-helix. The shorter helix is responsible for dimer formation. Superposition of two 12-bp B-DNA molecules on the Ocr dimer shows that multiple aspartic acid and glutamic acid residues in the three long α-helices lie in positions that approximate the helical turns of the DNA, providing a comparable electrostatic appearance. The two monomeric units that comprise the Ocr dimer are at ~34° to one another, approximating the B-DNA bend. The very recently characterized protein DMP12 (also known as NMB2123) from a variety of *Neisseria* species also presents a surface replete with aspartic acid and glutamic acid residues [27]. DMP12 binds to histone-like proteins with high affinity. DMP12 is proposed to control the stability of the nucleoid in *Neisseria*. DMP12 contains five α-helices and four anti-parallel β-strands assembled in a novel fold. As in the case of Ocr, the spacing and conformational arrangement of the aspartic acid and glutamic acid residues in DMP12 closely emulate the helicity and charge distribution found in DNA. A notably different structural template for a DNA mimic protein can be found in ArdA [33]. Interestingly, from a functional standpoint, ArdA is “Ocr-like”, also operating as an anti-restriction/modification agent. Restriction modification systems are generally used by bacteria to protect themselves from foreign DNA. ArdA has a particularly unique structure, consisting of three domains each with a different fold. The N-terminal domain of ArdA consists of a three-stranded β-sheet and one short α-helix. The central domain is a four-helix bundle. The C-terminal domain comprises a three-stranded β-sheet and three α-helices packed together in a fashion that creates a groove in the structure. The resulting full ArdA structure is quite long and does not exhibit a highly organized central hydrophobic core that anchors the three domains. Despite this very distinctive assembly, ArdA once again provides a negative electrostatic surface consisting of an array of aspartic acid and glutamic acid residues that imitate the phosphate backbone of a polynucleotide. The AbbA protein from *B. subtilis* represents a new member of the DNA mimic genre.

We determined that the functional AbbA dimer binds to AbrB with an affinity that is comparable to the affinity that AbrB binds to its DNA targets and that AbbA binds solely to the N-term DNA-binding domain of AbrB. Residues R8, R15, and R23 in AbrB, all involved in DNA binding, are also involved in AbrB binding to AbbA. These observations suggest that AbbA's mechanism of action likely involves displacement of DNA from AbrB. We elucidated the high-resolution NMR structure of the AbbA dimer and used chemical

