The structures of transcription factor CGL2947 from *Corynebacterium glutamicum* in two crystal forms: A novel homodimer assembling and the implication for effector-binding mode

YONG-GUI GAO, MIN YAO, HIROSHI ITOU,[1] YONG ZHOU, AND ISAO TANAKA
Faculty of Advanced Life Sciences, Hokkaido University, Sapporo 060-0810, Japan

(RECEIVED April 30, 2007; FINAL REVISION June 16, 2007; ACCEPTED June 18, 2007)

Abstract

Among the transcription factors, the helix-turn-helix (HTH) GntR family comprised of FadR, HutC, MocR, YtrA, AraR, and PlmA subfamilies regulates the most varied biological processes. Generally, proteins belonging to this family contain an N-terminal DNA-binding domain and a C-terminal effector-binding/oligomerization domain. The members of the YtrA subfamily are much shorter than other members of this family, with chain lengths of 120–130 residues with about 50 residues located in the C-terminal domain. Because of this length, the mode of dimerization and the ability to bind effectors by the C-terminal domain are puzzling. Here, we first report the structure of the transcription factor CGL2947 from *Corynebacterium glutamicum*, which belongs to the YtrA family. The monomer is composed of a DNA-binding domain containing a winged HTH motif in the N terminus and two helices (α4 and α5) with a fishhook-shaped arrangement in the C terminus. Helices α4 and α5 of two monomers intertwine together to form a novel homodimer assembly. The effector-accommodating pocket with 2-methyl-2,4-pentanediol (MPD) docked was located, and it was suggested to represent a novel mode of effector binding. The structures in two crystal forms (MPD-free and -bound in the proposed effector-binding pocket) were solved. The structural variations have implications regarding how the effector-induced conformational change modulates DNA affinity for YtrA family members.

Keywords: transcription factor; winged HTH; CGL2947; YtrA family; *Corynebacterium glutamicum*

The interactions between *trans*-acting transcription factors and *cis*-acting DNA elements are the key to DNA transcriptional regulation. By binding or dissociation of a molecular signal (also called an effector), the transcription factor undergoes a conformational change, which results in either enhanced or weakened protein–DNA interaction, and this impedes or strengthens transcriptional initiation (Mathews et al. 2000; Torrents et al. 2004; Tootle and Rebay 2005). Several groups of transcription factors have been identified based on conserved motifs and DNA-binding modes, including helix-turn-helix (HTH), zinc finger, leucine-zipper, and others (Pabo and Sauer 1992). Among these groups, the HTH group is now considered a reference for understanding the general rules that govern protein–DNA interactions and has also become a favorite subject of evolutionary studies (Haydon and Guest 1991; Nguyen and Saier 1995). As a major subdivision of the HTH group, the HTH GntR family is comprised of about 270 members distributed

[1]Present address: Structural Biology Center, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan.

Reprint requests to: Min Yao, Faculty of Advanced Life Sciences, Hokkaido University, Sapporo 060-0810, Japan; e-mail: yao@castor.sci.hokudai.ac.jp; fax: +81-011-706-4481.

Abbreviations: MPD, 2-methyl-2,4-pentanediol; IPTG, isopropylthiogalactoside; HTH, helix-turn-helix; RMSD, root mean square deviation; MAD, multiple wavelength anomalous dispersion; PDB, Protein Data Bank.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.072976907.

 Copyright © 2007 The Protein Society

among the most diverse bacterial groups and regulates the most varied biological processes (Rigali et al. 2002). In general, proteins of this family contain two domains, N-terminal DNA-binding (D-b) domain and C-terminal effector-binding/oligomerization (E-b/O) domain. The structures of the D-b domains are well conserved, whereas those of the E-b/O domains are more variable and consequently are used to define subfamilies, such as FadR, HutC, MocR, YtrA, AraR, and PlmA (van Aalten et al. 2000; Rigali et al. 2002; Lee et al. 2003; Franco et al. 2006; Gorelik et al. 2006).

As a representative of the YtrA family, the transcription factor YtrA contains 130 residues and negatively regulates the ytrABCDEF operon of *Bacillus subtilis* (Yoshida et al. 2000). This operon was deduced to encode an ATP-binding cassette (ABC) transport system participating in acetoin utilization. Bacterial ABC transport systems, functionally divided as ABC importers and extruders, play important roles in membrane transport (Fath and Kolter 1993). Most YtrA family members regulate these ABC transport systems (Rigali et al. 2002). However, no structure regarding these molecules has yet been reported. Unlike other members of the HTH GntR family, proteins belonging to the YtrA family range in length from 120 to 130 residues, as shown in Figure 1. Excluding the N-terminal D-b domain, the remainder is about 50 residues forming two α-helices as shown by structure prediction (Rigali et al. 2002). The modes of dimerization and effector binding by the two α-helices have yet to be determined (Yoshida et al. 2000; Rigali et al. 2002).

