THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 32, Issue of August 12, pp. 29073–29079, 2005

*Printed in U.S.A.*

# The Crystal Structure of Mlc, a Global Regulator of Sugar Metabolism in *Escherichia coli**

Received for publication, April 18, 2005, and in revised form, May 31, 2005
Published, JBC Papers in Press, June 1, 2005, DOI 10.1074/jbc.M504215200

**André Schiefner‡, Kinga Gerber, Sabine Seitz, Wolfram Welte, Kay Diederichs, and Winfried Boos**

*From the Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany*

**Mlc from *Escherichia coli* is a transcriptional repressor controlling the expression of a number of genes encoding enzymes of the phosphotransferase system (PTS), including *ptsG* and *manXYZ*, the specific enzyme II for glucose and mannose PTS transporters. In addition, Mlc controls the transcription of *malT*, the gene of the global activator of the *mal* regulon. The inactivation of Mlc as a repressor is mediated by binding to an actively transporting PtsG (EIICB$^{Glc}$). Here we report the crystal structure of Mlc at 2.7 Å resolution representing the first described structure of an ROK (repressors, open reading frames, and kinases) family protein. Mlc forms stable dimers thus explaining its binding affinity to palindromic operator sites. The N-terminal helix-turn-helix domain of Mlc is stabilized by the amphipathic C-terminal helix implicated earlier in EIICB$^{Glc}$ binding. Furthermore, the structure revealed a metal-binding site within the cysteine-rich ROK consensus motif that coordinates a structurally important zinc ion. A strongly reduced repressor activity was observed when two of the zinc-coordinating cysteine residues were exchanged against serine or alanine, demonstrating the role of zinc in Mlc-mediated repressor function. The structures of a putative fructokinase from *Bacillus subtilis*, the glucokinase from *Escherichia coli*, and a glucomannokinase from *Arthrobacter* sp. showed high structural homology to the ROK family part of Mlc.**

Mlc (makes large colonies) has been discovered as a regulator protein curbing the utilization of glucose in *Escherichia coli* (1, 2). Mlc, acting as a transcriptional repressor, controls the expression of *malT*, encoding the central transcriptional activator of the maltose system (3). In addition, Mlc controls the expression of two operons encoding PTS[1]-dependent transporters for glucose *ptsG* (4, 5) and mannose *manXYZ* (6) as well as the genes encoding the general components of the PTS (7–9). In contrast to the classical mode of repressor inactivation by a cognate inducer, Mlc is inactivated by the sequestrating interaction with the actively transporting glucose transporter, the EIICB$^{Glc}$ protein of the PTS (10–12). The interaction occurs at the EIIB$^{Glc}$ domain of the transporter encompassing a critical cysteine residue (Cys-421). This cysteine residue is phosphorylated in the resting transporter and becomes readily dephosphorylated during glucose transport by the transfer of the phosphoryl group onto the incoming glucose. Mlc binds only to the dephosphorylated form of EIIB$^{Glc}$ (13). The membrane-bound state of EIIB$^{Glc}$ is essential for Mlc inactivation. Soluble EIIB$^{Glc}$, even though able to interact with Mlc (12, 13), does not prevent Mlc from binding to its operator regions and from its repressing activity. However, EIIB$^{Glc}$ attached to the membrane by any lipophilic anchor, even unrelated to EIICB$^{Glc}$, binds Mlc in a fashion that prevents binding to the operator regions (13). This indicates that Mlc, when it is in close contact with the membrane, alters its conformation to suppress operator binding.

As judged by its amino acid sequence, Mlc belongs to the ROK family (repressors, open reading frames, and kinases) (14, 15) of transcriptional regulators encompassing xylose repressors, sugar kinases, and transcriptional regulators with the widely conserved C*X*CG*XX*GC*X*E motif (consensus sequence 2). They also harbor another consensus motif (consensus sequence 1) consisting of 28 amino acid residues, located 9 residues upstream from consensus sequence 2 (15). The DNA-binding motif of Mlc consists of a typical helix-turn-helix motif at its N terminus, and the protein behaves in dilute buffer solution as tetramer of a polypeptide of 44.3 kDa (12, 13). The removal of the 18 C-terminal residues leads to dimer formation, to the loss of EIICB$^{Glc}$ binding, as well as to the loss of operator interaction (13). Thus, most surprisingly, the C terminus, which is far from the helix-turn-helix motif in the primary sequence, must be involved directly or indirectly, possibly via a large conformational change in EIICB$^{Glc}$ binding as well as in operator recognition and subsequent repression.

Regarding its unusual mechanism of derepression, it was of interest to elucidate the crystal structure of this novel transcriptional regulator. Here we report the three-dimensional structure of dimeric Mlc R52H at 2.7 Å resolution.

## MATERIALS AND METHODS

*Structure Determination and Refinement*—Mlc was cloned, expressed, purified, and crystallized as described previously (16). Three data sets were collected on a selenomethionine-labeled Mlc crystal at the Swiss Light Source SLS Villigen (CH) beamline X06SA. The crystals of space group C2 with unit cell parameters of $a$ = 235.95 Å, $b$ = 74.71 Å, $c$ = 154.95 Å, $\beta$ = 129.15° diffracted to a maximum resolution of 2.7 Å. The raw data were reduced using XDS (17). Because of the radiation sensitivity of the crystals, only the peak and inflection data sets were used for structure solution and refinement (Table I). The

* This work was supported by the Deutsche Forschungsgemeinschaft. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*The atomic coordinates and structure factors (code 1Z6R) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

‡ To whom correspondence should be addressed: Dept. of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany. Tel.: 49-7531-882289; Fax: 49-7531-883183; E-mail: andre.schiefner@uni-konstanz.de.

[1] The abbreviations used are: PTS, phosphotransferase system; Ec-GlcK, glucokinase from *E. coli*; Bs-FrcK, putative fructokinase from *B. subtilis*; As-GMK, inorganic polyphosphate/ATP-glucomannokinase *Arthrobacter* sp.; ROK, repressors open reading frames and kinases; HTH, helix-turn-helix; r.m.s., root mean square; PDB, protein data bank; wt, wild type; MES, 4-morpholineethanesulfonic acid.