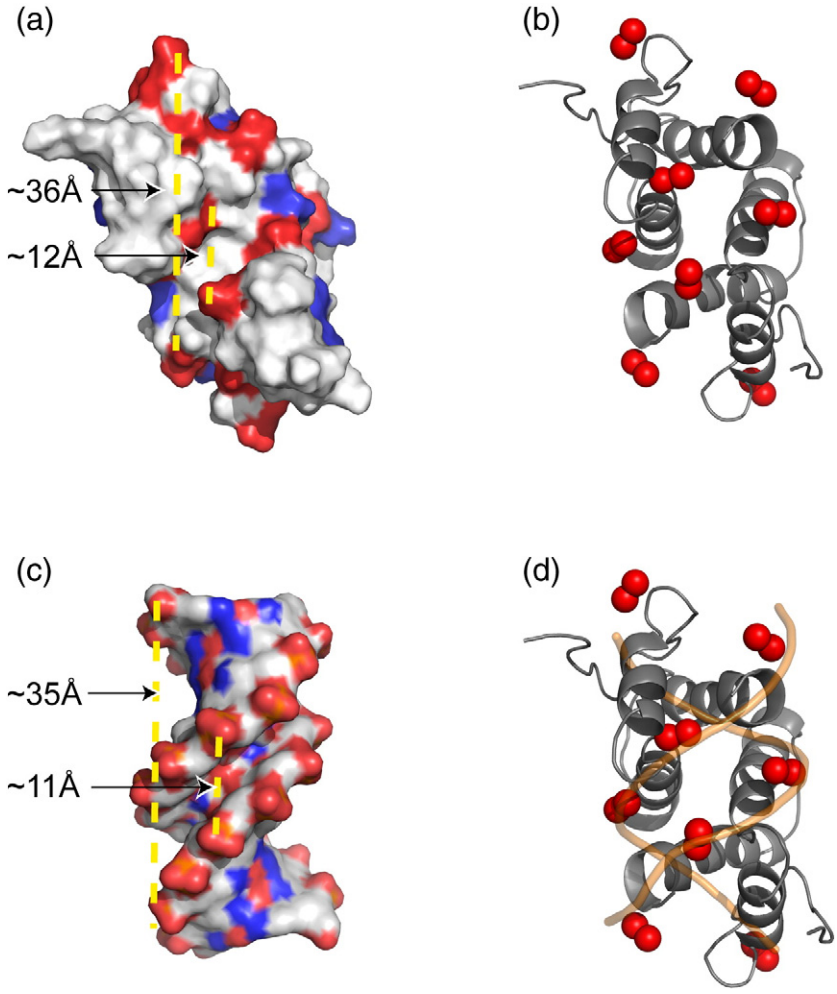

Fig. 4. AbbA is a DNA mimic. (a) Charge distribution shown in surface format of the unbound/free NMR characterized structure of AbbA presenting a DNA-like surface for the interaction with AbrB. Positive residues are shown in blue while negative residues are shown in red. (b) A cartoon format of the unbound/free NMR characterized structure of AbbA with side-chain oxygens of glutamic acid residues 16, 29, 33, and 67 shown as spheres. (c) B-DNA dodecamer (PDB code: 1BNA) shown as a surface plot according to charge distribution for comparison to (a). (d) Superposition of both the unbound/free NMR characterized structure of AbbA and 1BNA depicted in cartoon format with side-chain oxygens of glutamic acid residues 16, 29, 33, and 67 shown as spheres. Note how the side-chain oxygens line up to make a DNA-like phosphate backbone. For (a) and (b), the distance of one turn of DNA and minor grooves are noted with yellow broken lines and distances.

crosslinking and mass spectrometry to define AbbA:AbrB interactions. We also confirmed the binding stoichiometry between AbbA:AbrBN (dimer:dimer). This information provided restraints for subsequent computational docking efforts. These data allowed us to generate a detailed putative model for the AbbA:AbrBN complex. Of particular interest are any characteristics of AbbA that may help drive binding of AbrB and also any features that would make AbbA an appealing binding alternative to DNA.

The first 30 residues of AbbA, encompassing α1 and L1, are involved in the interaction with AbrB and also demonstrate the greatest propensity for conformational sampling. Additionally, in the AbbA dimer, these symmetrically disposed regions move in an opposing fashion. These data make functional sense when comparing the structure of the unbound AbbA dimer with the structure of the AbbA dimer bound to AbrBN. Upon binding, we note that α1/L1 in one monomeric AbbA unit moves away from α1/L1 in the other monomeric unit (α1′/L1′). This provides AbbA with the ability to reorient important elements in order to appropriately interact with AbrB.

From a purely structural standpoint, a particularly attractive incentive for AbrB's desire to bind to AbbA is apparent. Figure 4 shows that AbbA possesses structural and electrostatic traits suggesting that it is, to some degree, a “DNA mimic”. In Fig. 4a, the charge distribution of the surface AbbA uses to interact with AbrB is shown (negative residues are

red; positive residues are blue). A ribbon structure of AbbA is seen in Fig. 4b with side-chain oxygen atoms of glutamic acid residues 16, 29, 33, and 67 (16′, 29′, 33′, and 67′) as red spheres. It is instructive to compare the structures shown in Fig. 4a and b with those presented in Fig. 4c and d. In Fig. 4c, the charge distribution surface of a B-DNA dodecamer is shown (PDB code: 1BNA; negative charge in red and positive charge in blue) [34]. It is immediately evident that there are striking similarities between the charge distributions of the AbbA interaction surface and the DNA (including distances between adjacent grooves). In Fig. 4a, we note that the distance between glutamic acids 16 and 67′ in the AbbA dimer is ~36 Å and the distance between glutamic acids 29 and 29′ is ~12 Å. These distances compare very favorably with one turn of B-DNA (~35 Å) and the minor groove width (~11 Å), see Fig. 4c. Figure 4d shows AbbA (displaying the glutamic acids noted above) superimposed on the double-helical backbone of 1BNA (yellow). With these clear structural and charge similarities between DNA and AbbA in mind, it is not surprising that AbrB views AbbA as an excellent binding partner and remarkably competitive with DNA. In fact, despite the conformational change seen in AbbA upon binding AbrBN, AbbA retains nearly identically positioned glutamic amino acids between the bound and unbound forms, thereby presenting a similar DNA mimic interface in the bound and unbound forms (Supplementary Fig. 10).

In summary, in this work, we have determined the NMR structure of the anti-repressor protein AbbA from *B. subtilis*. In addition, using a combination of ITC, SELC, AUC, crosslinking mass spectrometry, and computational modeling, we have generated the first detailed model for how and why AbbA binds to the transition state regulator protein AbrB. In order to carry out its functional role, we have demonstrated that AbbA acts as a direct DNA mimic.

