Structural and functional characterization of IclR transcription regulators

by

Alexandra Ezersky

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Medical Biophysics
University of Toronto

© Copyright by Alexandra Ezersky (2009)
ABSTRACT

This work is a part of a large project in our laboratory that is aimed toward characterization of prokaryotic transcription regulators from different families and their interactions with small-molecule effectors. My study was focused of IclR family of transcription regulators, specifically on its founding member Isocytrate Lyase Regulator (IclR) from *E.coli* and AllR regulator from *E.coli*, which share 42% sequence identity with IclR. I used a combination of biophysical, biochemical and structural biology techniques to explore the mechanisms by which IclR and AllR interact with their effectors.

I performed site-directed mutagenesis experiments in order to research the role of individual amino acids in interaction of AllR regulator with its previously identified effector glyoxylate and to test whether oligomerization plays a role in effector-induced signal transduction by AllR. Using differential light scattering, which allows high-throughput screening of small molecules for thermostabilization of proteins, I identified potential effectors for the IclR regulator. The physiological relevance of these candidate molecules was tested *in-vitro* and *in-vivo* and their interaction with IclR was characterized by Isothermal Titration Calorimetry and X-ray Crystallography.
ACKNOWLEDGEMENTS

I would like to thank Dr. Aled Edwards for his professional guidance, his infinite enthusiasm and his patience as my supervisor.

I am thankful to my Supervisory Committee members Dr. David Rose and Dr. Alexey Bochkarev for valuable discussions.

I would like to gratefully acknowledge Dr. Alexei Savchenko for his expertise and guidance and Dr. Graciela Lorca and Dr. Svetlana Altamentova for their great contribution to this project, their help and collaboration.

I am thankful to Dr. Vladimir Lunin for his professionalism in X-ray crystallography and for his time, patience, and being always ready to help.

I am grateful to Marina Kudritska, Elena Evdokimova and Tatiana Skarina for sharing with me their expertise in protein purification and crystallization, and all the members of Edwards lab for their support, which was always available, and for warm and cheerful atmosphere in the lab.

I thank Dr. John Walker for the structure of AllR regulator, Dr. Masoud Vedadi for help with Stargazer assay and Dr. Linda Ball for analytical ultracentrifugation experiments.

I thank my parents Tatiana and Vladimir Ezersky and my husband Roman Bershadsky for their love, support and understanding.
# TABLE OF CONTENTS

## INTRODUCTION .............................................................................................................. 1
  - Transcription regulation in bacteria ........................................................................... 1
  - Transcription regulators families ................................................................................ 2
    - Lactose Repressor Protein (LacI) and the GalR/LacI family ...................................... 4
    - LysR family ............................................................................................................. 5
    - AraC/XylS family .................................................................................................... 5
    - MarR family ............................................................................................................ 6
    - TetR family ............................................................................................................. 8
    - The IclR family ....................................................................................................... 9
  - General characteristics .............................................................................................. 9
  - Oligomerization ........................................................................................................ 11
  - Small-molecule effectors (ligands) .......................................................................... 12
  - Proteins in this study ............................................................................................... 14
    - Isocitrate Lyase Regulator from E.coli ................................................................. 14
    - Allantoin utilization regulatory protein AllR from E.coli ....................................... 17

## STRUCTURAL AND BIOCHEMICAL STUDY OF EFFECTOR MOLECULE RECOGNITION BY THE E. COLI GLYOXYLATE AND ALLANTOIN UTILIZATION REGULATORY PROTEIN ALLR. ....................................................... 21
  - My Contribution to This Study: .............................................................................. 21

## ABSTRACT .................................................................................................................. 22

## RESULTS ..................................................................................................................... 25
  - Glyoxylate interaction with AllR binding pocket ..................................................... 35
  - Comparative structural analysis between AllR and TM0065 ligand binding pockets .......................................................... 37
  - Mutagenesis analysis of AllR glyoxylate binding pocket ........................................ 39
  - Effect of glyoxylate binding on AllR conformation ................................................. 42

## DISCUSSION ............................................................................................................... 44

## MATERIALS AND METHODS .................................................................................. 50
  - DNA manipulations and cloning .............................................................................. 50
  - Protein purification ................................................................................................... 50
  - Crystallization and data collection ........................................................................... 51
  - Construction of allR deletion mutant ........................................................................ 53
  - qRT-PCR studies ....................................................................................................... 54
  - Electrophoretic Mobility Shift Assays (EMSA) ....................................................... 55
  - Circular dichroism (CD) spectroscopy .................................................................... 56
  - Size-exclusion chromatography ............................................................................. 56
  - Analytical ultracentrifugation ............................................................................... 56
  - Protein Data Bank accession number ..................................................................... 57

## REFERENCES ............................................................................................................. 57

## GLYOXYLATE AND PYRUVATE ARE ANTAGONISTIC EFFECTORS OF THE ESCHERICHIA COLI ICLR TRANSCRIPTIONAL REGULATOR.................................................. 63
  - My Contribution to This Study: .............................................................................. 63
  - ABSTRACT .............................................................................................................. 64
FIGURES AND TABLES

Table 1 Prokaryotic regulator families (adapted from Ramos et al., 2005). ......... 3
Table 2: Examples of small-molecule effectors of IclR-type transcription regulators. ..... 13
Table 3 Data collection and refinement statistics. ......................................................... 88
Table 4 Summary of the effect of glyoxylate and pyruvate on DNA binding and on the oligomeric state of IclR or its mutant variants. ......................................................... 95

Figure 1 One dimer of TM0065 – an IclR family member from Thermotoga maritime. From Zhang et al., 2002(b). .......................................................... 10
Figure 2 The glyoxylate bypass in E. coli .................................................................... 15
Figure 3 - Modes of repression mediated by Escherichia coli IclR protein. (From Yamamoto & Ishihama, 2003). ................................................................. 17
Figure 4 (a) Purine degradation pathway in E. coli; the enzymes (shown as gene product names) involved in each step of the pathways are described. (b) Transcriptional organization of the genes encoding the enzymes in the allantoin degradation pathways. Adapted from Hasegawa et al., 2008. ................................................................. 19
Figure 5 AllR ligand binding site. ............................................................................... 36
Figure 6 Mutagenesis analysis of AllR ligand-binding site. ........................................ 42
Figure 7 Sedimentation velocity (SV). ......................................................................... 43
Figure 8. Homology between AllR and PAS/GAF domains. .................................... 47
Figure 9 Ligand binding pocket superimposition of AllR and IclR. ......................... 92
Figure 10 Gel filtration analysis of IclR oligomerization in presence of glyoxylate or pyruvate. .............................................................................................. 93
Figure 11 Analysis of IclR L143A and M146A oligomerization in presence of glyoxylate and pyruvate ................................................................. 96
Figure 12 Functional analysis of key amino acids involved in interdomain interactions and effector binding. ........................................................................ 97
## NAMES AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>Analytical Ultracentrifugation</td>
</tr>
<tr>
<td>E.chrisan temi</td>
<td><em>Erwinia chrysanthemi</em></td>
</tr>
<tr>
<td>E.coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic Mobility Shift Assay</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>HTH</td>
<td>Helix-Turn-Helix</td>
</tr>
<tr>
<td>IclR</td>
<td>Isocytrate Lyase Regulator</td>
</tr>
<tr>
<td>IPTG</td>
<td>β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal Titration Calorimetry</td>
</tr>
<tr>
<td>Kdg</td>
<td>2-keto-3-deoxygluconate</td>
</tr>
<tr>
<td>P.aeruginosa</td>
<td><em>Pseudomonas aeruginosa</em></td>
</tr>
<tr>
<td>P.putida</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>SDM</td>
<td>Site Directed Mutagenesis</td>
</tr>
<tr>
<td>wHTH</td>
<td>Winged helix–turn–helix</td>
</tr>
</tbody>
</table>
INTRODUCTION

Transcription regulation in bacteria

Bacteria are exposed to a variety of conditions in their environment, such as varying temperatures, availability of different nutrients, exposure to toxins, including products of their own metabolism. As a result, they need to be able to rapidly adjust to the changing conditions. The adaptive responses are usually mediated through regulatory proteins, mainly transcription regulators. Other types of response to environmental triggers include variations in levels of RNA polymerase sigma factors and such phenomena as cell differentiation to form spores in response to harsh conditions (Molina-Henares et al., 2006).

A type of bacterial transcription regulators known as one-component systems includes proteins that serve both as chemical sensors and transcription regulators. Those proteins include an “input domain” and an “output domain” in their structure. One-component systems of control of gene expression are more common and more diverse in bacteria and archea than two-component systems of transcription regulation. The latter includes two proteins: one has an “input domain” that receives the environmental signal and activates the other, that has an “output” domain and regulates transcription of genes through DNA-binding. One-component systems are older than two-component systems and might have served as their evolutionary precursors (Ulrich et al., 2005).

The mechanism of action of many families of one-component transcription regulation systems has been characterized on structural level. Their general mechanism of action is conserved: interaction of the “input domain” with small molecule ligand modulates the protein-DNA or protein-protein interaction activity of the “output domain”.

