Conformational change of the *Bordetella* response regulator BvgA accompanies its activation of the *B. pertussis* virulence gene fhaB

David Kim\(^{a,1,2}\), John Tracey\(^{a,1,2}\), Manuel Becerra Flores\(^{b,1}\), Kanita Chaudhry\(^{a,2}\), Rafae Nasim\(^{a,2}\), Abraham Correa-Medina\(^{a,2}\), Leslie Knipling\(^{a}\), Qing Chen\(^{c}\), Scott Stibitz\(^{c}\), Lisa M.M. Jenkins\(^{d}\), Kyung Moon\(^{a,2}\), Tim Cardozo\(^{b,\ast}\), Deborah M. Hinton\(^{a,\ast}\)

\(^a\) Gene Expression and Regulation Section, Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

\(^b\) Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York University Grossman School of Medicine, 455 First Avenue, New York, NY 10016, USA

\(^c\) Division of Bacterial, Parasitic, and Allergenic Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA

\(^d\) Collaborative Protein Technology Resource, Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

**Abstract**

The BvgAS two-component system regulates virulence gene expression in *Bordetella pertussis*. Although precise three-dimensional structural information is not available for the response regulator BvgA, its sequence conservation with *E. coli* NarL and previous studies have indicated that it is composed of 3 domains: an N-terminal domain (NTD) containing the phosphorylation site, a linker, and a DNA-binding C-terminal domain (CTD). Previous work has determined how BvgA\(^{CTD}\) dimers interact with the promoter (P\(_{ntd}\)) of fhaB, the gene encoding the virulence adhesin filamentous hemagglutinin. Here we use molecular modeling, FeBABE footprinting, and crosslinking to show that within the transcription complex of phosphorylated BvgA (BvgA \(\sim \) P), *B. pertussis* RNAP, and P\(_{ntd}\), the NTDs displace from the CTDs and are positioned at specific locations relative to the three BvgA \(\sim \) P binding sites. Our work identifies a patch of the NTD that faces the DNA and suggests that BvgA \(\sim \) P undergoes a conformational rearrangement that relocates the NTD to allow productive interaction of the CTD with the DNA.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**1. Introduction**

*Bordetella pertussis*, the causative agent of whooping cough, is a re-emerging infectious disease, which affects millions worldwide and has high infant mortality. Although an acellular vaccine against whooping cough is widely used today, the incidence of pertussis has been steadily rising since the 1980s with over 48,000 cases in the United States in 2012 alone [1]. The failure to prevent whooping cough resurgence in both developed and developing nations indicates the need for better insight and understanding into *B. pertussis* pathogenesis at the molecular level. The current acellular vaccines contain 2 to 5 different *B. pertussis* virulence factors, including pertussis toxin (Ptx), the adhesin filamentous hemagglutinin (Fha), pertactin, and fimbriae (Fim3 and Fim2). In *B. pertussis*, these and a broad spectrum of other virulence genes are regulated by the master two-component system (TCS) BvgAS, consisting of the histidine sensor kinase (SK), BvgS, and its cognate response regulator (RR), BvgA [2].
In the BvgAS system, increasing intracellular levels of phosphorylated BvgA (BvgA ~ P) coordinate regulation. This TCS is affected by changing growth conditions and has been classified into three distinct modes: BvgC, BvgS, and BvgE [3–5]. The non-modulated condition, BvgC, is observed at 37 °C, such as inside the human respiratory tract or at normal laboratory growth conditions. In this environment, BvgS autophosphorylates and then transfers a phosphate to BvgA via a phosphorelay. BvgA ~ P forms dimers, which bind to BvgA ~ P binding sites located upstream of various promoters, activating transcription. ChiP-seq analyses have demonstrated BvgA ~ P binding at 54 known virulence-activated promoters and at up to 91 more suspected promoters [6]. In the presence of millimolar concentrations of magnesium sulfate or nicotinic acid or when cells are grown at lower temperatures (ambient temperature 25 °C), B. pertussis enters the BvgC mode where BvgS no longer phosphorylates BvgA. Consequently, intracellular levels of BvgA ~ P are very low, and virulence genes remain inactive. The BvgS mode represents an intermediate state between the BvgC and BvgE modes with an intermediate concentration of BvgA ~ P. It has been suggested that the BvgAS TCS works like a rheostat with activation of different virulence genes occurring at different intracellular concentrations of BvgA ~ P, perhaps coordinated with the stage of infection [5]. However, the molecular mechanism of how BvgA ~ P dimers regulate transcription at specific promoters is still not fully understood.

Transcription initiation in prokaryotes is performed by RNA polymerase (RNAP) consisting of a core of proteins (σr[σ][σ]0) that interact with a sigma specificity factor. RNAP containing the primary σ factor, σ8 in B. pertussis, recognizes specific double-stranded promoter sequences at the −10 and −35 regions of DNA to form an initial, unstable closed complex (RPC) [7]. The RPC then converts to the more stable open complex, RPo, which is competent to begin transcription at the +1 transcription start site (TSS). At regulated promoters, activators are often required to achieve RPo. These proteins can affect initiation through various mechanisms, including Class I, Class II, and the combination Class I/Class II [8]. Class I activators bind to one or more sites upstream of the −35 region and directly interact with the C-terminal domains (CTDs) of the α subunits of RNAP. Class II activators bind to the promoter next to or sometimes overlapping the −35 region and interact with the σ factor and/or the RNAP α-subunits. Some activators, including BvgA ~ P, use a combination of Class I and Class II mechanisms. Like Class II activators, a downstream BvgA ~ P dimer is positioned immediately adjacent to or overlapping the −35 region while the other dimer(s) sit farther upstream [9]. However, the architecture of BvgA ~ P with RNAP at the promoter of fhaB (PfhaB) differs from the canonical Class II mechanism in that the αCTDs of RNAP are located at BvgA ~ P binding sites, on a different face of the DNA helix, rather than adjacent to the activator [10].