Materials and Methods

Expression and purification of AbbA

The *abbA* orf was cloned into expression vector pET-15b containing a thrombin cleavable N-terminal histidine tag. The purified plasmids were then transformed into competent BL21(DE3) cells (Novagen). AbbA was grown at 30 °C at 120 rpm in 1 L of LB broth containing 100 μg/ml ampicillin until an OD_{600nm} reached ~0.7. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.6 mM, and the culture was grown for 8 h at 30 °C. The cells were harvested by centrifugation and resuspended in 50 ml of lysis buffer containing 20 mM Tris (pH 7.5), 300 mM NaCl, 10 mM imidazole, and 0.02% NaN_3, supplemented with protease inhibitor and 0.01% Triton X-100. The cells were disrupted by sonication, and the resulting lysate was clarified by centrifugation and applied to Ni-NTA affinity resin (Qiagen). Purified AbbA fractions were pooled and extensively dialyzed against Q-Sepharose buffer containing 20 mM Tris (pH 7.5), 10 mM NaCl, and 0.02% NaN_3. The N-terminal histidine tag was removed using 200 U of thrombin. Cleaved AbbA was then applied to Q-Sepharose resin (Qiagen) and Ni-NTA affinity resin to capture any remaining His-tagged AbbA. The load and wash were collected and extensively dialyzed against NMR buffer containing 20 mM Tris (pH 6.5), 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, and 0.02% NaN_3. Throughout the protocol, the presence of AbbA was monitored by 12% *N*-[2-hydroxy-1,1-bis(hydroxymethyl) ethyl]glycine gel electrophoresis. Production of ^{1}H–^{15}N–^{13}C samples for NMR analysis followed the above mentioned protocol with the substitution of minimal medium (M9T) with 1 g of [^{15}N] ammonium chloride and 2 g of [*U*-$^{13}C_6$] glucose for expression. Protein concentrations in all experiments were determined using Bradford and BCA assays.

Expression and purification of AbrB, AbrBN, and AbrBC

Full-length AbrB from *B. subtilis* was expressed and purified as described previously [18]. AbrBN was expressed and purified as previously described [35]. AbrBC was cloned into expression vector pET-28a containing a thrombin cleavable N-terminal histidine tag. AbrBC was expressed and purified similarly to full-length AbrB with the following changes: After purification over the Q-Sepharose, fractions containing purified AbrBC were pooled and dialyzed into lysis buffer. AbrBC was then applied to Ni-NTA affinity resin equilibrated with lysis buffer to capture any remaining His-tagged AbrBC. The load and wash were collected and dialyzed into 10 mM KH_2PO_4 (pH 7.0) and 200 mM KCl.

Size-exclusion liquid chromatography

A protein molecular mass standard (Bio-Rad) containing bovine thyroglobulin (670,000 Da), bovine γ-globulin (158,000 Da), chicken ovalbumin (44,000 Da), equine myoglobin 17 (17,000 Da), and vitamin B12 (1350 Da) was used. Analysis was performed at room temperature using a Waters Breeze HPLC system equipped with a 4.6-mm ID Å ~ 30 cm TOSOH TSK-GEL SW2000 size-exclusion column. For the multimerization experiments, sample volumes of 10 μl with concentrations ranging from 10 μM to 1 mM were injected with a flow rate of 0.3 ml/min for 25 min. In order to detect complex formation, we mixed equimolar amounts of AbbA and AbrBN or AbrBC prior to injection. A mobile phase of 20 mM Tris (pH 7.5) and 200 mM NaCl was used. Chromatographs were analyzed with Waters Breeze software version 3.30.

Native mass spectrometry

All mass spectrometry analyses were performed on a Waters Quadrupole Time-of-Flight instrument (Waters, Milford, MA). Positive ion electrospray spectra of 20 μM AbbA were acquired in 10 mM ammonium acetate buffer (pH 7.5) using the standard electrospray ionization source

operated at a 3.5-kV capillary voltage. Samples were infused at a flow rate of 20 µl/min. The data were acquired at a scan rate of 1 scan per second over the range 200–5000 *m/z*. The instrument was operated at a resolution of 5000–6000. One hundred scans were collected and summed, and the multiply charged spectra were transformed into a calculated molecular mass using MassLynx 4.0 acquisition and data processing software. The spectra were smoothed twice using a moving average algorithm with a window width of 3 data points. Maximum entropy software, MaxEnt 1, was used for deconvolution of multiply charged electrospray envelopes.