**Transcription regulators families**

One-component transcription regulation systems that have been identified in sequenced genomes can be assembled into families, which are usually named after the best-characterized member. This assembly is based on sequence similarity, predominantly in the DNA-binding helix-turn-helix (HTH) domain (Gorelik et al, 2006). Proteins within a family tend to have similar size (number of amino acids) and to control genes that are involved in related functions. Among the largest of such families (“superfamilies”) are LysR, TetR, IclR, GntR and AraC (Perez-Rueda et al., 2004). These families often include proteins with significant similarity in DNA-binding domain and different types of the other domains (that may have several functions, such as allosteric regulation, dimerization, etc). For example, the GntR superfamily is divided into four major subfamilies based on the specific type of effector-binding domain, namely the FadR, HutC, MocR, and YtrA subfamilies (Gorelik et al, 2006). However, in case of the IclR family the C-terminal ligand binding domain is more conserved than the HTH domain (Perez-Rueda and Collado-Vides, 2000).

Only for a minority of transcription regulators have the small molecule effectors been identified. Following are examples of some representative prokaryotic transcription regulators families, regulation mechanisms and small molecule effectors.
<table>
<thead>
<tr>
<th>Family</th>
<th>Action</th>
<th>Some regulated functions</th>
<th>DBD motif</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>LysR</td>
<td>Activator/repressor</td>
<td>Carbon and nitrogen metabolism</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>AraC/XylS</td>
<td>Activator</td>
<td>Carbon metabolism, stress response and pathogenesis</td>
<td>HTH</td>
<td>C-terminal</td>
</tr>
<tr>
<td>TetR</td>
<td>Repressor</td>
<td>Biosynthesis of antibiotics, efflux pumps, osmotic stress, etc.</td>
<td>HTH</td>
<td>C-terminal</td>
</tr>
<tr>
<td>LuxR</td>
<td>Activator</td>
<td>Quorum sensing, biosynthesis and metabolism, etc.</td>
<td>HTH</td>
<td>C-terminal</td>
</tr>
<tr>
<td>LacI</td>
<td>Repressor</td>
<td>Carbon source utilization</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>ArsR</td>
<td>Repressor</td>
<td>Metal resistance</td>
<td>HTH</td>
<td>Central</td>
</tr>
<tr>
<td>IcIR</td>
<td>Repressor/activator</td>
<td>Carbon metabolism, efflux pumps</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>MerR</td>
<td>Repressor</td>
<td>Resistance and detoxification</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>AsnC</td>
<td>Activator/repressor</td>
<td>Amino acid biosynthesis</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>MarR</td>
<td>Activator/repressor</td>
<td>Multiple antibiotic resistance</td>
<td>HTH</td>
<td>Central</td>
</tr>
<tr>
<td>NtrC (EBP)</td>
<td>Activator</td>
<td>Nitrogen assimilation, aromatic amino acid synthesis, flagella, catabolic pathways, phage response, etc.</td>
<td>HTH</td>
<td>C-terminal</td>
</tr>
<tr>
<td>OmpR</td>
<td>Activator</td>
<td>Heavy metal and virulence (response regulator of a two-component system)</td>
<td>Winged helix</td>
<td>C-terminal</td>
</tr>
<tr>
<td>DeoR</td>
<td>Repressor</td>
<td>Sugar metabolism</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>Cold shock</td>
<td>Activator</td>
<td>Low-temperature resistance</td>
<td>RNA binding domain (CSD)</td>
<td>Variable</td>
</tr>
<tr>
<td>GntR</td>
<td>Repressor</td>
<td>General metabolism</td>
<td>HTH</td>
<td>N-terminal</td>
</tr>
<tr>
<td>Crp</td>
<td>Activator/repressor</td>
<td>Global responses, catabolite repression and anaerobiosis</td>
<td>HTH</td>
<td>C-terminal</td>
</tr>
</tbody>
</table>

Table 1 Prokaryotic regulator families (adapted from Ramos et al., 2005).
Lactose Repressor Protein (LacI) and the GalR/LacI family

The LacI repressor from *E.coli* was one of the first bacterial transcription regulators for which activity was studied (Jacob and Monod, 1961). It represses the expression of enzymes involved in lactose catabolism. When lactose is present in the bacteria’s environment the repression is released and enzymes for lactose catabolism are expressed. The structure of LacI has two domains: a helix-turn-helix (HTH) DNA binding domain and a ligand binding domain that includes two sub-domains connected by three flexible fragments. The helix-turn-helix motif makes the protein able to bind to the DNA operator of the lactose operon, so that it will interfere with the binding of RNA-polymerase and the transcription process. LacI binds its DNA operator as a homodimer, while its C-terminal domain allows tetramerization. The resulting tetramer binds two DNA binding sites which results in tighter repression. Binding of a small effector molecule called allolactose, which is a lactose metabolite, between the two sub-domains in the ligand binding domain produces a conformational change that involves the DNA-binding domain. The change in the helix-turn-helix reduces the affinity of the Lac repressor to the DNA operator, making transcription possible. Derepression can also be induced by isopropyl β-D-1-thiogalactopyranoside (IPTG), which is a gratuitous inducer (it cannot be metabolized by enzymes production of which it induces). The stoichiometry is two effector molecules per protein homodimer (Taraban *et al.*, 2008).

Other members of the LacI, such as PurR or TreR share similar structures and mechanisms. However, most LacI homologues are homodimers and only a few undergo tetramerization (Taraban *et al.*, 2008). The ligands that GalR/LacI family members bind include carbohydrates (such as D-galactose, D-fucose, D-ribulose and others),
nucleosides (and a purine derivative hypoxanthine), and opines, which are modified amino acids (Weickert and Adhya, 1992).

**LysR family**

Structure of proteins from another large family – LysR – is similarly composed of HTH DNA-binding domain and a ligand binding domain, which is made of two α/β sub-domains connected by two short polypeptide fragments (as was shown by a structure of CysB regulator) (Tropel and van der Meer, 2004). These connecting fragments form a hinge or cleft that accommodates the small molecule effector. Binding of the effector modulates the protein’s activity through conformational change, as in the case of LacI/GalR family. LysR family members act as activators or repressors and also regulate their own expression. DNA bending was shown to be involved in regulation (Maddocks and Oyston, 2008).

The effector molecules have been identified for some of LysR type regulators. They include substrates, products and intermediates of pathways under their control and related metabolites. BenM is an exceptional LysR type protein that controls pathways for benzoate utilization or degradation. BenM is unique because it can bind two different effectors: benzoate binds in one region (not the usual effector binding cleft) and induces a conformational change, and \textit{cis,cis}-muconate (product of catechol degradation), which binds at the regular site, producing a synergistic effect (Maddocks and Oyston, 2008).

**AraC/XylS family**

AraC/XylS family of transcriptional regulators includes mostly transcription activators. For these regulators, binding of the effector triggers a conformational change,
which changes the oligomeric state of the regulator (Gallegos et al., 1997). The mode of regulation in AraC/XylS family appears to be dimerization driven. AraC regulates the transcription of enzymes of the arabinose operon in E.coli. In the absence of arabinose its C-terminal DNA binding domain, which includes two HTH motifs, binds widely-spaced DNA half-sites. DNA bending can be induced by binding. Binding of arabinose to the N-terminal ligand-binding domain, which is not well conserved between the family members, blocks one of the two possible dimerization interfaces, enabling the protein to bind a different set of DNA half-sites, which are closely-spaced. The role of the N-terminal domain in effector binding was also confirmed by mutagenesis studies. For another regulator, XylS, binding of a regulator, m-toluate, induces a conformational change in the protein that results in change to the DNA at binding site that favours formation of open-complex and transcription (Tropel and van der Meer, 2004).

**MarR family**

This family is named after MarR (Multiple Antibiotic Resistance Regulator) from E.coli. The proteins in this family regulate a range of functions bacteria and archaea, from synthesis of pathogenic factors to resistance to antibiotics, organic solvents and household cleaning agents. MarR in E.coli represses transcription from marRAB operon that includes MarA regulator, transcription factor from AraC family that regulates a global network of over 60 genes (Alekshun et al, 2001; Wilkinson and Grove, 2006).

MarR proteins bind their DNA operators as homodimers. This binding is inactivated in presence of effector molecules which are mostly anionic lipophilic (usually phenolic) compounds. Some MarR regulators can be affected by more than one small molecule effector. Several compounds, including 2,4-dinitrophenol, menadione,
plumbagin and salicylic acid have been shown to antagonize *E.coli* MarR binding to DNA operators *in vitro* and *in vivo* (Wilkinson and Grove, 2006).

The structure of MarR regulators includes a conserved wHTH (winged helix-turn-helix) DNA-binding domain and a dimerization domain, while the effector-binding domain overlaps with the DNA-binding domain. Complex structure of *E.coli* MarR with salicylate has revealed two ligand-binding sites. One site is located inside the globular DNA-binding domain. The other ligand-binding site is exposed to the solvent (Alekshun *et al*, 2001). However, the physiological relevance of two binding sites is questionable, since the MarR crystals were soaked in very high concentration of salicylic acid, and experimental data that would support or disprove this claim is controversial (Wilkinson and Grove, 2006).

Two mechanisms of effector-induced regulation are proposed for MarR proteins. One states that binding of the ligand coordinates residues that participate in contacts with the DNA (Alekshun *et al*, 2000). The other states that ligand binding might stabilize MarR dimer in a conformation that does not allow DNA binding (Wilkinson and Grove, 2006).

