The CGL2612 Protein from Corynebacterium glutamicum Is a Drug Resistance-related Transcriptional Repressor

STRUCTURAL AND FUNCTIONAL ANALYSIS OF A NEWLY IDENTIFIED TRANSCRIPTION FACTOR FROM GENOMIC DNA ANALYSIS*

Received for publication, June 1, 2005, and in revised form, August 8, 2005. Published, JBC Papers in Press, September 15, 2005, DOI 10.1074/jbc.M505999200

Hiroshi Itou†1, Ui Okada1, Hiroaki Suzuki1, Min Yao1, Masaaki Wachi1, Nobuhsa Watanabe2, and Isao Tanaka1,2

From the 1Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan and the 2Department of Bioengineering, Tokyo Institute of Technology, Yokohama 226-8501, Japan

The emergence of antibiotic-resistant bacteria often causes serious clinical problems. The TetR family is one of the major transcription factor families that regulate expression of genes involved in bacterial antimicrobial resistance systems. CGL2612 protein is a transcription factor newly identified by genomic DNA analysis on Corynebacterium glutamicum, which belongs to the mycolic acid-containing Actinomycetales, including the well known pathogens Corynebacterium diphtheriae and Mycobacterium tuberculosis. Crystal structure analysis showed that the CGL2612 protein exhibits significant structural similarity to the multidrug resistance (MDR)-related transcription factor QacR from Staphylococcus aureus, despite poor amino acid sequence similarity between these proteins. Binding DNA sequence analysis of CGL2612 protein using the systematic evolution of ligands by the exponential enrichment (SELEX) method revealed that this protein is a new member of the TetR family, which regulates expression of the immediately upstream gene, cgl2611, probably encoding a major facilitator superfamily permease. Subsequent functional analyses confirmed a function of the CGL2612 as a transcriptional repressor responsible for the antimicrobial resistance system in C. glutamicum. The strategy used in the present study is one of the most convenient and powerful methods to analyze functionally unknown transcription factors, and the results obtained here will contribute to our understanding of the drug resistance mechanism not only in C. glutamicum but also in the related bacteria, C. diphtheriae and M. tuberculosis.

Many kinds of antibiotics have been developed to control infectious diseases. However, bacteria have developed various ways to survive when faced with these agents. One such bacterial mechanism of antibiotic resistance involves membrane proteins that expel toxic compounds from the cells (1). Two types of efflux system are known: protein systems that transport a specific toxic compound or class of toxins out of the cell, such as tetracycline efflux pumps (2), and those that transport numerously structurally dissimilar compounds. The latter type of transporter can actively export a wide range of compounds, and these are called multidrug transporters (1). Bacteria that acquire such transport systems can cancel the effects of many kinds of drugs (multidrug resistance (MDR)3). MDR has been shown to be widespread among both Gram-positive and Gram-negative bacteria (3), and the emergence of such bacteria causes nosocomial infection and often has adverse effects on health. Staphylococcus aureus is one of the bacteria that often develop MDR and can cause serious clinical problems. The TetR family of transcriptional factors are known to regulate expression of genes involved in bacterial efflux systems against antimicrobial compounds or drugs (4). Proteins that belong to this family are known to act as transcriptional repressors, and they bind constitutively to their regulatory sequence and repress expression of target genes that are related to drug detoxification or export. When the cellular level of toxic compounds increases, the repressor alters its DNA-binding ability and dissociates from the regulatory sequence, resulting in activation of antibiotic resistance of the bacterium.

Corynebacterium glutamicum is widely used for industrial production of amino acids. The genome of the C. glutamicum wild-type strain ATCC13032 has been sequenced by Kyowa Hakko and is now available in the public domain. The well known pathogenic bacteria, Corynebacterium diphtheriae and Mycobacterium tuberculosis, are phylogenetically related to C. glutamicum. Analysis of C. glutamicum (3.3 Mbp), C. diphtheriae (2.5 Mbp), and M. tuberculosis (4.4 Mbp) genomes using the GTOP protein structure prediction data base (available on the World Wide Web at spock.genes.nig.ac.jp/~gtop/gtop.html) (5) indicated that these genomes encode 16, 12, and 43 proteins that possess the TetR-type DNA-binding domain, respectively, but most of their functions are still unknown. Structural and functional analyses of such novel TetR family proteins found by genomic analysis promote the elucidation of drug resistance mechanisms in these bacteria and contribute to reduction of the threat of MDR. Here, we report the crystal structure and biological function of CGL2612 protein (177 residues, 20.3 kDa) from C. glutamicum as a transcriptional repressor of a putative major facilitator superfamily (MFS) drug efflux pump responsible for increasing resistance to antibiotics in bacteria.

MATERIALS AND METHODS

Protein Preparation

The full-length (amino acids 1–177) cgl2612 gene was amplified by PCR using the C. glutamicum wild-type strain ATCC13032 genome as a template.

1 Present address: Structural Biology Center, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan.
2 † Interdepartmental Program in Structural and Functional Genomic Biology, Department of Biological and Chemical Science, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan.
Structural and Functional Analysis of CGL2612 Protein

The structure of CGL2612 protein was determined using the single anomalous diffraction method. Its asymmetric unit contained two molecules of CGL2612 protein. The crystal volume per protein mass \( V_{m} \) (7) was 2.4 Å³ Da⁻¹ with a solvent content of 48.3%.

