Crystal Structure of the Caseinolytic Protease Gene Regulator, a Transcriptional Activator in Actinomycetes

Santina Russo, Jens-Eric Schweitzer, Tino Polen, Michael Bott, and Ehmke Pohl

Human pathogens of the genera Corynebacterium and Mycobacterium possess the transcriptional activator ClgR (clp gene regulator) which in Corynebacterium glutamicum has been shown to regulate the expression of the ClpCP protease genes. ClgR specifically binds to pseudo-palindromic operator regions upstream of clpC and clpP1P2. Here, we present the first crystal structure of a ClgR protein from C. glutamicum. The structure was determined from two different crystal forms to resolutions of 1.75 and 2.05 Å, respectively. ClgR folds into a five-helix bundle with a helix-turn-helix motif typical for DNA-binding proteins. Upon dimerization the two DNA-recognition helices are arranged opposite to each other at the protein surface in a distance of ~30 Å, which suggests that they bind into two adjacent major grooves of B-DNA in an anti-parallel manner. A binding pocket is situated at a strategic position in the dimer interface and could possess a regulatory role altering the positions of the DNA-binding helices.

Proteolysis in bacterial cells is a key cellular function. ATP-dependent self-compartmentalizing proteases, whose active sites are located in an inner cavity, play a dominant role in this process. Of these proteases, the serine protease caseinolytic protease (Clp)3 has been extensively studied both mechanistically and functionally. The Clp proteases are a crucial part of the cell’s protein quality control system disposing of truncated, misfolded, aggregated, or denatured proteins. In addition, Clp performs regulatory functions controlling the abundance of certain regulatory proteins in the cell. The activity of a number of regulatory proteins in the cell. The activity of a number of regulator proteins (e.g. ComK (1), SpoIIAB (2), Spx (3), and CtsR (4) in Bacillus subtilis, PopR in Streptomyces lividans (5)) is controlled through conditional proteolysis by Clp proteases.

The concentration of the Clp subunits in the cell must be strictly controlled to ensure the capability to respond adequately to stress conditions resulting in accumulation of truncated or misfolded proteins. In Gram-negative bacteria, including Escherichia coli, transcriptional activation of clp genes is mediated primarily by the action of the alternative sigma factor σ22, which binds to the RNA polymerase enhancing transcription (6). In contrast, in several Gram-positive bacteria with low G+C content, e.g. Bacillus subtilis, the transcriptional repressor CtsR controls the expression of the clpC, clpE, and clpP genes (7–9).

Recently, another transcriptional activator designated ClgR (for clp gene regulator) was discovered in Corynebacterium glutamicum, which activates expression of the clpP1P2 and clpC genes (10). According to genome-based searches, ClgR homologs are present in most members of the Actinomycetales, an order of high G+C Gram-positive bacteria that comprises prominent pathogenic species like Mycobacterium tuberculosis and Corynebacterium diphtheriae. Regulation of clp genes by ClgR was also demonstrated in S. lividans (11) and Bifidobacterium breve (12). The ClgR regulon has been analyzed in more detail in C. glutamicum (13). Besides clpP1P2 and clpC, ClgR binds to operator sequences of several genes that are involved in proteolysis and DNA repair. The binding sites of ClgR were determined by DNase I-footprint experiments, and a 18-bp consensus operator sequence (WNNWGGcYNANRGCGWW) has been proposed for the ClgR regulon in C. glutamicum. The pseudo-palindromic structure of the consensus sequence indicates that ClgR recognizes its target DNA as a dimer. In most cases, the ClgR binding site is located between position −39 and −67 relative to the transcriptional start site (+1). Only in the case of clpC the binding site lies further upstream stretching from position −121 to −138 (13). Thus, different intervals between the ClgR binding site and the core promoter appear to allow transcriptional activation by ClgR. In all bacteria containing ClgR the central consensus motif CGC-N5-GCG is found in the promoter upstream regions of clpP1P2 and clpC. Furthermore, it was shown that ClgR from M. tuberculosis is capable of replacing ClgR in C. glutamicum enhancing expression of the clpC and clpP1P2 genes as well (10). These results support the hypothesis that the regulation of clpC and clpP1P2 by ClgR is highly conserved within the actinomycetes. Similar to the clp gene regulators in Gram-negative and low G+C Gram-positive bacteria described above, convincing evidence was provided that the activity of ClgR is also controlled at the level of protein stability via ClpCP-mediated degradation (10, 14). However, neither the environmental stimuli nor the molecular mechanism leading to stabilization of ClgR are known at present.

ClgR of C. glutamicum consists of a polypeptide chain of 107 amino acids with a molecular mass of 11.3 kDa and a theoretical
pH of 5.4. The protein contains a helix-turn-helix type 3 (HTH 3) DNA-binding motif according to the Pfam database (15). It is most similar to the ClgR orthologs from *C. efficiens*, *C. diphtheriae*, and *C. jeikeium* with which it shares 91%, 75%, and 61% sequence identity, respectively. It shares 38% sequence identity with ClgR from *M. tuberculosis*.