Residues that participate in H-bond formation in one of the ligand-binding sites are strictly conserved in MarR homologue MexR from *P.aeruginosa*, but not the proline that mediates hydrophobic contact with salicylate ring and not the residues of the other ligand binding site. Therefore, MexR may differ from MarR in the nature of its effectors, mode of DNA binding and mechanism of regulation. However, MexR ligands remain unknown and only its apo-structure is available. Identification of the physiological ligand would
shed light not only on the details on interaction of MexR with it, but also on the physiological function of the multidrug efflux system under its control (Lim et al., 2002).

TetR family

The TetR family of transcription regulators, named after the protein that regulates genes of resistance to tetracycline, generally contains transcription repressors. Members of the family share high similarity in HTH DNA-binding region and control genes that are involved in multidrug resistance, catabolic pathways, biosynthesis of antibiotics, osmotic stress mechanisms and pathogenicity. No sequence conservation is observed outside the DNA-binding domain, which reflects diversity of the effectors that these family members bind.

The structures for complexes of TetR regulators with DNA and with tetracycline are available. TetR presents as homodimer. Each monomer consists of 10 α-helices. The N-terminal domain is a HTH DNA-binding motif, and the C-terminal domain is responsible for dimerization and effector binding. Binding of tetracycline changes relative orientation of the C-terminal domain α-helices, which is translated to the N-terminal domain. The recognition helices of each monomer in the homodimer are pulled apart and DNA contacts are disrupted.

The QacR regulator confers resistance to monovalent and bivalent cationic lipophilic antiseptics, disinfectants and other drugs. Structures of its complexes with different effectors revealed presence of expansive drug-binding pocket, so that different drugs bind in partially overlapping sub-pockets. The stoichiometry is one drug molecule for each homodimer (two binding pockets). A monomer that has a drug bound undergoes a major structural change resulting in release of DNA (Ramos et al., 2005).
The IclR family

General characteristics

The IclR family of transcription regulators is named after its founding member Isocitrate Lyase Regulator from E.coli (Sunnarborg et al., 1990). These regulators are an example of one-component signal transduction systems. The size of these proteins is typically in the range of 240-280 residues. More than 400 IclR family members have been detected in 61 genomes belonging to gram-positive, alpha-, beta- and gamma-proteobacteria and archaea. Among those there are pathogenic species, such as Salmonella, Yersinia or Bordetella, three strains of which contain particularly high number of IclR family members (18-40 per genome). (Molina-Henares et al., 2006)

The structure of an IclR family member from T. maritima (TM0065) revealed two domains: a HTH motif (output domain) at the N-terminal end and a characteristic ligand binding domain (input domain) at the C-terminal end, known as IclR domain (PF01614) in Pfam database (Zhang et al., 2002). Before the crystal structure of any of the family members was available, mutagenesis experiments of IclR family member PobR from Acinetobacter showed that mutations in the C-terminal part of the protein changed the signal specificity, but have no effect on DNA binding and transcriptional activation by the protein (Kok et al., 1998).
IclR family members in the different organisms control a variety of catabolic and anabolic pathways, as well as sporulation processes (such as SffR and SggR in Stroptomyces griseus and Stroptomyces coelicolor (Maloy and Nunn, 1982, Sunnarborg et al., 1990)), pathogenic processes (such as KdgR and RexZ that are involved in regulation of plant cell degradation by a plant pathogen Erwinia species (Nasser et al., 1992, Liu et al., 1999, Thomson et al., 1999)), and even drug resistance (TtgV in Pseudomonas putida (Guazzaroni et al., 2004, Rojas et al., 2003).

In contrast to other transcription regulator families, which usually include only activators or only repressors (such as XylS/AraC (Gallegos et al., 1997), or TetR (Ramos...
et al., 2005)), the IclR family includes proteins that act as repressors (i.e. the founding member IclR from *E.coli*, AllR from *E.coli*, KdgR from *E.chrisantemi* and others), activators (i.e. RexZ from *E.chrisantemi* (Thomson et al., 1999)), as well as proteins with dual function (i.e. PobR and PcaU in *Acinetobacter* sp. ADP1, that positively regulate catabolic pathways of aromatic compounds in this species, and negatively regulate their own expression (DiMarco et al., 1993, Gerischer et al., 1998, Kok et al., 1998)).

**Oligomerization**

DNA binding sites (operators) of several IclR regulators contain palindrome and pseudo-palindrome repeats. This feature is characteristic of many other helix-turn-helix transcriptional regulators, and usually implies that the regulator binds its DNA operator as dimer or oligomer (Potier et al., 1998; Donald et al., 1998). For some repressors from other families, binding of the effector induces a conformational change that changes the oligomeric state of the protein (Chen and Calvo, 2002; Schleif, 2004; Vilar and Saiz, 2005), therefore research of oligomerization is important to understand regulatory mechanisms.

The oligomeric state of various IclR family members has been assessed. The proteins are dimers or tetramers in solution and/or in crystal structure (full-length structure available for TM0065 in *T.maritima* (Zhang et al., 2002(b)). PcaU in *Acinetobacter* (Popp et al., 2002), Pir in *Erwinia* (Nomura et al., 1998) and PcaR in *Pseudomonas* (Guo and Houghton, 1999) are dimers in solution, however IclIR from *E.coli* was shown as a dimer by sedimentation equilibrium experiments (Negre et al., 1992), as a tetramer in sucrose density gradients earlier by the same authors (Negre et al.,
and eventually as mostly tertramers with presence of dimers by mass-spectrometry (Donald et al., 2001).

IclR from *E.coli* binds two sequences: *aceBAK* and *iclR*, both of which are variants of a 15 bp palindrome (Pan et al., 1996; Hosfield et al., 1996). This fits well to helix-turn-helix interaction. However, DNAase footprinting experiments have shown that 28-34 bp are protected (Cortay et al., 1991; Donald et al., 1996; Gui et al., 1996), suggesting either DNA bending or higher multimerization state than just a dimer (Donald et al., 2001). Donald et al used mass spectrometry to investigate IclR and AllR (or GclR, as they refer to it) conformation and their complexes with DNA operators. This study has shown that both proteins appear mostly as tetramers, however a smaller fraction appears as dimers, suggesting dimer-tetramer equilibrium. The study has also revealed an unexpected stoichiometry for protein-DNA complexes: 1 protein subunit : 1 dsDNA, unlike other transcription regulators with helix-turn-helix motifs, such as TrpR or LacI, that have shown 1:2 and 2:4 ratio, respectively. The authors suggest that in the case of IclR and AllR the other pair of helix-turn-helix motifs might be unavailable for binding due to steric reasons (Donald et al., 2001).

**Small-molecule effectors (ligands)**

Controversy exists regarding identity of the ligand of IclR repressor. Cortay et al attempted to identify the metabolite that is involved in control of acetate operon – the ligand for IclR. Using gel shift techniques, they showed that addition of phosphoenol pyruvate, but not acetate, acetyl-CoA, pyruvate or oxaloacetate disrupted the binding of purified IclR protein to a DNA fragment that contains its operator (Cortay et al., 1991). However, later this result could not be reproduced (Yamamoto & Ishihama, 2003).
Similarly, the ligands of only a few of the IclR family are known. Examples are presented in the following table.

<table>
<thead>
<tr>
<th>Regulator</th>
<th>Organism</th>
<th>Small molecule effector (name and structure)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllR</td>
<td>E.coli</td>
<td>Glyoxylate</td>
<td>Rintoul et al., 2002. See 1.2.2.2.</td>
</tr>
<tr>
<td>MhpR</td>
<td>E.coli</td>
<td>3-(3-hydroxyphenyl) propionic acid</td>
<td>Torres et al., 2003.</td>
</tr>
<tr>
<td>KdgR</td>
<td>E.chrysanthemi</td>
<td>2-keto-3-deoxygluconate</td>
<td>Reverchon et al., 1991.</td>
</tr>
<tr>
<td>PobR</td>
<td>Acinetobacter</td>
<td>p-Hydroxybenzoate</td>
<td>DiMarco et al., 1993.</td>
</tr>
<tr>
<td>PcaU</td>
<td>Acinetobacter</td>
<td>Protocatechuate</td>
<td>Popp et al., 2002.</td>
</tr>
<tr>
<td>AttJ</td>
<td>Agrobacterium tumefaciens</td>
<td>N-3-oxo-octanoyl homoserine lactone (3OC8HSL)</td>
<td>Zhang et al., 2002.</td>
</tr>
<tr>
<td>TtgV</td>
<td>Pseudomonas putida</td>
<td>1-naphthol</td>
<td>Guazzaroni et al., 2005.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2,3-dihydroxynaphthalene</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Examples of small-molecule effectors of IclR-type transcription regulators.

These small molecules show very substantial structural and chemical variability, while the structure of the proteins that belong to the IclR family and particularly of their
ligand-binding domains appears to be conserved. Therefore, discovering how specificity and selectivity are obtained in protein-ligand interactions in this family is of interest.