Corynebacterium glutamicum is a Gram-positive, nonpathogenic, and fast-growing soil bacterium. It has been used worldwide for producing not only various amino acids, including L-glutamic acid and L-lysine, but also additional substances, such as nucleotides and vitamins. Regulation of gene expression and metabolic pathways in this bacterium represents important issues for progressively improving the yield of amino acid production. Here, we report the structure of the transcription factor CGL2947 from *C. glutamicum*, which consists of 121 residues and belongs to the YtrA family (Fig. 1). This, the first structure of a member of the YtrA family, revealed a novel homodimer assembly and

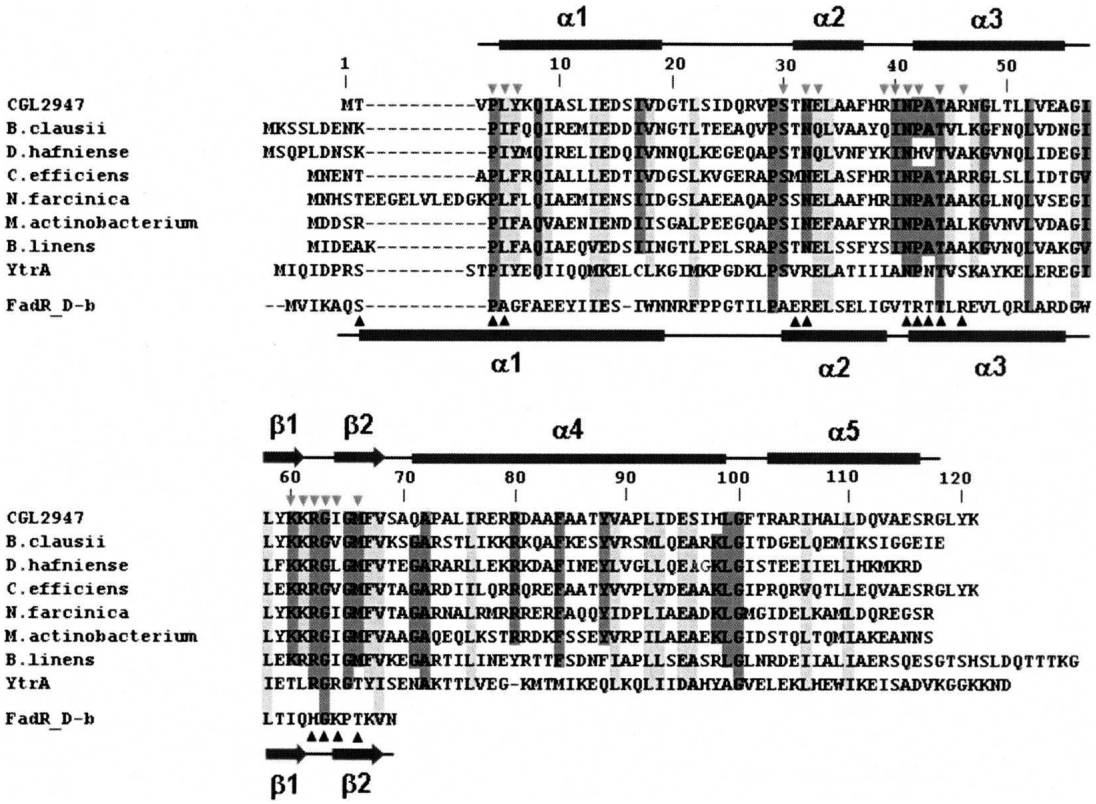

Figure 1. Sequence alignment of YtrA family proteins with CLUSTALW (Thompson et al. 1994). Sequences are numbered according to CGL2947. The conserved and conserved-changed residues are indicated by heavy and light gray shaded boxes, respectively. FadR_D_b is the N-terminal D-b domain of FadR, the representative FadR family protein. The black triangles indicate the residues of FadR_D_b (residues 1–73) that interact with DNA, and the gray triangles indicate the residues of CGL2947 suggested to contact DNA. The secondary structures (arrows represent β-strands; bars represent α-helices) of CGL2947 and FadR_D_b are drawn on the *top* and the *bottom* of the sequence, respectively.

effector-binding mode. Furthermore, the crystal structures in two forms were compared, and the results provided insight into how YtrA family proteins and other transcription factors with relatively few residues bind effectors and introduce conformational changes.

Results

Overall structure of the CGL2947 monomer

The structure of CGL2947 was solved from two forms of crystal. For Form I, the asymmetric unit contained a monomer, and a dimer was tightly formed by crystallographic symmetry. In Form II, four protein molecules formed two dimers in an asymmetric unit, and the structures of the four copies were similar, with a total average RMSD of 0.9 Å for Cα atoms. The two sets of atomic coordinates have been deposited in the Protein Data Bank with the identification numbers 2DU9 and 2EK5, respectively.

The final model of CGL2947 for Form I is composed of one protein molecule of 116 residues, one molecule of MPD, and 47 water molecules. Because of poor electron density, the N-terminal residues Met1–Thr2 and C-terminal residues Leu119–Lys121 could not be built. The overall structure is composed of a single α/β N-terminal domain (residues 3–70) and two C-terminal helices (residues 71–118), as shown in Figure 2A. The N-terminal domain, with the topological order α1, α2, α3, β1, and β2, contains a typically prokaryotic winged HTH D-b motif (Huffman and Brennan 2002). The motif is formed by helices α2, α3, and their connecting turn, together with the two antiparallel β strands (β1 and β2) designated as the wing. The arrangement of the two helices (α4 and α5) in the C terminus resembles a fishhook, with the long helix α4 (29 residues) forming the shank. Furthermore, there is a bend at Tyr88 in helix α4, which changes the direction of the helix axis by about 30°.