Co-purification of AbbA with His_6-AbrB

Unmodified AbbA was purified as previously described [8]. His_6-AbrB and derivatives were purified from *E. coli* BL21 (DE3) as follows: Overnight cultures were grown at 37 °C in LB supplemented with 100 µg/ml ampicillin, diluted 1:50 into 500 ml fresh LB/ampicillin, and grown at 30 °C until an OD_{600nm} reached ~0.5. IPTG was then added to a final concentration of 1 mM to induce expression. The culture was grown with IPTG for 2 h at 30 °C. Harvested cells were resuspended in 25 ml ice-cold buffer A (50 mM Tris at pH 7.5 and 150 mM NaCl) and disrupted by sonication. Cell debris was removed by centrifugation at 32,000*g* for 15 min. The clarified supernatant was applied to 500 µl (bed volume) Ni-NTA agarose beads (Qiagen). The column was washed with 50 ml buffer B (buffer A plus 20 mM imidazole). The bound protein was eluted with three 1-ml fractions of buffer C (buffer A plus 500 mM imidazole). Purity was assessed by 15% SDS-PAGE. Purified protein was stored at 4 °C.

Highly purified AbbA (2 µM) was added to binding buffer [25 mM Tris (pH 8.0), 100 mM KCl, 5 mM $MgCl_2$, 0.2 mM DTT, and 10% (vol/vol) glycerol] and incubated with highly purified wild-type His_6-AbrB, His_6-$AbrB^{R8A}$, His_6-$AbrB^{R15A}$, His_6-$AbrB^{R23A}$, or His_6-$AbrB^{R24A}$ (20 µM) (for a total volume of 1 ml) at 37 °C for 10 min. One hundred microliters (bed volume) of Ni-NTA agarose was added to each binding reaction and incubated for 10 min at room temperature. The Ni-NTA agarose was pelleted, and the supernatant was removed and set aside. Each reaction was then washed with 30 ml of binding buffer containing 20 mM imidazole and eluted at 50 °C with 200 µl of SDS sample buffer. AbbA and His_6-AbrB were detected in the various fractions by Coomassie staining.

Isothermal titration calorimetry

For the ITC measurements, purified AbbA and AbrB were extensively buffer exchanged into ITC buffer [10 mM KH_2PO_4 (pH 7.0) and 200 mM KCl]. The AbbA:AbrB ITC experiments were performed at 26 °C using an AutoITC200 microcalorimeter (GE MicroCal Inc., USA) by injecting 2 µl of 300 µM AbbA into a 200-µl sample cell containing 15 µM AbrB. A total of 20 injections were performed with a spacing of 180 s and a reference power of 7 µcal/s. The binding isotherms were fit to a one-site binding model using Origin 7 Software (MicroCal, Inc., USA).

NMR spectroscopy and structure calculations

NMR experiments were carried out on 1.0 mM samples in 90%:10% or 1%:99% H_2O:D_2O, 20 mM Tris (pH 6.5), 200 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.02% NaN_3. All experiments were performed at 305 K on a Varian INOVA 600 or Bruker Avance 700 both equipped with triple-resonance cryoprobes. Sequential assignments were made from HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments [36–40]. Side chains were assigned from H(CCO)NH, (H)C(CO)NH, and HCCH-TOCSY experiments [36,38,39]. Side-chain aromatic chemical shifts were identified through a combination of (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, aromatic CHSQC, and aromatic CNOESY (50 and 120 ms mixing times) [41]. Side-chain amines of glutamine and asparagine were identified through the CB(CACO)NH and HNCACB. HNHA, CSI, and TALOS experiments were used to determine coupling constants for assigning backbone ψ and ϕ angles and hydrogen bonds [42]. $^1D_{NH}$ RDCs (*r*esidual *d*ipolar *c*ouplings) were measured on 1.0 mM uniformly labeled ^{15}N samples of AbbA using IPAP-HSQC experiments recorded on a Bruker Avance 700 from 6.0 mm to 4.2 mm radially compressed 7% polyacrylamide gels [43]. NOEs were obtained from 3D ^{15}N-NOESY-HSQC experiments with 120 ms and 150 ms mixing times, 3D ^{13}C-NOESY-HSQC experiment with 120 ms mixing time, aromatic 3D ^{13}C-NOESY-HSQC experiments with 50 and 120 ms mixing times, and 3D ^{13}C-NOESY-HSQC isotope-filtered experiment with 120 ms mixing time to obtain intermolecular NOEs [44–46]. The spectra were processed with NMRPipe and analyzed with NMRView on LINUX or MAC workstations [47,48].