Proteins in this study

This study focuses on two representative transcription regulators from the IclR family: the founding member Isocytrate Lyase Regulator (IclR) and allantoin utilization regulatory protein AllR, also known as GclR. Both proteins act as repressors of catabolic pathways in *E. coli* and share 42% sequence identity (Walker *et al*., 2006). The pathways under regulation of IclR and AllR share common intermediates. While there is controversy regarding the identity of the physiological ligand of IclR, the effector molecule of AllR is known. This effector is glyoxylate – an intermediate of the pathways under AllR regulation (Rintoul *et al*., 2002).

*Isocitrate Lyase Regulator from E.coli*

The Isocitrate Lyase Regulator was discovered in *E.coli* by C. B. Brice and H. L. Kornberg in 1968 by interruption of conjugation method. Mutants in the *iclR* gene constitutively expressed isocitrate lyase. IclR was proposed to be the regulator that controls expression of isocitrate lyase and later of all three enzymes that belong to the glyoxylate bypass (Brice, Kornberg, 1968).
The glyoxylate bypass (or glyoxylate shunt) in *E. coli* includes three enzymes: isocitrate lyase (encoded by *aceA*), malate synthase (encoded by *aceB*) and isocitrate dehydrogenase (IDH) kinase/phosphatase (encoded by *aceK*). The glyoxylate bypass allows *E. coli* to grow on acetate or fatty acids. It is needed for growth on these carbon sources, since it prevents loss of carbon as two CO$_2$ molecules in the Krebs cycle when isocitrate is being converted to ketoglutarate which in turn is then converted to succinyl-CoA. However, when the glyoxylate bypass pathway is activated, isocitrate is converted to glyoxylate and succinate by glyoxylate lyase. Succinate reenters the Krebs cycle, and glyoxylate is being converted to malate by malate synthase (Kornberg, 1966). The third enzyme, isocitrate dehydrogenase kinase/phosphatase, has a dual function and controls flow of isocitrate through the pathway by phosphorylation and dephosphorylation of
isocitrate dehydrogenase, which competes with isocitrate lyase for isocitrate (LaPorte et al., 1989).

The enzymes of the glyoxylate bypass pathway are transcribed together as aceBAK operon and expressed from a single promoter (Chung et al., 1988). It was suggested that IclR regulator, which is encoded by iclR gene, negatively controls expression of this operon, since strains that had mutations in iclR show constitutive expression of aceBAK (Maloy, Nunn, 1982). Binding sites for IclR (IclR boxes) are located in aceB promoter region. One of these sites overlaps the -35 region. The other site is located further upstream in the -100 region (Chung et al., 1988; Donald et al., 1996; Pan et al., 1996).

The iclR gene is located downstream from aceBAK operon. IclR binds its own promoter and regulates its own expression (Gui et al., 1996).

The glyoxylate bypass pathway is tightly regulated on multiple levels – transcriptional and post-transcriptional – which indicates its high importance. The activation of this bypass is dependent on the metabolic state of the bacterial cell. When a preferred carbon source (such as glucose or pyruvate) is present the aceBAK operon is repressed (Molina-Henares et al., 2006).

Yamamoto and Ishihama suggested that the regulation of aceBAK by IclR is obtained through competition between IclR and RNA polymerase, whose binding site on aceB promoter corresponds to IclR box II, and the one on iclR promoter also corresponds to the IclR binding site. They based this suggestion on the observation that binding of IclR to IclR box II or to iclR promoter prevented RNA polymerase from binding to the corresponding regions. However, once RNA polymerase has bound to aceB promoter,
IclR is still able to bind distally located IclR box II. This binding destabilized aceB promoter open complexes, as shown by shorter half-life of the complexes in presence of IclR after addition of heparin. Yamamoto and Ishihama therefore suggested that in addition to competition with RNA polymerase for binding to aceB promoter, IclR also induces dissociation of already formed open-complexes, which further contributes to repression of aceB transcription. These two modes of repression allow regulation of the glyoxylate bypass by IclR on different levels.

![Diagram](image)

**Figure 3 - Modes of repression mediated by Escherichia coli IclR protein. (From Yamamoto & Ishihama, 2003).**

Left, repression mode I involves IclR bound to its target site overlapping RNA polymerase binding site, so that occupancy of the site by IclR prevents the entry of the polymerase. Right, repression mode II involves IclR bound to a distal site with respect to the RNA polymerase site. The interaction of IclR with the a-subunit of the promoter-bound RNA polymerase destabilizes and disassociates the open complex (Yamamoto & Ishihama, 2003).

**Allantoin utilization regulatory protein AllR from E.coli**

The AllR protein in E. coli (also known as GclR (Donald *et al.*, 2001) is a negative regulator of allantoin regulon. Allantoin, which is a product of purine catabolism, is utilized by E. coli as well as other enterobacteriaceae as a nitrogen source.
while growing in anaerobic conditions. Under these conditions allantoin can be used as a sole nitrogen source, but not as a sole carbon source (Cusa et al., 1999).

The twelve genes that encode for the enzymes that participate in allantoin and glycerate catabolism in *E.coli* are found as a cluster that contains five transcriptional units. *allR*, which encodes the AllR protein, is one of the genes in the cluster. It is constitutively expressed under both aerobic and anaerobic conditions (Cusa et al., 1999; Donald et al., 2001).

Cusa *et al* have shown that AllR negatively regulates the allantoin regulon through repression of transcription of *allS* and *allA* genes and the *gcl* operon, which includes another 7 genes that are involved in allantoin and glyoxylate catabolism (Cusa *et al.*, 1999; Rintoul *et al.*, 2002). AllR recognises and binds an almost palindromic sequence that overlaps the promoters of *allS*, *allA* and *gcl*. Due to the proximity of AllR binding sites to transcription initiation sites, the repression of transcription by AllR is probably carried out by steric hindrance or by contact inhibition of RNA polymerase (Müller-Hill, 1998; Rintoul *et al.*, 2002).

The expression of *allS*, *allA* and *gcl* is induced in presence of glyoxylate under both aerobic and anaerobic conditions. When glyoxylate is present, AllR’s binding of its DNA operator is inhibited, resulting in de-repression of transcription. This action is exclusive to glyoxylate and is not produced by structurally similar compounds, such as D-lactate or glyconate. Rintoul *et al* have proposed that glyoxylate is a selective inhibitor of AllR repressor. They also characterized another regulator of the same pathway, AllS that acts as an activator and together with AllR participates in control of allantoin regulon (Rintoul *et al.*, 2002).
Figure 4 (a) Purine degradation pathway in E. coli; the enzymes (shown as gene product names) involved in each step of the pathways are described. (b) Transcriptional organization of the genes encoding the enzymes in the allantoin degradation pathways. Adapted from Hasegawa et al., 2008.

STRUCTURAL AND BIOCHEMICAL STUDY OF EFFECTOR MOLECULE RECOGNITION BY THE E. COLI GLYOXYLATE AND ALLANTOIN UTILIZATION REGULATORY PROTEIN ALLR.


My Contribution to This Study:

I designed the primers for the Site Directed Mutagenesis (SDM) study of AllR regulator and performed the SDM procedure. The mutations affected residues in the suggested ligand-binding pocket or at the tetramerization interface of the protein. I expressed and purified the wild-type and the mutated AllR proteins, and studied the effect of the mutations in the ligand binding pocket on AllR-glyoxylate interaction \textit{in-vitro} by Electrophoretic Mobility Shift Assay (EMSA). Together with Dr. Altamentova I studied the effect of mutation at AllR tetramerization interface by EMSA and by FPLC Size-Exclusion Chromatography.
STRUCTURAL AND BIOCHEMICAL STUDY OF EFFECTOR MOLECULE RECOGNITION BY THE *E. COLI* GLYOXYLATE AND ALLANTOIN UTILIZATION REGULATORY PROTEIN ALLR

John R. Walker\(^{a,b}\), Svetlana Altamentova\(^{a}\), Alexandra Ezersky\(^{a,c}\), Graciela Lorca\(^{d}\), Tatiana Skarina\(^{a}\), Marina Kudritska\(^{a}\), Linda J. Ball\(^{e}\), Alexey Bochkarev\(^{b,d}\) and Alexei Savchenko\(^{a,d}\)

\(^a\)Ontario Center for Structural Proteomics, Best Institute, 112 College St., Toronto, Ontario, Canada.

\(^b\)Structural Genomics Consortium, 112 College St., Toronto, Ontario, Canada.

\(^c\)Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada.

\(^d\)Banting and Best Department of Medical Research, 112 College St., Toronto, Ontario, Canada.

\(^e\)Structural Genomics Consortium, Botnar Research Centre, University of Oxford, OX3 7LD, UK.