The N terminus of helix α4 contacts the whole of helix α1 through side-chain–side-chain interactions. These interactions include salt bridges and hydrophobic interactions, which are formed by the residues in α1 (Lys7, Ser11, Glu14, Asp15, Ile17, Val18) and α4 (Ala72, Pro73, Ile76, Arg77, Arg79, Arg80), respectively (Fig. 2A).

D-b domain

A search for structural homologs of the D-b domain of CGL2947 with Dali (Holm and Sander 1998) identified several matches (with Z scores ≥ 5.2), which are primary transcription factors. The best match is a fatty acid responsive transcription factor (FadR, PBD 1E2X; Z score 9.5), which was used to classify a subfamily of the GntR family (van Aalten et al. 2000). The results of

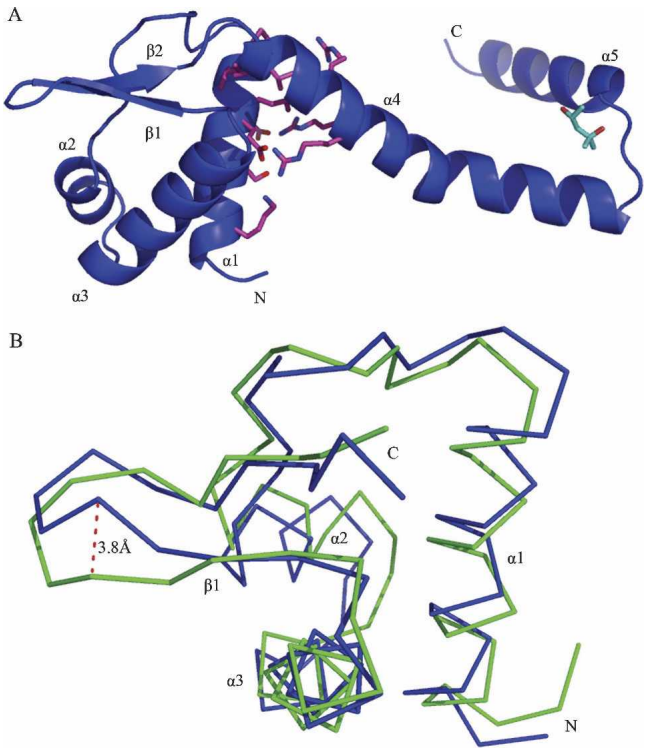

Figure 2. (*A*) Overall structure of the CGL2947 monomer. The protein is shown in ribbon representation. The residues contacting α1 and α4 are shown as a stick model in rose, and the bound MPD as a stick model in cyan, with oxygen and nitrogen atoms in red and blue, respectively. The figure was drawn using PyMOL (DeLano Scientific LLC, http://pymol.sourceforge.net). (*B*) Structural comparison of the D-b domain (residues 3–71) of CGL2947 with that (residues 5–73) of FadR. CGL2947 and FadR are colored blue and green, respectively.

structural comparison of the DNA-binding domains between CGL2947 and FadR are shown in Figure 2B. The HTH motif ranging from Thr31–Val55 (helices α2 and α3) in CGL2947 can be superposed on the corresponding HTH motif residues Glu34–Asp58 in FadR. For the two winged β-hairpins (Leu58–Val68 of CGL2947 and Leu61–Val71 of FadR), the lengths are equal, and the start and end parts can be superposed well (Figs. 1, 2B). However, the central parts of the two β-hairpins (the tips of the β-hairpins) differ in conformation, giving the maximum C_α distance of 3.8 Å between Lys61 (CGL2947) and Gln64 (FadR). The sequence alignment based on the structure covers residues Val3–Gln71 in CGL2947 and residues Ala5–Asn73 in FadR. Correspondingly, the RMSD of atoms is 1.6 Å.

Homodimer assembly

The results of gel filtration experiments indicated that CGL2947 exists as a homodimer in solution (data not shown). This was further confirmed from the crystal

structure of Form I: A dimer interaction was generated by a crystallographic twofold axis. The C-terminal helices with fishhook shape are intertwined, and consequently a novel homodimer assembly is formed (Fig. 3). The dimer contacts are made along the faces of helix α5 and the C-terminal part of helix α4. The significant inter-monomer side-chain–side-chain contacts include Phe84-Leu92′/Leu99′/Phe101′/Ile106′/Leu109′ (Phe interacting with the five residues), Tyr88-Glu95′, Val89-Leu110′/Val113′, Leu92-Leu92′, Ile93-Leu110′/Val113′, Asp94-Arg117′, Arg103-Asp111′, and His107-His107′, where the primes indicate the other monomer (same as below). A hydrophobic core was formed through van der Waals interactions and salt bridges made by these residues, burying ~2374 $Å^2$ of accessible surface area. Upon dimerization, the long helix α4 makes bilateral interactions; the N terminus shows intermolecular interaction with the D-b domain, and the C terminus interacts with helices α4′ and α5′. The bilateral interactions result in a bend at Tyr88 in helix α4, allowing contact between the DNA-binding domain of the monomer itself and the hydrophobic core of the dimer, which is helpful to fix the compact structure and provide scaffold support for the two helices (α4 and α5) in the C terminus. Thus, the structure becomes stable after dimerization.