Like many RRs, BvgA is a two-domain protein. The CTD is thought to bind DNA through a conserved helix-turn-helix DNA binding motif [10], and the N-terminal domain (NTD) contains the phosphorylation site at residue D54 [5]. These two domains are connected by a linker. Previous work has modeled BvgA CTD on the homologous E. coli NarL structure and demonstrated how the BvgATTD dimers sit on the primary binding site of PfhaB centered at −88.5 relative to the +1 TSS [10]. However, this structure does not include the BvgANTD. Consequently, the molecular details of how multiple BvgA ~ P dimers occupy the promoter and how the BvgATTDs are spatially oriented during transcription initiation remain unknown.

As the structure of BvgA or a BvgA-activated transcription complex has yet to be solved, we investigated these questions by coupling structural modeling and biochemical techniques. First, we generated a full monomer predicted structure of BvgA and an extended model of BvgA CTD binding at PfhaB. We then generated a map of the positions of various BvgANTD residues relative to the DNA within the transcriptional complex using Fe-BABE footprinting. Based on these data, we postulate that within the monomer form of BvgA, the NTD and CTD domains are relatively close and that they undergo a conformational change relative to each other in the formation of the activated transcription complex. This movement then results in discrete new positions for BvgANTD relative to the DNA. We hypothesize that this protein rearrangement is needed for BvgA ~ P activation at PfhaB and perhaps will be observed for other RRs.

2. Results

2.1. 3D structural model of full-length BvgA

No experimental 3D structures have been obtained for BvgA with or without DNA. However, BvgA exhibits similar domain architecture to the E. coli RR NarL: an N-terminal activation domain (NTD), which bears a conserved aspartic acid for phosphorylation; a linker; and a C-terminal domain (CTD), which exhibits a helix-turn-helix motif for DNA binding (Fig. 1A). Furthermore, the NTDs and CTDs of the two proteins exhibit structurally-significant, sequence similarity (CTD = 31 %, p = 10^−4.4; NTD = 31 %, p = 10^−10.7) [11] and exhibit only four N-terminal amino acids and one internally inserted amino acid (M28) in NarLNTD relative to BvgANTD (Fig. S1). However, although the linkers have similar lengths (NarL: 28 amino acids, residues 127–154; BvgA: 26 amino acids, residues 122–147), they are poorly related in sequence (Fig. S1).

Two NarL structures are available: a full length NarLFN monomer [12], PDB:1RNL and a dimer NarLNTD/DNA complex [13], PDB:1J88. In NarLFN, the linker is partly disordered, suggesting it is highly flexible and can easily accommodate movements of its NTD relative to its CTD. Given the overall similarity of NarL and BvgA, we used the NarFLN monomer structure as the template to construct a full length BvgA (BvgFLN) homology model (see Materi-
als and Methods). The secondary structure elements within the NTDs and CTDs of NarL and BvgA are highly conserved in sequence (Fig. S1). Thus, the nearly identical corresponding elements allow us to create highly accurate 3D models of respective CTD and NTD domain structures. (Fig. 1B, BvgA<sup>NTD</sup> in cyan, linker in magenta, BvgA<sup>CTD</sup> in green vs NarL in tan). The loop connecting two helices and containing M28 in NarL necessarily has a different conformation in BvgA compared to the NarL crystal structure, since it contains fewer amino acids in BvgA. The same is true of the linker between the CTD and NTD, as it also contains two fewer amino acids in BvgA compared to NarL. Accordingly, we searched all possible conformations of these two segments and calculated the biological energy (van der Waals, electrostatics, etc.) of each conformation to find the lowest energy conformations, which are the most likely biologically relevant ones. As an ordered, short helix within the linker observed in the NarL crystal structure (amino acids 134–145) is fairly well conserved in sequence in BvgA, conformations resembling this helix were favored in the search for the lowest energy conformation by lightly tethering each corresponding atom in this helix in BvgA to its matched atom location in the NarL crystal structure. The two lowest energy conformations of the linker (conformations 1 and 2, Fig. S2A) were very close in energy, suggesting structural flexibility within the linker. Generic structural validation scores for these 3D models were in the acceptable range: ERRAT<sup>[14]</sup> > 92.7 %; Ramachandran plot/Procheck<sup>[15]</sup> > 85 % of residues in the core areas, 13 % in additional allowed areas, and no residues in disallowed areas of the plot; no irregularities detected by Whatcheck<sup>[16]</sup>.

As orthogonal validation, we also predicted the BvgA<sup>NTD</sup> and BvgA<sup>CTD</sup> structures using AlphaFold2, which is an artificial intelligence algorithm that uses amino acid sequence to predict 3D structures of proteins <em>ab initio</em> with accuracy rivaling experimental methods<sup>[17,18]</sup>. The AlphaFold2 structures were nearly identical to our models (Fig. S2B). However, AlphaFold2 predicted an α-helix for the disordered segment of the linker. The combined observations of a disordered linker in NarL and multiple conformations of the linker in our model and in the AlphaFold2 models strongly suggest that the linker is highly flexible, which would mean that it could easily accommodate movement of the NTD relative to the CTD in both BvgA and NarL.

### 2.2. Protein-protein crosslinking is consistent with the predicted BvgA<sup>FL</sup> model.

The modeled BvgA<sup>FL</sup> predicted that in the absence of DNA, a face of the NTD is immediately adjacent to the DNA binding CTD (Fig. 1B and S2A, NTD in cyan and CTD in green). In fact, some residues in the assigned DNA binding helix, residues N177 to K190 (shown in dark blue in Fig. S2A), are within 5 Å of the NTD. Thus, to investigate the model, we performed protein crosslinking experiments using the crosslinking reagent 1,4-bismaleimidobutane (BMB), which crosslinks available sulphydryl groups within a distance of 10.44 ± 2.04 Å<sup>[19]</sup>. We started with a clone encoding a ‘cys-less’ BvgA protein in which the two native cysteines, C93 and C103, had been converted to alanines; previous work has indicated that this ‘cys-less’ BvgA is active <em>in vivo</em> and <em>in vitro</em><sup>[10]</sup>. Within this background, we constructed a BvgA S96C/T194C variant. S96 is present within the BvgA<sup>NTD</sup> region that faces T194C, which is present within the BvgA<sup>CTD</sup> (Fig. S2A, S96 shown as black sphere, T194 shown as lemon yellow sphere). <em>In vitro</em> transcription assays indicated that the variant was active and dependent on phosphorylation, as expected (Fig. S3A). After BMB treatment of phosphorylated and non-phosphorylated S96C/T194C and WT BvgA (which contains the natural C93 and C103), we observed products consistent with both monomer and dimer forms of the proteins (Fig. S3B). These products were isolated and analyzed by mass spectrometry to determine the location of any crosslinks. When using either the non-phosphorylated or phosphorylated variant after isolation as either the monomer or dimer species, we found a crosslink consistent with S96C-BMB-T194C.