ABSTRACT

The interaction of *Escherichia coli* AllR regulator with operator DNA is disrupted by the effector molecule glyoxylate. This is a general, yet uncharacterized regulatory mechanism for the large IclR family of transcriptional regulators to which AllR belongs. The crystal structures of the C-terminal effector-binding domain of AllR regulator and its complex with glyoxylate were determined at 1.7 and 1.8 Å, respectively. Residues involved in glyoxylate binding were explored \textit{in vitro} and \textit{in vivo}. Altering the residues Cys217, Ser234 and Ser236 resulted in glyoxylate independent repression by AllR. Sequence analysis revealed low conservation of amino acids participating in effector binding among IclR regulators, which reflects potential chemical diversity of effector molecules, recognized by members of this family. Comparing the AllR structure to that of *T. maritima* TM0065, the other representative of the IclR family that has been structurally characterized, indicates that both proteins assume similar quaternary structures as a
dimer of dimers. Mutations in the tetramerization region, which in AllR involve the Cys135-
Cys142 region resulted in dissociation of AllR tetramer to dimers in vitro and were functionally
inactive in vivo. Glyoxylate does not appear to function through the inhibition of tetramerization.
Using sedimentation velocity, glyoxylate was shown to conformationally change the AllR
tetramer as well as monomer and dimer resulting in altered outline of AllR molecules.

**Abbreviations used:** MPD – 2-methyl-2,4-pentanediol, TEV protease – Tobacco etch virus
protease, wHTH – winged helix-turn-helix, SV- sedimentation velocity, qRT-PCR – quantitative
real time polymerase chain reaction, FPLC – Fast performance liquid chromatography, TCEP-
Tris (2-carboxyethyl) phosphine hydrochloride.

**INTRODUCTION**

During anaerobic growth *Escherichia coli* and other *enterobacteriaceae* are able to use allantoin
as a nitrogen source\(^1\). Allantoin is first converted to ureidoglucolate, which is further processed
to either glyoxylate or oxaluric acid by ureidoglyconate hydrolase or ureidoglyconate
dehydrogenase, respectively\(^1\). Glyoxylate enters central metabolism via the glycerate pathway\(^3\)-\(^4\),
while oxaluric acid is converted to oxamate and carbamoyl phosphate by oxamate
transcarbamoylase\(^1\).

In *Escherichia coli* twelve of the genes coding for the enzymes involved in allantoin and
glycerate catabolism are clustered together and arranged in five transcriptional units: *allS, allA,*
*allR, gcl-hyi-glxR-0484-allB-o433-glxK* and *allD-f411-f261*\(^2\)-\(^5\). The expression of the genes *allS,*
*allA* and *gcl* is induced in the presence of allantoin and glyoxylate during aerobic and anaerobic
growth\(^6\), while *allD* expression is induced by these compounds only in anaerobic conditions\(^2\).

The induction of the *allS, allA* and *gcl* operons is dependent on the product of the *allR* gene\(^2\)-\(^7\),
which is constitutively expressed in aerobic and anaerobic growth conditions and serves as a
negative regulator for the allantoin regulon\(^2\)-\(^8\). AllR, also called GclR\(^8\), recognizes an almost
perfect palindrome sequence overlapping the promoter regions of the *allS, allA* and *gcl* genes.
The proximity of the AllR binding site and the start site of transcription suggests that this regulator prevents the initiation of transcription by the RNA polymerase by steric hindrance or by contact inhibition.

In the presence of glyoxylate, binding of the AllR regulator to the operator DNA is inhibited and transcription of the allantoin regulon is de-repressed. Inactivation of the AllR gene product causes constitutive expression of the allS, allA and gel promoters. Glyoxylate appears to be a selective effector; structurally related compounds such as glyconate or D-lactate do not have any effect on the ability of the protein to bind DNA in vitro.

AllR is a member of the large IclR family of transcriptional regulators and shares 42% sequence identity with the founding member of this family, IclR (E. coli Isocitrate lyase regulator). The E. coli K12 genome contains eight members of IclR family, while almost 450 bacterial and archaeal members of this family have been identified in other sequenced genomes (and Pfam1614 at http://www.sanger.ac.uk/Software/Pfam/).

The IclR family members have a conserved domain architecture illustrated by high sequence similarity with signature helix-turn-helix DNA binding motive in the N-terminus. Small molecule binding was proposed as a main function for C-terminal part of the IclR regulators. Effectors ranging from 2-keto-3-deoxy-glyconate (E. chrysanthemi KdgR) to p-hydroxybenzoate (Acinetobacter PobR regulator) , protocatechuate (Acinetobacter PcaU regulator) and homogentisate (Pseudomonas putida HmgR) are identified as stimuli for members of this family. The IclR proteins are also highly selective; p-hydroxybenzoate and protocatechuate are structurally similar ligands, yet protocatechuate has no effect on the family member PobR, which responds to p-hydroxybenzoate.

We provided first insights on domain composition and the ligand-binding mechanisms of IclR family members with the crystal structure of the IclR protein family representative, T. maritima TM0065 regulator. The structure confirmed the presence of an N-terminal winged helix-turn-helix (wHTH) DNA-binding domain and also revealed the presence of a large C-terminal domain.
with significant structural similarity to the well-characterized, small molecule-binding GAF\textsuperscript{27; 28; 29} and PAS\textsuperscript{30; 31; 32} domains. However, because the ligand of the TM0065 protein has not yet been identified, details of ligand binding by this protein could not be deduced from the structure. Here, we present the structure of the ligand-binding domain of the \textit{E. coli} AllR protein together with its natural ligand, glyoxylate, and investigate the effects of ligand binding on the interaction of the protein with DNA.

**RESULTS**

**Characterization of \textit{E. coli} AllR C-terminal domain**

The structure of Tm0065 regulator revealed the presence of two distinct domains: An N-terminal \textit{wHTH} DNA-binding domain (Fig. 1, amino acids 1 to 63) and a large C-terminal domain (Fig. 1, amino acids 76 to 246), which are linked together by an $\alpha$-helix (Fig. 1, amino acids 63 to 74). The structural architecture of the C-terminal domain of TM0065 regulator\textsuperscript{26} is similar to that of GAF/PAS domains, which bind small molecules in a large variety of signaling proteins\textsuperscript{27}. To determine if this is a common feature of IclR regulators and whether there is functional homology between the GAF/PAS domains and the IclR regulator C-terminal domain we purified the C-terminal domain of \textit{E. coli} AllR for structural and functional studies.

Three different versions of the AllR C-terminal domain, differing by a few amino acids at the N-terminus, were prepared (see Materials and Methods), based on multiple sequence alignments (Fig. 1), secondary structure predictions and insights from the TM0065 structure. In each case the first residue of the domain was selected to be positioned between the connecting $\alpha$-helix and the first structural element of C-terminal domain (Fig. 1). All three constructs led to the expression of soluble polypeptides of the expected sizes (data not shown). An AllR fragment corresponding to amino acids 97 to 271 (Fig. 1), was selected for further study and called C-AllR.
Both full length AllR and C-AllR were purified to homogeneity and submitted to crystallization trials. The crystals obtained for the full-length AllR regulator were not suitable for structure determination. The structure of the C-AllR was determined in the presence (PDB Code 1T9L) and absence (PDB Code 1TF1) of its effector molecule, glyoxylate, to resolutions of 1.7 Å and 1.8 Å, respectively (Fig. 2 and Table 1). The structure of C-AllR in complex with glyoxylate was determined by single anomalous diffraction (SAD) from selenomethionine-containing crystals. Like the TM0065 structure, the C-AllR domain is an αβ domain, with a centrally located 6-stranded anti-parallel β-sheet, surrounded on one side by two long α-helices (α5 and α9), and the other by three shorter α-helices (α6-α8) (Fig. 2A). The β-sheet is strongly curved with the shortest helix (α6) fitting on the inside of the half-barrel. The secondary structure elements are numbered according to that of the Thermotoga maritima 0065 structure (PDB code 1MKM).
Figure 1 Multiple sequence alignment of AllR against other E. coli members of the IclR family

