Crystal Structure of the Virulence Gene Activator AphA from Vibrio cholerae Reveals It Is a Novel Member of the Winged Helix Transcription Factor Superfamily*

Rukman S. De Silva‡, Gabriela Kovacicova§, Wei Lin§, Ronald K. Taylor§, Karen Skorupski‡, and F. Jon Kull¶

From the ‡Department of Chemistry, Dartmouth College and §Department of Microbiology and Immunology, Dartmouth Medical School, Hanover, New Hampshire 03755

AphA is a member of a new and largely uncharacterized family of transcriptional activators that is required for initiating virulence gene expression in Vibrio cholerae, the causative agent of the frequently fatal epidemic diarrheal disease cholera. AphA activates transcription by an unusual mechanism that appears to involve a direct interaction with the LysR-type regulator AphB at the tcpPH promoter. As a first step toward understanding the molecular basis for tcpPH activation by AphA and AphB, we have determined the crystal structure of AphA to 2.2 Å resolution. AphA is a dimer with an N-terminal winged helix DNA binding domain that is architecturally similar to that of the MarR family of transcriptional regulators. Unlike this family, however, AphA has a unique C-terminal antiparallel coiled coil domain that serves as its primary dimerization interface. AphA monomers are highly unstable by themselves and form a linked topology, requiring the protein to partially unfold to form the dimer. The structure of AphA also provides insights into how it cooperates with AphB to activate transcription, most likely by forming a heterotetrameric complex at the tcpPH promoter.

Cholera is a frequently fatal epidemic diarrheal disease caused by oral ingestion of food or water contaminated with the bacterium Vibrio cholerae. The two primary virulence factors responsible for the disease are the toxin-coregulated pilus (TCP), a critical colonization factor (1), and cholera toxin, which causes a copious diarrhea that can quickly lead to severe dehydration and death (2). The expression of these genes from the Vibrio pathogenicity island (3) and the lysogenic cholera toxin phage (4), respectively, are dependent upon a pair of transcriptional regulators, AphA and AphB, which are encoded by genes not physically associated with each other or with these pathogenicity islands in the V. cholerae genome (5, 6).

* This work was supported by National Institutes of Health Grants AI060031 (to F. J. K.), AI41558 (to K. S.), and AI39654 (to R. K. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1YG2) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

† To whom correspondence should be addressed: Dept. of Chemistry, 6125 Burke Laboratory, Dartmouth College, Hanover, NH, 03755. Tel.: 603-646-1552; Fax: 603-646-3946; E-mail: fjon.kull@dartmouth.edu.

¶ The abbreviations used are: TCP, toxin-coregulated pilus; MES, 2-(N-morpholino)ethanesulfonic acid; CNS, crystallography and NMR system.

AphA is a member of a new and largely uncharacterized regulator family comprising at least 30 proteins with mostly unknown functions that show homology to PadR, a repressor that controls the expression of genes involved in the detoxification of phenolic acids (7). AphA activates the transcription of the tcpPH promoter on the Vibrio pathogenicity island by an unusual mechanism that appears to require a direct interaction with the LysR-type regulator AphB (8). Upon binding of AphA to a region of partial dyad symmetry at the tcpPH promoter (between −101 and −71, relative to the transcriptional start site), the protein enhances the binding of AphB to an adjacent and proximal site that lies between −78 and −43 (9, 10). Under the appropriate environmental conditions, this results in activation of the tcpPH promoter and initiates a transcriptional cascade that culminates in the production of TCP and cholera toxin. An important aspect of virulence gene regulation is that the expression of aphA is controlled by a quorum-sensing system that decreases its intracellular levels at high cell density (11, 12). This regulation reduces virulence gene expression at high cell density and is thought to contribute to the self-limiting nature of V. cholerae infections.