![Fig. 2. Sequence of P<sub>fhAB</sub> and positions of FeBABE cut sites. The sequence of P<sub>fhAB</sub> from –110 to +6 is shown. The 3 BvgA – P binding sites (primary, secondary, tertiary) are boxed, and the position of each BvgA<sup>CTD</sup> dimer is indicated in green. The cut sites obtained for each residue that was substituted with cysteine and conjugated with FeBABE are shown for the top (non-template) strand and bottom (template) strand. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image-url)
Table S1). These intermolecular crosslinks can only form between S96C-BMB-S96C, when using the variant isolated as a dimer. When using WT isolated as a dimer, and a crosslink consistent with a C93-BMB-C93, which is predicted interface of the BvgA CTD and BvgA NTD. However, the protein, S96 and T194 are relatively close, consistent with the previous model results which suggested a flexible linker, the distance observed between the modeled and experimental S96 and T194 distance is not surprising. We conclude that within the BvgAFL model, the BvgACTD and BvgANTD are indeed relatively close. The distance observed between S96 and T194 in the BvgAFL model is 4.6 or 4.9 Å for conformations 1 or 2, respectively, while the crosslink distance for BMB is between 8.40 and 12.48 Å. Seen in the context of our previous model results which suggested a flexible linker, the difference observed between the modeled and experimental S96 and T194 distance is not surprising. We conclude that within the BvgAFL model, the BvgACTD and BvgANTD are indeed relatively close. The space between them is either ~5 Å larger than predicted from the models or more likely, the flexible linker generates a distance that the crosslinker can capture. Thus, it seems likely that in the free protein, there is flexibility in the space between the NTD and the CTD. 

2.3. Modeling a BvgACTD dimer on P_{phaB} DNA.

P_{phaB} contains three BvgA ~ P binding sites, centered at positions –88.5 (primary), –66.5 (secondary), and –44.5 (tertiary) (Fig. 2A). Previous work has developed a metric to determine the strength of a BvgA binding site based on the similarity of each half site to a consensus sequence. In this metric, scores less than zero indicate more deviation from the consensus [20,21]. For P_{phaB}, the primary site has the best score for each half (0/0), indicating a strong site, and the tertiary site, which is located adjacent to the –35 region of the promoter is the weakest of the three sites (score of –13/–4).

To model the BvgACTD dimer on its binding site, we began with the NarLCTD dimer/DNA complex [13], PDB:1JE8. A homology model of the helix-turn-helix portion of the BvgACTD based on this structure was nearly identical to that modeled on NarLCL or predicted by AlphaFold2. Consequently, two BvgACTD’s were placed on the DNA by superimposing them on the NarLCTD’s in the NarLCTD dimer/DNA complex. The DNA in the complex structure was then modified to bear the sequence of the primary BvgA ~ P binding site within P_{phaB} (Fig. 2A), for the purposes of designing our footprinting experiments. The final structure (Fig. 3A) is similar to a previously reported model of the BvgACTD dimer/DNA complex [10].

A comparison of the modeled BvgACTD dimer/DNA structure (Fig. 3A) with the NarLCTD dimer/DNA structure (Fig. 3B) shows a difference in a key contact for dimerization. For NarL the nonpolar residues V208 and H211 (shown in tan and grey spheres, respectively) constitute an important hydrophobic contact for dimerization [13], while for BvgA, the corresponding residues D201/K204 are positioned to assist dimerization via a salt-bridge.

As shown in Fig. 3C, the presence of a BvgANTD at its modeled location within the monomer protein would preclude the BvgACTD dimer from binding to the DNA. Thus, the positions of the NTDs within the transcription complex cannot be predicted from the position of an NTD within BvgAFL. Consequently, we hypothesized that each BvgANTD must move relative to its adjoining CTD to accommodate DNA binding. This hypothesis is supported by the disordered linker in the crystallographic structure of NarLCL and our models of BvgAFL, which strongly suggest a flexible linker between the CTD and NTD in both transcription factors. With a flexible linker, the NTD could thus move to random positions or to specific positions within the transcription complex, relative to the CTD. To test this hypothesis, we performed iron bromoacetamidobenzyl-EDTA (FeBUBE) footprinting on the BvgA ~ P/P_{phaB}/B. pertussis RNAP transcription complex as detailed below.
2.4. Selection of BvgA residues for conjugation with FeBABE

FeBABE footprinting is a powerful tool to investigate the architecture of a protein/DNA complex. The FeBABE moiety, which can be conjugated to a cysteine via a sulfhydryl bond, generates hydroxyl radicals in the presence of H$_2$O$_2$ and sodium ascorbate (NaAsc), which will then cleave DNA or protein within ~20 Å [22,23]. The resolution of this technique is ~3 bp, and importantly, the conjugated residue does not have to directly interact with the DNA. Consequently, substituting specific residues individually with cysteine, followed by conjugation with FeBABE and FeBABE footprinting can reveal the location of a residue relative to a position within the DNA [9,10,22,24–31]. Previously, this technique has been performed using BvgA with conjugated residues at positions V148 and T194, residues present within the BvgA$^{CTD}$ DNA binding domain of BvgA$^{CTD}$ [10,31]. This analysis demonstrated that in a transcription complex with either P$_{fhaB}$ or another BvgA-dependent B. pertussis promoter, P$_{phab}$, the positions of these residues were consistent with a previous structural model of the BvgA$^{CTD}$ dimer/DNA [5,10] as well as our recent model (Fig. 3A).