The proteolytic core subunits of the Clp protease ClpP1 and ClpP2 are essential in *C. glutamicum* (10) and possibly also in other members of the actinomycetes. We ultimately seek to understand the regulation mechanisms controlling the expression of the corresponding genes at a level that enables us to interfere with it and thus control viability of the bacteria.

Here, we present the first crystal structure of a *clp* gene regulator, ClgR, from *C. glutamicum* solved from two different crystal forms to 1.75- and 2.05-Å resolution, respectively.

### EXPERIMENTAL PROCEDURES

**Protein Expression and Purification**—The cloning, expression, and purification of the ClgR construct used for crystallization was performed as described previously by Engels et al. (10). ClgR was expressed in *dnaK*-deficient *E. coli* BB1553 cells and purified by StrepTag-II affinity chromatography using a StrepTactin-Sepharose column (IBA). An additional washing step was carried out with a buffer containing 100 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 500 mM NaCl. Elution fractions containing ClgR were pooled, and buffer was exchanged against 20 mM Tris- HCl, pH 8.0, using Vivaspin concentrators with a cut-off of 5 kDa (Sartorius). The resulting protein possesses ten additional amino acids at the C terminus representing a linker and the StrepTag-II peptide. Protein concentrations were determined using a BCA protein assay kit (Pierce) and bovine serum albumin as standard. For the incorporation of selenomethionine by metabolic inhibition, the cells were cultivated in M9 minimal medium (42.3 mM Na$_2$HPO$_4$, 22.0 mM KH$_2$PO$_4$, 18.7 mM NH$_4$Cl, 8.6 mM NaCl, 1 mM MgSO$_4$, 22.2 mM glucose, 1.5 mM thiamine) additionally containing 15 mM FeSO$_4$ at 30 °C until the $A_{600}$ reached a value of 0.3. At this point the following amino acids were added to the cultures; 100 mg/liter l-lysine, 100 mg/liter l-phenylalanine, 100 mg/liter l-threonine, 50 mg/liter l-isoleucine, 50 mg/liter l-leucine, 50 mg/liter l-valine, and 50 mg/liter l-selenomethionine. After 15 min the expression was induced by addition of 3 mM isopropyl-$\beta$-d-thiogalactopyranoside and allowed to continue for 4 h. The selenomethionine-labeled ClgR was purified in the same way as the native protein (10). The integrity of the purified proteins was confirmed by mass spectrometry and gel electrophoresis.

**Crystallization and Data Collection**—Crystal screens of native ClgR were set up by a Phoenix crystallization robot (Art Robbins Instruments) using the sitting drop vapor diffusion method. 200 nl of protein solution (5 mg/ml) were mixed with 100 nl of reservoir solution (0.1 M NaCl, 23% 2-methyl-2,4-pentanediol, 15% glycerol, 0.085 M sodium-acetate, pH 4.6) and equilibrated against 50 μl of reservoir solution at 20 °C. Crystals with an average size of 50 × 20 × 20 μm$^3$ were obtained after 3 days. They belong to space group P2$_1$, with cell constants of $a = 65.7$, $b = 85.1$, $c = 71.4$ Å, and $β = 95.6°$.

Crystals of selenomethionine-labeled ClgR were obtained after 5 days by mixing 2 μl of protein with 1 μl of reservoir solution (0.0085 M CoCl$_2$, 0.85 M 1,6-hexanediol, 15% glycerol, 0.085 M sodium acetate, pH 4.6). The largest crystals had dimensions of 30 × 30 × 15 μm$^3$ and, as the native crystals, belonged to space group P2$_1$, displaying slightly different cell constants ($a = 66.9$, $b = 84.6$, $c = 72.6$ Å, and $β = 96.6°$).

While this report was in the final stage of preparation, a second ClgR crystal form was obtained from a crystallization experiment designed to crystallize a ClgR-DNA complex. The crystals with an average size of 60 × 40 × 40 μm$^3$ grew after 1 month from a solution containing 0.085 M HEPES, pH 7.5, 8.5% polyethylene glycol 8000, and 15% glycerol. They belong to space group P4$_3$2$_1$2$_1$ with cell constants of $a = b = 55.1$, $c = 129.6$ Å.

Crystals were mounted directly from the mother liquor and flash-cooled in a nitrogen stream at 100 K (16). Native data from the monoclinic crystals were collected to 2.05-Å resolution on beamline X06SA at the Swiss Light Source using the PILATUS 6M hybrid pixel detector (Dectris Ltd.) (17). Phases to 2.7-Å resolution were determined from a multiwavelength anomalous dispersion experiment carried out on a selenomethionine-labeled crystal on beamline X10SA at the Swiss Light Source using a Mar225 detector (Marresearch) (18). Diffraction data from the tetragonal crystal form were collected at beamline X10SA on a Mar225 detector to a resolution of 1.75 Å. All datasets were integrated and scaled with XDS (19). The data from the monoclinic crystal form were corrected for radiation damage using zero-dose extrapolation as implemented in XSCALE (20). Data collection statistics are summarized in Table 1.