Structure calculations

The automated NOE assignment and the structure calculation were performed with CYANA 3.0 [49]. Monomeric structures were first calculated through dihedral constraints, J-coupling constraints, and carbon and nitrogen NOESY experiments (120 ms, 120 ms, and 150 ms mixing times, respectively). The resulting monomeric structure was then run through the DALI server [17]. Through the DALI analysis, a model of the AbbA dimeric sequence was constructed through MODELLER and scored by the DOPE scoring method [50]. The model with the best DOPE score was used as a starting structure for the structural calculations. Dimeric structure calculations included the calculation of 100 structures with the top 20 structures being used for analysis, 20,000 torsion angle dynamics steps per conformer, and used standard RDC weight and cutoff values. RDC D_a and R values within the CYANA protocol were −1.843 and 0.444. RDC restraints for the dimer structures were fit and analyzed with the MODULE program using SVD, where the measured RDC value was assigned to the N–H vectors of a given residue in each of the two monomeric subunits [51]. Structures were calculated with NOEs, J-coupling restraints (determined from the HNHA), and RDCs and ψ and ϕ angles (TALOS predictions). Ninety-nine CYANA conformers with the lowest final target function values were subjected to restrained energy minimization in implicit solvent in AMBER [52]. The lowest 20 energy structures as calculated through AMBER were than analyzed, minimized for further violations, and

ultimately submitted to the PDB. Molecules were visualized and aligned with PyMOL [53]. The structure was further analyzed by PSVS and WHATCHECK to confirm stereochemical quality of a protein structure [15,16].

Sample preparation and nano-flow liquid chromatography electrospray ionization tandem mass spectrometry analysis

Purified AbbA and AbrBN in 10 mM KH_2PO_4 (pH 7.5), 15 mM KCl, 1 mM EDTA, and 0.02% NaN_3 were mixed in a ratio of 1.5:1 and incubated at 4 °C overnight. The amine-to-amine crosslinkers BS3 and DSG (Pierce) were added to 100 µl of the reaction mixture at a 1:50 (protein-to-crosslinker) ratio. The reaction proceeded for 1.5–2 h and was quenched with Tris. For the EDC and sulfo-NHS (*N*-hydroxysulfosuccinimide) reaction, 8 mM EDC and 10 mM sulfo-NHS were added to 500 µM AbbA in 4-morpholineethanesulfonic acid buffer [100 mM 4-morpholineethanesulfonic acid (pH 6.0) and 500 mM NaCl] for 15 min. β-Mercaptoethanol was added to deactivate excess EDC. Next, 400 µM AbrBN in phosphate-buffered saline buffer [10 mM KH_2PO_4 (pH 7.5) and 200 mM KCl] was added to the reaction mixture. The reaction proceeded for 2 h before quenching with ethanolamine. The reaction mixtures were then separated on a 1D-SDS-PAGE gel. Following separation, the molecular mass region corresponding to that of an AbbA and AbrBN complex (~20–28 kDa) was excised and subjected to an in-gel reduction, iodoacetamide alkylation, and trypsin digestion as previously described [54]. Extracted peptides were lyophilized to dryness and then resuspended in 20 µl of 2% acetonitrile and 0.1% formic acid prior to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. Chromatographic separation was performed on a Waters NanoAcquity UPLC equipped with a 1.7 µm BEH130 C_{18} 75 µm ID × 250 mm reversed-phase column. The mobile phase consisted of (a) 0.1% formic acid in water and (b) 0.1% formic acid in acetonitrile. Following a 4-µl injection, peptides were trapped for 5 min on a 5-µm Symmetry C_{18} 180 µm ID × 20 mm column at 5 µl/min in 99.9% A. The analytical column was then switched in-line and a linear elution gradient of 5% B to 40% B was performed over 60 min at 400 nl/min. The analytical column was connected to a fused silica PicoTip emitter (New Objective, Cambridge, MA) with a 10-µm tip orifice and coupled to a Waters Synapt G2 Quadrupole Time-of-Flight mass spectrometer through an electrospray interface. The instrument was operated in data-dependent mode of acquisition with precursor mass spectrometry scans from *m*/*z* 50–2000 and the top three most abundant precursor ions being subjected to MS/MS fragmentation. For all experiments, charge-dependent CID energy settings were employed and a 120-s dynamic exclusion was employed for previously fragmented precursor ions.