TABLE I
*Crystal data and x-ray data collection statistics for a single SeMet-Mlc crystal*

Numbers in parentheses refer to the highest resolution shell. $R_{meas}$, the redundancy independent $R_{merge}$ and $R_{mrgd\text{-}F}$, a measure for quality of the reduced amplitudes, were calculated according to Diederichs and Karplus (46). The Friedel pairs were merged in the merged data set also.

| Data set | Peak | Inflection | Merged |
|---|---|---|---|
| Wavelength (Å) | 0.9786 | 0.9787 | |
| Total rotation range (°) | 180 | 180 | 2 × 180 |
| Resolution limits (Å) | ∞-2.9 (3.0–2.9) | ∞-3.0 (3.1–3.0) | ∞-2.7 (2.8–2.7) |
| No. of reflections | 174,722 | 157,372 | 398,481 |
| Unique reflections | 89,905 | 80,926 | 57,176 |
| Completeness (%) | 98.5 (99.7) | 98.1 (99.6) | 98.4 (93.6) |
| $I/\sigma(I)$ | 9.9 (2.3) | 9.5 (2.1) | 12.2 (2.0) |
| $R_{sym}$ | 5.4 (36.6) | 5.8 (41.3) | 10.5 (59.8) |
| $R_{meas}$ | 7.5 (50.4) | 8.0 (56.7) | 11.4 (71.2) |
| $R_{mrgd\text{-}F}$ | 13.1 (67.8) | 13.9 (71.6) | 12.2 (69.4) |
| *B*-factor (Wilson plot) (Å$^2$) | 68.3 | 73.1 | 66.0 |
| Phasing power | 0.952 | 0.510 | |
| Figure of merit overall | 0.57 | | |

selenium substructure was determined with SHELXD (18), and phases were calculated with SHARP (19) using the data sets collected at the peak and the inflection wavelength. Density modification was done in RESOLVE (20), resulting in interpretable electron density maps up to 3 Å resolution. Model building was done manually with the programs "O" (21) and COOT (22). Refinement of the model was done with the program REFMAC5 (23). To use the best data for refinement, the peak and the inflection data sets were merged resulting in a better overall quality of the data (Table I). Refinement statistics and quality indicators of the resulting Mlc model are listed in Table II.

*Spectroscopy*—The zinc content of the Mlc protein was determined by atomic absorption spectroscopy using a Varian AA240. UV-visible spectra were obtained with a Lambda 16 spectrophotometer (PerkinElmer Life Sciences). EPR spectra were recorded at 10 K on a Bruker Elexsys 500 spectrometer equipped with an ER049 X microwave bridge and an ESR 900 helium cryostat. The sample concentrations for the UV-visible and EPR measurements were 15 mg/ml Mlc.

*Site-directed Mutagenesis*—Single point mutations were carried out using the QuickChange multikit from Stratagene according to the manufacturer's protocol using the plasmid template pQE60mlc (16) and the following phosphorylated oligonucleotide primers: 5′-cca gta tca cta aaa ttg tcc gtg aga tgc tgc aag c-3′ for constructing wild type Mlc; 5′-ccg tat ggg aaa cgc gct tat gcc ggg aat cac ggc tgc-3′ for the C257A/C259A double mutant labeled "AYA," and 5′-ccg tat ggg aaa cgc tct tat tcc ggg aat cac ggc tgc-3′ for the C257S/C259S double mutant labeled "SYS." Final vector products were analyzed by sequencing (GATC, Germany).

*β-Galactosidase Assay (ptsG-lacZ assay)*—The assay was performed according to the method of Miller (24) with the following changes. Cells of *E. coli* K12 strain *JM101λRSIptsG-lacZ mlc*::*tet* (called *JM-G2* cells (6)) were grown for 12 h at 37 °C in minimal medium A (25) supplemented with 1% casamino acids or with 0.5% casamino acids and additional 0.2% glucose, as indicated in the figure legend. Specific activity was calculated as units per mg of total protein.

*Chemicals*—Chemicals were purchased from Fluka unless otherwise stated.

*Programs Used for Structural and Sequence Analyses*—The program DSSP (Definition of Secondary Structure of Proteins) (26) was used for buried surface calculations and SIM (27) for sequence alignments. The quality of the resulting Mlc model was checked with the program PROCHECK (28).

*PDB Accession Code*—The coordinate data sets of the structure of SeMet-Mlc R52H is available in the Protein Data Bank (29) with the accession code 1Z6R.

## RESULTS

*Structure of the Mlc Monomer*—The structure of Mlc represents the first described structure of an ROK family member. An Mlc molecule shown in Fig. 1 consists of three domains as follows: (*a*) a helix-turn-helix (HTH) domain (30, 31) from amino acid residues 1 to 81 + 395 to 406 (domain 1, *green*); (*b*) a smaller α/β-domain from residues 82 to 194 + 381 to 394 (domain 2, *yellow*); and (*c*) a larger α/β-domain from residues 195 to 380 (domain 3, *blue*). The final Mlc model contains 382 of 406 residues. Two segments in domain 1, residues 1–11 and 64–76, were structurally disordered in all four molecules within the asymmetric unit and were therefore not included in the final model. The structurally disordered region 64–76 of the HTH domain is known to be very flexible from HTH motifs described previously. It adopts its destined conformation, the so-called hinge helix, only upon binding to the operator DNA (32–34). Domains 2 and 3 are common to all ROK family members (14, 15) (see Fig. 1, *yellow* and *blue*). Both α/β-domains (domains 2 and 3) consist of a central β-sheet flanked by a pair of α-helices on one side and a single α-helix on the other side. Between domains 2 and 3 the polypeptide chain switches twice, so that domain 3 is formed by a continuous polypeptide, whereas the fold of domain 2 is completed by the returning C terminus from domain 3, packing as a C-terminal helix against the β-sheet of domain 2 (*bright yellow* in Fig. 1). The interface between domains 2 and 3 is mainly formed by the two single α-helices flanking the β-sheet in each domain. However, the packing of both domains toward each other is not very tight, allowing the domains to adopt different conformations with respect to each other. In addition to being part of domain 2, the C-terminal helix bends and is also part of domain 1 (*bright-green* in Fig. 1), thereby stabilizing the orientation of the HTH domain (domain 1) with respect to domain 2. This stabilization might be the reason why the HTH part of domain 1 is structurally ordered, whereas the connecting segment, including the hinge helix in domain 1, is not.