To extend this work, we used our BvgA$^{LS}$ homology model (Fig. 4A) to select specific residues predicted to lie on the surface of BvgA$^{NTD}$, the linker, and BvgA$^{CTD}$ (Fig. 4B). We then generated BvgA variants using the ‘cys-less’ BvgA with individual cysteine substitutions at the following positions: K4, K41, L60, R68, G73, and S96 within BvgA$^{NTD}$ (Fig. 4B, left); Y122, S127, E132, R134, and S141 within the linker (Fig. 4B, middle); and F174 as well as V148 and T194, which were used in the previous studies, within BvgA$^{CTD}$ (Fig. 4B, right). We also constructed variants at positions R19, P75, S176, Y183, R186, and R205; however, these proteins could not be used because of difficulties with purification and/or loss of activity.

The BvgA variants were treated with FeBABE, and the percent conjugation was determined by SDS-PAGE, since the unconjugated 23 kDa BvgA migrates slightly faster than the conjugated species, which contains the ~570 Da FeBABE moiety (Fig. 5A). From these gels, we estimated that the percent conjugation ranged from approximately 25 % to 100 %. To test for activity, the conjugated BvgA variants were assayed in single round transcription assays using B. pertussis RNAP and a supercoiled plasmid template containing P$_{phab}$ and the control promoter for RNAI (Fig. 5A). Most of the proteins retained phosphorylation-dependent transcriptional activity of ~50 % or greater relative to WT BvgA, while K41C, L60C, Y122C, and S127C demonstrated lower activity (~20–35 %) (Fig. 5B).

To generate stable transcription complexes for the FeBABE reactions, we incubated B. pertussis RNAP, the FeBABE conjugated BvgA ~ P, and [5'-32P] end-labeled P$_{phab}$ DNA at 37 °C, challenged the complexes with heparin to eliminate unstable species, and then added H$_2$O$_2$/NaAsc to generate hydroxyl radicals via the Fenton reaction. The DNA was separated on denaturing, sequencing gels to determine the cleavage positions. Control reactions using complexes of BvgA T194C-FeBABE or V148C-FeBABE (Fig. 6A, B) with B. pertussis RNAP and P$_{phab}$ DNA yielded cleavage patterns (Fig. 2) as were seen previously when using transcription complexes made with E. coli RNAP and BvgA ~ P [10], consistent with the BvgA$^{CTD}$ dimer/DNA modeled structure (Fig. 3A, T194C-FeBABE or V148C-FeBABE spheres in blue and yellow, respectively).

We observed the cleavage pattern obtained with F174C-FeBABE (Fig. 2 and Fig. 6A and B). In the modeled BvgA$^{CTD}$ dimer/DNA structure (Fig. 3A), residue 174 is predicted to yield cleavages that lie between the outside edges of the dimer, as defined by the cut sites with V148C-FeBABE, and the interior cut sites, as defined by T194C-FeBABE. F174C-FeBABE did yield these intermediate positions of cleavage (Fig. 2). This result is consistent with the modeled structure and confirms this position of the CTDs on the DNA within the transcription complex.

2.5. FeBABE footprinting reveals a face of BvgA$^{NTD}$/linker that is located across from the DNA within the P$_{phab}$ transcription complex

There is no evidence that the BvgA$^{NTD}$ or the BvgA linker directly interacts with the DNA, but we hypothesized that FeBABE analyses might reveal the positions of the NTDs within the transcription complex since FeBABE cleavage can extend to ~20 Å. Consequently, we determined the cleavage patterns made by each of the 10 BvgA-FeBABE proteins with an NTD or linker substitution (Fig. 4: Fig. 6A and B). Of these residues, FeBABE at K4C failed to

![Fig. 4](image-url) Location of FeBABE-conjugated residues within BvgA. (A) The modeled structure of BvgA$^{LS}$ (conformation 1). NTD is in cyan, linker is in magenta, and CTD is in green. (B) Residues that were substituted with cysteine and conjugated with FeBABE are shown as red spheres within the NTD, linker, and CTD modeled structures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
generate cuts. However, each of the other nine conjugated residues generated specific cuts that were dependent on the presence of H$_2$O$_2$ and NaAsc. This observation is consistent with there being a specific position for the NTD/linker within the transcription complex. Consequently, we eliminated the possibility that the NTDs simply move away from the CTDs into a variety of random locations. Furthermore, we found that a face along a surface of the modeled BvgA, including NTD residues K41, R68, S96, and linker residues E132 and R134, generated a set of cleavages at each of the BvgA binding sites that are positioned increasingly farther upstream as the FeBABE moiety is placed from the more N-terminal K41C to the more C-terminal R134C. This is best seen by observing the downward shift in the DNA cleavage sites relative to the red line that indicates the same position of the G + A ladder on each side of the gel in Fig. 7. It should be noted that the additional minor cuts observed with some of these residues (T194, K41, R68, E132, R134) are typically seen in FeBABE analyses and are thought to arise from the position of the residue relative to the major and minor grooves of the DNA [23].

2.6. Speculative model of BvgA dimers at P$_{fhaB}$

To determine how the BvgA$^{NTD}$s could be located relative to the BvgA$^{CTD}$s, we first used the results of our FeBABE analyses to position the farthest upstream BvgA$^{NTD}$ relative to the primary site DNA. We generated a two site DNA model containing the primary and secondary sites with the intervening DNA modeled as B form DNA (See Materials and Methods). Our studies strongly suggest that BvgA$^{NTD}$ adopts a different orientation relative to the CTD when BvgA $\sim$ P is bound to P$_{fhaB}$ DNA. Consequently, using the two site model, we systematically rotated and translated the farthest upstream BvgA$^{NTD}$, as an isolated domain, relative to the upstream half of the primary site in steps of a few degrees or Angstroms for each step in the space. We did this within a 3D envelope bounded by the intersection of a) the face of BvgA$^{NTD}$ with FeBABE footprinting residues to be within 20 Å of their FeBABE cleavage sites (Fig. 2) on the DNA and b) the space allowing the BvgA linker to connect the C-terminal amino acid of BvgA$^{NTD}$ (residue 121) to the N-terminal amino acid of BvgA$^{CTD}$ (residue 148) from either our model or the AlphaFold2 model, making sure that neither the linker nor the BvgA$^{NTD}$ clashed with the DNA and that the linker was in a relaxed low energy conformation. Importantly, it was not possible to generate a model without clashes unless the positions between S96 and T194 were greater than the maximum crosslink distance of 12.5 Å. Consequently, we hypothesize that within the transcription complex, the NTD moves even farther away from the CTD than is observed by the crosslink using BvgA alone.