Qualitative and crosslinked peptide identification from raw LC-MS/MS

Raw LC-MS/MS data files were processed in Mascot distiller (Matrix Science) and then submitted to independent Mascot database searches (Matrix Science) against a SwissProt (taxonomy *B. subtilis*) database (4290 forward sequences, updated December 2012) appended with the reverse sequence of all of the forward entries. Search tolerances were 10 ppm for precursor ions and 0.04 Da for product ions using trypsin specificity with up to two missed cleavages. Carbamidomethylation (+57.0214 Da on C) was set as a fixed modification, whereas oxidation (+15.9949 Da on M) and hydrolyzed DSG (+114.031694 Da on K) or hydrolyzed BS3 (+156.0786 Da on K) were considered as variable modifications. All searched spectra were imported into Scaffold (Proteome Software), and protein confidence thresholds were set using a Bayesian statistical algorithm based on the PeptideProphet and ProteinProphet algorithms that yielded a peptide and protein false discovery rate of <1% [55,56]. To identify crosslinked species, we generated Mascot distiller. MGF files were submitted to MassMatrix (v 2.4.2, February 2012) searches against a forward/reverse SwissProt database (taxonomy *B. subtilis*) of AbrBN and appended with the N-terminal thrombin cleavage site modified sequence of AbbA [57]. Search mass tolerances and modifications were as described for Mascot searches, with the "advanced search" option enabled to allow for interpeptide or intrapeptide crosslinking of DSG (+96.0211 Da), BS3 (+138.0681 Da), or EDC (−18.0106 Da). Specificity of the BS3 and DSG crosslinkers was initially confined to lysine–lysine residues and specificity of the EDC crosslinker was initially confined to lysine–aspartate or lysine–glutamate residues. A secondary search was then performed for each crosslinking reagent replacing lysine with glycine to allow for the mapping of crosslinked sites to the protein N-terminal primary amine. Trypsin rules were set to not allow cleavage at a crosslinked modified residue and only one crosslink per peptide pair was allowed. A peptide match within MassMatrix was only considered if peptide scoring thresholds were above that required for a matching probability less than *p*-value of <0.05. All crosslinked MS/MS spectra were manually inspected for adequate fragment ion coverage.

Sedimentation equilibrium AUC

AbbA and AbrBN were purified and concentrated to 2 mM and subjected to further purification using Sephacryl-100 High-Resolution gel-filtration media. Purified fractions of AbbA and AbrBN were diluted to 170 µM. For complex stoichiometry, purified AbbA and AbrBN were mixed at a 1:1 ratio at 170 µM. Sedimentation equilibrium experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) equipped with absorbance optics and an An60 Ti rotor. Data were collected at 280 nm at rotor speeds of 16,000 and 24,000 rpm at 20 °C. The rotor was run at each speed for 20 h, and then scans were obtained at 20, 22, and 24 h.

AbbA:AbrBN modeling with HADDOCK

Default HADDOCK parameters were used throughout the docking procedure [22] with the following exceptions: Non-crystallographic and *C*2 symmetry restraints were used for each dimer (AbbA and AbrBN, individually) and *C*2 symmetry restraints for the complex were used between residues 1–53 (and 1′–53′) of AbrBN and

residues 8–60 (and 8′–60′) of AbbA. Semi-flexible residues were defined as the DNA-binding interface of AbrBN (residues 9–32 for each monomer) and DNA mimic interface of AbbA (residues 15–33 and 66–68). Fully flexible residues were set to accommodate the mass spectrometry crosslinking data, BS3 has a 11.4-Å spacer arm and DSG has a 7.4-Å spacer arm. For defining the appropriate HADDOCK flexibility parameters, the following criteria were taken into account: the large distances of the crosslinking moieties, the dynamic nature of both proteins (from tCONCORD, THESEUS, and published studies [4]), and the need to retain dimeric structures of both AbrBN and AbbA at all concentrations (from SELC and AUC data above). This resulted in the following residues being defined as fully flexible: 1–53 for AbrBN and 1–68 for AbbA on both monomers. Lastly, ambiguous restrains were not randomly excluded as in default HADDOCK parameters.

Active residues for AbrBN were determined from mutational studies showing that R8, R15, and R23 mutations knocked out the ability for AbbA to bind to AbrB and the mass spectrometry crosslinking studies showing that K9, D11, E12, E30, and D40 had crosslinks to AbbA (Supplementary Table 5). Passive residues for AbrBN were defined as charged residues around the active residues (the DNA-binding interface) with solvent accessibility of >50% calculated by NACCESS [23]. Active residues for AbbA were determined from mass spectrometry crosslinking studies showing that G1, E11, E15, and E67 had crosslinks to AbrBN. Passive residues for AbbA were defined as charged residues around the active residues (the DNA mimic interface) with solvent accessibility of >40% calculated by NACCESS (Supplementary Table 5). One thousand structures were generated for the first iteration (rigid docking), 200 were generated for the second iteration (semi-flexible docking), and 200 lowest energy structures were water refined. The C^{α} r.m.s.d. values of the complexes were calculated using ProFit [58]. A cluster analysis was performed on the final docking solutions using a minimum cluster size of 4. The r.m.s.d. cutoff for clustering was manually determined to be 5.5 Å (lower than the default 7.5 Å). The r.m.s.d. matrix was calculated over the backbone atoms of the interface residues of the proteins. The lowest energy structure from the molecular docking within the highest populated cluster was further analyzed by PSVS to confirm stereochemical quality of a protein structure [15].