**Structure Solution, Refinement, and Analysis**—Initially, retrieval of phase information was attempted using heavy atom labeling of the first obtained monoclinic crystals. The crystals were soaked in mother liquor containing different heavy atom compounds in concentrations ranging from 1 to 10 mM. The most promising result was obtained with a crystal soaked with 10 mM Na$_2$WO$_4$ overnight. A three-wavelength multiwavelength anomalous dispersion experiment was carried out across the tungsten absorption edge. Three tungsten positions per asymmetric unit could be determined with SHELXC/D (21) using the $F_A$ values calculated from all three datasets. The positions were refined, and phases were calculated using SHARP (22). The resulting map showed clear helical features; however it was not possible to trace the amino acid chain from this map. In parallel, selenomethionine-substituted protein was prepared and crystallized as described above. Apart from the N-terminal methionine, ClgR contains one methionine residue at position 95 of the amino acid sequence. Six selenium positions per asymmetric unit were located from a multiwavelength anomalous dispersion experiment to 2.7-Å resolution using SHELXD. Phases were calculated and extended to 2.1 Å with SHARP (22), refined by solvent flattening using SOLOMON (23) and further improved by likelihood-based density modification as implemented in RESOLVE (24). Initially, the ARP/wARP Quick Fold tool (25) was used to build most helical parts of the structure. The six molecules embodied in the asymmetric unit were then built manually using COOT (26). The structure was completed by iterative cycles of model building and crystallographic refinement with REFMAC5 (27) using medium main chain and

---

**Crystal Structure of the clp Gene Regulator ClgR**
loose side chain non-crystallographic symmetry restraints. Final refinement was performed against the native data to 2.05-Å resolution without the use of non-crystallographic symmetry restraints. The structure of the tetragonal crystal form was subsequently solved by molecular replacement employing the program PHASER (28) and a ClgR monomer from the solved monoclinic crystal structure as search model. The two molecules enclosed in the tetragonal asymmetric unit were refined with REFMAC5 to 1.75-Å resolution. The quality of the structures were verified with PROCHECK (29) and Molprobity (30). Comparisons of three-dimensional structures were performed with DALI (31). Multiple sequence alignments were generated using ClustalW (32) and ESPRIPT (33). Structure figures were created using the program PyMOL.4

EMSAs—EMSAs were performed as described previously (35). Different concentrations of purified ClgR were mixed with promoter DNA fragments of clpP1P2 and clpC (309 bp and 290 bp, respectively, final concentration 25 nM) containing the ClgR binding site (10, 13) in 50 mM Tris-HCl, 10% (v/v) glycerol, 50 mM KCl, 10 mM MgCl2, 0.5 mM EDTA, pH 7.5, in a total volume of 20 μl. A non-target promoter fragment from cg2311 (190 bp, final concentration 40 nM) served as negative control and was not shifted by ClgR.

RESULTS

Overall Structure—To confirm the DNA binding activity of ClgR as it was crystallized, EMSAs were performed with the promoter regions of the clpP1P2 and clpC genes containing the pseudo-palindromic ClgR binding sites. As shown in Fig. 1, ClgR is able to bind to the clpP1P2 promoter fragment in vitro without the requirement of an additional activator molecule or protein cofactor. A complete shift was observed at a 15-fold molar excess of ClgR, whereas the cg2311 promoter fragment serving as negative control was not shifted even at a 63-fold molar excess. Similar results as with the clpP1P2 promoter fragment were obtained with the clpC promoter fragment (data not shown).

The structure of the ClgR monoclinic crystal form was determined to a resolution of 2.05 Å. Six ClgR monomers are present in the asymmetric unit containing 4–7 additional residues of the C-terminal StrepTag-II. Chain A comprises residues 21–114, chain B residues 19–111, chain C residues 23–112, chain D residues 20–112, chain E residues 19–112, and chain F residues 21–113. Altogether, the final model consists of 557 amino acid residues, 3 MPD molecules, 1 acetate ion, and 188 water molecules. The side chains of some polar residues at the protein surface are invisible in the electron density and were therefore refined as alamines. All residues are in the allowed region of the Ramachandran plot (36), 99.5% of those are in the most favorite regions. The structure of the tetragonal crystal form was refined to a resolution of 1.75 Å. Here, the asymmetric unit contains two monomers (chains A and B) comprising residues 22–98 and 19–96, respectively, 5 glycerol molecules and 156 water molecules. Except for Lys-19 at the N terminus of subunit B, which was refined as an alanine, all side chains could be built in the electron density. All residues are in the most favorite region of the Ramachandran plot. Refinement statistics of the two models are summarized in Table 1.