The structure of the homodimer resembles a bidentate rake with the intertwined helices in the center and two HTH motifs as ended dentations, which putatively wedge into the major groove of the double-stranded DNA to make interactions. A conserved residue, Asn41 (Fig. 1), is located at the turn of HTH motif (a D-b region for a HTH family). The distance between the two residues Asn41 in one homodimer is 64 Å, corresponding to about two turns of the double-stranded DNA. This suggests that the homodimer would span two turns of the double-stranded DNA while binding. YtrA, the homolog of CGL2947, was also shown to exist as a dimer in solution by gel filtration experiments (Y.-G. Gao, M. Yao, and I. Tanaka, unpubl.). Moreover, predicted structure showed that YtrA has an N-terminal D-b domain containing a winged HTH motif and two C-terminal α-helices (Rigali et al. 2002). The combination of the sequence alignment (Fig. 1) and the

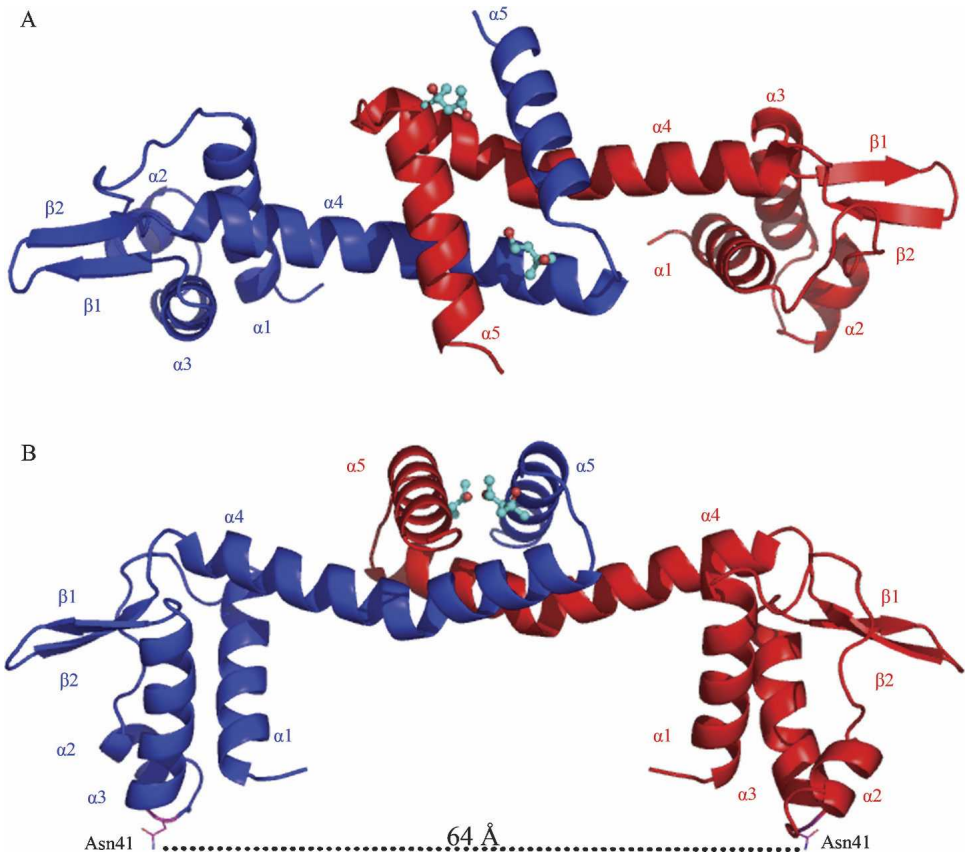

Figure 3. Ribbon representation of the novel assembly of the homodimer shown from *top* (*A*) and *side* (*B*). The two structures generated by a crystallographic twofold axis are colored blue and red, respectively. The distance between the two residues Asn41 (shown in wire model) in the homodimer is 64 Å, shown in the side-view diagram (*B*).

conservation of resides involved in dimer formation (described below) predicts that YtrA has a similar homodimer structure. YtrA negatively regulates the ytrABCDEF operon, and its binding site overlaps the transcriptional start site with partial dyad symmetry: $^{-2}$AGTGTACTAATTG AAGTAATACACT^{23} (Yoshida et al. 2000). Interestingly, the length of this sequence is in accordance with the structural characterization of the CGL2947 homodimer, further confirming the similarity of the homodimer structures of YtrA and CGL2947.

Family sequence alignment and the putative effector-binding mode

The sequence alignment for YtrA family members is shown in Figure 1; the percentages of the conserved and conserved-changed residues are 8.8% and 16.3%, respectively. There is a large gap at the N terminus (residues 9–11) due to extra sequence of the homolog from *Nocardia farcinica*. The conserved region involved in DNA binding is located after the gap. The degree of sequence conservation of the N-terminal D-b domain is higher than that of the whole molecule. Moreover, the conserved residues are mainly distributed in the winged HTH motif (indicated by the gray triangles). In the C-terminal region corresponding to α4 and α5 of CGL2947, most of the conserved/conserved-changed residues, such as Phe84, Val89, Leu92, Ile93, Phe101, Ile106, and Leu110, are hydrophobic, which plays a very significant role in dimer formation through hydrophobic interactions of their side chains. Furthermore, the residues Glu95 and Arg105, conserved in charged side chain, are involved in the formation of side-chain–side-chain contacts (Glu95–Tyr88′, Arg105–Asp81′) that may improve stabilization of the dimer. In addition, the residues Glu14, Ile17, Val18, Ala72, and Ile76 participating in the contacts of helices α4 and α1 (Fig. 1) remain relatively highly conserved among YtrA family proteins.