Accession numbers

Protein Data Bank: The AbbA coordinates have been deposited with the accession code 2LZF. Biological Magnetic Resonance Data Bank: The AbbA chemical shifts have been deposited with the accession code 18753.

Acknowledgements

The work is dedicated to the memory of Mark A. Strauch, whose invaluable contributions to our understanding of AbrB we acknowledge. AUC work was performed at the Wake Forest University AUC Facility with Dr. Roy Hantgan. ITC work was performed at the University of North Carolina Macromolecular Interactions Facility. Kevin Blackburn (North Carolina State University) assisted with native mass spectrometry studies. This work was supported in part by the following: National Institutes of Health Grant RO1 GM55769 (J.C.); a grant from the V Foundation for Cancer Research (J.C.); a grant from the Kenan Institute for Engineering, Technology and Science (J.C.); and National Institutes of Health Grant GM18568 (R.L.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.02.010.

Received 7 January 2014;
Received in revised form 7 February 2014;
Accepted 7 February 2014
Available online 15 February 2014

Keywords:
transition state regulator;
AbbA;
DNA mimic;
molecular docking;
NMR

Present address: A. V. Banse, Institute of Molecular Biology, 297 Klamath Hall, 1229 University of Oregon, Eugene, OR 97403, USA.

A.T.T. and B.G.B. contributed equally to this work.

Abbreviations used:
SELC size-exclusion liquid chromatography; ITC isothermal titration calorimetry; HSQC heteronuclear single quantum coherence; NOE nuclear Overhauser enhancement; NOESY NOE spectroscopy; AUC analytical ultracentrifugation; PCA principal components analysis; EDTA ethylenediaminetetraacetic acid; 3D three-dimensional; LC-MS/MS liquid chromatography tandem mass spectrometry

References

[1] Aertsen A, Michiels CW. Stress and how bacteria cope with death and survival. Crit Rev Microbiol 2004;30:263–73.
[2] Hecker M, Volker U. General stress response of *Bacillus subtilis* and other bacteria. Adv Microb Physiol 2001;44:35–91.
[3] Strauch MA, Hoch JA. Transition-state regulators: sentinels of *Bacillus subtilis* post-exponential gene expression. Mol Microbiol 1993;7:337–42.
[4] Sullivan DM, Bobay BG, Kojetin DJ, Thompson RJ, Rance M, Strauch MA, et al. Insights into the nature of DNA binding of AbrB-like transcription factors. Structure 2008;16:1702–13.