The ClgR monomer folds into a five-helix bundle (Fig. 2), of which helix2 (H2) and helix3 (H3) build a helix-turn-helix (HTH) motif typical for DNA-binding proteins comprising residues 44–107 and 84–107 and continued by the additional C-terminal residues representing the two linker amino acids (Leu-108 and Glu-109) and part of the StrepTag-II (residues 110–114). Of the six monomers embodied in the asymmetric unit, molecules A and E and molecules B and C form two physiological dimers, respectively, and molecules D and F each form half of a dimer, which is completed by their symmetry equivalents. In the crystal structure, the C-terminal H5 helices are involved in crystal contacts forming extensive hydrophobic interactions with each other. The H5 helices of subunits A and B build a coiled-coil as do the H5 of subunits C and D, and E and F, respectively. Remarkably, the C-terminal StrepTag-II residues (residues 108–114) contribute to the crystal contacts by composing short four-helix bundles with the H5 helices of the neighboring dimer and its coiled-coil partner.

The subsequently solved tetragonal crystals contain two monomers per asymmetric unit, which form a physiological dimer. Overall, the structure is the same five-helix bundle as seen in the monoclinic crystals showing <1 Å root mean square deviations (r.m.s.d.s) of Cα atoms. However, substantial differences occur at the C termini of the polypeptides. Here, the C-terminal helices H5 and H5’ are considerably shorter, only comprising residues 84–95 (Fig. 2, C and D). The C-terminal residues 99–107 of chain A and 97–107 of chain E are invisible in the electron density.

Dimer Structure—The previous finding that ClgR binds to pseudo-palindromic DNA sequences (13) already suggested that ClgR binds to DNA as a dimer. Gel-filtration experiments

4 DeLano, W. L. (2002) PyMOL, DeLano Scientific, San Carlos, CA.
crystals (subunits B and C). The view is approximately perpendicular to the 2-fold axis of the dimer. The N and C termini are labeled light red, respectively. The view is approximately perpendicular to the 2-fold axis of the dimer. C. ClgR dimer from tetragonal crystals (subunits A and B). The view is approximately perpendicular to the 2-fold axis of the dimer. D. same molecule as in C. The view is approximately along the 2-fold axis of the dimer.

FIGURE 2. Ribbon schematic of the physiological ClgR dimers. One subunit is shown in dark colors, and all secondary structure elements are annotated. The second subunit is shown in light colors, and the corresponding DNA-recognition helix H3 is annotated. The DNA recognition helices H3 and H3′ are depicted in red and light red, respectively. The N and C termini are labeled N and C, respectively. A. ClgR dimer from monoclinic crystals (subunits B and C). The view is approximately perpendicular to the 2-fold axis of the dimer. B. same molecule as in A. The view is approximately perpendicular to the 2-fold axis of the dimer. C. ClgR dimer from tetragonal crystals (subunits A and B). The view is approximately perpendicular to the 2-fold axis of the dimer. D. same molecule as in C. The view is approximately along the 2-fold axis of the dimer.

confirmed that ClgR as used for the gel-shift assay shown in Fig. 1 is a dimer in solution (data not shown). In the crystal structures, the physiological dimers are immediately recognizable by their tight dimer interface burying 740 Å² of solvent-accessible surface per monomer, of which 67% is of hydrophobic nature. Furthermore, the relative positioning of the two DNA-binding helices H3 and H3′ in a distance of ~30 Å from each other indicates that H3 and H3′ could well bind into two adjacent major grooves of slightly bent canonical B-DNA (Fig. 2). An electrostatic potential calculation shows that the surface of H3 and H3′ displays a positive electrostatic potential whereas the rest of the molecule predominantly exhibits a negative electrostatic potential (data not shown). The dimer interface is composed of a number of polar and non-polar contacts (indicated in Fig. 3). On the dimer surface opposite to the C-terminal H5 and H5′ helices, several hydrophilic interactions are formed by residues of the N-terminal regions of H4 and of the loops between helices H3 and H4, and by