To more fully understand the molecular mechanisms involved in virulence gene activation in V. cholerae, we have determined the 2.2-Å x-ray crystal structure of the AphA dimer (Protein Data Bank code 1YG2). Each AphA subunit consists of an N-terminal DNA binding domain that adopts a winged helix fold similar to that of the multiple antibiotic resistance repressor MarR (13) and a distinctive C-terminal dimerization domain comprising an extensive antiparallel coiled coil. The structure of AphA allows a number of predictions to be made regarding how the protein binds to DNA and how it interacts with AphB to initiate transcriptional activation of the virulence cascade.

MATERIALS AND METHODS

Protein Preparation and Crystallization—The purification of AphA using the IMPACT-CN protein fusion and purification system (New England Biolabs) has been described previously (14). Briefly, Escherichia coli strain ER2566, expressing pWEL18, was grown in LB at 30 °C to an A600 of ~0.8, induced with 0.5 mM isopropyl β-D-1-thio-galactopyranoside, and incubated at 16 °C for 20 h. AphA was purified from sonicated, clarified supernatant using a chitin column (New England Biolabs), dialyzed overnight in 20 mM Tris-HCl, pH 7.9, 1 mM EDTA, 10 mM sodium Iodoacetate, and concentrated to 5 mg/ml using Centricon filters (Millipore). AphA crystals were obtained by mixing equal volumes of the protein, as described above, with 0.1M MES, pH 6.3, and 1.3M magnesium sulfate. Crystals appeared after 3–4 days of incubation at room temperature. A useful heavy atom derivative was obtained by soaking the native crystals in 0.01 mM methyl mercuric chloride for 12 h. For both native and derivative crystals, 20% glycerol was used as the cryoprotectant.

Data Collection and Structure Solution—Native and anomalous de-
rivative data were collected on a Mar345db image plate detector mounted on a Rigaku rotating anode generator operating with a CuK$_\alpha$ target at 50 kV and 100 mA. The data were processed using the program XDS (15) and resulted in a complete native data set to 3.0 Å and a derivative data set to 2.6 Å (Table I). Patterson maps, calculated with CNS (16), clearly indicated two mercury peaks, and single isomorphous replacement and single anomalous dispersion phases were combined using the CNS program to produce an initial 3.0 Å electron density map. The quality of the map was excellent, allowing a complete polyalanine chain to be traced without difficulty. The presence of a number of tryptophan side chains allowed placement of the AphA sequence into the map. A 2.2 Å native data set was collected at National Synchrotron Light Source beamline X8A, and the data were processed in space group C222$_1$. Following molecular replacement using the program CNS to place the existing model into the new data set, multiple rounds of refinement were carried out with CNS, resulting in a final structure having an R-factor of 23.3% and a free-R of 28.0%. Analysis of the structure using the program PROCHECK showed 94.1% of residues in the most favored regions, 5.2% in additional allowed regions, and no residues in generously allowed or disallowed regions. The model was also checked by a composite omit map calculated in CNS, which showed good density for all amino acids, with the exception of the two N-terminal residues, the two C-terminal residues, the side chains of Tyr-19 and Lys-23, and all of residues 58–67, located in the wing, which are disordered.

**RESULTS AND DISCUSSION**

**Overall Structure and Comparison with Other Winged Helix Proteins**—The AphA protein forms a dimer in which the two subunits are related by a crystallographic two-fold rotation axis. Each molecule of AphA is 70 Å long × 30 Å wide × 40 Å deep, resulting in a dimer with overall dimensions of 90 × 30 × 40 Å. Each α/β subunit contains seven helices and two β-strands and folds into two domains: a globular N-terminal DNA binding domain made up of amino acids 1–86, and a C-terminal coiled coil dimerization domain including amino acids 98–179 (Fig. 1A). These two domains are connected between helices α$_4$ and α$_5$ by an extended, eleven-amino-acid polypeptide linker.