[5] Xu K, Strauch MA. *In vitro* selection of optimal AbrB-binding sites: comparison to known *in vivo* sites indicates flexibility in AbrB binding and recognition of three-dimensional DNA structures. Mol Microbiol 1996;19:145–58.
[6] Strauch M, Webb V, Spiegelman G, Hoch JA. The SpoOA protein of *Bacillus subtilis* is a repressor of the *abrB* gene. Proc Natl Acad Sci USA 1990;87:1801–5.
[7] Molle V, Fujita M, Jensen ST, Eichenberger P, Gonzalez-Pastor JE, Liu JS, et al. The Spo0A regulon of *Bacillus subtilis*. Mol Microbiol 2003;50:1683–701.
[8] Banse AV, Chastanet A, Rahn-Lee L, Hobbs EC, Losick R. Parallel pathways of repression and antirepression governing the transition to stationary phase in *Bacillus subtilis*. Proc Natl Acad Sci USA 2008;105:15547–52.
[9] Bobay BG, Andreeva A, Mueller GA, Cavanagh J, Murzin AG. Revised structure of the AbrB N-terminal domain unifies a diverse superfamily of putative DNA-binding proteins. FEBS Lett 2005;579:5669–74.
[10] Xu K, Clark D, Strauch MA. Analysis of abrB mutations, mutant proteins, and why abrB does not utilize a perfect consensus in the −35 region of its sigma A promoter. J Biol Chem 1996;271:2621–6.
[11] Bobay BG, Benson L, Naylor S, Feeney B, Clark AC, Goshe MB, et al. Evaluation of the DNA binding tendencies of the transition state regulator AbrB. Biochemistry 2004;43:16106–18.
[12] Strauch MA. *In vitro* binding affinity of the *Bacillus subtilis* AbrB protein to six different DNA target regions. J Bacteriol 1995;177:4532–6.
[13] Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003;50:437–50.
[14] Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. J Mol Biol 1993;231:1049–67.
[15] Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–95.
[16] Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graphics 1990;8:29.
[17] Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–80.
[18] Olson AL, Liu F, Tucker AT, Goshe MB, Cavanagh J. Chemical crosslinking and LC/MS analysis to determine protein domain orientation: application to AbrB. Biochem Biophys Res Commun 2013;431:253–7.
[19] Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. J Mass Spectrom 2003;38:1225–37.
[20] Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J Struct Biol 2011;173:530–40.
[21] Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. Mol Cell Proteomics 2010;9:1634–49.
[22] Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–7.
[23] Hubbard SJ, Thornton JM. NACCESS. London: Department of Biochemistry and Molecular Biology, University College; 1993.
[24] Seeliger D, De Groot BL. tCONCOORD-GUI: visually supported conformational sampling of bioactive molecules. J Comput Chem 2009;30:1160–6.
[25] Theobald DL, Wuttke DS. Accurate structural correlations from maximum likelihood superpositions. PLoS Comput Biol 2008;4:e43.
[26] Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, et al. Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol Cell 2002;9:187–94.
[27] Wang HC, Wu ML, Ko TP, Wang AH. *Neisseria* conserved hypothetical protein DMP12 is a DNA mimic that binds to histone-like HU protein. Nucleic Acids Res 2013;41:5127–38.
[28] Putnam CD, Tainer JA. Protein mimicry of DNA and pathway regulation. DNA Repair (Amst) 2005;4:1410–20.
[29] Dryden DT, Tock MR. DNA mimicry by proteins. Biochem Soc Trans 2006;34:317–9.
[30] Banos-Sanz JI, Mojardin L, Sanz-Aparicio J, Lazaro JM, Villar L, Serrano-Heras G, et al. Crystal structure and functional insights into uracil-DNA glycosylase inhibition by phage Phi29 DNA mimic protein p56. Nucleic Acids Res 2013;41:6761–73.
[31] Wang HC, Ko TP, Lee YM, Leu JH, Ho CH, Huang WP, et al. White spot syndrome virus protein ICP11: a histone-binding DNA mimic that disrupts nucleosome assembly. Proc Natl Acad Sci USA 2008;105:20758–63.
[32] Hegde SS, Vetting MW, Roderick SL, Mitchenall LA, Maxwell A, Takiff HE, et al. A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. Science 2005;308:1480–3.
[33] McMahon SA, Roberts GA, Johnson KA, Cooper LP, Liu H, White JH, et al. Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. Nucleic Acids Res 2009;37:4887–97.
[34] Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, et al. Structure of a B-DNA dodecamer: conformation and dynamics. Proc Natl Acad Sci USA 1981;78:2179–83.
[35] Benson LM, Vaughn JL, Strauch MA, Bobay BG, Thompson R, Naylor S, et al. Macromolecular assembly of the transition state regulator AbrB in its unbound and complexed states probed by microelectrospray ionization mass spectrometry. Anal Biochem 2002;306:222–7.
[36] Grzesiek S, Bax A. Measurement of amide proton exchange rates and NOEs with water in $^{13}C/^{15}N$-enriched calcineurin B. J Biomol NMR 1993;3:627–38.
[37] Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 1990;29:4659–67.
[38] Logan TM, Olejniczak ET, Xu RX, Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. J Biomol NMR 1993;3:225–31.
[39] Logan TM, Olejniczak ET, Xu RX, Fesik SW. Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. FEBS Lett 1992;314:413–8.
[40] Montelione GT, Emerson SD, Lyons BA. A general approach for determining scalar coupling constants in polypeptides and proteins. Biopolymers 1992;32:327–34.
[41] Slupsky CM, Gentile LN, McIntosh LP. Assigning the NMR spectra of aromatic amino acids in proteins: analysis of two Ets pointed domains. Biochem Cell Biol 1998;76:379–90.
[42] Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302.

[43] Chou JJ, Gaemers S, Howder B, Louis JM, Bax A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J Biomol NMR 2001;21:377–82.
[44] Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry 1989;28:6150–6.
[45] Stuart AC, Borzilleri KA, Withka JM, Palmer AG. Compensating for variations in ^{1}H-^{13}C scalar coupling constants in isotope-filtered NMR experiments. J Am Chem Soc 1999;121:5346–7.
[46] Zuiderweg ER, Fesik SW. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. Biochemistry 1989;28:2387–91.
[47] Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 2004;278:313–52.
[48] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–93.
[49] Guntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol 2004;278:353–78.
[50] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 2006. http://dx.doi.org/10.1002/0471140864.ps0209s50.
[51] Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 2000;122:3791–2.
[52] Case TAD, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, et al. Amber 11. San Francisco: University of California; 2010.
[53] The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
[54] Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 1996;379:466–9.
[55] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75:4646–58.
[56] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74:5383–92.
[57] Xu H, Freitas MA. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. BMC Bioinforma 2007;8:133.
[58] Martin ACR. ProFit V3.1 edit. SciTech Software; 1992.