Se-Met-substituted crystals were also obtained under conditions similar to those used for the native crystals, and the best crystal was obtained under the following conditions: 0.08 M Tris-HCl (pH 8.3), 0.16 M MgCl₂, 23% polyethylene glycol 400, and 20% glycerol. The dimensions of the crystal were 0.3 × 0.3 × 0.4 mm³, and this crystal belonged to the space group \( P2_1 \), with cell dimensions of \( a = 56.9 \AA, b = 67.0 \AA, \) and \( c = 101.2 \AA, \) and the asymmetric unit contained two molecules of CGL2612 protein. The crystal volume per protein mass \( V_{m} \) (7) was 2.6 Å³ Da⁻¹ with a solvent content of 52.1%.

### Data Collection and Structure Determination

Data collection statistics are listed in Table ONE. Native data were collected at a wavelength of 0.9000 Å at 100 K by flash cooling on beamline BL38B1, SPring-8 (Hyogo, Japan). Single anomalous diffraction data were collected at a wavelength of 0.9800 Å from a Se-Met-substituted crystal at 100 K on beamline BL38B1, SPring-8 (Hyogo, Japan). Single anomalous diffraction data were collected at a wavelength of 0.9800 Å from a Se-Met-substituted crystal at 100 K on beamline BL41XU in SPring-8. These data were processed using the HKL2000 program (8).

The structure of CGL2612 protein was determined using the single anomalous diffraction method. Its asymmetric unit contained two molecules of the protein, and six of the eight selenium sites were found using the program SOLVE (9). Using the initial phase calculated by SOLVE, phase improvement and automated model building were performed using the RESOLVE program (10, 11). The program successfully built a model except for some residues in terminal parts or loop regions. The missing residues were built manually using the graphic program O (12).

### Identification of the CGL2612 Binding Site

Random SELEX Selection—To generate an initial random sequence DNA library, single-stranded 62-mer oligonucleotides (5′-CGGAATT-CCGATCCAGCAGAAGNN16TATGTGCGTCTACATGGATCCTC-3′), containing a random core of 16 nucleotides, were constructed. Both ends (the first 22 nucleotides and the last 24 nucleotides) of the oligonucleotide were primer binding sites. A double-stranded oligonucleotide DNA pool was amplified by PCR using primers (primer PRS5 and PRS3) complementary to the 5′- and the 3′-primer binding sites, respectively. PCR amplification was performed with ExTag DNA polymerase (Takara Bio Inc., Otsu, Shiga, Japan). The PCR products, a random sequence DNA library, were mixed with Ni²⁺-nitrilotriacetic acid resin (Qiagen Inc., Valencia, CA) preadsorbed with 50 µg of purified CGL2612 protein in binding buffer (20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 10 mM imidazole). The mixture was incubated at room temperature for 1 h. After washing of nonspecific binding sites with binding buffer, protein-DNA complexes were eluted with elution buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 500 mM imidazole). DNA fragments were extracted from the eluted fraction by phenol/chloroform extraction and ethanol precipitation. Eluted fragments were amplified by PCR

### Table ONE

| Summary of data collection and phasing | Native | Se-SAD (peak) |
|--------------------------------------|-------|--------------|
| Wavelength (Å)                       | 0.9000| 0.9800       |
| Resolution (Å)                       | 40.0–1.90 (1.97–1.90) | 40.0–1.85 (1.92–1.85) |
| Space group                          | \( P2_1 \) | \( P2_1 \) |
| Unique reflections                   | 32,643 (3,198) | 36,311 (3,411) |
| Completeness (%)                     | 99.7 (99.7) | 99.5 (95.1) |
| Averaged redundancy                  | 4.3 (4.1) | 9.9 (9.1) |
| Averaged \( I/\sigma(I) \)           | 4.5    | 26.7         |
| \( R_{merge} \) (%)                  | 6.6 (29.8) | 6.9 (24.1) |
| Selenium atom sites                  | 6 of 8 |              |
| Resolution range of data used (Å)    | 40–1.9 |              |
| Figure of merit                      | 0.42   |              |

\( ^a \) Values in parentheses are for the outermost resolution shell.

\( ^b \) \( R_{merge} = \Sigma_{h} \Sigma_{j} (I_{hj} - I_{h0}) / \Sigma_{h} \Sigma_{j} I_{hj} \), where \( I_{h0} \) is the mean intensity of symmetry-equivalent reflections.

4 M. Yao, Y. Zhou, and I. Tanaka, manuscript in preparation.

---

Template DNA and was inserted into the vector pET-26b (Novagen Inc., Madison, WI) into the NdeI and XhoI restriction enzyme sites. The 3′-end primer for cloning was designed for direct attachment of a His₆ tag to the C-terminal end of CGL2612 protein via a two-amino acid linker, consisting of Leu and Glu (pET/CGL2612-His). The resulting plasmid was transformed into *Escherichia coli* BL21-STAR (DE3) host cells (Stratagene, La Jolla, CA).

Protein expression and purification of both native and Se-Met-substituted proteins for crystallization experiments were performed using methods similar to those described previously (6).