Although AphA does not have any homologs of known structure, its N-terminal end has a conserved domain architecture similar to that of the Vibrio cholerae Virulence Activator AphA. Consistent with this, a DALI (18) search of known crystal structures shows MarR (Fig. 1B) to be the closest structural relative to AphA (Z = 11.4), with 91 α carbons superposing with an root mean square deviation of 4.0 Å. Although the overall sequence identity between AphA and MarR is low (11%), an alignment of the two proteins (Fig. 1C) reveals analogous secondary structure in the DNA binding domain that produces a similar fold in the crystal structures (compare Fig. 1, A with B). These analogous structural elements include AphA α$_1$ and MarR α$_2$. AphA α$_2$ and MarR α$_3$, AphA α$_3$ and MarR α$_4$, as well as the two β-strands that form the wing between helices α$_3$ and α$_4$ in AphA and between helices α$_4$ and α$_5$ in MarR. The similarity between AphA and MarR appears to be limited to the DNA binding domain, however, because AphA lacking the N-terminal helix α$_1$ that is important for MarR dimerization and MarR lacking the C-terminal helix α$_7$ that is important for AphA dimerization. In addition, the MarR helix α$_5$, which is analogous to the AphA helix α$_4$, is much longer and contributes to the dimerization interface, whereas the shorter MarR α$_6$ has only a minor role in dimerization.

**Phasing**

**Resolution range (Å)***

- Native (in-house): 20–30
- Native (synchrotron): 20–2.2

**Completeness (%)***

- Native (in-house): 99.8 (100.0)
- Native (synchrotron): 99.6 (99.2)

**R$_{meas}$ (%)***

- Native (in-house): 11.0 (13.0)
- Native (synchrotron): 8.1 (26.1)

**Root mean square deviation (Å)**

- Bond (Å): 0.0061
- Angles (°): 1.01

**Average B-factors (Å$^2$)**

- 51.4

* $R_{meas}$ = $100 \times \sum_h |I_h| - I_o|/\sum_h |I_o|$, where $h$ is unique reflection indices.
In forming a dimer, the two AphA chains literally wrap around each other (Fig. 2C). The extended linker from one subunit leaves one side of the DNA binding domain and then runs along the underside of the dimerization domain of the second subunit until reaching helix 5. The helix 5 (yellow) is much longer than AphA α4 and forms the main dimerization interface. The shorter MarR helix α6 (orange) has only a minor role in dimerization. Secondary structural elements, as determined from the crystal structures, are shown above and below the sequences (the first six residues of MarR were disordered). H, α helices; S, β-strands; ., residues not visualized in the crystal structure. Identical residues between the two are shown in red. Bold characters in the AphA sequence show positions of mutations described in the text that influence its DNA binding and dimerization activities. Homologous α-carbons within the DNA binding domain that superpose according to the program DALI (18) include: AphA (amino acids 1–14) with MarR (Amino acids 33–46), AphA (amino acids 15–26) with MarR (amino acids 48–59), AphA (amino acids 35–83) with MarR (amino acids 60–108), AphA (amino acids 84–87) with MarR (amino acids 111–114), and AphA (amino acids 88–91) with MarR (amino acids 117–120).

In forming a dimer, the two AphA chains literally wrap around each other (Fig. 2C). The extended linker from one subunit leaves one side of the DNA binding domain and then runs along the underside of the dimerization domain of the second subunit until reaching helix 5. The 5 helix of the first subunit packs against the top center of the DNA binding domain of the second subunit and leads into the coiled coil formed by helices 6 and 7. This pair of helices then extends back toward the first subunit DNA binding domain, where there is a minor hydrophobic interaction between Tyr-31 of the DNA binding domain and Leu-154, Tyr-129, and Ile-132 of the dimerization domain. Given the rather open and extended structure of the monomer, it is unlikely to be stable by itself.

The AphA dimerization domain differs substantially from that observed in the MarR dimer (Fig. 2D) in which the ends of helices α1 and α5 pack together so that the helical pairs from the two subunits cross in an almost perpendicular orientation, rather than the four relatively flat and parallel helices in the AphA dimerization domain (Fig. 2A). In addition, unlike AphA, the subunits in the MarR dimer are not topologically entwined. Comparison of the AphA dimerization domain with all of the known structures of winged helix proteins listed at the structural classification of proteins website (20) shows the AphA dimerization domain topology to be unique in the winged helix superfamily. A DALI (18) search also indicates that the dimerization domain of AphA is not similar to any other members of this protein superfamily.