| Table 1: Data collection and refinement statistics |
|--------------------------------------------------|
| **Native** | **Peak** | **Inflection point** | **Remote** | **Native** |
| Crystal parameters | P<sub>2</sub><sub>1</sub>; a = 65.4, b = 84.8, c = 71.4 Å, β = 95.9° | P<sub>2</sub><sub>1</sub>; a = 66.9, b = 84.6, c = 72.6 Å, β = 96.6° | P<sub>4</sub>2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>; a = b = 55.1, c = 129.6 Å |
| Data collection (XDS) | 6 molecules/a.s.u. | 6 molecules/a.s.u. | 2 molecules/a.s.u. | X10SA, SLS |
| Beamline | X06SA, SLS | X10SA, SLS | X10SA, SLS |
| Wavelength (Å) | 1.000 | 0.9790 | 0.9717 | 0.9717 |
| Resolution range (Å) | 46–2.05 (2.15–2.05) | 30–2.70 (2.93–2.70) | 30–2.55 (2.72–2.55) | 30–2.40 (2.54–2.40) |
| No. of reflections | 179,294 | 167,884 | 200,184 | 232,866 |
| No. of unique reflections | 48,625 | 43,447<sup>a</sup> | 51,594<sup>a</sup> | 60,426<sup>a</sup> |
| Completeness (%) | 99.7 (99.8) | 99.9 (99.9) | 99.9 (99.9) | 97.6 (85.5) |
| R<sub>merge</sub> (%) | 5.2 (69.9) | 1.361 | 11.8 (2.3) | 12.8 (2.5) |
| R<sub>r.m.s.d.</sub> (%) | 6.1 (81.6) | 13.2 (75.0) | 9.0 (60.7) | 8.1 (62.1) |
| Ramachandran favored (%) | 99.6 | 100.0 | 99.9 | 100.0 |
| Ramachandran outliers (%) | 0.0 | 0.0 | 0.0 | 0.0 |
| Solvent water molecules | 187 | 187 | 187 | 187 |
| No. non-hydrogen atoms | 4,284 | 1,320 | 171.4 Å, 184.6, 184.8, 187.4, 190.1 Å |
| No. reflections test set | 51,594 | 51,594 | 51,554 | 51,554 |
| No. reflections working set | 200,184 | 200,184 | 202,184 | 202,184 |
| Resolution (Å) | 46.0–2.05 | 42.0–1.75 | 42.0–1.75 |
| R<sub>work</sub> (electrons Å<sup>-1</sup>) | 4.7 | 4.7 | 4.7 | 4.7 |
| R<sub>free</sub> (%) | 21.2 | 23.8 | 18.3 | 21.3 |

<sup>a</sup> a.s.u, asymmetric units.
<sup>b</sup> The values in parentheses of resolution range, completeness, and I/σ(I) correspond to the outermost resolution shell.
<sup>c</sup> Friedel pairs were treated as different reflections.
<sup>d</sup> R<sub>merge</sub> = Σ<sub>h</sub> |Σ<sub>i</sub> I(h)<sub>i</sub> − |I(h)<sub>j</sub>| / Σ<sub>i</sub> |I(h)<sub>i</sub>|, where I(h)<sub>i</sub> is the jth measurement of the intensity of the unique reflection h, and |I(h)<sub>j</sub>| is the mean over all symmetry-related measurements.
<sup>e</sup> R<sub>r.m.s.d.</sub> = Σ<sub>h</sub> √(|n<sub>h</sub>|/|n<sub>h</sub>|) |I(h)<sub>j</sub>|/Σ<sub>i</sub> |I(h)<sub>i</sub>|, where |n<sub>h</sub>| is the multiplicity and |I(h)<sub>j</sub>| is the mean over all symmetry-related measurements (34).
the N-terminal ends of the first helices H1 (Fig. 4). The side chain of Glu-67 forms two hydrogen bonds with Ser-70, one with the main-chain amide group and one with the side-chain hydroxyl group. The Ser-70 side chain is in addition the acceptor of a hydrogen bond with the main-chain amide nitrogen of Arg-26. The side chain of Arg-26 forms a salt bridge with the carboxyl group of Glu-71. Additionally, the carbonyl moiety of Glu-22 forms two hydrogen bonds with the main-chain amino group of Val-84 and Ala-85 located at the beginning of the C-terminal helix H5. These interactions are formed by both subunits and follow the internal 2-fold symmetry of the dimer.

Moreover, there is a strong network of hydrophobic contacts in the core of the dimer involving residues of helices H1 (Leu-25 and Leu-29), H4 (Leu-73 and Ala-74), and H5 (Val-84, Leu-88, and Ile-89). These interactions between physiological dimers are present in all three dimers of the monoclinic crystal form as well as in the tetragonal structure solved from the monoclinic crystal form. The only exceptions are the hydrogen bonds between Glu-22 and the main-chain nitrogens of residues 84 and 85, which in the tetragonal structure are formed only on one side of the dimer. A multiple sequence alignment of ClgR and homologous proteins (Fig. 3) shows that, except Glu-22, which is an alanine in M. tuberculosis ClgR (Ala-2), all polar amino acids involved in dimer contacts are strictly conserved among ClgR orthologs. Also the hydrophobic residues that are part of the dimer interface display a high degree of conservation. It is therefore likely that the dimer interface and the resulting orientation of the dimer subunits are conserved among the ClgR proteins.