TABLE II
*Refinement statistics*

| Resolution limits (Å) | 20–2.7 |
|---|---|
| Total no. of reflections | 54,299 |
| Reflections in working set | 51,443 |
| Reflections in test set | 2,856 |
| *R* (%) | 20.3 |
| $R_{free}$ (%) | 26.3 |
| No. of amino acid residues | 1,528 |
| No. of protein atoms | 11,724 |
| No. of ions | 4 |
| *B*-factor for all atoms (Å$^2$) | 71.5 |
| R.m.s. deviation bonds (Å) | 0.02 |
| R.m.s. deviation angles (°) | 1.55 |

The three domains of an Mlc monomer behave as rigid groups, but the pairwise arrangement of the domains is different in the four molecules (*A–D*) within the asymmetric unit. The differences in the domain orientations were analyzed with the program DYNDOM (35) using domain 3 of the Mlc molecule A as the reference. Several rotational axes do not coincide in all possible domain pairs of the four Mlc molecules. Domains 1 and 2 are rotated as single units in molecules B and C by 18 and 5°, respectively. In molecule D, however, domain 2 is rotated by 22°, and domain 1, with respect to domain 2, is rotated by 12° around another axis at the same time. In addition, the comparison of molecule A with molecule B shows that residues 244–

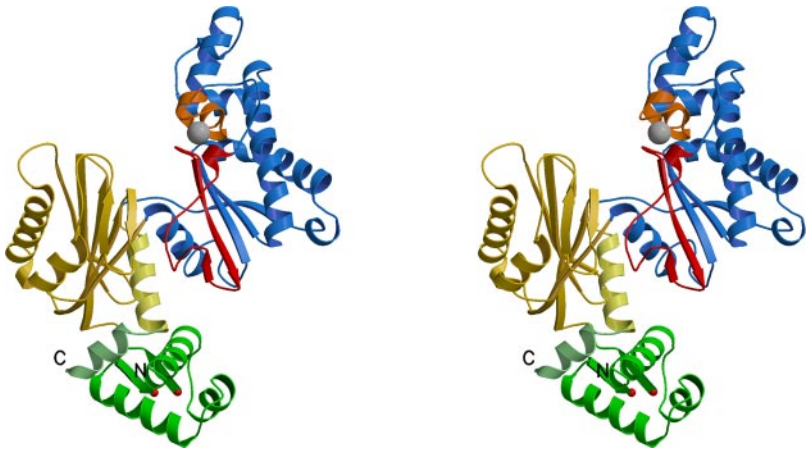


FIG. 1. **The Mlc monomer.** Stereo view of the Mlc monomer in ribbon presentation. The three distinct domains are colored *green* (domain 1), *yellow* (domain 2), and *blue* (domain 3). N and C termini are labeled with *capital letters,* and the missing segment, including the hinge helix, in domain 1 is indicated by the *blunt ends* marked with *small red spheres*. In addition, the two C-terminal helices that complete domain 1 and 2 are highlighted by brighter colors (*bright green* in domain 1 and *bright yellow* in domain 2). The zinc ion bound by domain 3 is depicted as a *gray sphere.* The consensus motifs that have been found to determine the ROK family members according to Hansen *et al.* (15) are highlighted as *red* and *orange ribbons* within domain 3. Consensus motif 1 (*red*) contains the zinc coordinating histidine His-247, whereas the consensus motif 2 (*orange*) contains the three zinc coordinating cysteines Cys-257, Cys-259, and Cys-264.

270 of domain 3, harboring parts of the two ROK motifs, are able to rotate separately by 14° with respect to the rest of domain 3.

*The ROK Signature Forms a Zinc-binding Site*—Domain 3 contains consensus motifs 1 and 2 that characterize the ROK family members (14, 15). Both motifs are highlighted in Fig. 1 as *red* and *orange ribbons*. Consensus motif 1 (Fig. 1, *red*) forms part of the central $\beta$-sheet in domain 3, leading into a loop followed by a short $3_{10}$ helix that ends with the invariant histidine His-247. Nine residues downstream, consensus motif 1 is followed by consensus motif 2 (Fig. 1, *orange*), starting with the conserved cysteine residues Cys-257 and Cys-259, followed by the conserved cysteine residue Cys-264. The structural explanation for the conservation of these residues is the tetrahedral coordination of a zinc ion by the four residues His-247, Cys-257, Cys-259, and Cys-264 (see Fig. 1, highlighted as a *gray sphere*). The presence of the zinc ion ($0.9 \pm 0.1$ zinc per protein) was confirmed by atom absorption spectroscopy, EPR, and UV-visible spectroscopy.

