Published in final edited form as:

J Struct Biol. 2007 September ; 159(3): 424–432.

Crystal structure of a transcriptional regulator TM1030 from *Thermotoga maritima* solved by an unusual MAD experiment

Katarzyna D. Koclega^{a,b,d}, Maksymilian Chruszcz^{a,d}, Matthew D. Zimmerman^{a,d}, Marcin Cymborowski^{a,d}, Elena Evdokimova^{c,d}, and Wladek Minor^{a,d,*}

^aDepartment of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA ^bInstitute of Technical Biochemistry, Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Lodz, Poland ^cDepartment of Medicinal Biophysics, University of Toronto, and Ontario Center for Structural Proteomics, Ontario Cancer Institute, Toronto, Ontario M5G 2C4, Canada ^dMidwest Center for Structural Genomics

Abstract

The crystal structure of a putative transcriptional regulator protein TM1030 from *Thermotoga maritima*, a hyperthermophilic bacterium, was determined by an unusual multi-wavelength anomalous dispersion method at 2.0 Å resolution., in which data from two different crystals and two different beamlines were used. The protein belongs to the tetracycline repressor TetR superfamily. The three-dimensional structure of TM1030 is similar to the structures of proteins that function as multidrug-binding transcriptional repressors, and contains a large solvent-exposed pocket similar to the drug-binding pockets present in those repressors. The asymmetric unit in the crystal structure contains a single protein chain and the two-fold symmetry of the dimer is adopted by the crystal symmetry. The structure described in this paper is an *apo*-form of TM1030. Although it is known that the protein is significantly overexpressed during heat shock, its detailed function cannot be yet explained.

Keywords

Transcriptional regulator; TM1030; DNA-binding; MAD

1. Introduction

To adapt to changes in the environment, a living organism must react very fast in order to survive. The reaction of the organism in many cases is connected with dynamic changes in gene expression. *Thermotoga maritima* is a hyperthermophilic bacterium that grows optimally at temperatures of 80°C or higher, and has to quickly respond to temperature changes in the environment. Several genes in *T. maritima* were demonstrated to undergo significant changes in expression level upon heat shock from 80°C to 90°C, as measured by cDNA microarray (Pysz *et al.*, 2004). One of these genes, TM1030, encodes for a 200 amino acid, 23.8 kDa protein. The N-terminal domain of the TM1030 protein shows significant sequence similarity to the N-terminal helix-turn-helix (HTH) domain of the TetR family of transcriptional

*Correspondence e-mail: wladek@iwonka.med.virginia.edu, University of Virginia, Department of Molecular Physiology and Biological Physics, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA, Phone: +1-434-243-0033, Fax: +1-434-982-1616.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

regulators. The TM1030 gene is located in an operon together with two genes that encode for ABC transporter proteins: TM1028 produces an ATP-binding protein, while TM1029 produces a putative permease protein (Nelson *et al.*, 1999). While the exact function of TM1030 protein is not yet known, expression data and sequence comparisons suggest that the protein may serve as a transcriptional regulator.

Proteins belonging to the TetR family control genes whose products are involved in multidrug resistance, specific catabolic pathways, biosynthesis of antibiotics, response to osmotic stress, and pathogenicity (Ramos *et al.*, 2005). For example, TetR, from which the name of the protein superfamily is taken, is responsible for the regulation of a resistance mechanism against the antibiotic tetracycline (Tc) in Gram-negative bacteria. In these bacteria, the resistance mechanism actively transports the drug out of the cell. The protein TetA acts as an energy-dependent antiporter across the cell membrane, coupling the efflux of a complex between Tc and a divalent cation, M^{2+} (MTc) to the influx of a proton. The TetR repressor regulates the expression of TetA (Grkovic *et al.*, 2002). The repressor forms a homodimer *in vivo* containing two HTH DNA binding motifs. TetR binds to two specific DNA operators and blocks the expression of the associated genes, one encoding for TetA and other for TetR. If a MTc complex binds to TetR, a conformational change occurs and the induced TetR is unable to bind to DNA (Kisker *et al.*, 1995).