### The Putative Binding Pocket

In all three dimers present in the monoclinic asymmetric unit a 2-methyl-2,4-pentanediol (MPD) molecule from the crystallization buffer is bound in a pocket formed by the dimer interface (Fig. 5A). The pocket is 6 Å wide and could thus accommodate a larger molecule. A 3 Å wide opening connects the pocket with the protein surface in the center between H3 and H3'. The hydrophobic core part of the pocket is formed by Leu-25, Leu-29, Val-84 of both subunits. In addition, at the surface of the dimer both subunits form hydrogen bonds to the hydroxyl groups of the MPD molecule via the side-chain carboxyl group of Glu-67 and the main-chain amino group of Val-68. Fig. 5B shows the binding pocket formed by the tetragonal dimer structure. Here, the loop between H3 and H4 of subunit A is slightly shifted along the dimer interface and toward the dimer partner compared with the monoclinic structure. Especially the different positions of Val-68 and Ser-69, which are shifted by ~1.5 Å, result in a reduction of the cavity volume to ~6 × 6 Å. Along with two ordered water molecules, the pocket contains strong positive difference density indicating the presence of another
molecule that we could not identify until now. The position of the binding pocket directly in the dimer interface and the observed shift of the H3H4 loop between the two crystal forms may suggest a regulating role that will be discussed in detail below.

The ClgR monomer structure resembles the 434 cro family of DNA-binding proteins. It is most similar to C.BclI (37) (PDB code 2b5a) and C.AhdI (38) (PDB code 1y7y), two bacterial restriction modification controller proteins. Like ClgR, these proteins are presumably functional dimers. The monomer folds are conserved showing larger deviations only at the termini of the polypeptides. Especially the long C-terminal helix 5 comprising 24 amino acid residues is unique to ClgR. However, as mentioned above, the long H5 could only be observed in the monoclinic crystal form and is stabilized by crystal contacts. According to DALI (31) C.BclI and C.AhdI both share 26% structure based sequence identity with ClgR. Both monomer structures superimpose well with ClgR, showing r.m.s.d.s of C atoms of 1.7 Å for 74 and 68 equivalent atoms, respectively. However, the relative position of the second subunit in the ClgR dimer differs considerably compared with C.BclI and C.AhdI. It is shifted by ~8 Å along the dimer interface and rotated by ~10° along the 2-fold axis of the dimer compared with the dimer arrangements of C.BclI and C.AhdI (Fig. 4). The different dimer conformation results from the polar interactions at the dimer interface of ClgR that contribute to the adopted dimer orientation. Salt bridges and hydrogen bonds are shown as black dashed lines. Arg-26, Glu-67, and Glu-22 of subunit A interact with Glu-71, Ser-70, and main-chain nitrogens of Val-84 and Ala-85 of subunit E. The same interactions are present between Arg-26, Glu-67, and Glu-22 of subunit E and Glu-71, Ser-70, Val-84, and Ala-85 of subunit A.

FIGURE 4. Superposition of ClgR and C.BclI. The HTH motif of subunit A of ClgR (residues 44–63) has been superimposed to the HTH motif of one subunit of C.BclI (residues 26–45). A, stereo view of the superposed dimers. The ClgR dimer (subunits A and E) is depicted in blue; the C.BclI dimer is shown in red. The Cα-trace schematic shows the overall differences in orientation of the subunits. ClgR subunit E, which was not used to calculate the least-square superposition matrix, is shifted by ~8 Å and rotated by ~10° compared with the second C.BclI subunit. B, close-up stereo view showing the amino acids involved in polar interactions at the dimer interface of ClgR that contribute to the adopted dimer orientation. Salt bridges and hydrogen bonds are shown as black dashed lines. Arg-26, Glu-67, and Glu-22 of subunit A interact with Glu-71, Ser-70, and main-chain nitrogens of Val-84 and Ala-85 of subunit E. The same interactions are present between Arg-26, Glu-67, and Glu-22 of subunit E and Glu-71, Ser-70, Val-84, and Ala-85 of subunit A.

FIGURE 5. Close-up view of the binding pockets. A, surface representation of the binding pocket containing MPD as seen in the monoclinic crystals. The MPD molecule (carbon atoms, black) and the amino acid side chains forming the pocket (carbon atoms, gray) are shown as sticks. Residues 23–28 from chain A and residues 69–71 from chain E were omitted for clarity. Hydrogen bonds are depicted as dashed lines. The position of the opening leading to the surface of the dimer is indicated with a black star. The electron density is unbiased weighted difference density contoured at 2.8 above the mean (Fourier coefficients: DELFWT and PHDELWT as calculated by REFMACS (27)). B, surface representation of the binding pocket containing positive difference density (Fstructure − |Fcalc|; 3.0σ) as seen in the tetragonal crystals. The residues forming the cavity are shown as yellow sticks. Water molecules are shown as red spheres. Residues of the monoclinic ClgR dimer superposed to the tetragonal ClgR dimer (residues 25–95 were used to calculate the superposition matrix) are shown as gray sticks.
Crystal Structure of the clp Gene Regulator ClgR