### Crystallization

Crystallization experiments were performed using the same strategy as described previously (6). The best native crystal was obtained under the following conditions: 0.08 M Tris-HCl (pH 8.3), 0.16 M MgCl₂, 21% polyethylene glycol 400, and 20% glycerol. The dimensions of the crystal were 0.3 × 0.3 × 0.4 mm³, and this crystal belonged to the space group \( P2_1 \), with cell dimensions of \( a = 56.9 \AA, b = 67.0 \AA, \) and \( c = 101.2 \AA, \) and the asymmetric unit contained two molecules of CGL2612 protein. The crystal volume per protein mass \( V_{m} \) (7) was 2.4 Å³ Da⁻¹ with a solvent content of 48.3%.

Se-Met-substituted crystals were also obtained under conditions similar to those used for the native crystals, and the best crystal was obtained under the following conditions: 0.08 M Tris-HCl (pH 8.3), 0.16 M MgCl₂, 23% polyethylene glycol 400, and 20% glycerol. The dimensions of the crystal were 0.3 × 0.3 × 0.4 mm³, and this crystal also belonged to the space group \( P2_1 \), with cell dimensions of \( a = 58.2 \AA, b = 67.9 \AA, \) and \( c = 105.5 \AA, \) and the asymmetric unit contained two molecules of CGL2612 protein. The crystal volume per protein mass \( V_{m} \) (7) was 2.6 Å³ Da⁻¹ with a solvent content of 52.1%.

### Data Collection and Structure Determination

Data collection statistics are listed in Table ONE. Native data were collected at a wavelength of 0.9000 Å at 100 K by flash cooling on beamline BL38B1, SPring-8 (Hyogo, Japan). Single anomalous diffraction data were collected at a wavelength of 0.9800 Å from a Se-Met-substituted crystal at 100 K on beamline BL41XU in SPring-8. These data were processed using the HKL2000 program (8).

The structure of CGL2612 protein was determined using the single anomalous diffraction method. Its asymmetric unit contained two molecules of the protein, and six of the eight selenium sites were found using the program SOLVE (9). Using the initial phase calculated by SOLVE, phase improvement and automated model building were performed using the RESOLVE program (10, 11). The program successfully built a model except for some residues in terminal parts or loop regions. The missing residues were built manually using the graphic program O (12). The phasing statistics are summarized in Table ONE.

Structure refinement was performed on the native data. The orientation and the position of molecules in the native crystal were determined by a molecular replacement method using the AMoRe program (13). The monomer model built on the single anomalous diffraction data was used as a search model. Model fitting and refinement were carried out automatically using the Laref program with the CNS program (15). Finally, manual model checking and fitting were carried out using O. Water molecules were found using CNS, and the final model had an \( R \)-factor of 20.8% and a free \( R \)-factor of 24.1% for data between 10 and 1.9 Å resolution. The final refinement statistics are given in Table TWO.

---

Identification of the CGL2612 Binding Site

Random SELEX Selection—To generate an initial random sequence DNA library, single-stranded 62-mer oligonucleotides (5′-CGGAATT-CCGATCCAGCAGAAGNN₁₆TATGTGCGTCTACATGGATCCTC-3′), containing a random core of 16 nucleotides, were constructed. Both ends (the first 22 nucleotides and the last 24 nucleotides) of the oligonucleotide DNA pool was amplified by PCR using primers (primer PRS5′ and PRS3′) complementary to the 5′- and the 3′-primer binding sites, respectively. PCR amplification was performed with ExTag DNA polymerase (Takara Bio Inc., Otsu, Shiga, Japan). The PCR products, a random sequence DNA library, were mixed with Ni²⁺-nitrilotriacetic acid resin (Qiagen Inc., Valencia, CA) preadsorbed with 50 µg of purified CGL2612 protein in binding buffer (20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 10 mM imidazole). The mixture was incubated at room temperature for 1 h. After washing of nonspecific binding sites with binding buffer, protein-DNA complexes were eluted with elution buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 500 mM imidazole). DNA fragments were extracted from the eluted fraction by phenol/chloroform extraction and ethanol precipitation. Eluted fragments were amplified by PCR

---

4 M. Yao, Y. Zhou, and I. Tanaka, manuscript in preparation.
using the primers PRS5’ and PRS3’, and the PCR products were used as a DNA pool for the next cycle of selection. This process was repeated for a total of eight cycles. Selected DNA fragments after these cycles were cloned into the pGEM-T Easy Vector (Promega, Madison, WI), and the DNA sequences of their random core regions were analyzed.

Genomic SELEX Selection—Chromosomal DNA of the C. glutamicum wild-type strain ATCC13032 was fragmented into short fragments (~100–300 bp) by sonication (Astrason XL2020; Misonix Inc., Farmingdale, NY) on ice, and both ends of the fragments were blunted using DNA polymerase (DNA Blunting kit, Takara Bio Inc., Otsu, Shiga, Japan). The blunt-ended fragments were cloned into EcoRV-digested pBR322 (New England Biolabs Inc., Beverly, MA) using T4 DNA ligase (Takara Bio). Aliquots of 120 ng of labeled probes, 10 μg of salmon testis DNA (Sigma), and 150 μg of purified CGL2612 protein were mixed in 20 μl of reaction buffer (40 mM Tris–HCl, pH 7.5, 200 mM NaCl, 8 mM MgCl2, 5 mM dithiothreitol) and incubated for 20 min at 37 °C. After incubation, 0.25 units of DNase I (Takara Bio) in 0.5 μl of reaction buffer were added and mixed, and digestion was allowed to continue for 1 min at 37 °C. Digestion was terminated by the addition of phenol/chloroform, and reaction mixtures were extracted. DNA was precipitated with ethanol, and the resulting pellets were resuspended in 8 μl of loading dye. After heating at 90 °C for 1 min, aliquots were analyzed by electrophoresis on an 8 M urea-containing 6% polyacrylamide gel. Sequence ladders were generated using a Thermo sequence cycle sequence kit (Amersham Biosciences) with the same primer as was end-labeled for probe preparation.