An interesting and rare interaction observed in the AphA dimer is an edge-on polar bond between His-89 in the middle of the linker of one monomer and the electron ring of Trp-164 of the other monomer. Another feature of the dimer is the presence of a pocket that is formed by helices α1 and α2 from one subunit and helix α5 of the second subunit. The protein chains are poorly packed in this region and there appears to be a solvent-accessible tunnel leading −15 Å into the dimer. The interface between the two monomers in this region is nonspecific. As it is the only one forming direct contacts between the DNA binding region of one monomer and the dimerization domain of the other, this loose interaction may allow the DNA binding domains to swing away from the dimerization domain to form a tight binding interface with the DNA. It could also serve as a potential site of interaction with AphB.
DNA Binding Domain—As discussed in the first section, the DNA binding domain of AphA forms a winged helix fold. The winged helix superfamily contains at least 36 structural families of proteins (20) that all contain at least one winged helix domain but have highly divergent structures in other regions of the protein. These families include: the MarR (13) family, containing members such as MexR (13), SlyA (21), SarR, and SarS (22); the CAP family; and the LysR family (23, 24), among many others. As the winged helix superfamily is divided into families by structural similarities in regions of the protein outside of the DNA binding domain, it appears that AphA is the founding member of the 37th structural family of this large superfamily. In all winged helix proteins, the helix-turn-helix motif is followed by a wing composed of two antiparallel β-strands. In AphA, helices α2 and α3 form the helix-turn-helix with the wing following helix α3. The loop linking helices α2 and α3 contains a stretch of three amino acids in an α-helical geometry, labeled as α2’ in Fig. 1C. Although a third β-strand is often present in winged helix proteins, AphA contains only a small interaction of two β-strand-like hydrogen bonds between the wing strand β2 and Ala-16 of the loop connecting helices α1 and α2.

The importance of helices α2, α3, and the wing for the DNA binding activity of AphA has been confirmed by mutational analyses (8). For example, dominant negative mutations, which typically define amino acids important for DNA binding activity, have been isolated in α2 (G18R, Y19C, and G30D), α3 (H37E, Q39R, and Y41H), and the wing (R65E, R66E, and K67E) (see Fig. 1C). Dominant negative mutations in MarR have also been isolated within the helix analogous to AphA α3 (MarR α4) and the wing (19), supporting the functional similarity of these two proteins. In addition, this region of AphA contains a number of positively charged residues as has been observed in other winged helix proteins (19, 21, 22).

Interactions with DNA—Using existing structures of winged helix protein-DNA complexes, a model of the AphA structure in complex with DNA was generated (Fig. 3A). In this model, the wing packs along the side of the DNA, spanning the phosphate backbones of both strands on either side of the minor groove. This arrangement allows both nonspecific contacts with the phosphate backbone as well as specific interactions between side chains from the wing and bases in the minor groove. Major groove interactions would primarily involve helix α3. In our

FIG. 2. Structure of the AphA and MarR dimers. A, the AphA dimer viewed along the crystallographic two-fold axis. In this top view, the DNA binding face is underneath the dimer. For one molecule, the N-terminal DNA binding domain is green, and the C-terminal dimerization domain is red. For the second molecule, the DNA and dimerization domains are blue and orange, respectively. Unresolved residues in the wing motif are shown as a dashed line. B, electrostatic surface potential of the AphA dimer, oriented as in A. The complementary positive (blue) and negative (red) charge clusters that stabilize the dimer interface can be seen at both ends of the α7 helices. C, the AphA dimer viewed from the side, colored as in A. The DNA binding surface is at the bottom in this orientation. D, side view of the MarR dimer. The DNA binding and dimerization domains are colored the same as for AphA.