Several unclashed conformations were found that satisfied the crosslinking data, demonstrating that conformations consistent
with our data can exist. The model exhibiting predicted FeBABE cleavage sites that agreed best with those determined experimentally (Table 1) was selected as the final model (Fig. 8, conjugated residues and their corresponding cut sites are color coded similarly). As seen in Table 1, in most cases, the closest experimentally determined cut sites were the same as those predicted by the model. It is possible that unknown distortions in the DNA helix are responsible for those few cases in which the match was not the same (e.g., residues 132 and 134 on the template strand).

For illustrative purposes, we then extended our model to include both the primary and secondary sites present in P_{fhaB} by replacing all four BvgACTDs in our model of the DNA with CTDs bound to both primary and secondary sites with the new full length BvgA model (Fig. 9A). Using Alphafold2 predicted domains in both the primary and primary + secondary site models did not result in noticeable differences (data not shown). This model shows how the bend in the DNA, imposed by the binding of each BvgACTD dimer, will result in a large bend in the DNA when all three sites are filled.

3. Discussion

The B. pertussis two component system encoded by bvgAS is the master regulator of virulence genes [32]. Upon phosphorylation by the sensor kinase BvgS, the response regulator BvgA binds to sequences upstream of multiple promoters, activating transcription. RNA-seq analyses have shown that BvgA directly or indirectly regulates 15% of the pathogen’s genome [33], and recent ChIP-seq analyses indicate that BvgA binds to 54 known virulence activated promoters and up to 91 more suspected promoters [6]. Consequently, understanding the mechanism of BvgA activation is crucial to understanding the biology of B. pertussis.

BvgA-mediated regulation is accomplished by dimers of BvgA binding sites, whose CTDs sit upstream of the promoter at imperfect 14 bp palindromes with centers separated by 22 bp (2 helical turns) [reviewed in [432]]. BvgA binds to P-regulated promoters contain 2 or more BvgA binding sites with the primary site bearing a better match to consensus than the others. In P_{fhaB}, there are three sites, centered at −88.5, −66.5, and −44.5 (Fig. 2). The farthest...
downstream tertiary site is positioned similarly to a Class II activa-
tor [8], lying immediately adjacent to the –35 region of the core
promoter, the sequence that is bound by the major specificity sub-
unit of RNAP, in this case B. pertussis σA. The farther upstream pri-
mary and secondary sites are located at positions like those for a
Class I activator, which typically sits on the DNA helix adjacent
to a CTD of one of the two α subunits of RNAP (αCTD) [8]. However,
analyses of the transcription complex with either PfhaB or Pfim3,
BvgA/P, and RNAP, containing FeBABE-conjugated αCTD, revealed a surprising finding. Instead of being adjacent to the reg-
ulator, in BvgA regulation, an αCTD and a BvgA ∼ P occupy the same
sequence, just on different faces of the DNA helix [9,10].

A further complication arises because the number and loca-
tions of BvgA/C24P sites vary relative to the TSS. Unlike at PfhaB,
at Pfim3 the farthest downstream site is centered at –32.5, placing
this binding site directly over the /C035 region [31]. Nevertheless,
in nearly every case that has been studied, the site nearest the
/C035 is not the primary BvgA/C24P site, and in fact, in Pfim3 and Pptx,
mutation of that site to an excellent /C035 sequence for σ, elimi-
nates regulation [20,34]. For some promoters, like PfhaB and Pfim3,
3 sites are filled, while at the promoter for the pertussis toxin,
there are 6 sites \[20\]. Consequently, it has been proposed that a BvgA\(_{\text{C24P}}\) dimer binds to the primary site first, which then results in the filling of the other sites through cooperativity, generating the overall transcription complex and resulting in BvgA\(_{\text{C24P}}\) activation.

There is no direct structural information for how BvgA\(_{\text{C24}}\) binds to the DNA. However, previous FeBABE analyses using transcription complexes with RNAP and BvgA variants conjugated at position 148 or 194 within the BvgA\(_{\text{C24}}\) have revealed the inside and outside edges of binding at P\(_{\text{fhaB}}\) and P\(_{\text{fim3}}\). It was possible to determine these locations because of the sequence similarity between BvgA and the E. coli response regulator NarL and the presence of a crystal structure of a NarL\(_{\text{CTD}}\)/DNA complex \[13\]. Our work here has provided an additional FeBABE cleavage site from conjugation at the BvgA\(_{\text{CTD}}\) position 174. Together these analyses are consistent with a structure in which the binding of BvgA\(_{\text{CTD}}\) dimers mirrors what was observed in the NarL\(_{\text{CTD}}\)/DNA structure (Fig. 3), strongly suggesting that the DNA binding helix (shown in dark blue, Fig. 3C...
and Fig. 7B) resides within the major groove of the DNA, bending the BvgA binding site as seen in Fig. 3A.