Construction of a cgl2612 Disruptant

For disruption of the cgl2612 gene, the C. glutamicum wild-type strain ATCC31831 was used. A 2.1-kb DNA fragment encompassing the cgl2612 gene was amplified by PCR using the primers, 5’-AG-GAGCCAATTCCATCGTCC-3’ and 5’-GGACATACCTTGCA- CAGATGTC-3’. The amplified DNA fragment was digested with XbaI and PstI and then ligated with the vector plasmid pK18mobsacB (16) digested with XbaI-PstI. From the resulting plasmid, a 440-bp Xhol fragment encoding an internal part of cgl2612 was removed by Xhol digestion followed by self-ligation. The constructed plasmid, named pK18Δcgl2612, was introduced directly into ATCC31831 cells by electroporation, and kanamycin-resistant transformants were selected. As pK18Δcgl2612 did not have a replication origin functional in C. glutamicum cells, only cells in which homologous recombination occurred between the chromosomal DNA and plasmid DNA could grow on kanamycin-containing plates. The kanamycin-resistant transformants

### TABLE TWO

| Parameter                     | Value  |
|-------------------------------|--------|
| Resolution range (Å)         | 10.0–1.90 |
| No. of reflections            | 30,289 |
| Completeness (%)              | 97.8 |
| Total number of non-hydrogen atoms | 2,799 |
| Protein                       | 154  |
| Solvent                       | 18.2  |
| R-factor (%)                  | 20.8  |
| R-free factor (%)             | 24.1  |
| r.m.s. deviation from standard values | 0.006  |
| Bonds (Å)                     | 1.136 |
| Dihedral angles (degrees)     | 18.33 |
| Average B-factor (Å²)         | 21.87 |
| Residues in most favored regions (%) | 96.9 |
| Residues in additional allowed regions (%) | 95.6 |
| Residues in generously allowed regions (%) | 96.2 |
| Residues in disallowed regions (%) | 92.1 |

The Ramachandran plot was calculated by PROCHECK (19).

**DNase I Footprint Analysis**

The probes for DNase I footprinting experiments were prepared by PCR from C. glutamicum wild-type strain ATCC13032 genomic DNA using the specific primer pair PDF5’-PDF3’. These primers (PDF5’, 5’-AAGTGAACCTCACAAGATCGC-3’; PDF3’, 5’-TGCGGTGTA-GGAATCGAGT-3’) were designed for 110 bp upstream and downstream from positions −36 to −6 of the cgl2611 gene, respectively. Prior to PCR amplification, the 5’ termini of the primers of either the coding strand or the noncoding strand were labeled with [γ-32P]ATP (PerkinElmer Life Sciences) using T4 polynucleotide kinase (Takara Bio). Aliquots of 120 ng of labeled probes, 10 μg of salmon testis DNA (Sigma), and 150 μg of purified CGL2612 protein were mixed in 20 μl of reaction buffer (40 mM Tris–HCl, pH 7.5, 200 mM NaCl, 8 mM MgCl2, 5 mM dithiothreitol) and incubated for 20 min at 37 °C. After incubation, 0.25 units of DNase I (Takara Bio) in 0.5 μl of reaction buffer were added and mixed, and digestion was allowed to continue for 1 min at 37 °C. Digestion was terminated by the addition of phenol/chloroform, and reaction mixtures were extracted. DNA was precipitated with ethanol, and the resulting pellets were resuspended in 8 μl of loading dye. After heating at 90 °C for 1 min, aliquots were analyzed by electrophoresis on an 8 M urea-containing 6% polyacrylamide gel. Sequence ladders were generated using a Thermo sequence cycle sequence kit (Amersham Biosciences) with the same primer as was end-labeled for probe preparation.

**Construction of a cgl2612 Disruptant**

For disruption of the cgl2612 gene, the C. glutamicum wild-type strain ATCC31831 was used. A 2.1-kb DNA fragment encompassing the cgl2612 gene was amplified by PCR using the primers, 5’-AG-GAGCCAATTCCATCGTCC-3’ and 5’-GGACATACCTTGCA-CAGATGTC-3’. The amplified DNA fragment was digested with XbaI and PstI and then ligated with the vector plasmid pK18mobsacB (16) digested with XbaI-PstI. From the resulting plasmid, a 440-bp Xhol fragment encoding an internal part of cgl2612 was removed by Xhol digestion followed by self-ligation. The constructed plasmid, named pK18Δcgl2612, was introduced directly into ATCC31831 cells by electroporation, and kanamycin-resistant transformants were selected. As pK18Δcgl2612 did not have a replication origin functional in C. glutamicum cells, only cells in which homologous recombination occurred between the chromosomal DNA and plasmid DNA could grow on kanamycin-containing plates. The kanamycin-resistant transformants
thus carried the intact cgl2612 gene and the deleted cgl2612 gene in addition to the vector plasmid sequence on their chromosome. One of the kanamycin-resistant transformants was grown in L liquid medium without kanamycin for 16 h, and the cells were then spread on sucrose-containing L agar plates (20% sucrose). As cells carrying the sacB gene encoded by the vector plasmid were killed in the presence of sucrose, only those cells in which the sacB gene was excised from the chromosome by the second homologous recombination between the intact and deleted cgl2612 could grow on sucrose-containing plates. The resultant sucrose-tolerant recombinants should have the wild-type or the deleted cgl2612 gene, depending on the recombination point. Among sucrose-tolerant recombinants, the desired deletion mutant strain was selected by PCR amplification with the primers used for cloning of the cgl2612 gene. Chromosomal deletion of cgl2612 was further confirmed by Southern hybridization (data not shown). The constructed deletion mutant strain was named D2612.