FIG. 3. Models of protein-DNA complexes. A, a model of the AphA dimer (one chain in blue and the other in green) bound to its DNA site with the complete AphA wing modeled in, based on the wing structure observed in MarR. The primary DNA binding interface is predicted to involve helix α3 from each monomer (orange), which fit into adjacent major grooves on the same face of the DNA. As the helices are not optimally spaced, some conformational change in the protein and/or DNA likely takes place. B, model of the AphA-AphB heterotetramer bound to DNA. The AphB dimer structure was modeled by threading the AphB sequence into the crystal structure of CbnR and then placing it next to the AphA-DNA model shown in A. In this orientation, the C-terminal domain of one molecule of AphB (red) would be able to interact with the AphA dimer (green and blue). Interaction between AphA and AphB could also form between the tips of the wings. This model predicts a linear mode of binding that would not significantly distort the DNA in which four adjacent major grooves are bound by the AphA dimer and the AphB dimer.
model, the second molecule in the dimer is oriented such that it sits above the second DNA binding site, related to the first molecule by a two-fold rotation axis. In this orientation, the AphA dimer would sit on one face of the DNA in a manner similar to the model proposed for MexR (25), with the dimer spanning ~26 nucleotide base pairs, consistent with the results of DNaseI footprinting (9).

In order for both subunits to pack against the DNA in a manner identical to that observed for other winged helix proteins, it appears that a small amount of structural rearrangement is necessary, as the recognition helices of the two AphA subunits are too far apart (by ~6–8 Å) to fit symmetrically into the DNA binding sites. This rearrangement likely involves either a slight bending of the DNA, movement of the AphA DNA binding domains slightly toward each other, or both. It is also possible that neither the DNA nor the AphA dimer is distorted, in which case AphA would bind in an asymmetric fashion, as has been observed in the E. coli Rob transcription factor (26).

**Implications for Transcriptional Activation with AphB**—In its ability to dimerize and bind to DNA, AphA must cooperate with the LysR-type regulator AphB to activate tcpPH transcription (6). A variety of evidence suggests that AphA enhances the ability of AphB to bind to DNA by an unusual mechanism that involves direct interaction with it (8, 9, 12). For example, certain mutants of AphA unable to bind to DNA on their own have been found to be rescued for DNA binding in the presence of AphB. Additionally, insertion of half a helical turn between the AphA and AphB binding sites blocks this interaction by shifting the proteins to opposite faces of the DNA and prevents transcriptional activation. Once stably bound, it appears that AphB has the primary role in activation. The AphA structure presented here, together with the recent report of the first crystal structure of a full-length LysR-type transcriptional regulator CbnR from Ralstonia eutropha (27), suggests a model for this interaction (see Fig. 3B).

The DNA binding domain of CbnR is composed of three α helices and two β-strands that form a winged helix-turn-helix motif. Modeling studies indicate the recognition helix (α3) of CbnR contacts the major groove of the DNA. The primary dimerization interface of CbnR, which forms a biological tetramer, involves helix α4 from each subunit. This CbnR dimer of dimers also forms complex interactions between the four C-terminal domains and is predicted to bend DNA upon binding. Because a variety of evidence suggests that AphB is a biological dimer, its apparent interaction with AphA at tcpPH raises the possibility that the AphA-dimer-AphB-dimer complex forms a heterodimer of dimers when bound to DNA analogous to the homodimer of dimers formed by CbnR. In this model (Fig. 3B), the C-terminal domain of at least one subunit of the AphB dimer would interact with an AphA dimer bound at the adjacent promoter site. A putative interaction site could include the solvent-exposed tunnel on AphA that is oriented toward AphB. Such an interaction with AphA would greatly stabilize AphB binding by providing a second binding site, in effect anchoring it on the DNA in a ternary complex. This dimer-of-dimers would be a novel evolutionary variation on the theme of tetrameric LysR-type regulators.

**Concluding Remarks**—AphA is the first of a new family of winged helix transcriptional regulators to be crystallized. Its structure contains a distinctive C-terminal antiparallel coiled coil that is involved in self-dimerization. Although AphA is capable of binding to DNA on its own, it is unusual in its requirement that it interact with a second protein to activate transcription. As it appears that the Vibrio pathogenicity island was acquired during the evolution of V. cholerae to the pathogenic form of the organism, AphA may have evolved to interact with and stabilize AphB on the DNA, thereby promoting virulence gene expression. Future structural investigations will shed light on the nature of the interactions between AphA and AphB that facilitate transcriptional activation and the initiation of virulence gene expression.