Members of the HTH GntR family, including FadR, HutC, MocR, YtrA, AraR, and PlmA subfamilies, typically consist of an N-terminal D-b domain and C-terminal E-b/O domain (Rigali et al. 2002). Members of the YtrA family are generally comprised of 120–130 residues (Fig. 1). Excluding the D-b domain, the average length is about 50 residues forming two α-helices (Rigali et al. 2002). It is expected that the C-terminal residues are too short to accommodate effector binding (Yoshida et al. 2000). Although the structure of CGL2947 was solved in the absence of effector, an effector-binding mode can be proposed to unravel this problem. The two monomers of the CGL2947 protein are intertwined with the two α-helices in their C termini to form a homodimer, and then a pocket to accommodate effector binding is formed (Fig. 4). The side chains of the following residues in the homodimer are involved in formation of this pocket: Leu92, Ile93, Ser96, Ile97, Phe101, Thr102, Arg103, His107, Leu110′, Asp111′, and Ala114′. According to the sequence alignment, five of these residues are conserved or conserved-changed, and are mainly hydrophobic (Fig. 1). For a homodimer, there are two pockets that maintain crystallographic symmetry. Two molecules of MPD in the crystallization buffer are bound in the pockets, providing indirect support for this prediction (Fig. 4). In previous studies, the biochemical results implied that the effector of YtrA may be acetoin ($H_3CCOCH(OH)CH_3$); however, the authors could not detect the induction of the reporter gene present in the strain BFS47 by the addition of 5 mM acetoin to DSM medium (Yoshida et al. 2000). Acetoin is similar in size to MPD ($CH_3CH(OH)CH_2C(CH_3)_2OH$), which is in agreement with the capacity of the pocket. Hence, the present structure suggests that YtrA may have the ability to accommodate acetoin binding and provides support for the role of acetoin as the effector. Further experiments are required to determine whether acetoin is the effector, including varying the acetoin concentration, growth conditions (e.g., medium), reconstruction of reporter gene plasmid, etc., based on the experiments reported previously (Yoshida et al. 2000).

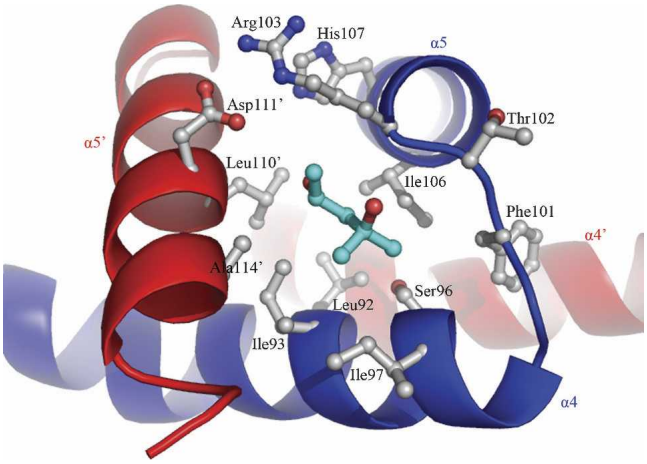

Figure 4. The putative pocket to accommodate effector binding. The bound MPD and the residues involved in pocket formation are shown as a ball-and-stick model, colored cyan and gray, respectively. The intertwined helices in homodimer are colored blue (one monomer) and red (another monomer). The primes indicate helices and residues of the other monomer (as in Fig. 5).

Structural variations between the two forms

The structures of transcription factor CGL2947 were solved in two forms, with the total average RMSD of Cα atoms being 1.2 Å. For Form I, the MPD in the

crystallization buffer was bound in the pocket, which was formed through homodimer assembly (Fig. 4). In the case of Form II, the crystal structure of the apo form was obtained. To explore the variations in the two D-b domains between the two forms of homodimers, a structural comparison was performed by fitting the Cα of the C termini, as shown in Figure 5. Without MPD (Form II), residue Arg103 partially occupies the pocket. Upon binding MPD (Form I), it undergoes a dramatic conformational change to broaden the pocket. As a result, residues Leu99, Phe101, and Arg105 are forced to shift toward helix α4′. The shifts of the three residues push helix α4′ through the interactions between them and the residues of helix α4′, leading the N terminus of helix α4′ to shift up to 3.9 Å, resulting in motion of the entire D-b domain. For example, the shift of the tip of the β1′–β2′ hairpin (a portion of the winged HTH motif) is 9.0 Å (Fig. 5B). Because of the (pseudo) twofold axis relating the two monomers in the homodimer, the shifts of the two DNA-binding domains would occur in opposite directions, which would cause variations in their distance, that is, for the two forms of homodimer, the distances between the two residues Asn41 proposed to contact DNA are 64 and 56 Å, respectively (Fig. 5A).

The side chain of residue Arg103 is of particular importance to trigger the conformational changes. Although residue Arg103 is less highly conserved, the residues at or neighboring this position possess similar side chains (Fig. 1). The side chain–side chain contact Arg105–Asp81 is also crucial for the structural shift. Notably, the similar side chain–side chain contacts are thought to be retained in YtrA family members according to the sequence alignment. Other residues important to trigger or transmit the conformational changes include Ile76, Arg77, Glu78, Arg79, Arg80, Leu99, and Phe101, most of which are conserved.