Primer Extension Analysis

The transcriptional initiation site of the cgl2611 gene was determined by primer extension analysis. Total cellular RNA was isolated from exponentially growing cells of ATCC31831 or D2612 using an RNA...
isolation kit (RNeasy, Qiagen, Chatsworth, CA). Aliquots of 40 µg of total RNA were used for primer extension with Superscript II reverse transcriptase (Invitrogen) and the biotinylated oligonucleotide (5'-AGAAAGAGACCCGCTGTATAACGC-3'), which was complementary to a region within cgl2611. The primer extension products were separated on an 8 M urea-containing gel together with the sequencing ladders obtained using the same primer. Separated products were detected by chemiluminescence using a Phototope-Star detection kit (New England Biolabs Inc., Beverly, MA).

Antimicrobial Susceptibility Analysis

Minimal inhibitory concentrations (MICs) were determined using a standard 2-fold serial dilution format on L agar plates. Aliquots of 10 µl of stationary culture diluted in 0.85% NaCl to 10^6 cells/ml were spotted onto each plate. Following incubation at 30 °C for 24 h, the MIC was defined as the lowest concentration that resulted in >99.9% inhibition of colony formation.

RESULTS AND DISCUSSION

Overall Structure—A view of the overall structure of the CGL2612 dimer is shown in Fig. 1. Atomic models were successively built for each subunit (subunit A, Met1–Ser175; subunit B, Thr2–Pro174). The subunit was composed of nine helices (α1, Lys5–Glu19; α2, Tyr27–Thr34; α3, Lys38–His44; α4, Arg48–Ile70; α5, Pro72–Leu88; α6, Arg84–Ser105; α7, Pro107–Trp120; α8, Leu127–Tyr151; α9, Lys159–Glu171). In both native and Se-Met-substituted crystals, two subunits in the asymmetric unit were found to exist as a dimer.

The N-terminal three helices of CGL2612 protein form a compact DNA-binding domain that contains the typical helix-turn-helix motif (α2–α3), in which the latter helix was assumed to be a DNA recognition helix. The C-terminal portion of this protein consists of six helices and was assumed to act as a regulatory domain (RD). CGL2612 subunits form stable dimers related by 2-fold symmetry with this domain. C-terminal helices α2 and α6 from each subunit form a helix bundle structure, and hydrophobic interactions among residues in this bundle mainly take part in stabilization of the dimer. The RD and N-terminal DNA-binding domain of CGL2612 are connected by the α6 helix, and these two domains are associated with each other through a common hydrophobic internal core. Residues located at the edge of the hydrophobic core formed in the DNA-binding domain (Ile8, Leu10, and Ala13 in the α1 helix and Leu22 and Leu25 in the α3 helix) also participate in hydrophobic interaction with hydrophobic residues from RD (α2 and α6 helices). In addition to these hydrophobic interactions, a salt bridge was also observed between these domains (Arg14 in α1 and Asp62 in α6 helix).

Comparison with Other TetR Family Proteins—DALI structural similarity search (17) revealed that the overall structure of CGL2612 protein was similar to those of the TetR family proteins. Structural homologs of CGL2612 were EthR (Protein Data Bank number 1t5e; z-score = 11.7, root mean square differences = 3.6 Å for 162 Ca atoms; Protein Data Bank number 1u9n; z-score = 11.0, r.m.s. deviation = 4.0 Å for 160 Ca atoms) (18, 19), QacR (Protein Data Bank number 1jt0; z-score = 11.1, r.m.s. deviation = 4.0 Å for 159 Ca atoms) (20), YcdC (Protein Data Bank number 1bp6: z-score = 9.8, r.m.s. deviation = 4.0 Å for 164 Ca atoms), TetR (Protein Data Bank number 2ctt: z-score = 9.6, r.m.s. deviation = 4.5 Å for 151 Ca atoms) (21), and CprB (1ui5: z-score = 8.9, r.m.s. deviation = 3.6 Å for 145 Ca atoms) (22). This confirmed that the CGL2612 protein also belongs to the same family of proteins. The amino acid sequences of CGL2612 and its homologous proteins from the related bacteria, C. diphteriae and M. tuberculosis, are aligned in Fig. 2 (16–30% identity). As shown in the figure, DNA-binding domains of these proteins are well conserved and exhibit 24–61% identities in their primary structures. In particular, the residues around α9, a putative DNA recognition helix, are highly conserved. These conserved residues have already been shown to participate in the interaction with DNA molecules in DNA-protein complex structures of the other TetR family proteins, QacR (Protein Data Bank number 1jt0) (20) and TetR (Protein Data Bank number 1qpi) (23) (Fig. 2, residues marked d), and residues that may be involved in DNA-protein interactions in CGL2612 and its homologues can also be predicted (Fig. 2, residues marked d).