Despite this understanding of how BvgA<sup>NTD</sup> interacts with the DNA, there has been no information on the location of BvgA<sup>NTD</sup> within the transcription complex. Consequently, we used the NarL<sup>FL</sup> structure to model BvgA<sup>FL</sup>, providing the information needed to select BvgA<sup>NTD</sup> and linker positions for FeBABE conjugation. This work has several conclusions. First, the AlphaFold2 structure and our independently modeled structure of BvgA<sup>FL</sup> are quite similar, suggesting that the modeled BvgA<sup>FL</sup> structure is highly probable. In both models the NTD and CTD are close and are separated by a flexible linker that could accommodate movement of the NTD relative to the CTD. Using the BvgA variant S96C/T194C, we could generate a BMB crosslink of between ~8 and 12 Å when using the free protein. This distance, which is somewhat greater than that predicted by the models, is consistent with linker flexibility. In addition, the relative association of the NTD and the CTD in solution may be in equilibrium between that seen in the predicted structure and the one that we have inferred. Consequently, the solution structure of BvgA might be at least partially in a more open form. Interestingly, this crosslink is observed both in the presence or absence of acetyl phosphate (Ac~P), indicating that phosphorylation is not required for this conformation. The second conclusion is that the BvgA<sup>FL</sup> model predicts that the position of the NTD precludes DNA binding by the CTD, even if the distance between S96 and T194 is the maximum of ~12 Å. A similar conclusion was reached from the analysis of the NarL<sup>FL</sup> structure, leading to a ‘telephone-receiver’ model in which large conformational changes must occur to enable DNA binding[12,13]. Thus, a BvgA<sup>CTD</sup> dimer cannot interact with the DNA without movement of the NTDs to either specific or random locations. Finally, footprinting analyses using FeBABE-conjugated NTD residues indicate that within the transcription complex the BvgA<sup>NTD</sup>s move to specific, not random, locations and reveal a side of the NTD that faces the DNA. If the NTD locations were either random or flexible enough to allow for several locations, the FeBABE analyses would have yielded a ‘smear’ of cuts or multiple cuts compatible with the various locations, respectively. Consequently, our data are consistent with fixed positions for the NTDs on the biological timescale. Given that the NTDs are not thought to interact with the DNA, it seems likely then that protein–protein interactions within the transcription complex, which have not yet been determined, are responsible for the NTD placements.

Prior homology modeling [35] and AlphaFold2 [36] validation studies have established that our approaches should yield 3D models of BvgA<sup>CTD</sup> and BvgA<sup>NTD</sup> and the BvgA<sup>CTD</sup>/DNA complex, in isolation, rivaling atomic crystallographic accuracy. The same cannot be said of the linker region or the two-site model of P<sub>phad</sub> that we have suggested. Consequently, these are speculative models. However, several analyses have previously used FeBABE footprinting to correctly reveal how a protein is positioned relative to DNA [9,10,22,24–31]; in fact, our FeBABE-derived prediction for how the T4 transcriptional regulator MotA interacts with its site on the DNA [30] was fully validated by the subsequent crystal structure [37]. It is highly likely then that the NTD patch that we have identified here by FeBABE analysis is the region that faces the DNA. Furthermore, the combination of poor sequence conservation in the linker between NarL and BvgA, the FeBABE observations on the NTD, and our finding of unclashed novel conformations of the NTD relative to the CTD in single and two-site models that are consistent with the FeBABE data strongly suggests that the linker is flexible and that the NTD rearranges to a new location within the transcription complex.

Thus, from this work, we can propose a speculative model for how binding at all three P<sub>phad</sub> binding sites might work (Fig. 5B). In this scenario, a dimer of BvgA<sup>CTD</sup>s occupies each of the three binding sites, requiring the movement of the NTDs to new locations in the transcription complex as compared to their locations in NarL<sup>FL</sup>. Because our analyses cannot specify which NTD produces which FeBABE cuts, there is more than one possibility for which NTD is connected to which CTD by the linker; all of these are acceptable given the length and flexibility of the linker. Furthermore, because of the predicted bending of the DNA from BvgA binding, we posit that the upstream sites could possibly bend back toward the RNAP/core promoter region. It is also possible that the positions of the RNAP σ<sup>CTD</sup>s, at the primary and secondary sites, may interact with the BvgA<sup>NTD</sup>s. Although the accuracy of this transcription model awaits experimental visualization, our work further demonstrates how FeBABE footprinting can provide useful information for future attempts to visualize the architecture of protein/DNA complexes.

4. Materials and Methods

4.1. BvgA homology models

The initial alignment between BvgA and NarL was obtained from ICM global alignment. The alignment was subsequently modified to allow relaxation of segments with no secondary structures as well as sections with low sequence similarity. Soft tethers were used in segments with low similarity, but retention of the overall location in space of the segment was restricted. Segments found as indels in the final alignment were considered loops and were fully modeled with Montecarlo with the number of calls depending on the number of residues heuristic. Verification was done using the Structural Analysis and Verification Server (SAVS; [https://www.doe-mbi.ucla.edu/saves/]) running the programs ERRAT, Procheck, and Whatcheck. Procheck and Whatcheck were used to analyze the overall geometry of the whole structure and on a residue-by-residue basis. ERRAT was used to analyze the non-bonded interactions of each residue and average the score on a window of nine residues.

Using the NarL monomer structure [112]PDB:1RLN] as the template, the BvgA full length homology model was then constructed by modifying the initial alignment between BvgA and NarL obtained from ICM global alignment to allow the relaxation of segments with no secondary structure as well as section with low sequence similarity. Soft tethers were used in segments with low similarity but retention of the overall location in space of the segment was desired. At the end of the modeling, the largest loop (residues Y122 to N151) resulted in 2 conformations close in energy, indicating that the loop is found in transition between these two conformations. Both loop conformations are shown in Fig. S2A.

The initial model of the BvgA<sup>CTD</sup>/DNA complex was generated using the NarL<sup>CTD</sup> dimer/DNA structure [113], PDB:1EBB] as a template. Distance restraints were put to the carbon backbone of the template and these were minimized until the model sequence was fully overlapping the template. Subsequently, a minimization of the structure was made to relax the strain introduced with the mutations in the sequence. Another model was made, this time using an extended DNA sequence. B form DNA was used as the template (D. Hinton, personal communication). Given that the extended segments were not identical in sequence to any segments in this B form DNA, the closest matches were used and the non-identical nucleotides were restrained only in the carbon backbone. Similarly, the structure was minimized to relax the molecule to account for the non-identical nucleotides from the B form DNA. The mutant sequence and the extended sequence were attached to make a single sequence. Finally, the BvgA<sup>CTD</sup> dimers
were modeled onto the extended DNA sequence of \( P_{\text{fhaB}} \) as shown in Fig. 9.