Proteins from the TetR superfamily with known structures have α -helical architecture and each monomer consist of two domains. The smaller N-terminal domain with the HTH motif contains 3 helices (which is the domain conserved in TM1030), while the C-terminal regulatory domain typically contains 6–7 helices. Sequence comparison of TetR family members revealed that the DNA-binding domains have a higher degree of sequence similarity than the regulatory domains. Because of the twofold symmetry of the homodimer, the regulators bind palindromic DNA fragments, and each HTH motif binds to a major groove of the palindrome (Saenger *et al.*, 2000). Conversely, the C-terminal regulatory domains are less conserved. In order to recognize structurally diverse ligands, the regulatory domains are likely to be more tolerant to mutations in the process of evolution. The regulatory domains are not only responsible for ligand binding, but also for dimerization (Kisker *et al.*, 1995).

We present here the crystal structure of a putative transcriptional regulator protein TM1030 from *Thermotoga maritima*, that was determined by an unusual approach to the MAD technique. The peak and inflection datasets were collected using two different crystals and two different beamlines. The data sets were combined and both data sets were treated as a single MAD experiment, while in standard MAD technique (Hendrickson, 1991; Hendrickson *et al.*, 1990) data are collected on the same beamline and usually from a single crystal.

2. Materials and methods

2.1 Protein cloning, expression and purification

TM1030 was cloned using the standard protocol developed at Midwest Center for Structural Genomics (MCSG), as described previously (Zhang *et al.*, 2001). Small (10 mL) seed cultures of cells containing the TM1030 gene cloned into a modified pET-15b plasmid were grown overnight (37°C) in the *E. coli* methionine auxotrophic strain B834(DE3)pLysS and M9 minimal media plus 50 mg/L methionine (Met). The seed cultures were used to start full-size cultures containing Met at a concentration of 8mg/L, and their growth was monitored by absorbance at 595 nm. When the growth of cells leveled off as the supply of Met was exhausted, selenomethionine (Se-Met) was added to the flask at a concentration of 50mg/L. When the cells reentered log phase, they were induced with isopropyl-1- β -D-thiogalactopyranoside (final concentration of 1mM) and allowed to express overnight at a temperature of 16°C. Bacterial cells were resuspended and sonicated in a buffer containing 500mM NaCl, 50mM HEPES

pH=7.5, and 5% glycerol. Protein and cell solutions were kept at 4°C and protease inhibitors were added to prevent protein degradation. After centrifugation, supernatant was applied to a nickel chelate affinity resin (Ni-NTA, QIAGEN) previously equilibrated with binding buffer (500mM NaCl, 50mM HEPES pH=7.5, 5% glycerol, 5mM imidazole). The resin was washed with binding buffer containing 30mM imidazole, and protein was eluted with binding buffer containing 250mM imidazole. Protein was dialyzed into a buffer containing 500mM NaCl and 10mM HEPES pH=7.5, and concentrated with Millipore concentrators. The N-terminal His-tag was removed by cleavage with recombinant tobacco etch virus (rTEV) protease (Kapust *et al.*, 2001). After the affinity chromatography step, eluted and concentrated protein samples were additionally purified using a gel filtration column (HiLoad 6/16 Superdex 200) on an AKTA FPLC system (GE Healthcare).

2.2 Crystallization

Crystallization was performed by hanging drop vapor diffusion using Crystool (NEXTAL) crystallization plates. Drops were created by mixing 1 µl of screen solution and 1 µl of a solution of 7.2 mg/ml of TM1030 in 500mM NaCl and 10mM HEPES pH=7.5. The plates were stored at 20°C. An initial hit was obtained with condition #95—0.1M KSCN and 30% w/v polyethylene glycol monomethyl ether 2000 (PEG 2K MME)—of Hampton Research's Index Screen. Further optimizations led to the following crystallization conditions: 0.05M Na citrate pH=4.5, 30% w/v PEG 2K MME and 0.1M KSCN. Optimization strategy, crystal and drop tracking, and analysis of intermediate results were performed using the crystallization database Xtaldb (Zimmerman and Minor, 2007). Prior to data collection, crystals were transferred to a cryoprotectant, composed of a 7:3 mixture of crystallization solution and ethylene glycol, and flash-frozen in liquid nitrogen.