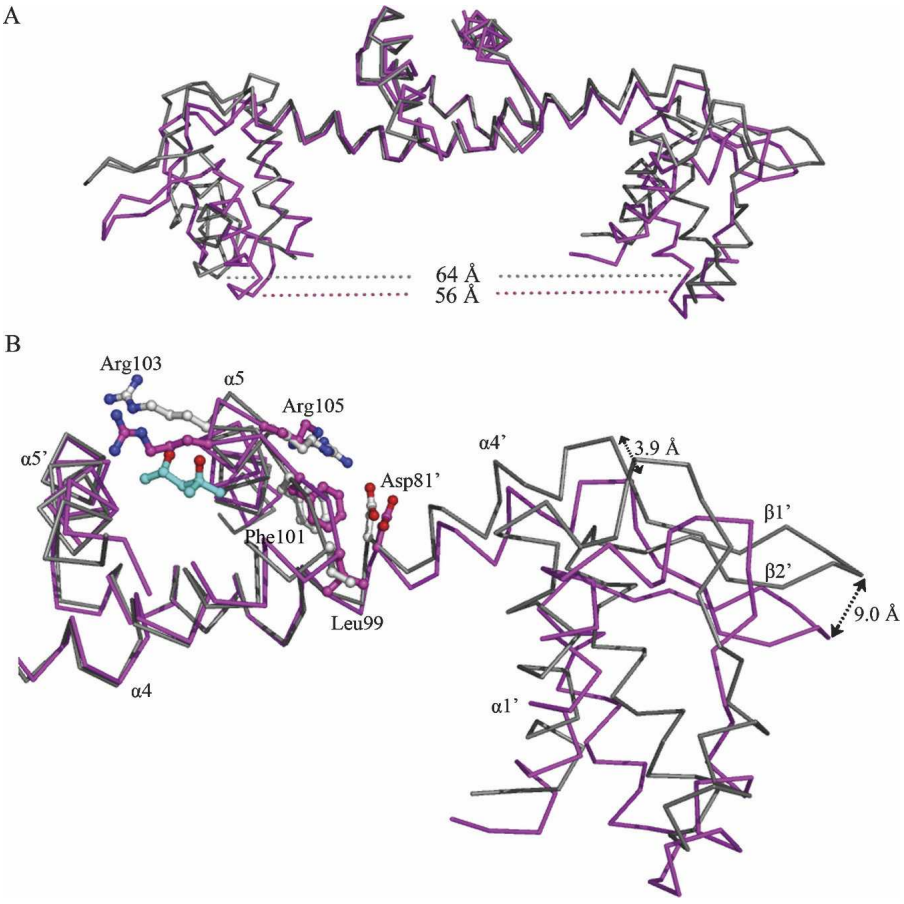

Figure 5. Structural comparison of the homodimers in two forms (MPD-bound and -free). By fitting the C terminus (Cα 89–115), the homodimer structures in Form I (MPD-bound) and Form II (MPD-free) crystals were superposed, with Form I and Form II shown in gray and pink, respectively. (*A*) Overall view of the homodimer structures in two forms. (*B*) Enlarged view showing the detailed structural variations for the D-b domain. The important residues for conformational changes are shown as ball-and-stick models and are colored gray and pink for Form I and Form II, respectively. The bound MPD is shown as a ball-and-stick model with carbon atoms colored cyan. All oxygen and nitrogen atoms in ball-and-stick models are colored red and blue, respectively.

Discussion

D-b mode

On the basis of the interactions of FadR–DNA complex (van Aalten et al. 2001; Xu et al. 2001), the residues contacting the DNA are marked by black triangles in Figure 1. As the D-b domains of CGL2947 and FadR can be superposed well (Fig. 3), it is likely that CGL2947 contacts DNA in a manner similar to that adopted by FadR. As the FadR–D-b mode, CGL2947 may make interactions with DNA in four distinct regions that possess highly conserved residues among members of the YtrA family, indicted by the gray triangles in Figure 1. The four regions are the N terminus (Pro4–Tyr6), the start of helix α2 (Ser30–Glu33), the turn (between α2 and α3) and the helix α3 (Ile40–Arg46), and the winged β1–β2 (Lys60–Met66). Helix α3, designated the recognition helix of the HTH motif, enters into the major groove of the DNA helix to make interactions that are critical for specificity (Huffman and Brennan 2002). In the case of CGL2947, the residues from the turn (between α2 and α3) and the tip of helix α3 seem to contribute a great deal to specific recognition. To discuss further details regarding the contacts of CGL2947 with DNA and the dimer-binding model, it will be necessary to solve the structure of the CGL2947–DNA complex.