hydrogen bonds formed by Glu-22 with the N-terminal part of H5 seem to stabilize the orientation adopted by ClgR. As mentioned above, the C-terminal parts of the H5 helices are involved to a large extent in crystal packing contacts. H5 and H5′ laterally protrude out of the dimer body on opposite sides almost perpendicularly by 18 Å each not including the residues introduced by cloning (Fig. 2, A and B). In the tetragonal crystal form, the C termini adopt a different conformation with a shorter H5 and 9–11 flexible C-terminal residues. Additionally, in both crystal forms the N termini are not visible in the electron density and probably adopt multiple conformations.

DISCUSSION

The most striking feature of the ClgR monomer structure presented here compared with similar proteins is the long C-terminal helix 5 seen in the monoclinic crystal form. As mentioned above, the C-terminal parts of the H5 helices are involved to a large extent in crystal packing contacts. H5 and H5′ laterally protrude out of the dimer body on opposite sides almost perpendicularly by 18 Å each not including the residues introduced by cloning (Fig. 2, A and B). In the tetragonal crystal form, the C termini adopt a different conformation with a shorter H5 and 9–11 flexible C-terminal residues. Additionally, in both crystal forms the N termini are not visible in the electron density and probably adopt multiple conformations.

DNA Binding—The crystal structure of the ClgR dimer clearly suggests that the putative DNA-recognition helices H3 and H3′ bind into two adjacent major grooves of B-DNA. This notion is confirmed by the positive electrostatic potential displayed by the protein surface of H3 and H3′ and the loops connecting them to H2 and H2′, the N-terminal helices of the helix-turn-helix motifs. Compared with transcriptional regulators for which a protein-DNA complex structure has been determined, the ClgR monomer is structurally most similar to the bacteriophage 434 repressor (39, 40) and the 434 cro repressor (41, 42) with which it shares its basic five-helix bundle fold. However, both the 434 repressor and the 434 cro repressor possess a dimer interface that, similar to C.BclI (Fig. 4), positions the second subunits slightly tilted and substantially shifted by ~13 Å compared with ClgR. Interestingly, the spatial arrangement of the ClgR helix-turn-helix motifs in the dimer is most similar to the λ phage cro repressor bound to a 19-bp DNA operator fragment (43) (PDB code 6cro). The λ cro repressor consists of three helices, which are the equivalents to ClgR helices 1–3, and a β-sheet dimerization domain. In the N-terminal three helices ClgR and λ cro only share three identical residues in a structure-based sequence alignment (Fig. 3). The conserved residues are involved in the hydrophobic core stabilizing the helices. The amino acid sequences building the HTH motifs of the two regulators show no similarity.

Although the dimerization interfaces of ClgR and λ cro are completely different, the position and orientation of the helix-turn-helix domains are quite similar. Fig. 6 shows a superposition of the ClgR dimer with the structure of λ phage cro repressor-operator complex. The HTH motif of subunit A of ClgR was superimposed to the HTH motif of one subunit of λ cro. It is remarkable that, although subunit E, which was not used to calculate the least-square superposition matrix, is slightly shifted and turned compared with the λ cro dimer, the DNA-binding helices of the two regulators are positioned very similarly. Contrary to similar superpositions with the 434 repressor and 434 cro repressor, there is no displacement at all along the dimer interface. However, H3 of ClgR subunit E is shifted by ~3 Å along the dimer 2-fold axis away from the DNA. In addition, the angle between DNA recognition helices is different. In λ cro this angle measures ~53°, whereas in ClgR H3 and H3′ are rotated ~75° relative to each other.

In the λ cro structure the DNA-recognition helices are separated by ~30 Å, which results in bending of the bound DNA by 40° toward the protein. In addition, there are substantial perturbations throughout the operator fragment relative to standard B-DNA (43). The minor groove at the center of the complex becomes sharply compressed and deepened. Because the positions of the DNA-binding helices are alike in λ cro and ClgR, we conclude that ClgR probably binds to DNA similar as λ cro inserting the DNA recognition helices H3 and H3′ in an anti-parallel way into the major grooves of the DNA fragment. Accordingly, it is likely that upon binding to ClgR the DNA undergoes considerable bending toward the dimer and also compression of the central minor groove. Presumably, this distortion is an important feature for the mechanism of transcription activation.