*The Mlc Dimer and Its Binding to DNA*—The four molecules within the asymmetric unit are arranged as two homodimers AB (chains A and B) and CD (chains C and D). Fig. 2*A* shows AB perpendicular to the 2-fold axis, and Fig. 2*B* shows AB along the 2-fold axis of the dimer. The dimerization occurs via domain 3 (Fig. 2*B, blue*) of each monomer, burying a surface area of 1378 Å$^2$ (in the case of AB) and 1393 Å$^2$ (in the case of CD). In the CD dimer there is an additional contact of 303 Å$^2$ between the HTH domains resulting from the above-mentioned Mlc flexibility and the crystal packing, but it is apparently not relevant for the dimer formation. Both dimers found in the asymmetric unit show different conformations. Although the superposition of chains A and C shows almost identical molecules, the superposition of chains B and D reveals two conformationally distinguishable dimers (see superposition in Fig. 3*A*). The conformational flexibility results in different distances between the two recognition helices in each dimer. Fig. 3, *B–D*, shows the isolated HTH domains of both dimers (AB in *yellow* and CD in *blue*). In dimer AB, the distance between the recognition helices is ~31.7 Å, very close to the helical pitch of B-DNA, which is 34 Å (Fig. 3, *B* and *C*). On the other hand, the CD dimer is much narrower with a distance of ~26.7 Å between the two recognition helices (Fig. 3, *D* and *E*), excluding its participation in DNA binding in this conformation. In addition, Fig. 3, *C* and *E,* shows that the recognition helices are not completely parallel, as one would assume from Fig. 3, *B* and *D*. The DNA would have to bend upon Mlc binding in order to accommodate both recognition helices of the AB dimer.

*A Tetramer of Mlc*—Earlier studies using size exclusion chromatography indicated that Mlc forms tetramers *in vitro* (12, 13). In order to determine whether the biochemically described tetramer is present in the Mlc crystals as well, we investigated all intermolecular contacts of the two Mlc dimers (AB and CD) within possible asymmetric units. The most symmetric arrangement relates both Mlc dimers by a pseudo 2-fold axis via domains 3. The single contacts are relatively weak with only 600 Å$^2$ between chains A and D and 636 Å$^2$ between chains B and C. Both contacts taken alone are not significant for a stable multimer formation (36). On the other hand, the sum of both contacts in a dimer of dimers with 1236 Å$^2$ could be relevant for a tetrameric structure. Nevertheless, we do not consider the crystallographic tetramer to be physiologically relevant.

*Comparison of Mlc with Related Structures*—A similarity search using the DALI server (37) revealed three bacterial kinases having the same structure as the ROK part of Mlc. In Table III the alignment lengths and the r.m.s. deviations of the identical C-$\alpha$ positions are listed. The highest structural homology to Mlc shows a putative fructokinase from *Bacillus subtilis* (Bs-FrcK), PDB code 1XC3, which has not been published so far. Bs-FrcK belongs to the ROK family as well, a fact explaining the structural homology. Its zinc ion is coordinated by two histidines and two cysteines spaced by a different number of amino acids as compared with Mlc. The second and third most similar structures are the glucokinase from *E. coli* (Ec-GlcK), PDB codes 1Q18 and 1SZ2 (38), and a bacterial inorganic polyphosphate/ATP glucomannokinase from *Arthrobacter* sp. (As-GMK), PDB code 1WOQ (39). Both proteins, Ec-GlcK and As-GMK, are not regarded as ROK family proteins but are in fact very similar to the ROK part of Mlc consisting of domains 2 and 3.

Most surprisingly, four of the five residues directly involved in glucose binding in Ec-GlcK and As-GMK are identical in Mlc and Bs-FrcK (Asp-195, Glu-244, His-247, and Glu266; Mlc numbering). The fifth residue, an asparagine in Ec-GlcK and in As-GMK is a histidine (His-194) in Mlc and a threonine in Bs-FrcK. However, despite the high structural similarity of the corresponding region in Mlc to the binding site for glucose in Ec-GlcK and As-GMK, Mlc does not bind glucose or glucose 6-phosphate as measured by the ammonium sulfate precipita-

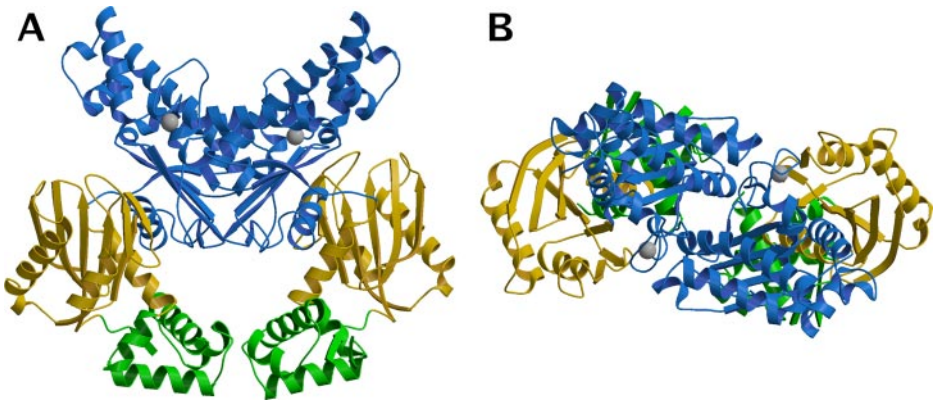


FIG. 2. **The Mlc dimer.** *A*, Mlc dimer is shown perpendicular to the 2-fold axis. The dimer is formed via domains 3 (*blue*) of two Mlc molecules. *B,* Mlc dimer (rotated by 90°) shown along the 2-fold axis with domain 3 (*blue*) in front.

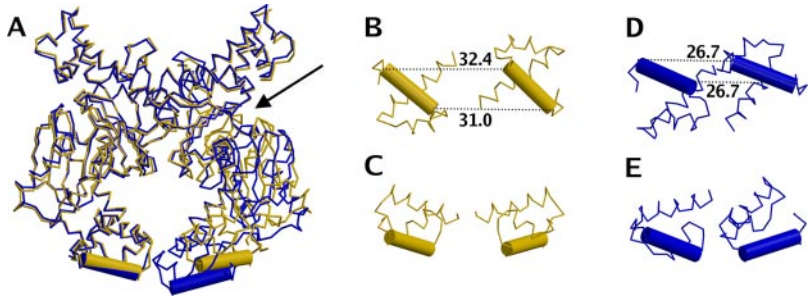


FIG. 3. **Comparison of the two Mlc dimers and the orientation of their recognition helices.** *A*, superposition of the two Mlc dimers AB (*yellow*) and CD (*blue*) as found in the asymmetric unit. Chains A and C (*left*) are almost identical in their conformation and fit well onto each other, whereas chains B and D have different conformations. The *arrow* indicates the hinge region that is the main source of flexibility. *B--E,* dimers reduced to the three N-terminal helices of the HTH domain, dimer AB shown in *yellow* (*B* and *C*) and dimer CD in *blue* (*D* and *E*). In the *upper panels* the HTH domains are shown along the dimer 2-fold axis, and in the *lower panels* they are shown perpendicular to it. The recognition helices are highlighted as *solid cylinders,* and the distances between them are given in Å.