For final modeling, we used our extended model of BvgA\(^{\text{CTD}} \) dimers on \( P_{\text{fhaB}} \). We aligned the cut sites derived from the FeBABE footprinting onto the DNA to match the corresponding residues on BvgA. Our model suggests that the BvgA dimers asymmetrically deviate and flank either side of the DNA during transcription initiation.

4.2. Bacterial strains and cell culture

E. coli BL21(DE3)/pLysE (Novagen, Madison, WI, USA) was used to overexpress WT BvgA and BvgA variant proteins. Cells were grown in LB at 37°C in LB broth with 100 \( \mu \)g/ml of carbenicillin and 25 \( \mu \)g/ml chloramphenicol.

4.3. DNA

The construction of a `BvgA\(^{\text{cysless}} \)` plasmid, which encodes BvgA\(^{\text{C93A, C103A}} \) in the pET21a (Novagen) background, has been described [10]. The BvgA-single-cysteine derivatives were synthesized using this BvgA\(^{\text{cysless}} \) plasmid and the Q5 site-directed mutagenesis kit and protocol therein (New England Biolabs, Inc., Ipswich, MA, USA). The plasmid containing the S96/T194 variant was constructed starting with the S96 variant plasmid, again using Q5 site-directed mutagenesis.

DNA sequence analyses were performed by Macrogen (Rockville, MD, USA) and pTE103-P\(_{fhaB}\) DNA was sequenced on the extended DNA sequence of F. coli fhaB as shown in Fig. 9.

4.4. Proteins

B. pertussis RNAP core, primary B. pertussis \( \sigma^{(\alpha)} \), WT BvgA, and the BvgA variants were purified as described [5,34]. When indicated, BvgA was phosphorylated by incubation in the presence of 20 mM acetyl phosphate (lithium potassium acetyl phosphate; Sigma-Aldrich, made as a 200 mM solution in 20 mM Tris-Cl (pH 8)) for 30 min at room temperature.

Non-phosphorylated BvgA was incubated similarly in the presence of Tris-Cl (pH 8) alone.

4.5. Crosslinking and mass spectrometry

After dialysis into BvgA Storage Buffer lacking 1,4-dithiothreitol (DTT) [20 mM HEPES-NaOH, 50 mM KCl, 1 mM MgCl\(_2\), 45 % (vol/vol) glycerol (pH 7.4)], WT BvgA or BvgA\(^{\text{cysless}} \) S96C/T194C (875 pmol in 360 \( \mu \)L) was incubated at room temperature for 30 min in the presence of 40 \( \mu \)L 20 mM HEPES-NaOH (pH 7.4) with and without 200 mM acetyl phosphate to generate phosphorylated and non-phosphorylated protein, respectively. For each 40 \( \mu \)L of the protein solution, 5 \( \mu \)L of dimethyl sulfoxide (DMSO) with or without 2 mM BMB (Thermo Fisher) was added, and the solution was incubated for 2 h at room temperature. Reactions were quenched by the addition of 5 \( \mu \)L 0.2 M DTT.

After separation of monomer and dimer species by SDS-PAGE on a 4–12 % Bis-Tris gel (Thermo Fisher), specific crosslinked protein bands were in-gel digested with trypsin for 16 h at 37°C as described [39]. The extracted peptides were dried and lyophilized prior to mass spectrometry analysis. Each sample was separated on a 75 \( \mu \)m \( \times \) 15 cm, 2 \( \mu \)m Acclaim PepMap reverse phase column (Thermo Fisher) using an UltiMate 3000 RSLCnano HPLC (Thermo Fisher) at a flow rate of 300 nL/min, followed by online analysis by tandem mass spectrometry using a Thermo Orbitrap Fusion mass spectrometer. Peptides were eluted into the mass spectrometer using a 120 min gradient. Parent full-scan mass spectra were collected in the Orbitrap mass analyzer set to acquire data at 120,000 FWHM resolution; ions were then isolated in the quadrupole mass filter, fragmented within the HCD cell (HCD normalized energy 35 %, stepped \( \pm 3 \) %), and the product ions analyzed in the Orbitrap at 30,000 resolution. Raw files were converted to MGF format in Proteome Discoverer 2.4 (Thermo Fisher), and then cross-linked peptides were identified using the StavroX tool [40] in the MeroX (v. 2.0.1.4) package [41].

4.6. In vitro transcriptions

Single round in vitro transcriptions were performed as follows using previously published procedures [31]. BvgA \( \sim P \) (5 pmol) was phosphorylated by treatment of BvgA with 20 mM acetyl phosphate (previously dissolved as 200 mM in 20 mM Tris-Cl, pH 8.0) for 30 min at room temperature. RNAP was reconstituted at 37°C for 10 min using a \( \sigma^{(\alpha)} \) core ratio of 2:1. Reaction transcriptions were assembled (9 \( \mu \)L total) with 0.05 to 0.1 pmol DNA, 5 pmol BvgA \( \sim P \), 20 mM acetyl phosphate, 0.75 mM RNAP core, 1.5 mM \( \sigma^{(\alpha)} \), and 1.1 X Transcription Buffer [1 X concentration: 50 mM NaCl, 50 mM Tris-HCl (pH 8.0), 3 mM magnesium acetate, 0.1 mM EDTA (pH 8.0), 0.1 mM DTT, and 25 \( \mu \)g/mL BSA], and the solutions were incubated at 37°C for 15 min. Single round transcription was initiated by the addition of a 1 \( \mu \)L solution containing 2.5 mM ATP, CTP, and GTP; 0.25 mM [\( \gamma^{32} \)P] UTP (\( \sim 4000 \) dpm/pm mole); and 2 mg/mL heparin. Reactions were incubated for 10 min at 37°C, quenched by the addition of loading solution, and collected on dry ice; RNA products were electrophoresed on 4 % wt/vol polyacrylamide, 7 M urea denaturing gels as described [9]. Gels were imaged by autoradiography followed by scanning with a GS800-calibrated Densitometer (Bio-Rad). Quantification was performed using Quantity-One software (Bio-Rad).