Yagl (GI1786468), KgdR (GI1736468), Icll (GI1790449), YiaJ (GI1789999), Yjhl (GI1790752), YiaX (GI1799601), MhpR (GI1786541), Actinobacter PobR (GI3172124) and PcaU (GI3264839) and T. maritima Tm0065 regulators. The Tm0065 and C-AllR secondary structure elements are represented as arrows for β-strands and cylinders for α-helices, shown in blue and red respectively. The TM0065 N-terminal DNA-binding domain and inter-domain helix (α5) elements are coloured in light blue, while C-terminal domain elements are coloured dark blue. The first residue of each of the three versions (corresponding to Ala91, Asn95 or Glu97 to Pro271) of C-AllR domain is indicated with an arrow. AllR residues that have been mutated for biochemical studies have been boxed. The amino acids in YfaX, YiaJ, MhpR and PobR regulators, corresponding to AllR mutations are underlined. Residues that are within 4.2 Å from glyoxylate in AllR structure are highlighted in green. Residues that are involved in inter-subunit interactions are highlighted in grey: dark grey represents residues that make contacts within 4.2 Å in all subunits, light grey indicates residues that make contact in a subset of the four subunits. Amino acid sequence alignments were carried out using the program ClustalW71.
The C-AllR structure superimposes well with C-terminal domain of the full-length *Thermotoga maritima* 0065, the only full-length structurally characterized AllR homolog, with a mean rmsd of 1.8 Å over 173 Cα positions (Fig. 2B). Each of the secondary structural elements is preserved with the exception of a 3-10 helix observed in the TM0065 structure, and the intersubunit anti-parallel β-sheet β4a:β4a; residues 139-141 observed in the C-AllR structure (Fig. 1 and 2). Thus the properties of the C-AllR domain probably reflect its properties in the context of the full-length transcriptional factor.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met141Ala</td>
<td>AGCTTAAAATCGACTGCGGCGGCCCTGGGGCCCAGTC</td>
</tr>
<tr>
<td>Cys142Trp</td>
<td>AATCGATGCTAGAATGTCGGGGCAACGTGAGTC</td>
</tr>
<tr>
<td>Cys142, Cys153</td>
<td>GGTATATATATGCGGTACGGGCGGCCCTGGGGCCCAGTC</td>
</tr>
<tr>
<td>Leu149Met</td>
<td>GCCCCCTGGGGCCAGCGGCGCAATC</td>
</tr>
<tr>
<td>Asp207Asn</td>
<td>GAACCTCAGCTACCCCTAAGCTGAAGACAGTGGTCGCTGACT</td>
</tr>
<tr>
<td>His211Asp</td>
<td>TACCGTAGATTAAAAGAGAGTGAATGCTGACTTGAATT</td>
</tr>
<tr>
<td>Leu215Ala</td>
<td>GAGGAGCATGTTTGGAGGCGCGCGCGCGAGTGGTCAAG</td>
</tr>
<tr>
<td>Cys217Ala</td>
<td>GACGATACATCTAGGCTAGTGTGCATCGACAGATTTATGAT</td>
</tr>
<tr>
<td>Cys217Ser</td>
<td>GACGATACATCTAGGCTAGTGTGCATCGACAGATTTATGAT</td>
</tr>
<tr>
<td>Ser234Ala</td>
<td>TAGAGTTTGTGCGCCTAAGCGCAGCTGCCCCTCATC</td>
</tr>
<tr>
<td>Ser234Asn</td>
<td>AGTGGTTTGTGCGCCTAAGCGCAGCTGCCCCTCATC</td>
</tr>
<tr>
<td>Ser236Ala</td>
<td>GCTGCGGTCTCTCGCTCGGTGCGGTGCGGTGCGGTG</td>
</tr>
<tr>
<td>Ser236Met</td>
<td>GTGCGGTCTCTCGCTCGGTGCGGTGCGGTGCGGTG</td>
</tr>
</tbody>
</table>

Mutagenesis primers listed are designated for non-coding strand. Primers for the coding strand are the reverse complements of the corresponding non-coding strand primers. Both strand primers were used for each PCR reaction. The sequence for the mutant amino acid is underlined. Modified nucleotides are in italic.

**Table 1 AllR site-directed mutagenesis primers**
Figure 2 Overall structure of C-AllR.

A. Ribbon diagram of a monomer of the AllR ligand-binding domain in a complex with glyoxylate (PDB code 1T9L). Loops are colored yellow, α helices red and β strands blue. Glyoxylate is represented as a stick figure, with oxygens coloured red, and carbon atoms white. Secondary structural elements and the amino and carboxyl termini are labeled; B. Superimposition of a monomer of the AllR ligand-binding domain (PDB code 1TF1) with a monomer of the full length *T. maritima* TM0065 regulator (PDB code 1MKM.) 1TF1 is depicted as a yellow tube, while 1MKM is colored cyan. The amino and carboxyl termini are labelled.
C-AllR interdomain interaction – functional implications

The palindromic nature of the AllR binding site suggests that the functional unit of AllR is an oligomer. To determine the oligomeric state of AllR, we subjected full-length and C-AllR preparations to gel filtration analysis. Based on calculated masses, the purified full-length AllR regulator was predominantly a tetramer, while C-AllR was a dimer (data not shown). While the TM0065 protein was found in predominantly dimeric form in solution it formed tetramers in crystal structure\textsuperscript{26}. In the Tm0065 structure, dimers were formed through interactions in N-terminal DNA-binding domains and inter-domain \( \alpha \)-helixes (\( \alpha \)-helix 5 in Fig. 1). In contrast, TM0065 tetramers formed as a dimer of dimers connected through C-terminal domains. Thus in Tm0065 N-terminal and C-terminal domains are involved in separate intersubunit interactions. In the C-AllR structure, AllR C-terminal domains dimerise through the \( \beta_4 \) strand and the loops adjacent to this strand that covers 1282 Å\(^2\) (Fig. 3A). This corresponds to the interface observed between the C-terminal domains of TM0065 (1209 Å\(^2\)). Therefore although the C-AllR structure lacks an AllR N-terminal domain, we can infer that the full-length AllR oligomerization arrangement is similar to that of Tm0065.

The dimerization interface within AllR ligand-binding domain comprises the region between residues Cys135 to Cys142, which extends toward the ligand-binding pocket of the other monomer (Fig. 3A and B). The interface is stabilized by residues Val139-Met141, forming an intersubunit antiparallel \( \beta \)-sheet. Interestingly, the Met141 side chain also contacts the glyoxylate molecule (Fig. 3B). To test the importance of the C-terminal inter-domain contacts in AllR function, we mutated the dimerization interface in the context of the full-length AllR by site-directed mutagenesis of Cys142 to tryptophan and by replacing the region between Cys135 – Cys142, which make main chain contacts between dimers, by a shorter sequence Ala-Pro-Ala.
Figure 3 AllR inter-domain interface.

A. Ribbon diagram of the subunit interface between two monomers in the 1T9L structure. Secondary structural elements are colored as in Fig. 2B. Glyoxylate is represented as a stick figure. B. Close-view of the interface between two monomers in the 1T9L structure. Glyoxylate is represented as CPK. Residues that make contact between the two monomers are shown as stick figures, and, in the case of one monomer as a light violet surface area. Hydrogen bonds between residues of β-strand 4a are depicted as green spheres.
The two mutant proteins (Cys142Trp and Cys135 – Cys142/AlaProAla) were purified, analyzed by circular dichroism (CD) to ensure that they were properly folded, and tested by gel filtration for their oligomerization properties. The CD spectra and gel filtration profiles were compared to those of the wild type AllR. The CD analysis showed that both mutants displayed the same CD spectra, as did the wild type AllR (data not shown). The gel filtration on the other hand showed that, contrary to the wild type AllR, the Cys142Trp and Cys135 – Cys142/AlaProAla mutants were now dimers (data not shown).

The Cys142Trp and Cys135 – Cys142 dimerization mutants were then tested for their ability to repress the expression of the gcl and allA genes in vivo. For these studies an allR E. coli strain was constructed and complemented by either the wild type or mutant AllR regulators expressed from a low copy number vector, pBAD33, using an arabinose-inducible promoter. The transcript levels of two known target genes, allA and gcl, were used as the read-out. Transcript levels were measured by qRT-PCR, in cells grown on the minimal medium containing xylose as a carbon source with or without added glyoxylate. The expression profiles for both allA and gcl genes had the same pattern and thus the profile of the gcl gene only is presented. The allR E. coli strain carrying the pBAD33 vector alone was used as a control.

As summarized in the Fig. 4A, expression of gcl in the BW25113 E. coli increases 10-fold in the presence of glyoxylate compared to growth on xylose alone. Deletion of allR (allR- graphs) in this strain caused dramatic derepression (5000 fold) of gcl, confirming a major role of AllR in transcription regulation of this gene. This deletion experiment highlighted that glyoxylate induction only partially releases AllR-mediated repression of gcl.

As further evidence of a glyoxylate effect on the AllR regulator, growth on glyoxylate containing medium had no effect on the gcl gene expression in a allR- E. coli strain.

gcl repression in the allR- strain was restored by complementation with plasmid-born wild type AllR (Fig. 4A, WT graphs). Similar to the results obtained for the BW25113 strain, a 10-fold
increase in gcl expression was observed during the growth of allR- strain harboring wild type AllR on glyoxylate. On the other hand, complementation of allR- strain with Cys135 – Cys142 deletion and Cys142Trp AllR mutants was severely diminished compared to the wild type AllR. The Cys135 – Cys142 deletion mutant lacked the ability to repress gcl, with a notable effect on gcl expression compared to the allR- strain. While the Cys142Trp AllR mutant caused 50-fold repression of gcl expression compared to allR- strain, that nevertheless was significantly less than the effect of wild type AllR regulator (5000 fold repression during the grown in absence of glyoxylate). No considerable change in gcl expression was observed in allR- strains carrying either of mutant AllRs during growth on glyoxylate.

The ability of Cys135 – Cys142 deletion and Cys142Trp mutants to bind operator DNA was tested in vitro by electrophoretic mobility shift assays (EMSA). Purified Cys135 – Cys142 deletion and Cys142Trp AllR mutants were compared with wild type AllR in a binding assay with a 26 bp operator found in the promoter region of the gcl gene (Fig. 4B). In agreement with in vivo studies EMSA results (Fig. 4B) confirmed a significant (7 fold at the 25 nM protein concentration) decrease in the binding ability of Cys142Trp mutant protein. No significant mobility shift was observed in the presence of up to 100 nM Cys135 – Cys142 deletion mutant, suggesting a dramatic loss of DNA binding ability by this mutant protein.

Despite the fact that Cys142Trp mutant protein was predominantly dimeric in solution in contrast to wild type AllR, which formed tetramer, both proteins complexes with operator DNA had similar migration rate in EMSA (Fig. 4B). This similarity indicated that Cys142Trp mutant protein might regain the ability to form tetramers when bound to the operator DNA.