TABLE III

*Structural comparison of Mlc, Bs-FrcK, a putative fructokinase from B. subtilis (PDB code 1XC3), Ec-GlcK, the glucokinase from E. coli (PDB code 1Q18), and As-GMK, the inorganic polyphosphate/ATP-glucomannokinase from Arthrobacter sp. (PDB code 1WOQ)*

R.m.s. deviations and alignment lengths were determined by using the DALI structural alignment server. Crystal contacts equivalent to the Mlc dimer contact are given by their buried surfaces between two monomers as calculated by the program (Definition of Secondary Structure of Proteins). Except for the structure of Bs-FrcK, where this contact was formed by a crystallographic 2-fold axis, this putative dimeric arrangement was found within the asymmetric units of the described structures. Oligomeric states were determined either by size exclusion chromatography (SEC) for As-GMK (47) or by dynamic light scattering (DLS) for Ec-GlcK (38). As-GMK forms only monomers, and Bs-FrcK has not yet been investigated for oligomeric states. ND, not determined.

| | R.m.s. deviation | Alignment length | No. of residues | Crystal contact | *n*-Mer in solution |
|---|---|---|---|---|---|
| | *Å* | | | *A*$^2$ | |
| Mlc | | | 406 | ~1400 | Tetramer (SEC) |
| Bs-FrcK | 2.7 | 276 | 295 | ~2050 | ND |
| Ec-GlcK | 3.6 | 262 | 320 | ~1650 | Dimer (DLS) |
| As-GMK | 3.2 | 243 | 253 | ~1200 | Monomer (SEC) |

tion technique (40). The same technique readily revealed glucose binding of Ec-GlcK (data not shown). This suggested that the fifth position (His-194 in Mlc) either discriminates between different sugars or between sugar binding and non-binding. However, altering His-194 to Asn did not result in glucose binding or glucokinase activity.[2]

The structural homology between the monomeric forms of Mlc, Bs-FrcK, Ec-GlcK, and As-GMK indicated similar quaternary structures. Our Mlc structure clearly shows two dimers within the asymmetric unit. In Table III the contact interfaces within the crystal packings are listed, showing the same 2-fold symmetry as the Mlc dimer. According to their buried surfaces, Bs-FrcK, Ec-GlcK, and As-GMK could be able to form stable homodimers of similar architecture as well (see Table III).

*Structural Localization of Mlc Mutants*—All mutations in Mlc characterized so far are shown in Fig. 4, highlighted by different colors. The mutant R52H has been accidentally selected on plates containing Luria Bertani (LB) medium during the cloning step for the structural analysis (16). Although the mutation is located in the recognition helix of the HTH domain, the protein still showed full repression of a *ptsG-lacZ* fusion (Fig. 5, *gray histograms*). In the presence of glucose, both the wild type Mlc and the R52H mutant show derepression of Mlc regulated genes (Fig. 5, *black histograms*) demonstrating that the R52H mutation neither affects repression nor induction. The latter is equivalent to the ability of Mlc to be bound by EIICB$^{Glc}$.

To study the role of the bound zinc ion in more detail, we constructed two double mutants by changing Cys-257 and Cys-259 into alanine or serine, respectively, resulting in Mlc C257A/C259A and Mlc C257S/C259S. Both mutant proteins showed only residual ability to repress the *ptsG-lacZ* fusion, pointing to a structural role of the zinc ion necessary for DNA binding (Fig. 5, *gray histograms*).

Seitz *et al.* (13) found that C-terminal deletions of Mlc influence its ability to tetramerize *in vitro* as well as its activity as a transcriptional repressor and its capacity to bind EIICB$^{Glc}$ *in vitro*. Although the deletion of the last nine residues (MlcΔC9) does not influence the activity of Mlc, the deletion of the last 18 residues (MlcΔC18) produces a protein no longer able to tetramerize *in vitro* and unable to bind its operator sites or EIICB$^{Glc}$. Fig. 4 shows the parts deleted in MlcΔC9 and MlcΔC18 as *yellow* and *yellow* + *orange ribbons*, respectively.

Furthermore, four point mutations of Mlc have been charac-

[2] M. Erhard, unpublished results.

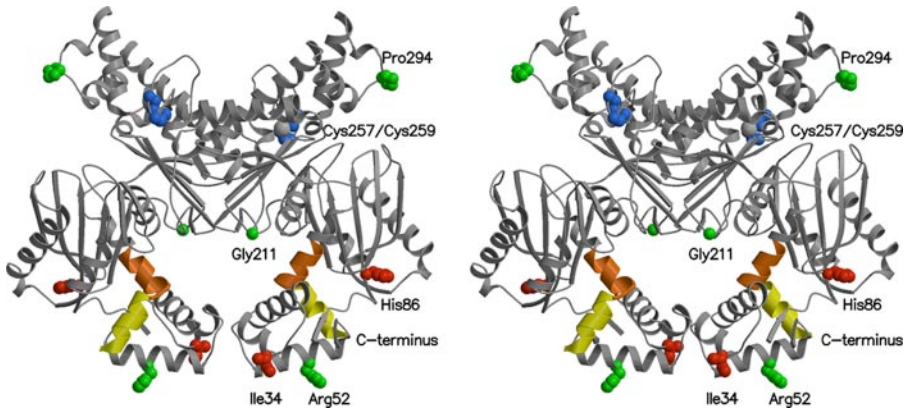


FIG. 4. **Characterized mutants of Mlc.** The Mlc dimer is shown as a *gray* stereo ribbon model, and the point mutations are highlighted with space-filled, colored side chains. Mutants known to impair $EIIB^{Glc}$ binding (I34V and H86R are indicated in *red*); mutants that show raised expression levels but are apparently not altered in their function (R52H, G211R, and P294S are shown in *green*), and the mutations of the ROK motif (C257A, C257S, C259A, and C259S) are highlighted in *blue*. In addition, the C-terminal deletion mutants are presented as *yellow* (MlcΔC9) and *yellow* + *orange* (MlcΔC18) *ribbons*.