2.3 Data collection, structure solution and refinement

The crystal structure of TM1030 (MCSG code: APC4542, PDB code: 1z77) was determined by an unusual multiple-wavelength anomalous dispersion (MAD) method, in which datasets from two different crystals collected on two different beamlines were treated as a single two-wavelength MAD dataset. Initially, data at the Se absorbance peak wavelength (0.97932 Å) were collected on Structural Biology Center beamline 19-BM at the Advanced Photon Source (APS) at Argonne National Laboratory. The data were indexed, integrated and scaled (Table 1) with *HKL-2000* (Otwinowski and Minor, 1997), *HKL-3000* (Minor *et al.*, 2006), *SHELXD* (Schneider and Sheldrick, 2002), *SHELXE* (Sheldrick, 2002), *MLPHARE* (Otwinowski, 1991), *DM* (Cowtan and Main, 1993; Cowtan and Zhang, 1999), *CCP4* (CCP4, 1994), *O* (Jones *et al.*, 1991), *ARP/wARP* (Perrakis *et al.*, 1999), *SOLVE* (Terwilliger and Berendzen, 1999) and *RESOLVE* (Terwilliger, 2004; Terwilliger, 2002) were used to phase the structure by single-wavelength anomalous dispersion (SAD) and gave an initial model that contained ~31% of all α-helix forming residues (Table 2), but attempts to extend that model by automatic building failed. A higher resolution dataset from another TM1030 crystal was collected on a second crystal from a similar crystallization condition on APS beamline 19-ID (Rosenbaum *et al.*, 2006) at a wavelength of 0.9193 Å (equivalent to the far remote energy of Se MAD). This dataset was combined with the Se peak data as second wavelength, resulted in successful MAD phasing that gave clearly interpretable electron density map that was used to complete model building.

This unusual MAD experiment resulted in the initial model, built by *RESOLVE* as implemented in *HKL-3000*, having more than 50% of α-helix forming residues traced. This model was improved by manual rebuilding with *COOT* (Emsley and Cowtan, 2004) and later used as a starting point for automatic model rebuilding with *ARP/wARP*. Final refinement was performed with *REFMAC* (Murshudov *et al.*, 1997). *MOLPROBITY* and *KING* (Lovell *et al.*, 2003) were used for model validation and slight model adjustments. Coordinates and structure factors for

the structure have been deposited at the Protein Data Bank (RCSB, <http://www.rcsb.org>) with accession code 1z77.

Figures were made using *MOLSCRIPT* (Kraulis, 1991), *RASTER3D* (Merritt and Murphy, 1994), *EsPript* (Gouet *et al.*, 2003) and *PyMOL* (DeLano, W.L. The PyMOL Molecular Graphics System; <http://www.pymol.org>).

3. Results and discussion

3.1. Non-standard MAD technique

The peak dataset collected on 19-BM was used for initial structure solution and model building. Although it was possible to build a part of the TM1030 model, its extension turned out to be difficult. The failure of complete model building was caused by a combination of relatively low solvent content (42%), the lack of non-crystallographic symmetry (NCS) and the low resolution of the data (2.8 Å). A second data set collected on 19-ID from a different crystal was combined with data from 19-BM and both data sets were treated as an unusual MAD experiment. The dataset from 19-ID contains higher-resolution data collected at a wavelength that was not optimal for a SAD experiment with selenium, but could be used as high energy remote data. Such an unusual combination resulted in successful MAD phasing, and *HKL-3000* produced an interpretable electron density map of better quality than the one obtained by SAD phasing alone (Table 2). In the classical approach for MAD phasing (Hendrickson, 1991; Hendrickson *et al.*, 1990), it is assumed that all datasets are collected from a single crystal on the same beamline. In the MAD case reported here, it seems that the improvement of the initial phases compared to SAD is significant, despite the errors introduced by non-isomorphism of the data (two different crystals on two different beamlines). The unit-cell parameters indicate that non-isomorphism between crystals was small (Table 1), and both beamlines have similar, high quality characteristics (Rosenbaum *et al.*, 2006).