TetR family proteins are transcriptional repressors, and there are ligand molecules that induce dissociation of proteins from their conjugate DNA. Tertiary structures of related proteins showed that TetR family proteins commonly possess cavities in their RDs (18, 21, 22, 24, 25), and the structures of EthR-hydrophobic ligand, TetR-tetracycline Mg2+, and complexes of QacR with different compounds demonstrated that these cavities are the inducer-binding sites for the TetR family of proteins (19, 21, 24–26). The present structure of CGL2612 protein indicated that there is also a deep cavity in its RD, and the cavity is assumed to act as an inducer-binding site in this protein. This tentative binding site has two openings to bulk solvent in subunit A; one opening is framed by the helices α9, α9, and α9 whereas the other is framed by helices α9, α9, and the α9–α9 loop from another subunit (Fig. 3a). Around the latter opening, the α9–α9 loop covers the entrance of the cavity and narrows its opening. Main chain trajectories of subunits A and B are slightly different from each other (r.m.s. deviation for residues 3–174 is 1.00 Å). Especially, in subunit B, the main-chain trajectory of
the α₅-α₆ loop and α₆ helix is different from another subunit, and the latter opening is closed by the loop (Fig. 3b). This suggests that the corresponding region is rather flexible, and the former larger opening is likely to work as an inducer entrance of this cavity. The interior of the inducer-binding site of CGL2612 is surrounded by residues Leu59, Trp63, and Glu66 in α₄, Thr87, Leu88, Glu90, and Val92 in α₅ and α₅-α₆ loop, Glu96 and Leu100 in α₆, Trp113, Asn117, and Ile121 in α₇, Gln140, Asp144, and Phe147 in α₈, and residues from the other subunit (Ile152, His153, and Asp154) (Fig. 2, residues marked with a number symbol). As shown in Fig. 2, the inducer-binding site of this protein consists mainly of hydrophobic and acidic residues, and the site exhibits a significantly acidic nature (Fig. 3). Inducer molecules will dock into the binding site using affinity with this biased electrostatic surface potential.

Despite the similarities in their tertiary structures, TetR family proteins display poor amino acid sequence identity in their RDs, and also the tertiary structures of the RDs are slightly different from each other to adopt their divergent inducers. Tertiary structures of TetR family proteins showed that each protein has an inducer binding cavity of various shapes and sizes. Ligand complex structures of EthR, QacR, and TetR proteins have already revealed residues that interact with the bound ligands (19, 21, 24–26). These residues were spread on all helices framing the binding cavities, and those were particularly clustered around the region of α₅ and α₆ and the loop between these two helices (Fig. 2). In QacR and TetR proteins, it was shown that the binding of ligand molecule induces conformational changes around the area, and the structural changes cause relocation of DNA-binding domain to alter its DNA binding ability (20, 21, 24). Although the RDs of TetR family proteins show poor amino acid sequence similarity, primary structure alignment based on their secondary and tertiary structures revealed that the binding site-forming residues of CGL2612 and ligand-interacting residues for other homologues (marked with a number symbol and hatched in black, respectively) are located in similar positions in their primary structures, and especially, CGL2612 and QacR exhibit rather similar propensities (Fig. 2). TetR proteins act as dimer, and structural comparison in their dimer, the biological functional form, revealed that the QacR was the best homolog to CGL2612 (QacR, Protein Data Bank number 1jt0, r.m.s. deviation 3.8 Å for 317 C atoms; EthR, Protein Data Bank number 1t56, r.m.s. deviation 5.2 Å for 309 C atoms). At present, certain inducer-binding residues of CGL2612 protein are not known, and detailed mechanisms for DNA binding ability alternation after the binding of inducer molecule are also still unclear; nevertheless, such strong commonality observed in their structures implies that these proteins may share similar functional mechanisms. To discuss further details, ligand-complexed structure analysis of CGL2612 will be necessary.

**Identification of CGL2612 Binding Site**—The tertiary structure of CGL2612 showed that this protein is a member of the TetR family that is structurally quite similar to EthR and QacR. This raises the question of what types of genes are regulated by this protein. To identify the CGL2612 binding site on DNA, two different selections were examined. First, a random SELEX method was attempted, and this selection showed that the 16-bp random nucleotide core from isolated DNA frag-

**FIGURE 4.** Top, schematic representation of the structures of cgl2611 and cgl2612 genes and the cgl2611 promoter region. Bottom, sequence representation of the DNA fragments isolated by genomic and random SELEX analysis. In genomic SELEX, the consensus sequence composed of all isolated fragments is represented in boldface type. In random SELEX, positions homologous to the consensus sequence are highlighted in **boldface** type. The opposing two arrows indicate the half-site of the pseudoinverted-repeat sequence observed here. *, the symmetry center of the palindrome.
A structural and functional analysis of the CGL2612 protein was conducted. DNase I footprint analysis demonstrated that CGL2612 binds to the DNA at the position of IR1. Through SELEX, a 30-bp sequence was identified as the high-affinity binding site for CGL2612. The footprint analysis showed that CGL2612 binds to its binding site. The half-sites of inverted repeat and DNase I-sensitive bases in the center part of the binding site indicate the presence of a DNase I-accessible gap. It is expected that the CGL2612 molecule does not bind to this center part or interacts only weakly. The CGL2612 binding site is composed of an inverted repeat pair and a 4-bp spacer sequence between the repeated sequence sites. The DNase I footprint analysis provided another molecular basis for considering how CGL2612 dimers bind to its binding site and are separated each other, and a structural gap will be found in the symmetric center.