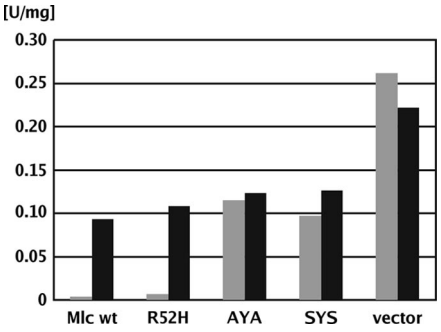


FIG. 5. **Analysis of Mlc activity.** The activity of Mlc mutants was determined by a ptsG-lacZ fusion as described under "Materials and Methods." The Mlc mutants Mlc R52H labeled *R52H*, Mlc C257A/C259A labeled *AYA*, and Mlc C257S/C259S labeled *SYS* are compared with the wild type Mlc "*wt*," and the vector control is pQE60. DNA repression activity is shown in *gray*, and derepression in presence of glucose is shown in *black*.

terized by Tanaka *et al.* (41). Two of these mutants, H86R and I34V, are impaired in $EIICB^{Glc}$ binding (see Fig. 4). This observation indicates that Mlc might bind to unphosphorylated $EIICB^{Glc}$ with both the HTH domain and domain 2. On the other hand, Mlc mutants G211R and P294S show raised expression levels, but they neither influenced repression nor binding to $EIICB^{Glc}$ (41).

## DISCUSSION

*The Structure of Mlc*—Overall, the Mlc molecule consists of three domains. Only domain 3, which contains both ROK consensus motifs, is composed of one continuous polypeptide, whereas domains 1 and 2 are completed by the back folding of the C terminus of the molecule. This structural arrangement results in a defined orientation of domain 1 with respect to domain 2. The asymmetric unit of the Mlc crystals contains four molecules that are clearly arranged as two dimers with a buried surface of ~1400 $Å^2$ in each dimer. The dimer formation between the two Mlc monomers occurs only via domain 3. Both domains 3 seem to form a stable scaffold with domains 1 and 2 flexibly attached to them. The hinge between domains 2 and 3 allows the movement of both domains 1 in an Mlc dimer with respect to each other. In this way the Mlc dimer is able to adopt different conformations. We conclude that the dimer contact is of biological relevance for three reasons. 1) The buried surface within the dimer is much larger than expected for an artificial crystal contact. 2) Mlc needs to be a dimer with the recognition helices being in proximity to bind to palindromic DNA. 3) The structurally very similar molecules Bs-FrcK and Ec-GlcK apparently form dimers of similar architecture. Furthermore, the structural similarity to bacterial sugar kinases suggests that Mlc represents a former kinase reused as a transcriptional repressor by the fusion of an HTH domain at its N terminus.

*What Determines the ROK Family Proteins?*—Mlc represents the first structure of an ROK family protein described to date. Two non-overlapping consensus motifs characteristic for ROK family proteins have been described by sequence comparison (15). Based on the Mlc structure, these two consensus motifs can be merged into a single one forming a zinc-binding site. From structural and sequence data, two similar zinc-binding motifs, $G\underline{\mathbf{H}}X_{9-11}\mathbf{C}X\mathbf{C}GX_2G(\mathbf{C/H})X\underline{E}$ and $G\underline{\mathbf{H}}X_{11-17}\mathbf{C}X_2\mathbf{H}X_2\mathbf{C}X\underline{E}$, can be distinguished in ROK family proteins (metal-binding residues highlighted in boldface). Mlc and most of the published protein sequences of ROK family proteins contain the first zinc-binding motif, whereas the second one is found only in a minority of the investigated sequences. The underlined residues are in the same position in the ROK proteins Mlc, Bs-FrcK, and in the non-ROK proteins Ec-GlcK and As-GMK. These two residues are involved in glucose binding in Ec-GlcK and As-GMK and may serve as sugar-binding residues in ROK proteins as well. Apparently, ROK proteins need a structural zinc ion to keep these two residues in place, whereas non-ROK sugar kinases without a metal-binding motif found another way to stabilize those residues, *e.g.* by a helical structure as found in Ec-GlcK.

Structure-based sequence alignments of the four structures Mlc, Bs-FrcK, Ec-GlcK, and As-GMK, show that these molecules resemble each other much more than expected by pure sequence comparison analysis (data not shown). The fold of both group II sugar kinases (Protein Families Data base (42) accession number PF02685) and group III (ROK) sugar kinases (Protein Families Data base accession number PF00480) is basically the same. Thus, we believe that both types of sugar kinases as well as repressors, with an HTH domain fused to the N terminus, evolved from a common ancestor in two lineages, one with a structural zinc ion and the other one without.

*Interaction of Mlc with DNA*—The dimeric structure of Mlc explains its ability to bind to single palindromic operator sites. The adaptation of the Mlc dimer to its operator sites is not achieved by flexible HTH domains but by the hinge within the ROK part of the molecule, with domains 1 and 2 moving as one rigid group with respect to domain 3. From band shift assays and DNase digestion experiments, it is known that only *ptsG* has two operator sites upstream of the coding region, whereas all others have only one (43). Mlc has not been observed to form a DNA loop known from experiments with other tetrameric repressors such as LacI or NagC (32, 44, 45), suggesting that Mlc only needs to be a dimer for DNA binding even though it has always been found as a tetramer in dilute buffer solutions *in vitro* (12, 13).