3.2. Description of TM1030 structure

The overall fold of TM1030 is very similar to those observed for other members of the TetR superfamily (Fig. 1a), containing 9 alpha helices (Fig. 2). The N-terminal domain consists of helices $\alpha 1$, $\alpha 2$ and $\alpha 3$, whereas the C-terminal domain is comprised of helices $\alpha 4$ – $\alpha 9$ (Fig. 1a). The protein crystallized in the orthorhombic space group $P2_12_12$ with one monomer in the asymmetric unit. However, gel filtration retention volumes (not shown) and comparison to other TetR proteins indicate that the oligomeric state in solution is a dimer. The two-fold symmetry of the TM1030 dimer was adopted by crystallographic symmetry, and the second molecule of the dimer can be generated from the monomer present in the asymmetric unit by applying the operator $-x, -y+1, z$. The interface area between monomers forming the dimer is approximately 1400 Å² (Henrick and Thornton, 1998). The dimer is stabilized by hydrophobic interactions between symmetry related residues (Phe153, Trp156, Met190, Ile193 and Leu194) and by 3 pairs of charge-charge interactions (6 in all) formed between residues Asp22...Arg23, Lys152...Glu186, and Glu163...Arg167 across the dimer interface. The gel filtration experiments showed that the dimer of TM1030 was stable in crystallization buffer at a pH range of 5–7, although small amounts of higher oligomeric states were also observed (data not shown). The surface curvature and charge distribution of the probable dimer, as well as a ribbon representation of one polypeptide chain, are shown in Figure 1b.

Approximately one month after our deposition (PDB code: 1z77), the Joint Center for Structural Genomics (JCSG) also deposited a structure of the same protein (PDB code: 1zkg). The structure reported by JCSG was determined in a different space group ($P2_1$), containing a dimer in the asymmetric unit. Most interestingly, in this structure of TM1030 an unidentified ligand is modeled by the assembly of unknown atoms. The ligand in 1zkg is represented by a

set of atoms that form a linear C-shaped structure. Moreover, comparison of the two structures reveals significant changes in conformation of the regulator protein, apparently caused by binding of the unknown ligand (Fig. 3). The fact that this unknown ligand is not found in the structure reported here indicates it is an *apo*- form of the protein.

A search for homologous structures was performed using both *DALI* (Holm and Sander, 1995) and *PROFUNC* (Laskowski et al., 2003; Laskowski et al., 2005) servers. The most similar structures (excluding 1zkg) are reported in Table 3. The overall fold of TM1030 is similar to proteins from the TetR family that is found in many pathogenic organisms like *Mycobacterium tuberculosis*, *Salmonella typhimurium*, and *Staphylococcus aureus*. Despite high similarity of the overall architecture of the TetR family members, sequence identity between TM1030 and other proteins from this group is lower than 30% (Table 3).

3.3 Ligand binding site

The binding site in the regulatory domain of TM1030 is composed mainly of hydrophobic residues (Leu62, Phe66, Trp85, Ile86, Phe128, Phe158, Phe161 and Tyr165) with only a few of a more hydrophilic character (Lys89, Ser124, Glu125 and Glu162). A significant number of hydrophobic residues may suggest that the TM1030 ligands also have hydrophobic characteristics. The presence of many aromatic residues could be an indicator that the ligand is not only hydrophobic, but it could also have an aromatic system. Accordingly, we tried to co-crystallize TM1030 with several hydrophobic and aromatic compounds: tetracycline, oxytetracycline, rhodamine 6G, proflavine hemisulfate, dequalinium chloride, malachite green and crystal violet. However, none of the crystals produced during co-crystallization had ordered ligands in complex with the regulator.

One of the structures similar to that of TM1030 is the EthR protein (Frenois *et al.*, 2004) from *Mycobacterium tuberculosis* in complex with hexadecyl octanoate (PDB code: 1u9o). EthR is a transcriptional repressor involved in the ethionamide resistance mechanism in mycobacteria. Ethionamide is used as a second-line antitubercular drug for the treatment of patients infected with multidrug-resistant *Mycobacterium tuberculosis* who have developed resistant to the first-line drugs, such as isoniazid and rifampicin (Dover *et al.*, 2004; Engohang-Ndong *et al.*, 2004). In the structure of EthR, hexadecyl octanoate is present in the binding site in extended conformation. The linear shape of the hexadecyl octanoate is somewhat similar to the shape of the unknown ligand in the structure of TM1030 reported by JCSG, but it has a different conformation (Fig. 4). On such a basis, it could be assumed that the TM1030 ligand is a molecule with a long, probably aliphatic chain. It is not clear if the ligand observed in the structure reported by JCSG is an unknown physiologically relevant substrate, or rather a molecule that was introduced during protein purification and/or crystallization. There is a possibility that the unknown ligand is a PEG molecule (or a fragment of a PEG molecule), which would be in agreement with the shape of the unknown ligand and the reported crystallization conditions (30% PEG 8000, 0.2M Mg(NO₃)₂, 0.1M citrate pH=4.5).