In λ cro the amino acid residues involved in specific hydrogen bond interactions with bases in the major groove of the DNA are Gln-27, Ser-28, Asn-31, and Lys-32. The side chains of Gln-16, Tyr-26, His-35, Arg-38, and Lys-56 additionally interact with the phosphate backbone. A superposition with the cro HTH reveals equivalent amino acids in ClgR. The positions of Asn-31 and Lys-32 are taken by Ser-59 and Glu-60 in ClgR.
Additionally, Ser-54 (positioned equivalent to Tyr-26 in cro) and Tyr-57 are positioned appropriately for interacting directly with bases. An array of four arginine residues (Arg-34, Arg-37, Arg-45, Arg-63, and Arg-65) as well as Thr-43 at the N-terminal end of H2 is prepared to interact with the DNA phosphate backbone. Except of Arg-34, which is a lysine (Lys-35) in C. jejikeium ClgR, all amino acids proposed to be involved in DNA binding are strictly conserved in ClgR orthologs (Fig. 3). Nevertheless, to define the specific interactions in detail the high resolution structure of a ClgR-DNA complex is required.

**Putative Binding Pocket**—As mentioned above, a putative binding pocket is formed at the dimer interface of ClgR, which in the monomeric structure is occupied by a solvent molecule from the crystallization buffer (Fig. 5). This is not unusual; many crystal structures contain bound molecules from the crystallization buffer, and MPD is frequently found in crystals. Nevertheless, the MPD binding mode in ClgR reveals that the pocket provides a large hydrophobic area inside the dimer as well as hydrogen bond partners from the protein surface. It also shows that the pocket provides more space than used by MPD and that it could easily accommodate a bulkier molecule. Moreover, the tetragonal structure contains a bound molecule in the cavity as well. Superpositions show that the tetragonal dimer is slightly rotated outwards compared with the first structure, so that, although the cavity volume is decreased, the distance between the DNA-recognition helices H3 and H3' is increased by ~1.5 Å. Sequence alignments indicate that the residues involved in MPD binding are conserved among the ClgR orthologs (Fig. 3). Other homologous transcriptional regulators of known structure do not possess such a binding pocket. Because ClgR binds its target DNA sequence without the addition of any activating small molecule (Fig. 1) it is highly likely that the binding pocket is not a direct activation site. Still, the strategic location of the pocket in the dimer interface together with the observed subunit movements may suggest a regulatory role, because the orientation of the dimer subunits, and with this the orientation of the two DNA-recognition helices, is certainly crucial for the DNA-binding mode. Therefore, it is possible that the binding pocket provides a fine tuning site for DNA binding, which may alter the affinity of the ClgR dimer to specific DNA sequences.

It has been shown previously that ClgR activity in C. glutamicum is controlled by degradation through the ClpCP protease (10). However, the degradation signal in C. glutamicum ClgR has not yet been identified. In the homologous proteins ClgR and PopR from Streptomyces species two alanine residues at the C terminus have been identified to be the degradation signal (5, 11), however this motif is not present in C. glutamicum ClgR. It is therefore not known by what mechanism ClgR is protected from degradation and thus allowed to bind to DNA. The different conformations of the C termini seen in the two crystal structures might indicate that the C terminus of ClgR possesses an intrinsic flexibility maybe needed for the controlled degradation of ClgR.

Clearly, more investigations are needed to understand how the activity of ClgR is regulated. Also, the mechanism by which ClgR acts as transcriptional activator remains to be unraveled.
Crystal Structure of the clp Gene Regulator ClgR

31. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
32. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
33. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) Bioinformatics 15, 305–308
34. Diederichs, K., and Karplus, P. A. (1997) Nat. Struct. Biol. 4, 269–275
35. Wennerhold, J., Krug, A., and Bott, M. (2005) J. Biol. Chem. 280, 40500–40508
36. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Proteins 50, 437–450
37. Sawaya, M. R., Zhu, Z., Mersha, F., Chan, S. H., Dabur, R., Xu, S. Y., and Balendiran, G. K. (2005) Structure 13, 1837–1847
38. McGeehan, J. E., Streeter, S. D., Papapanagiotou, I., Fox, G. C., and Kneale, G. G. (2005) J. Mol. Biol. 346, 689–701
39. Shimon, L. I., and Harrison, S. C. (1993) J. Mol. Biol. 232, 826–838
40. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) Science 242, 899–907
41. Mondragon, A., and Harrison, S. C. (1991) J. Mol. Biol. 219, 321–334
42. Wolberger, C., Dong, Y. C., Ptashne, M., and Harrison, S. C. (1988) Nature 335, 789–795
43. Albright, R. A., and Matthews, B. W. (1998) J. Mol. Biol. 280, 137–151
Crystal Structure of the Caseinolytic Protease Gene Regulator, a Transcriptional Activator in Actinomycetes
Santina Russo, Jens-Eric Schweitzer, Tino Polen, Michael Bott and Ehmke Pohl

J. Biol. Chem. 2009, 284:5208-5216.
doi: 10.1074/jbc.M806591200 originally published online November 19, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M806591200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 43 references, 8 of which can be accessed free at http://www.jbc.org/content/284/8/5208.full.html#ref-list-1