Seitz *et al.* (13) demonstrated that the deletion mutant MlcΔC18 is no longer able to bind to its operator DNA *in vivo*. MlcΔC18 is still able to form dimers and therefore should be able to recognize its operator sites because the DNA-binding HTH domain is located at the N terminus. We assume that the

loss of the DNA binding activity of MlcΔC18 is because of increased flexibility of the HTH domain with respect to domain 2, preventing the effective binding of the Mlc dimer to DNA. On the other hand deletion of MlcΔC9 seems not to be impaired in operator binding; it shows the same repressor activity as the wild type (13). Apparently, the contact between residues 394 and 397 and the first helix in domain 1 is still sufficient to keep domain 1 in a orientation that allows the recognition of the operator DNA. We expect that a stepwise deletion of more than 9 residues at the C terminus would gradually decrease the affinity of the Mlc dimer to its operator DNA.

Mutations within the zinc-binding motif (C257A/C259A called AYA and C257S/C259S called SYS) dramatically impair the ability of Mlc to repress its operator site (Fig. 5, *gray histograms*). The mutant proteins were not degraded, indicating structural stability. Although the zinc-binding motif is located in domain 3 without direct contact to domains 1 or 2, the coordination of the zinc ion apparently plays an important structural role for the correct orientation of the HTH domain.

*The Problem of Tetramerization*—Native Mlc forms tetramers *in vitro* (12, 13). The two Mlc dimers in our crystals can be arranged as a tetramer; however, mutational data argue against the tetramer configuration in the crystals. Seitz *et al.* (13) demonstrated that the MlcΔC18 deletion results in dimer formation. A tetrameric arrangement as observed in the crystals should not be influenced by the deletion because the contacts occur only via the domains 3, which appears to be properly folded as long as Mlc forms stable dimers. We believe that there exists an equilibrium between the dimeric and the tetrameric form of Mlc and that the crystallization buffer (1.6 M $MgSO_4$, 100 mM MES, pH 6.5 (16)) shifts the equilibrium toward the dimeric state of Mlc, whereas the buffer used in size exclusion chromatography (300 mM NaCl, 50 mM Tris-HCl, pH 7.5 (13)) favors the tetrameric state. Furthermore, the MlcΔC18 deletion indicated that Mlc might tetramerize in a similar way as LacI, forming a dimer of dimers via an α-helical bundle composed of the C-terminal helix of each monomer (32). From our structural data, it seemed rather unlikely that the tetramerization of Mlc could happen via the C terminus as positioned in the dimeric Mlc structure shown in Fig. 2*A*. To achieve a similar tetramerization via the C termini, the C-terminal helices would either have to fold out or the HTH part of domain 1 would have to rotate, to expose the C-terminal amphipathic helix allowing the interaction between two dimers.

*Interaction of Mlc with EIICB$^{Glc}$*—So far nothing is known about the stoichiometry of an Mlc-EIICB$^{Glc}$ complex. It is unclear whether Mlc binds to the EIICB$^{Glc}$ complex as a tetramer or as a dimer. The MlcΔC18 is no longer tetrameric and is unable to bind to EIICB$^{Glc}$, suggesting a role of the tetramerization for EIICB$^{Glc}$ binding. The two point mutants of Mlc, I34V and H86R, have been demonstrated to impair EIICB$^{Glc}$ binding (41). Both residues, Ile-34 and His-86, are separated from each other by a distance of ~30 Å within an Mlc monomer. Because the largest distance between two residues in soluble EIIB$^{Glc}$ (PDB code 1IBA) is only ~35 Å, it is unlikely that both residues contribute to the binding interface between Mlc and EIICB$^{Glc}$ as long as Mlc adopts the conformation we find in our crystals. On the other hand, if we assume that the mutation of these residues does not cause folding defects within the Mlc, the HTH domain would have to move with respect to domain 2, bringing both residues close enough to each other to form a composite binding site for EIICB$^{Glc}$. Such a conformational change should be possible because the HTH domain is flexibly connected to domain 2 by a single polypeptide that shows no secondary structure. Furthermore, a conformational rearrangement of the latter kind would expose the amphipathic C-terminal α-helices that stabilize the HTH domains in our dimer structure, thereby allowing these C-terminal helices coming from two Mlc dimers to interact with each other, *e.g.* forming an α-helical bundle in between the two dimers. Although this proposal is not in line with the tetrameric arrangement in our crystals, it can provide an explanation as to how a C-terminal helix buried in the dimeric structure could be involved in tetramerization. However, this mode of tetramerization should be impaired in the MlcΔC9 deletion mutant, but this deletion neither influences tetramer formation nor EIICB$^{Glc}$ binding.

*A Model for the Function of Mlc in Controlled Gene Expression*—We propose the exposure of the amphipathic helix at the very C terminus of Mlc to be the underlying mechanism by which Mlc switches from its active state (being a transcriptional repressor bound to the operator sites of Mlc-regulated genes) to its inactive state (being sequestered by binding to the EIICB$^{Glc}$ transporter). Thus, in the repressor mode, the HTH motif in the dimeric Mlc (see Fig. 2*A*) is stabilized by the C-terminal amphipathic helix with respect to domain 2. The hinge region between domains 2 and 3 helps the two recognition helices to achieve the correct distance for effective interaction of an Mlc dimer with the major groove of the palindromic operator sites (see Fig. 3). In its inactive mode, the HTH motif rotates, forming the EIICB$^{Glc}$-binding site and at the same time exposing the C-terminal amphipathic helix, which we propose to be the basis for tetramerization.

There has to be an equilibrium between the two states of Mlc. The finding that Mlc forms a tetramer in dilute buffer solutions at pH 7.5 indicates that Mlc is mainly in its tetrameric form ready to be bound by dephosphorylated EIICB$^{Glc}$ when EIICB$^{Glc}$ is transporting glucose. Our hypothesis that tetrameric Mlc binds to dephosphorylated EIICB$^{Glc}$ is consistent with the high number of EIICB$^{Glc}$ molecules present in the bacterial cell *versus* the very small number of operator sites, the potential targets of the dimeric Mlc. Apparently, the low proportion of dimeric Mlc is sufficient to shut down transcription *in vivo*. The effective dimer concentration of Mlc drops (by reforming tetramers) only when the bulk of Mlc tetramers is removed by binding to dephosphorylated, glucose-transporting EIICB$^{Glc}$, thus allowing the release of Mlc from its specific operator sites.