A second structure similar to TM1030 is that of QacR. QacR is a member of the TetR/CamR family of transcriptional regulators, which share highly homologous N-terminal DNA-binding domains connected to non-homologous ligand-binding domains (Schumacher *et al.*, 2002). The *Staphylococcus aureus* multidrug-binding protein QacR represses transcription of the *qacA* multidrug transporter gene and is induced by structurally diverse cationic lipophilic compounds, such as rhodamine 6G, crystal violet, ethidium, dequalinium, chlorhexidine, and pentamidine (Murray *et al.*, 2004; Schumacher *et al.*, 2004; Schumacher *et al.*, 2001). It cannot be excluded, taking into account the similarity of TM1030 and QacR, that TM1030 may be a multi-ligand binding regulator.

3.4 DNA binding

Structural comparison of the DNA-binding domain between TM1030 of transcriptional regulator and other structures from the TetR superfamily reveals structural conservation of the DNA binding domain, as predicted by the conservation of sequence in this domain. Despite relatively low sequence identity, the overall structure of TM1030 (1z77) is very similar to the structure of QacR (Schumacher *et al.*, 2002) from *Staphylococcus aureus* determined in complex with DNA (1jt0), as shown in Fig. 5. The changes in the conformation of TM1030 after ligand binding (Fig. 3) causes the distance between DNA-binding domains to increase from approximately 30 Å to 50 Å, which is larger than the distance between two neighboring major grooves in double stranded B-DNA. This suggests that this conformational change prevents strong DNA-binding from a sterical point of view. A similar change of dimer conformation after ligand binding is observed for TetR (Orth *et al.*, 2000; Orth *et al.*, 1998). On this basis, we suggest that the structural form of TM1030 reported by us should be able to bind DNA, while the structural form of the protein reported by JCSG cannot bind DNA, assuming that binding does not induce a large conformational change in the DNA. It is probable that TM1030 binds DNA in the same mode as QacR, but other modes of DNA binding cannot be excluded. QacR binds DNA with a stoichiometry of two QacR dimers per operator, whereas TetR binds as a single dimer (Schumacher *et al.*, 2002). There are 17 invariant and 12 conserved residues in the N-terminal domain of TM1030 compared to the sequence of QacR (Fig. 3). Only four invariant residues (Thr25, Thr26, Lys37 and Gly38) and one conserved residue (Ala36) are in close contact with DNA.

3.5 Two crystal forms of TM1030

It is not clear what influence, if any, the crystallization conditions of TM1030 has on the generation of two different crystal forms of the protein. The crystallization conditions in both cases (1zkg vs. 1z77) are similar in terms of chemical composition and concentrations, as both contain 0.1 M citrate pH=4.5 and 30% w/v of PEGs. The most significant difference between the crystallization conditions is caused by the presence of different salts, 0.2M Mg(NO₃)₂ in 1zkg and 0.1M KSCN in 1z77, respectively. It is possible that the ionic strength of the crystallization solutions influences the equilibrium between *apo*- and ligand-bound forms of the protein, with the higher ionic strength promoting or stabilizing the formation of the ligand-bound form. A change of ionic strength can easily influence the charge-charge interactions involved in dimer formation and in this way induce conformational changes in the dimer. The packing of the two crystal forms of TM1030 are completely different and different residues are involved in the crystal contacts. The only exception is Asp27, which not only mediates crystal contacts in different crystal forms, but also is involved in contacts formed by each of the two polypeptide chains present in the asymmetric unit of 1zkg. It is also worth mentioning that the more compact dimers of the *apo*- form are packed more tightly in the crystal than the dimer with ligands in the binding sites, as the difference in solvent content between those two forms is quite significant (42% vs. 52%). The difference in crystal form could also be due to differences in protein construct or expression or purification protocols, but due to a lack of published detailed information on these protocols, it is not possible to compare them directly.