**FIGURE 5.** a, DNase I footprinting of CGL2612 protein in the cgl2611 promoter region (left, coding strand; right, noncoding strand). Areas protected by CGL2612 are indicated on the right of each lane. #, digestion bands observed in the protection area. Sequences in white characters represent DNase I-sensitive positions. Summary of the DNase I footprint assay is also represented below the gel image. The shaded box shows the protected area, and other marks are the same as in Fig. 4. b, schematic diagram of a CGL2612-DNA interaction model and a QacR-IR1 complex structure (20). The half-sites of inverted repeat and DNase I-sensitive bases on each strand are colored in red and blue, respectively. In the case of the QacR-IR1 complex, the symmetry center of IR1 was bound by two protein dimers from both sides and surely protected. On the other hand, taking into consideration the result obtained from footprint analysis, CGL2612 dimers bind to its binding site and are separated each other, and a structural gap will be found in the symmetric center.
half-site of the long inverted repeat sequence. Each equivalent subunit
in a dimer recognizes different DNA sequences comprised in the half-
site. To determine the details of how four CGL2612 subunits recognize
its binding DNA sequence, structural analysis of CGL2612-DNA com-
plex will be necessary.

Functional Role of CGL2612 Protein as a Transcriptional Repressor of
drug efflux MFS Permease—The CGL2612 binding site determined in
this study contains a putative promoter sequence, "−35" TTGTA and
"−10" TACAGT, and is located just upstream of the cgl2611 gene,
which probably encodes an MFS permease (Fig. 4). Thus, it was
expected that CGL2612 regulates the expression of the cgl2611 gene by
binding to its promoter region. However, the putative initiation codon
ATG (genomic map position 2,778,966−2,778,968) of cgl2611 is posi-
tioned very close to the putative promoter sequence, with an 8-bp spac-
ing sequence, and moreover, no SD-like sequences were found in this
region. To determine whether CGL2612 regulates expression of
cgl2611, primer extension analysis of cgl2611 transcripts was carried out
using the wild-type C. glutamicum strain, ATCC31831, and its cgl2612
deletion mutant derivative, D2612. D2612 was constructed by the dou-
bble-crossover method as described under "Experimental Procedures."

As shown in Fig. 6, a primer extension product initiated from the third
G of the putative initiation codon ATG was detected in the wild-type
strain ATCC31831 as well as the Δ2612 mutant strain D2612. This
indicated that the real initiation codon is located further downstream.
CTG (2,779,050−2,779,052) or ATG (2,779,161−2,779,163) could be
the real initiation codon of cgl2611, but this remains to be determined
experimentally. Primer extension analysis also showed that the amount
of cgl2611 transcript in the Δ2612 mutant cells was 2.2-fold higher than
that in the wild-type cells. This clearly indicated that CGL2612 func-
tions as a transcriptional repressor of cgl2611. The two genes, cgl2611
and cgl2612, are located in tandem, and they overlap by 10 bp, suggest-
ing that these genes form an operon (Fig. 4, top). Regulation of the
cgl2611 promoter by CGL2612 protein may suggest automodulation of
cgl2612.

CGL2611, the product of the cgl2611 gene, is composed of 494 amino
acid residues and is predicted to be a membrane protein based on its
amino acid sequence. CGL2611 protein has 14 α-helical segments that
traverse the cytoplasmic membrane, and a BLAST search (27) showed
that this protein has notable sequence similarity with tolin efflux
pumps that belong to the MFS family; MFS is one of the major families
to which a number of bacterial drug efflux pumps belong (1). CGL2611
exhibits 28.7% amino acid sequence identity with the S. aureus MFS
QacA multidrug efflux pump, which renders cells resistant to antimic-
robial compounds, such as quaternary ammonium compounds (Qacs),
via a proton motive force-dependent mechanism (28). Amino acid
inspection revealed that although residue Asp323 of QacA, which is the
key residue for the "high affinity" of QacA to divalent cations, is dis-
placed by serine in CGL2611 (Ser250), other intermembrane residues,
Asp34 and Arg114 in QacA, the crucial residues for drug export (28), are
perfectly conserved in CGL2611 protein (Asp34 and Arg114, respec-
tively), suggesting that CGL2611 also possesses strong potential to act as
a drug exporter. Since the Δ2612 mutant strain D2612 overproduces
CGL2611, it is expected that D2612 shows a drug-resistant phenotype.
To estimate the biological function of CGL2611, the antibacterial sus-
cceptibilities of the wild-type strain ATCC31831 and the Δ2612 mutant
strain D2612 were investigated. The sensitivities of ATCC31831 and
D2612 to several antibiotics, which are structurally and functionally
diverse and in common use, and substrates expelled by other homolo-
gous anti-drug systems EthA-EthR (ethidium bromide (Et)) and QacA-
QacR (Et, benzalkonium chloride, and malachite green) were examined.
As shown in TABLE THREE, D2612 showed increased resistance to
Norfloxacin, a quinolone-type DNA gyrase inhibitor. The increased rate
of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-
sion rate of Norfloxacin resistance (2-fold) was correlated well with the derepres-