The results summarized in the Fig. 4 clearly demonstrate that the ability of AllR to repress transcription through binding to operator DNA was severely weakened by the Cys142Trp mutation and Cys135 – Cys142 deletion. These results confirm the biological relevance of inter-domain contacts in C-AllR crystals and highlight the importance of AllR tetramerization for binding of operator DNA.
In the C-AllR crystal structure, another dimerization interface was also observed (Fig. 3A). This interface was dependent upon residues 89-96, which make up part of the N-terminal cloning tag, and thus probably is not biologically relevant.

**Figure 4 Mutational analysis of AllR inter-domain surface.**

*A. In vivo* studies of mutations in the inter-domain region. Relative expression levels of *gcl* gene was measured by qRT-PCR. Gcl expression level was compared between *E. coli* background strain BW25113 (*allR*+ graph), *allR* deletion strain (*allR*− graph) and *allR* deletion strain complemented with wild type AllR (WT) and inter-domain region mutant AllR proteins (*ΔC135-C142 and C142W*). All the experiments were conducted in absence (dark grey bar) or in the presence (light grey bar) of the glyoxylate (25 mM). *B. In vitro* studies of mutations in the inter-domain region by EMSA. The 5 nM 26 bp DNA fragment corresponding to the operator region and 0 to 100 nM of the purified wild type or mutant AllR protein were used for each binding assay.
Glyoxylate interaction with AllR binding pocket

In the structure of the C-AllR/glyoxylate complex, the glyoxylate molecule is located above the N-terminus of α6 (Fig. 2A), with the glyoxylate aldehyde oxygen positioned over the helix axis, and the carboxylate group proximal to the concave surface of central β-sheet. Along with these two secondary structural elements, the ligand-binding cavity is formed by residues from strand β4a to helix α6, and the loop between strands β5 and β6 (Fig. 1, 2A and 5). In all, 15 residues (Asn118, Leu129, Met141, Ala143, Leu149, Ser154, Gly155, Ala156, Asp207, His211, Val212, Leu215, Cys217, Ser234 and Ser236) make up the ligand binding pocket, covering 192.8 Å² of surface area, and enclosing a volume of 150.4 Å³. The position of the ligand-binding pocket correspond to that in PAS and GAF domains. Of the 15 residues comprising the ligand-binding pocket, only three residues, Leu129, Ala143, and Val212 do not either bond with or come within van der Waals distance of glyoxylate. The high resolution of the complex structure revealed the atomic details of the interactions between protein and the effector molecule (Fig. 5A). The aldehyde oxygen of glyoxylate makes hydrogen bonds with the main chain amide of Gly155 and the side chains of Cys217, Asp207 and, through an ordered water molecule (present in two of the four binding sites), with His211 (Fig. 5A). The aldehyde oxygen is also within van der Waals distance of Ser154 CB (average distance 3.3 Å (Table 1)). The carboxylate atom O3 of glyoxylate forms a hydrogen bond to the main chain amide of Ala156, while the other carboxylate oxygen forms hydrogen bonds to the hydroxyls of Ser234 and Ser236 (Fig. 5A). Binding of glyoxylate via the main chain amides of two sequential residues has been observed previously, such as in the case of malate synthase G, however in that structure the amides interact with the carboxylate oxygens, rather than one oxygen from the carboxylate and the other from the aldehyde.
Figure 5 AllR ligand binding site.

A. 2fo-fc omit map calculated at one σ, showing the electron density at the glyoxylate-binding site. Protein residues and ligand at the binding site are represented as stick figures, while potential hydrogen bonds are depicted as green spheres. B. Superimposition of the active site of AllR with that of TM0065. Protein residues and glyoxylate are represented as stick figures, water molecules and the zinc ion (blue) of TM0065 are represented as small spheres. Components of TM0065 are represented as semi-transparent objects, while components of AllR are colored with grey representing carbon atoms, blue nitrogen, and red oxygen. Lines of small green spheres represent potential hydrogen bonds. Analysis of the AllR ligand binding site has been assisted by CastP 72, CNS 1.1 66, and the program HBPLUS v 3.0 73.
Both Ser234 and Ser236 could make potential H-bonds to the carboxylate O2, but Ser234 is in a more favourable position. The main chain amides of 155 and 156 donate protons to make hydrogen bonds to glyoxylate O1 and O3 respectively. The proton of the carboxylate group will then more likely be found on O2, bound either to the anti, or more favorably, the syn lone pair of electrons. Ser234 OG is in position to accept the proton bound to the syn lone pair, as it is forward and approximately (within 15 degrees) coplanar with the carboxylate group, and furthermore this residue does not make H-bonds to other residues. In contrast, Ser236 OG would more likely interact with a proton in the less-stable anti-conformation, and as it already accepts a proton from the main-chain amide of Asn118, the second H-bond to glyoxylate O2 would be less favorable energetically.

The position of glyoxylate correlates well not only with the ligand-binding regions of GAF and PAS domains, but also with the metal ion and water molecules in the structure of the full-length TM0065. The partially charged C1 atom of glyoxylate is positioned very closely (1.2 Å) to that of the metal ion in TM0065 (Fig. 5B). These data support our previous suggestion\(^2^6\) that this cavity forms the ligand-binding pocket for the IclR transcriptional regulators.

**Comparative structural analysis between AllR and TM0065 ligand binding pockets**

With the TM0065 structure and the C-AllR-glyoxylate complex structure in hand, it was possible to compare the characteristics of the two ligand binding pockets (Fig. 1 and Fig. 5B) as well as to use multiple sequence alignment to compare the characteristics of the potential ligand-binding residues in the other IclR family members.

The closest interactions of the AllR protein with the glyoxylate are between Cys217 of AllR and C1 of the glyoxylate; the distance (averaged among the four molecules in the structure) between the sulfur atom and glyoxylate is 2.6 Å. The ligand interaction with Cys217 is mimicked in TM0065 by the interaction of the co-crystallized metal ion with the corresponding cysteine at position 196. The observation that in TM0065 the cysteine can bind a metal ion, and previous
mass spectroscopic analysis that shows that this residue is capable of making a covalent bond to β-mercaptoethanol, indicates that this residue is particularly reactive. A cysteine residue is found at the corresponding position in a large number of IclR regulators, suggesting that it is a key feature of the ligand-binding pockets.

The C1 of the glyoxylate also comes into van der Waals contact with Leu149 and Leu215. Multiple sequence alignments demonstrate that hydrophobic amino acids predominate at the corresponding positions in the members of IclR family: TM0065 has valine and isoleucine in the respective positions (Fig. 1).

Overall, the TM0065 ligand-binding pocket is slightly larger than that of C-AllR, with 245.5 Å² of surface area, and enclosing a volume of 234.7 Å³. Five amino acids (Met141, Asp207, Cys217, Ser234 and Ser236) that interact with glyoxylate in the C-AllR are conserved in the ligand-binding pocket of TM0065 (Fig. 1 and 5B). An additional four AllR residues in the ligand-binding pocket, Leu149, Ser154, His211, Leu215, are substituted by chemically similar residues in TM0065 (Val128, Thr133, Asp190 and Ile194, respectively). The amino acids Gly155 and Ala156 have been substituted in TM0065 with alanine and serine respectively, and the hydroxyl oxygens of Thr133 and Ser135 extend further into the ligand-binding pocket compared to their C-AllR counterparts. Leu129 is replaced in TM0065 with Tyr128, which also points its hydroxyl into the ligand-binding pocket. Ile99 also protrudes into the pocket, but Val128 is recessed compared to its C-AllR counterpart Leu149. Sequence analysis based on positions of AllR and TM0065 ligand binding cavity residues reveals several interesting preferences for these positions in the IclR family. Small or hydrophobic residues dominate the positions corresponding to AllR Met141, Leu149 and Leu215, while hydroxyl or amine amino acids occupy positions corresponding to Ser234 and Ser236. The amino acids corresponding to Asn118, Asp207 and His211 demonstrate the widest range of variation among IclR regulators.
Mutagenesis analysis of AllR glyoxylate binding pocket

The structural basis for ligand binding by AllR was investigated by mutagenesis of key binding residues and subsequent function analysis of the mutants both *in vivo* and *in vitro*. The qRT-PCR (Fig. 6A) and EMSA (Fig. 6B and C) studies were performed as described above.

Each of the conserved residues Met141, Leu215, Cys217, Ser234 and Ser236 were replaced individually by alanine (Table 1, Fig. 1). An *E. coli* strain expressing the wild type AllR demonstrated a seven and four fold increase in transcription of the *allA* and *gcl* genes, respectively in the presence of glyoxylate. All five mutants retained their ability to repress the transcription of these genes *in vivo*, showing that the proteins were both folded and functional. In contrast, there was no significant glyoxylate-triggered derepression in the presence of any of these mutants *in vivo* (Fig. 6A). DNA-binding assays confirmed that the alanine mutants bound the operator DNA even in the presence of 1 mM glyoxylate, whereas only 10 percent of the DNA remained bound to the wild type AllR under the same conditions (Fig. 6B). These five amino acids are clearly important for glyoxylate-binding and glyoxylate-mediated derepression.