4. Conclusions

Structural and bioinformatic analysis of TM1030 shows that it belongs to a superfamily of transcriptional regulator proteins (TetR). These proteins control genes whose products are involved in multidrug resistance, catabolic pathways, osmotic stress, and pathogenicity in Gram-negative and Gram-positive bacteria (Grkovic *et al.*, 2002; Ramos *et al.*, 2005). The reported structure of TM1030 exhibits a fold similar to proteins such as EthR, CprB and QacR, which are composed of 9 or 10 α -helices that form an N-terminal DNA-binding domain and a C-terminal regulatory core domain involved in dimerization and ligand binding. Comparison

of the TM1030 with other similar regulators suggests that most probably a ligand of the proteins contains an aliphatic chain or other chemical group forming a linear structure. Moreover, the comparison of the two forms of the protein suggests that the *apo*-form is able to bind DNA, while the ligand-bound form of TM1030 cannot bind DNA in a way that is characteristic of other TetR superfamily members. In addition, we used a non-standard protocol for MAD phasing, where due to a low solvent content and lack of NCS the SAD technique showed its limitations.

Acknowledgements

The authors would like to thank Andrzej Joachimiak and the members of the Structural Biology Center at the Advanced Photon Source and the Midwest Center for Structural Genomics for help and discussions. The authors also thank Alexander Wlodawer and Sri Krishna Subramanian for reading the manuscripts and making valuable comments. The results shown in this report are derived from work performed at Argonne National Laboratory, at the Structural Biology Center of the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. The work described in the paper was supported by NIH PSI grants GM62414 and GM074942.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410. [PubMed: 2231712]
- Baker N, Holst M, Wang F. Adaptive multilevel finite element solution of the Poisson-Boltzman equation II. Refinement at solvent accessible surfaces in biomolecular systems. *J Comput Chem* 2000;21:1343–1352.
- CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 1994;50:760–763. [PubMed: 15299374]
- Cowan KD, Main P. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr D* 1993;49:148–157. [PubMed: 15299555]
- Cowan KD, Zhang KY. Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* 1999;72:245–270. [PubMed: 10581970]
- Dover LG, Corsino PE, Daniels IR, Cocklin SL, Tatituri V, Besra GS, Futterer K. Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance. *J Mol Biol* 2004;340:1095–1105. [PubMed: 15236969]
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 2004;60:2126–2132. [PubMed: 15572765]
- Engohang-Ndong J, Baillat D, Aumercier M, Bellefontaine F, Besra GS, Loch C, Baulard AR. EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Mol Microbiol* 2004;51:175–188. [PubMed: 14651620]
- Frenois F, Engohang-Ndong J, Loch C, Baulard AR, Villeret V. Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Mol Cell* 2004;16:301–307. [PubMed: 15494316]
- Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 2003;31:3320–3323. [PubMed: 12824317]
- Grkovic S, Brown MH, Skurray RA. Regulation of bacterial drug export systems. *Microbiol Mol Biol Rev* 2002;66:671–701. [PubMed: 12456787]
- Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 1991;254:51–58. [PubMed: 1925561]
- Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* 1990;9:1665–1672. [PubMed: 2184035]
- Henrick K, Thornton JM. PQS: a protein quaternary structure file server. *Trends Biochem Sci* 1998;23:358–361. [PubMed: 9787643]

- Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 1995;20:478–480. [PubMed: 8578593]
- Holst M, Baker N, Wang F. Adaptive multilevel finite element solution of the Poisson-Boltzman equation I. Algorithms and examples. *J Comput Chem* 2000;21:1319–1342.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 1991;47:110–119. [PubMed: 2025413]
- Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* 2001;14:993–1000. [PubMed: 11809930]
- Kisker C, Hinrichs W, Tovar K, Hillen W, Saenger W. The complex formed between Tet repressor and tetracycline-Mg²⁺ reveals mechanism of antibiotic resistance. *J Mol Biol* 1995;247:260–280. [PubMed: 7707374]
- Kraulis PJ. Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Crystallogr* 1991;24:946–950.
- Laskowski RA, Watson JD, Thornton JM. From protein structure to biochemical function? *J Struct Funct Genomics* 2003;4:167–177. [PubMed: 14649301]
- Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res* 2005;33:D266–268. [PubMed: 15608193]
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by C α geometry: phi,psi and C β deviation. *Proteins* 2003;50:437–450. [PubMed: 12557186]
- Merritt EA, Murphy ME. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr D* 1994;50:869–873. [PubMed: 15299354]
- Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr D* 2006;62:859–866. [PubMed: 16855301]
- Murray DS, Schumacher MA, Brennan RG. Crystal structures of QacR-diamidine complexes reveal additional multidrug-binding modes and a novel mechanism of drug charge neutralization. *J Biol Chem* 2004;279:14365–14371. [PubMed: 14726520]
- Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 1997;53:240–255. [PubMed: 15299926]
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson LD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 1999;399:323–329. [PubMed: 10360571]
- Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* 2000;7:215–219. [PubMed: 10700280]
- Orth P, Cordes F, Schnappinger D, Hillen W, Saenger W, Hinrichs W. Conformational changes of the Tet repressor induced by tetracycline trapping. *J Mol Biol* 1998;279:439–447. [PubMed: 9642048]
- Otwinowski, Z. Isomorphous Replacement and Anomalous Scattering, CCP4 Study Weekend. Wolf, W.; Evans, PR.; Leslie, AGW., editors. Daresbury Laboratory; Warrington: 1991. p. 80-86.
- Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW.; Sweet, RM., editors. *Methods in Enzymology* 276, Macromolecular Crystallography, part A. Academic Press; New York: 1997. p. 307-326.
- Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999;6:458–463. [PubMed: 10331874]
- Pysz MA, Ward DE, Shockley KR, Montero CI, Connors SB, Johnson MR, Kelly RM. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. *Extremophiles* 2004;8:209–217. [PubMed: 14991425]