Implications for effector-induced conformational changes

Although MPD is not the effector for the protein CGL2947, the structural variations between MPD-bound and -free forms provide implications for how the effector-induced conformational changes regulate the affinity for DNA (Fig. 5). The effector binding to the homodimer would have an influence on the pocket and the dimer interface and trigger the shift of helix α4, resulting in a shift of the entire D-b domain. Consequently, the relative positions of the two D-b domains in the homodimer also change (Fig. 5). Such changes modulate the affinity of the homodimer for DNA. The crystal packings in the two forms are distinct, which may have some synergistic effects on the structural variations, and thus the aforementioned conformational changes induced by MPD are plausible. Moreover, the transmission mechanism employed by FadR supports this prediction. For FadR, the effector-induced conformational changes were transmitted from α4 to the winged HTH rigid body (van Aalten et al. 2001). Interestingly, both proteins have very similar D-b domains. In addition, the allosteric effects, triggered by IPTG for LacI (Bell and Lewis 2000) and by guanine for PurR (Schumacher et al. 1995) also support the results of our structural analysis regarding the conformational changes that are conducted in the dimer interface. In the cases of both LacI and PurR, effectors binding to the pockets at the junction of two C-subdomains influence the dimer interface, leading to the orientations of D-b domains, which affect their affinity for DNA.

In summary, the protein–D-b region, dimer interface, and pocket formation are mediated by residues that are highly conserved within the YtrA family, which implies that members of this family share common modes of DNA binding, homodimer assembly, effector binding, and transmission of conformational changes. The structures of CGL2947 in two crystal forms (MPD bound and free) provide implications concerning these modes, especially dimerization and effector binding by the C-terminal domain with only two α-helices.

Materials and Methods

Recombinant protein expression and purification

Recombinant CGL2947 protein with a C-terminal His-tag was expressed in *Escherichia coli* strain B834 (DE3) by addition of 1 mM IPTG to LB broth (298 K) at OD_{600} (optical density) 0.6. After induction for about 20 h, the cells were harvested and resuspended in buffer A (50 mM Na-Pi at pH 8.0, 0.5 M NaCl), then disrupted with a French press. The CGL2947 protein was captured with a HiTrap chelating HP column (Amersham Biosciences Inc.) and eluted with a linear gradient of 0.0–0.5 M imidazole in buffer A. The fractions containing the target protein were sequentially purified with HiLoad 26/60 Superdex 75 pg (Amersham Biosciences) chromatography with buffer A as the eluent. The target peak was collected and dialyzed overnight against 10 mM CAPS-NaOH at pH 10.2, 0.2 M NaCl. Finally, the CGL2947 protein was concentrated to 4.8 mg/mL. For production of selenomethionine-substituted CGL2947 (Se-Met CGL2947), cells were cultured in minimal medium containing Se-Met. The procedure for purification of Se-Met CGL2947 was the same as that of the native protein.

Crystallization and data collection

The sitting-drop method was used to perform crystal screening with crystallization screening kits. Each drop consisted of 1 μL of sample and 1 μL of reservoir solution, and was equilibrated against 100 μL of reservoir solution at 293 K. Initial crystals were obtained with Mem Fac (No. 40, 0.1 M Tris-HCl at pH 8.5, 12% MPD, 0.1 M Li_2SO_4) and Wizard I (No. 23, 0.1 M Imidazole-HCl at pH 8.0, 15% ethanol, 0.2 M $MgCl_2$) screening kits, with the approximate shapes of a rhombohedra (Form I) and a cube (Form II), respectively. Further optimization was performed to improve the quality of crystals using the hanging-drop method for Se-Met CGL2947. The drop contained 1.5 μL of sample and 1.5 μL of reservoir solution, equilibrated against 1 mL of reservoir solution. Large crystals were obtained under both conditions.

X-ray diffraction data were collected at BL41XU of SPring-8 (Hyogo, Japan). A single crystal was mounted in a loop and flash-cooled under a stream of nitrogen gas after soaking in mother liquor containing 20% glycerol. The two types of crystal belonged to the $P4_12_12$ and $C222_1$ space groups, respectively, with different unit cell parameters (see Table 1). A MAD data

Table 1. *Data collection statistics*

Crystal	Form I			Form II
Data set	Peak	Edge	Remote	Single wavelength
Space group		$P4_12_12$		$C222_1$
Unit cell (Å)		$a = b = 47.0$, $c = 152.5$		$a = 68.6$, $b = 74.5$, $c = 182.2$
Beamline			BL41XU (SPring-8)	
Resolution (Å)		50.00–2.28 (2.36–2.28)		48.62–2.20 (2.28–2.20)
Wavelength (Å)	0.9790	0.9793	0.9000	1.0000
Unique reflections	8413	8391	8392	24007
Completeness	99.6 (97.3)	99.4 (95.4)	99.3 (95.6)	99.2 (95.3)
R_{merge} (%)[a]	6.2 (38.1)	6.2 (39.8)	6.0 (38.7)	7.5 (29.9)
I/σ	13.0	12.8	10.6	13.5
Redundancy	12.0 (6.4)	11.8 (5.5)	11.7 (5.5)	5.2 (4.2)

Values in parentheses refer to the highest resolution shell.
[a] $R_{merge} = \sum_h \sum_j |\langle I \rangle_h - I_{hj}| / \sum_h \sum_j I_{hj}$, where $\langle I \rangle_h$ is the mean intensity of symmetry equivalent reflections.

set diffracted to 2.28 Å and single-wavelength data diffracted to 2.2 Å were collected, corresponding to Form I and Form II crystals, respectively. All data were processed using an HKL2000 package (Otwinowski and Minor 1997) with data statistics shown in Table 1.