### TABLE THREE

| Drug       | Targets          | MIC WT  | MIC D2612 |
|------------|------------------|---------|-----------|
| Norfloxacin| DNA synthesis    | 1.25    | 2.5       |
| Sparfloxacin| DNA synthesis    | 0.25    | 0.25      |
| Rifampin   | RNA synthesis    | 0.00125 | 0.00125   |
| Tetracycline| Protein synthesis| 5       | 5         |
| Chloramphenicol| Protein synthesis| 2.5     | 2.5       |
| Penicillin G | Peptidoglycan synthesis | 0.25  | 0.25      |
| Ethambutol | Mycolic acid layer formation | 5     | 5         |
| Benzalkonium chloride | Antiseptics | 0.5 | 1         |
| Ethenium bromide | Dye | 0.31 | 1.25     |
| Malachite green | Dye | 0.31 | 0.31     |

* C. glutamicum wild-type strain ATCC31831.
the protein. Our examination demonstrated that CGL2611 protein is also responsible for resistance to cationic compounds. Although D2612 showed no effect on malachite green resistance, the strain displayed 2-fold and 4-fold increased tolerance for benzalkonium chloride and Et, respectively (TABLE THREE). On the other hand, QacA does not exhibit resistance to anionic substrate such as Norfloxacin (29). These suggest that CGL2611 is a multidrug export protein whose spectrum of substrates is not exactly the same with QacA.

In this study, we identified CGL2612 protein, a hypothetical protein found by DNA analysis, as a transcriptional repressor of an upstream gene, cgl2611, encoding a probable MFS permease that has a nonnegligible effect on resistance for toxic compounds in C. glutamicum. These results will extend our understanding of the drug resistance mechanism not only in C. glutamicum and but also in other related bacteria. To identify the biological functions of this CGL2612 protein, we combined tertiary structure determination and DNA-binding sequence analysis using the SELEX method. It is worth noting that this strategy is one of the most convenient and powerful methods for analysis of functionally unknown transcription factors taking advantage of extensive genome information.

Acknowledgments—We thank Drs. M. Kawamoto and H. Sakai of SPring-8 (Hyogo, Japan) for help during data collection and K. Hatakeyama, N. Hirano, M. Takano, and N. Harada of the Frontier Research Center for Post-genomic Science and Technology, Hokkaido University for an introduction to random SELEX. We also thank Dr. N. Shi-mamoto and the members of his laboratory at the National Institute of Genet-

REFERENCES

1. Putman, M., van Veen, H. W., and Konings, W. N. (2000) Microbiol. Mol. Biol. Rev. 64, 672–693
2. Hillen, W., and Berens, C. (1994) Annu. Rev. Microbiol. 48, 345–369
3. Grkovic, S., Brown, M. H., Roberts, N. J., Paulsen, I. T., and Skurray, R. A. (1998) J. Biol. Chem. 273, 18665–18673
4. Aramaki, H., Yagi, N., and Suzuki, M. (1995) Protein Eng. 8, 1259–1266
5. Kawabata, T., Fukushi, S., Homma, K., Ota, M., Araki, J., Ito, T., Ichiyoshi, N., and Nishikawa, K. (2002) Nucleic Acids Res. 30, 294–298
6. Ito, H., Yao, M., Watanabe, N., and Tanaka, I. (2004) Acta Crystallogr. Sect. D 60, 1094–1100
7. Matthews, B. W. (1968) J. Mol. Biol. 33, 491–497
8. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
9. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D 55, 849 – 861
10. Terwilliger, T. C. (2000) Acta Crystallogr. Sect. D 56, 965–972
11. Terwilliger, T. C. (2002) Acta Crystallogr. Sect. D 59, 34–44
12. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A 47, 110–119
13. Novaza, J. (1994) Acta Crystallogr. Sect. A 50, 157–163
14. Laskowski, R. A., Mac Arthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
15. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., J., Kuszewski, J., Nilges, N., Pannun, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D 54, 905–921
16. Schafer, A., Tausch, A., Jager, W., Kalinowski, J., Thierbach, G., and Puhler, A. (1994) Gene (Amst.) 145, 69–73
17. Holm, L., and Sander, C. (1995) Trends Biochem. Sci. 20, 478–480
18. Dover, L. G., Corsino, P. E., Daniels, I. R., Cocklin, S. L., Tatituri, V., Besra, G. S., and Futterer, K. (2004) J. Mol. Biol. 340, 1095–1105
19. Frenois, F., Engo-hang-Ndong, J., Locht, C., Baulard, A. R., and Villerer, V. (2004) Mol. Cell 16, 301–307
20. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2002) EMBO J. 21, 1210–1218
21. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995) J. Mol. Biol. 247, 260–280
22. Natsume, R., Ohnishi, Y., Senda, T., and Horinouchi, S. (2004) J. Mol. Biol. 336, 409–419
23. Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Nat. Struct. Biol. 7, 215–219
24. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2001) Science 294, 2158–2163
25. Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., and Saenger, W. (1994) Science 264, 418–420
26. Murray, D. S., Schumacher, M. A., and Brennan, R. G. (2004) J. Biol. Chem. 279, 14365–14371
27. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997) Nucleic Acids Res. 25, 3389–3401
28. Brown, M. H., and Skurray, R. A. (2001) J. Mol. Microbiol. Biotechnol. 3, 163–170
29. Mitchell, B. A., Brown, M. H., and Skurray, R. A. (1998) Antimicrob. Agents Chemother. 42, 475–477
30. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
31. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524

Structural and Functional Analysis of CGL2612 Protein