- Ramos JL, Martinez-Bueno M, Molina-Henares AJ, Teran W, Watanabe K, Zhang X, Gallegos MT, Brennan R, Tobes R. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 2005;69:326–356. [PubMed: 15944459]
- Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang RG, Ginell SL, Duke N, Naday I, Lazarz J, Molitsky MJ, Keefe L, Gonczy J, Rock L, Sanishvili R, Walsh MA, Westbrook E, Joachimiak A. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J Synchrotron Radiat* 2006;13:30–45. [PubMed: 16371706]
- Saenger W, Orth P, Kisker C, Hillen W, Hinrichs W. The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed* 2000;39:2042–2052.
- Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D* 2002;58:1772–1779. [PubMed: 12351820]
- Schumacher MA, Miller MC, Brennan RG. Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *EMBO J* 2004;23:2923–2930. [PubMed: 15257299]
- Schumacher MA, Miller MC, Grkovic S, Brown MH, Skurray RA, Brennan RG. Structural mechanisms of QacR induction and multidrug recognition. *Science* 2001;294:2158–2163. [PubMed: 11739955]
- Schumacher MA, Miller MC, Grkovic S, Brown MH, Skurray RA, Brennan RG. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J* 2002;21:1210–1218. [PubMed: 11867549]
- Sheldrick GM. Macromolecular phasing with SHELXE. *Zeitschrift Fur Kristallographie* 2002;217:644–650.
- Terwilliger T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* 2004;11:49–52. [PubMed: 14646132]
- Terwilliger TC. Automated structure solution, density modification and model building. *Acta Crystallogr D* 2002;58:1937–1940. [PubMed: 12393925]
- Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D* 1999;55:849–861. [PubMed: 10089316]
- Zhang RG, Skarina T, Katz JE, Beasley S, Khachatryan A, Vyas S, Arrowsmith CH, Clarke S, Edwards A, Joachimiak A, Savchenko A. Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure* 2001;9:1095–1106. [PubMed: 11709173]

Figure 1.

(a) A stereo ribbon diagram showing monomers forming the dimer of TM1030. Protein monomers are shown in green and blue. The DNA binding domains and the regulatory domains of TM1030 are colored, respectively, in pale and bright colors. (b) The ribbon diagram of a monomer of TM1030 together with the surface of oligomeric form of the protein. The surfaces of apo (top) and ligand-bound (bottom) forms of the protein_ are colored with the electrostatic charge distribution as calculated by the APBS plugin (Baker *et al.*, 2000; Holst *et al.*, 2000) of PyMOL (blue-positive charge, red-negative charge), shown in two different orientations.

Figure 2.

Sequence alignment of TM1030 and the four representative sequences found with BLAST (Altschul *et al.*, 1990), aligned with the secondary structure identified from the structure of TM1030. The last sequence corresponds to QacR from *Staphylococcus aureus* (PDB code: 1jt0). THEMA—transcriptional regulator, TetR family from *Thermotoga maritima*; 9CLOT—regulatory protein, TetR family from *Alkaliphilus metalliredigens*; STAAU—Hypothetical protein from *Staphylococcus aureus*; CLOAB—Transcriptional regulator, AcrR family from *Clostridium acetobutylicum*; LISMF—Transcriptional regulator, TetR family from *Listeria monocytogenes* serotype 4b.