4.7. FeBABE conjugation

For FeBABE conjugation, 1 nmol of BvgA was first dialyzed against BvgA Storage Buffer lacking DTT at 4°C for 2 h. Samples were collected and FeBABE (Dojindo, Rockville, MD, USA; 20 mM in DMSO) was added as a ratio of 0.5 \( \mu \)L per 170 pmol of protein. Samples were incubated at 37°C for 1 h and then dialyzed against BvgA Storage Buffer containing 1 mM DTT at 4°C for 2 h. Protein concentration and % conjugation were determined by SDS-PAGE using a known amount of WT BvgA as the standard.

4.8. FeBABE Footprinting

Footprinting reactions were done as previously described [31]. Briefly, 0.05 pmol of labeled DNA, 5 pmol of FeBABE-conjugated protein (BvgA \( \sim P \) or BvgA), and reconstituted B. pertussis RNAP (0.75 pmol core, 1.5 pmol of sigma) were assembled as described for the in vitro transcription reactions (final volume of 9 \( \mu \)L) and incubated at 37°C for 15 min. Heparin (1 \( \mu \)L, 2 mg/mL) was added for 1.75 sec before rapid sequential addition of 2 \( \mu \)L 100 mM
sodium ascorbate and 2 μL 0.6 % hydrogen peroxide or 4 μL water as the control. Samples were returned to 37°C for 10 min and then quenched by the addition of 80 μL 100 mM thiourea (at room temperature) and placed on ice. After addition of 100 μL of room temperature 1 X TE (pH 8.0), 0.5 μL 1 mg/ml calf thymus DNA, and 2 μL GlycoBlue™ coprecipitant (Thermo Fisher), reactions were phenol extracted with phenol:chloroform:isoamyl alcohol (25:24:1: 200 μL), the aqueous phase was collected, and DNA was ethanol precipitated. DNA products were resolved by electrophoresis on denaturing 6 wt% vol polyacrylamide, 7 M urea sequencing gels. Ladder lanes (GA) were obtained from G + A reactions [42] using the labeled P
\text{shuttle} DNA. Gels were visualized using autoradiography, scanned using a GS800 calibrated Densitometer (Bio-Rad), and analyzed with Quantity-One software (Bio-Rad). Gels shown in Fig. 6 represent one of 2 or more independent reactions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (D.K., J.T., K.C., R.N., A.C.-M., L. K., K.M., and D.M.H.); by NIH grants DP2OD004631 from the Office of the Director and RC2LM010994 from the National Library of Medicine (M.B.F. and T.C.); by the Intramural Research Program of the National Institutes of Health, National Cancer Institute (L. M.M.); and by the Food and Drug Administration, Center for Biologies Evaluation and Research (Q.C. and S.S.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.10.042.

References

[1] Prevention, C. I. D. C. a. (2012) Pertussis (Whooping Cough): Pertussis Cases by Year (1922–2012). www.cdc.gov/pertussis/surv-reporting/cases-by-year.html.
[2] Cummings CA, Bootes MJ, Rabel DA, Miller JF. Species- and strain-specific phosphorylation dynamics of the BvgASR virulence regulon of Bordetella pertussis. Curr Opin Microbiol 2017;43:52–9.
[3] Lacey BW. Antigenic modulation of Bordetella pertussis. J Hyg (Lond) 1960;58:57–93.
[4] Decker KB, James TD, Meares CF, Fujita N, Ishihama A, Bussy SJ, et al. Verification of protein structures: patterns of nonbonded interactions. Protein Sci 2019;28:1387–409.
[5] Marr MT, Roberts JW, Brown SE, Klee M, Gussin GN. Interactions among CII protein, RNA polymerase and the lambda PRE promoter. contacts between RNA polymerase and the -35 region of PRE are identical in the presence and absence of CII protein. Nucleic Acids Res 2004;32:1083–90.
[6] King JM, Brutinel ED, Marsden AE, Schubot FD, Yahr TL. Organization of open complex at Escherichia coli promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem 1999;274:2263–70.
[7] Owos J, Chmura AJ, Murakami K, Fujita N, Ishihama A, Meares CF. Mapping the promoter DNA sites proximal to conserved regions of sigma 70 in an Escherichia coli RNA polymerase-lactV5 open promoter complex. Biochemistry 1998;37:7670–7.
[8] Owos J, Miyake R, Murakami K, Chmura AJ, Fujita N, Ishihama A, et al. Mapping the sigma70 subunit contact sites on Escherichia coli RNA polymerase with a sigma70-conjugated chemical probe. Proc Natl Acad Sci U S A 1995;92:5601–6.
[9] Marr DM, Dorayler SY, Meares CF, Roberts JW. Restructuring of an RNA polymerase holoenzyme elongation complex by lambda phage Q proteins. Proc Natl Acad Sci U S A 2001;98:8972–7.
[10] Chen Q, Stibitz S. The BvgASR virulence regulon of Bordetella pertussis. Curr Opin Microbiol 2019;47:74–81.
[11] Mook K, Bencorop RA, Kim DD, Chen Q, Wade JT, Stibitz S, et al. The BvgAS Region of Bordetella pertussis. MBio 2018;9:1156–72.
[12] Decker KB, Hinton DM. Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. Annu Rev Microbiol 2013;67:113–39.
[13] Lee DJ, Minchin SD, Busby SJ. Activating transcription in bacteria. Annu Rev Microbiol 2012;66:125–52.
[14] Decker KB, Chen Q, Hsieh ML, Boucher P, Stibitz S, Hinton DM. Different requirements for sigma Region 4 in BvgA activation of the Bordetella pertussis promoters P
\text{fim3} and P
\text{shuttle}. J Mol Biol 2011;409:692–709.
[15] Boucher PE, Maris AE, Yang MS, Stibitz S. The response regulator BvgA and RNA polymerase alpha subunit C-terminal domain bind simultaneously to different faces of the same segment of promoter DNA. Mol Cell 2003;11:163–73.
[16] Abagyan RA, Batalov S. Do aligned sequences share the same fold? J Mol Biol 2001;309:2755–68.
[17] Baikov I, Schroeder L, Kaczor-Grezskowiak M, Grezskowiak K, Gunsalus RP, Dickerson RE. Structure of the Escherichia coli response regulator NarL. Biochemistry 1996;35:11653–61.