**Acknowledgment**—In-house diffraction facilities were supported by HHMI award 76290-560801 to Dartmouth Medical School.

**REFERENCES**

1. Taylor, R. K., Miller, V. L., Furlong, D. B., and Mekalanos, J. J. (1997) **Proc. Natl. Acad. Sci. U. S. A.** 94, 2833–2837
2. Kaper, J. B., Morris, J. G., Jr., and Levine, M. M. (1995) **Clin. Microbiol. Rev.** 8, 48–86
3. Karaiskos, D. K., Johnson, J. A., Bailey, C. C., Bordeker, E. C., Kaper, J. B., and Reeves, P. H. (1998) **Proc. Natl. Acad. Sci. U. S. A.** 95, 3134–3139
4. Walder, M. K., and Mekalanos, J. J. (1996) **Science** 272, 1910–1914
5. Skorupski, K., and Taylor, R. K. (1999) **Mol. Microbiol.** 33, 763–771
6. Kovacikova, G., and Skorupski, K. (1999) **J. Bacteriol.** 181, 4250–4256
7. Bartelmez, L., Lecomte, B., Divies, C., and Cavin, J. F. (2000) **J. Bacteriol.** 182, 6724–6731
8. Kovacikova, G., Lin, W., and Skorupski, K. (2004) **Mol. Microbiol.** 53, 129–142
9. Kovacikova, G., and Skorupski, K. (2001) **Mol. Microbiol.** 41, 393–407
10. Kovacikova, G., and Skorupski, K. (2002) **Mol. Microbiol.** 44, 533–547
11. Miller, M. B., Skorupski, K., Lenz, D. H., Taylor, R. K., and Bassler, B. L. (2002) **Cell** 110, 305–314
12. Kovacikova, G., and Skorupski, K. (2002) **Mol. Microbiol.** 46, 1135–1147
13. Aleksunsh, M. N., Levy, S. B., Mealy, T. R., Seaton, A. B., and Head, J. F. (2001) **Nat. Struct. Biol.** 8, 710–714
14. Kovacikova, G., Lin, W., and Skorupski, K. (2003) **J. Bacteriol.** 185, 4825–4836
15. Rabie, W. W. (1999) **J. Appl. Crystallogr.** 26, 795–800
16. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) **Acta Crystallogr. D54**, 905–921
17. Deleted in proof
18. Holm, L., and Sander, C. (1996) **Science** 273, 595–603
19. Aleksunsh, M. N., Kim, Y. S., and Levy, S. B. (2000) **Mol. Microbiol.** 35, 1394–1404
20. Murphy, J. B., Brenner, S. E., Weinberg, A. L., and Levy, S. B. (1998) **J. Mol. Biol.** 272, 530–540
21. Wu, R. Y., Zhang, R. G., Zagnitko, O., Dementieva, I., Maltzev, N., Watson, J. D., Laskowski, R., Gornicki, P., and Joachimiak, A. (2000) **Acta Crystallogr. D53**, 4825–4836
22. Huffman, J. L., and Brennan, R. G. (2006) **Curr. Opin. Struct. Biol.** 16, 85–88
23. Huffman, J. L., and Brennan, R. G. (2002) **Curr. Opin. Struct. Biol.** 12, 95–100
24. Lim, D., Poole, K., and Strynadka, N. C. (2002) **J. Biol. Chem.** 277, 29253–29259
25. Kovacikova, G., Bennik, M. H., Demple, B., and Ellenberger, T. (2000) **Nat. Struct. Biol.** 7, 424–430
26. Kovacikova, G., and Skorupski, K. (2002) **Mol. Microbiol.** 46, 1135–1147
27. Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. (2003) **J. Mol. Biol.** 328, 555–566