*Acknowledgments*—We thank Günter Fritz for the metal analysis, the staff at the synchrotron beamline X06SA at the Swiss Light Source in Villigen/Switzerland for their support, and Jacqueline Plumbridge from the Institut de Biologic Physico-chimigue in Paris/France for helpful discussions.

*Note Added in Proof*—Recently, the coordinates of the Mlc homolog from *vibrio cholerae* have been deposited at the PDB (accession code 1Z05) by Minasov, G., Brunzelle, J. S., Shuvalova, L., Collart, F. R., Anderson, W. F., Midwest Center for Structural Genomics (2005). The structure is very similar to that of Mlc from *E. coli*, including the dimeric structure discussed here.

REFERENCES

1. Hosono, K., Kakuda, H., and Ichihara, S. (1995) *Biosci. Biotechnol. Biochem.* **59,** 256–261
2. Plumbridge, J. (2002) *Curr. Opin. Microbiol.* **5,** 187–193
3. Decker, K., Plumbridge, J., and Boos, W. (1998) *Mol. Microbiol.* **27,** 381–390
4. Kimata, K., Inada, T., Tagami, H., and Aiba, H. (1998) *Mol. Microbiol.* **29,** 1509–1519
5. Plumbridge, J. (1998) *Mol. Microbiol.* **29,** 1053–1063
6. Plumbridge, J. (1998) *Mol. Microbiol.* **27,** 369–380
7. Kim, S. Y., Nam, T. W., Shin, D., Koo, B. M., Seok, Y. J., and Ryu, S. (1999) *J. Biol. Chem.* **274,** 25398–25402
8. Plumbridge, J. (1999) *Mol. Microbiol.* **33,** 260–273
9. Tanaka, Y., Kimata, K., Inada, T., Tagami, H., and Aiba, H. (1999) *Genes Cells* **4,** 391–399
10. Lee, S. J., Boos, W., Bouché, J. P., and Plumbridge, J. (2000) *EMBO J.* **19,** 5353–5361
11. Tanaka, Y., Kimata, K., and Aiba, H. (2000) *EMBO J.* **19,** 5344–5352

12. Nam, T. W., Cho, S. H., Shin, D., Kim, J. H., Jeong, J. Y., Lee, J. H., Roe, J. H., Peterkofsky, A., Kang, S.-O., Ryu, S., and Seok, Y.-J. (2001) *EMBO J.* **20,** 491–498
13. Seitz, S., Lee, S. J., Pennetier, C., Boos, W., and Plumbridge, J. (2003) *J. Biol. Chem.* **278,** 10744–10751
14. Titgemeyer, F., Reizer, J., Reizer, A., and Saier, M. H., Jr. (1994) *Microbiology* **140,** 2349–2354
15. Hansen, T., Reichstein, B., Schmid, R., and Schönheit, P. (2002) *J. Bacteriol.* **184,** 5955–5965
16. Gerber, K., Boos, W., Welte, W., Schiefner, A. (2005) *Acta Crystallogr. Sect. F* **61,** 183–185
17. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26,** 795–800
18. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1772–1779
19. Fortelle, de la, E., and Bricogne, G. (1997) *Method Enzymol.* **276,** 472–494
20. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 965–972
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
22. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
23. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
24. Miller, J. H. (1972) *Experiments in Molecular Genetics,* pp. 352–355, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
25. Miller, J. H. (1992) *A Short Course in Bacterial Genetics,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
26. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
27. Huang, X., and Miller, W. (1991) *Adv. Appl. Math.* **12,** 337–357
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242
30. Harrison, S. C., and Aggarwal, A. K. (1990) *Annu. Rev. Biochem.* **59,** 933–969
31. Wintjens, R., and Rooman, M. (1996) *J. Mol. Biol.* **262,** 294–313
32. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271,** 1247–1254
33. Spronk, C. A., Slijper, M., van Boom, J. H., Kaptein, R., and Boelens, R. (1996) *Nat. Struct. Biol.* **3,** 916–919
34. Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R., and Kaptein, R. (2002) *EMBO J.* **21,** 2866–2876
35. Hayward, S., and Berendsen, H. J. C. (1998) *Proteins* **30,** 144–154
36. Janin, J. (1997) *Nat. Struct. Biol.* **4,** 973–974
37. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
38. Lunin, V. V., Li, Y., Schrag, J. D., Iannuzzi, P., Cygler M., and Matte, A. (2004) *J. Bacteriol.* **186,** 6915–6927
39. Mukai, T., Kawai, S., Mori, S., Mikami, B., and Murata, K. (2004) *J. Biol. Chem.* **279,** 50591–50600
40. Richarme, G., and Kepes, A. (1983) *Biochim. Biophys. Acta* **742,** 16–24
41. Tanaka, Y., Itoh, F., Kimata, K., and Aiba, H. (2004) *Mol. Microbiol.* **53,** 941–951
42. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acid Res.* **32,** D138–D141
43. Plumbridge, J. (2001) *J. Mol. Microbiol. Biotechnol.* **3,** 371–380
44. Borowiec, J. A., Zhang, L., Sasse-Dwight, S., and Gralla, J. D. (1987) *J. Mol. Biol.* **196,** 101–111
45. Plumbridge, J., and Kolb, A. (1993) *Mol. Microbiol.* **10,** 973–981
46. Diederichs, K., and Karplus, P. A. (1997) *Nat. Struct. Biol.* **4,** 269–275
47. Mukai, T., Kawai, S., Matsukawa, H., Matuo, Y., and Murata, K. (2003) *Appl. Environ. Microbiol.* **69,** 3849–3857