Structure solution and refinement

The asymmetric units contain one (Form I) and four (Form II) protein molecules, corresponding to solvent contents of 58.0% and 52.0%, respectively. Using MAD data collected from Form I crystals, one selenium site was located, and phase calculations were carried out using the program SOLVE/RESOLVE (Terwilliger and Berendzen 1999; Terwilliger 2003). The initial model was built automatically to 53.8%. The model building and refinement were performed semi-automatically using LAFIRE with CNS (Yao et al. 2006). Several iterations of LAFIRE running improved the crystal structure with an *R* factor of 30.6% at the resolution range 20.0–2.28 Å. On the basis of the results of a manual check, the MPD contained in the crystallization solution was located according to the corresponding maps of both $2F_o - F_c$ and $F_o - F_c$. Then, the model was refined and water molecules were located using the program CNS (Brünger et al. 1998), finally giving *R/R-free* factor to 26.3/28.6. The final model was examined using PROCHECK (Laskowski et al. 1993): 93.4% of the residues lie in the most favored regions, with the remaining 6.6% assigned to the additionally allowed regions. The final refinement statistics for the structure are listed in Table 2.

For the Form II crystals, no structure solutions were found by molecular replacement using the solved structure (Form I) as a search model. A MAD data set diffracted to 2.9 Å was collected again (data not shown), and the initial structure was solved. Once the model was built, refinement and water location were performed using the single-wavelength data at 2.2 Å resolution (as for Form I). The corresponding refinement statistics are shown in Table 2.

Table 2. *Refinement statistics*

Refinement	Form I	Form II
Resolution range (Å)	20–2.28	20–2.20
R[a]/*R-free*[b] (%)	26.3/28.6	22.5/24.0
RMSD bond length/angle (Å/°)	0.009/1.5	0.007/1.3
B factor (Å^2)		
Protein molecules	48.7	30.5
Water molecules	57.8	36.8
Others	61.8	-
Ramachandran plot (%)		
Most favored regions	93.8	94.1
Additionally allowed regions	6.8	5.4
Generously allowed regions	0	0.5
Disallowed regions	0	0

[a] $R = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$, where F_{obs} and F_{calc} are observed and calculated structure factor amplitudes.
[b] *R-free* was calculated for *R*, using a random 10% subset from all reflections.

Acknowledgments

We thank the staff of Beamline BL41XU, SPring-8, Hyogo, Japan, for their kind help with data collection. This work was supported by the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

Bell, C.E. and Lewis, M. 2000. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7:** 209–214.

Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Panns, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Fath, M.J. and Kolter, R. 1993. ABC transporters: Bacterial exporters. *Microbiol. Rev.* **57:** 995–1017.

Franco, I.S., Mota, L.J., Soares, C.M., and de Sa-Nogueira, I. 2006. Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.* **188:** 3024–3036.

Gorelik, M., Lunin, V.V., Skarina, T., and Savchenko, A. 2006. Structural characterization of GntR/HutC family signaling domain. *Protein Sci.* **15:** 1506–1511.

Haydon, D.J. and Guest, J.R. 1991. A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.* **63:** 291–295.

Holm, L. and Sander, C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26:** 316–319.

Huffman, J.L. and Brennan, R.G. 2002. Prokaryotic transcription regulators: More than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **12:** 98–106.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26:** 283–291.

Lee, M.H., Scherer, M., Rigali, S., and Golden, J.W. 2003. PlmA, a new member of the GntR family, has plasmid maintenance functions in Anabaena sp. strain PCC 7120. *J. Bacteriol.* **185:** 4315–4325.

Mathews, M.B., Sonenberg, N., and Hershey, J.W.B. 2000. Origins and principles of translational control. In *Translational control of gene expression* (eds. N. Sonenberg and J.W.B. Hershey), pp. 1–31. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nguyen, C.C. and Saier, M.H.J. 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS Lett.* **377:** 98–102.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Pabo, C.O. and Sauer, R.T. 1992. Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61:** 1053–1095.

Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. 2002. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J. Biol. Chem.* **277:** 12507–12515.

Schumacher, M.A., Choi, K.Y., Lu, F., Zalkin, H., and Brennan, R.G. 1995. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell* **83:** 147–155.

Terwilliger, T.C. 2003. Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr. D Biol. Crystallogr.* **59:** 45–49.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55:** 849–861.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Tootle, T.L. and Rebay, I. 2005. Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. *Proteins* **27:** 285–298.

Torrents, E., Roca, I., and Gibert, I. 2004. The open reading frame present in the nrdEF cluster of *Corynebacterium ammoniagenes* is a transcriptional regulator belonging to the GntR family. *Curr. Microbiol.* **49:** 152–157.

van Aalten, D.M., DiRusso, C.C., Knudsen, J., and Wierenga, R.K. 2000. Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J.* **19:** 5167–5177.

van Aalten, D.M., DiRusso, C.C., and Knudsen, J. 2001. The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.* **20:** 2041–2050.

Xu, Y., Heath, R.J., Li, Z., Rock, C.O., and White, S.W. 2001. The FadR.DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J. Biol. Chem.* **276:** 17373–17379.

Yao, M., Zhou, Y., and Tanaka, I. 2006. LAFIRE: Software for automating the refinement process of protein-structure analysis. *Acta Crystallogr. D Biol. Crystallogr.* **62:** 189–196.

Yoshida, K.I., Fujita, Y., and Ehrlich, S.D. 2000. An operon for a putative ATP-binding cassette transport system involved in acetoin utilization of *Bacillus subtilis*. *J. Bacteriol.* **182:** 5454–5461.