Figure 3. Superposition of two structures of TM1030, without ligand (PDB code: 1z77) in green and with ligand (PDB code: 1zkg) in light blue, shown in two different orientations. The unknown ligand in 1zkg is represented by red spheres.

Figure 4.
Superposition of 1zkg in light blue onto EthR (1u9o) in dark blue. The TM1030 ligand is shown as red spheres and the EthR ligand—hexadecyl octanoate—in yellow.

Figure 5. Superposition of the *apo*-form of the structure of TM1030 (1z77) in green onto QacR complexed with DNA (1jt0) in raspberry.

Table 1

Data collection, processing, and refinement statistics. Data for the highest resolution shells are given in parentheses.

PDB code	1z77		
Data Collection			
Beamline	SBC 19-BM		SBC 19-ID
Space group		P2 ₁ 2 ₁ 2	
Unit cell parameters (Å)	a= 55.8 b= 66.0 c = 55.6		a= 56.0 b= 65.7 c = 55.7
Wavelength (Å)	0.9793		0.9193
Resolution range (Å)	50.0-2.8		50.0-1.95
Highest resolution shell (Å)	2.9-2.8		2.02-1.95
Number of observations	38896		123954
Number of reflections	5010		15569
Completeness (%)	99.7(99.2)		98.3 (100.0)
Mean I/σ(I)	18.1 (2.8)		24.8 (3.3)
R _{sym} on I	0.091(0.346)		0.070(0.539)
Model and refinement statistics			
Resolution range (Å)			50.00-2.00
Highest resolution shell (Å)			2.05-2.00
No. of reflections (total)			13472
No. of reflections (test)			715 (61)
Completeness (% total)			98.3 (100.0)
R			0.229 (0.230)
R _{free}			0.267 (0.291)
Stereochemical parameters			
Restraints (RMS observed)			
Bond length (Å)			0.022
Bond angle (°)			1.970
Average isotropic B-value (Å ²)			37.8
ESU based on R _{free} value (Å)			0.190
Protein residues / atoms			200/1597
Water molecules / ligands			40/2

Correlation coefficient I (CC I) was calculated between the initial models for each combination of datasets and the experimental density map obtained from *HKL-3000*. The initial models were built using the fast mode of *RESOLVE*. Correlation coefficient II (CC II) was calculated between the final refined model (1z77) and the experimental density map obtained from *HKL-3000*. In all cases, for calculation of the maps we used the same Se substructure. The fraction of residues in α -helices automatically built with *RESOLVE* (fast mode) were calculated by comparing initial models with the final, refined model of TM1030 reported to the PDB.

Table 2

Data set(s)	Resolution (Å)	FOM _{MLPHARE}	FOM _{DM}	CC I (main/side chain)	CC II (main/side chain)	Fraction of secondary structure automatically built
19-BM	2.8	0.41	0.77	0.45/0.49	0.60/0.43	0.32
19-ID	2.0	0.21	0.78	0.23/0.51	0.48/0.41	0.17
19-BM + 19-ID	2.0	0.38	0.77	0.53/0.57	0.68/0.52	0.51

Table 3

Proteins in the PDB with similar structures to TM1030 (PDB code: 1z77), as identified by *PROFUNC* (Laskowski et al., 2003; Laskowski et al., 2005).

PDB entry	Name	Organism	Seq. id (%)	Number of aligned residues	Z-score	RMSD (Å)
1vi0	TetR	<i>Bacillus subtilis</i>	24	160	5.5	2.2
1t56	TetR	<i>Mycobacterium tuberculosis</i>	16	163	3.9	2.5
1u9o	EthR	<i>Mycobacterium tuberculosis</i>	15	162	3.1	2.7
1ui6	Arpa-like	<i>Streptomyces coelicolor</i>	23	137	3.6	2.5
1u9n	EthR	<i>Mycobacterium tuberculosis</i>	18	158	3.2	2.8
1t33	TetR/acr	<i>Salmonella typhimurium</i>	19	176	2.8	2.9
1rpw	QacR	<i>Staphylococcus aureus</i>	21	163	3.2	2.9
2hyj	TetR	<i>Streptomyces coelicolor</i>	17	162	2.6	2.9