A second series of mutagenesis studies was designed to test whether replacement of the residues involved in glyoxylate binding (Leu149, Asp207, His211, Cys217, Ser234 and Ser236) with the corresponding residues from *E. coli* YiaJ, YfaX and MhpR and *Acinetobacter* PobR regulators (Fig. 1) would affect binding. The amino acids were replaced by methionine, asparagine, aspartic acid, serine, asparagine and methionine (Fig. 1), respectively.

Similarly to the Cys217Ala mutation, the replacement of Cys217 by serine destroyed the ability of AllR to respond to glyoxylate *in vivo* (Fig. 6A). No dissociation of the Cys217Ser complex with operator DNA in the presence of glyoxylate could be detected in DNA-binding assays (Fig. 6B and C). These results underline the critical role of Cys217 in ligand binding.

Mutations of the other residues to their mates in other IclR family members were not as deleterious. In the case of the Leu149 mutation to methionine the glyoxylate derepression was
almost three times more efficient compared to the wild type AllR. This result was confirmed by the in vitro DNA binding (Fig. 6B). Modeling of methionine into the C-AllR - glyoxylate structure (data not shown) demonstrated an improved van der Waals interaction between methionine and glyoxylate, compared with that of leucine and glyoxylate. The distance between the glyoxylate C1 and methionine CE atoms was now 2.99 Å compared to 4.04 Å between C1 and leucine CD1.

The Asp207Asn mutation reduced the effects of glyoxylate on the derepression of the gcl gene two fold, while other mutants (His211Asp, Ser234Asn, Ser236Met) showed no response to glyoxylate in vivo (Fig. 6A). Interestingly, these last three mutants (Asp207Asn, His211Asp, Ser236Met) were able to respond to glyoxylate in assays of DNA-binding, releasing 67, 69 and 53 percent of the operator DNA in the presence of the ligand (Fig. 6B and C). The residual activity of the Asp207Asn mutant may arise from the fact that both aspartate and asparagine would be predicted to contribute to the formation of important hydrogen bonds.
Figure 6 Mutagenesis analysis of AllR ligand-binding site.

A. Relative expression levels of gcl gene in the absence (allR graph) and in the presence of wild type (WT) and mutant AllR proteins (M141A, L149M, L215A, D207N, H211D, C217A, S234A, S234N, S236A, S234M) measured by qRT-PCR. All the experiments were conducted in triplicates in the allR E. coli strain grown in absence (dark grey bar) or in the presence (light grey bar) of the glyoxylate (25 mM). B. Glyoxylate effect on the purified wild type and mutant AllR proteins binding to operator in the gcl promoter region tested by EMSA. The 5 nM 26 bp DNA fragment corresponding to the operator region and 0.1 μM of the purified AllR wild type or mutant proteins were used for each binding essay. 1 mM glyoxylate adjusted to the pH 7.5 was added when indicated. The full binding conditions are described in “Materials and Methods”. The name of the corresponding mutant protein is indicated above each image. C. EMSA results quantified by image analysis. Each graph represents the percentage (average of three experiments) of the bound DNA in the presence of the glyoxylate (1 mM) relative to the amount of the DNA bound to the same protein in the absence of the ligand.

Effect of glyoxylate binding on AllR conformation

The binding of glyoxylate reduces the affinity of AllR for its operator DNA sequence. In other prokaryotic inducer-regulator systems containing wHTH and inducer binding domains, the binding of the inducer in the signaling domain has been shown to trigger a conformational change that alters the DNA binding domains often through a change in the oligomeric state of the regulator protein. There was no gross conformational change in the C-terminal domain of AllR caused by glyoxylate binding; the structures are completely super-imposable. Since crystallographic studies may not have allowed us to observe ligand-induced changes in the AllR
oligomerization, we turned to analytical ultracentrifugation to investigate the effect of glyoxylate on AllR protein conformation in solution.

The sedimentation velocity (SV) studies were performed on the full-length AllR in the absence and presence of a 100-fold molar excess of glyoxylate (Fig. 7). The experiments were carried out at two different protein concentrations (25 and 50 μM) to assess whether AllR self-association was also concentration dependent. Fitting the data as a continuous distribution $c(s)^{42; 43}$ provided excellent sensitivity and resolution, enabling a clear distinction between different sedimenting species.

**Figure 7 Sedimentation velocity (SV).**

Continuous sedimentation distribution analysis of SV experiments acquired on 25 (A) and 50 (B) μM AllR protein, with and without 100-fold molar excess of glyoxylate ligand. Plots show overlays of AllR distribution in the presence (filled symbols) and absence (open symbols) of ligand. The residuals showed a good random spread of error in each case (data not shown).
At both tested protein concentrations, with and without inducer, three main AllR oligomeric species were identified (Fig. 7A). Their sedimentation coefficients (s) corresponded to expected values for monomeric, dimeric and tetrameric forms of AllR protein (29,607 Da per monomeric polypeptide chain as measured by mass spectrometry; data not shown). Not surprisingly, all peaks are less well resolved at the higher protein concentration, with or without glyoxylate present, due to higher molecular crowding and spontaneous formation of higher aggregates (Fig. 7B).

The addition of the ligand causes a dramatic increase (Fig. 7A and B) in the percentage of tetrameric AllR present in the sedimenting mixture and a corresponding decrease in the populations of monomer and dimer species, showing that binding to glyoxylate stabilizes and favors the tetrameric state. In addition, the presence of glyoxylate leads to sharper, more discrete peaks for all three oligomers, as well as small shifts towards higher s values for each. In the absence of ligand, calculated Svedberg constants (S) were: 1.3 S, 2.1 S and 3.3 S at 50 μM AllR concentration, and 1.3 S, 2.1 S and 3.3 S at 25 μM AllR concentration, for monomer, dimer and tetramer species, respectively. In the presence of ligand, these values increased slightly to: 1.4 S, 2.3 S and 3.6 S at 50 μM AllR concentration, and 1.4 S, 2.2 S, 3.5 S at 25 μM AllR concentration (1 S =10^{-13} s). These effects indicate that the conformations of each oligomer become more compact upon glyoxylate binding, and that this is most notable for the tetramer.

**DISCUSSION**

The AllR repressor belongs to the large but poorly investigated family of IclR bacterial transcriptional factors, which regulate transcription in response to a variety of metabolites. The structural scaffold of the family representatives was revealed in our earlier studies of the *Thermotoga maritima* IclR regulator - TM0065. Here, we investigate the molecular basis for ligand binding and its effects on regulator activity by defining the apo and ligand-bound structure of *E. coli* member of this family - AllR repressor.
Although the two proteins share relatively little sequence identity (26%), the structure of AlIR C-terminal domain is highly similar to that of TM0065. Three other IclR regulators (*E. coli* IclR, KdgR and YiaJ), whose C-terminal domains had been recently structurally characterized, (PDB codes 1TD5, 1YSP and 1YSQ, respectively) are equally related by sequence and share similar fold (Savchenko, unpublished data), suggesting that this general fold is adapted by large number of IclR family members.

The architecture of IclR ligand-binding domains resembles to that of versatile, small molecule-binding GAF/PAS domains found in all three kingdoms of life. The presence of cofactors such as 4-hydroxycinnamyl (PYP^{33}), heme (FixL^{44}) or flavin adenine dinucleotide (NifA^{45}) enables PAS domains to monitor changes in light, redox potential and oxygen level in the cell^{32, 46}. GAF domains are known primarily as 3', 5' cyclic guanosine monophosphate binding modules^{29, 47}, but they have also been reported to bind formate and 2-oxoglutarate^{48}.

Four main structural elements are recognized in the typical PAS domain: (i) the N-terminal cap or lariat, composed of two N-terminal α-helixes; (ii) the PAS core that usually includes first three β-strands of the central β-sheet as well as two subsequent α-helixes; (iii) the short helical connector represented by an α-helix connecting two periphery β-strands; and (iv) the β-scaffold, that include three β-strands that makes the second half of the central β-sheet (Fig. 8). In all known cases the cofactor is invariably located in the cavity (so-called “active site”) of the PAS core, thus making it most functionally relevant element of the PAS domain. The PAS core is also known to harbor residues involved in protein-protein interactions. The β-scaffold provides the structural support for the PAS core and completes the central β-sheet and contains a number of highly conserved residues, constituting a so called ‘PAC sequence motive’, while the N-terminal cap is the least conserved element of the PAS domain and can be replaced with other structural elements in different PAS domains^{49}. 
The architecture of both the PAS core and the β-scaffold of the PAS domain are clearly recognized in the C-terminal domains of AllIR and TM0065, with the location of glyoxylate contact corresponding to the PAS core “active site” (Fig. 8). The PAS N-terminal cap appeared to be replaced with helixes 5 and 9 in C-AllIR and TM0065. The C-AllIR α-helices (number 7 and 8) do not have an equivalent in the PAS domain fold. Interestingly, the representatives of GAF/PAS domains share no significant sequence similarity with any IclR regulator. The C-AllIR – glyoxylate complex structural analysis also demonstrates that the similarities between C-AllIR and PAS/GAF domains do not extend into the chemical nature and content of amino acids involved in the ligand binding. The equivalent amino acids involved in cofactor binding in the PAS/GAF domains are not found in the IclR regulators. While the strongest interaction of glyoxylate with C-AllIR is through the cysteine residue (Cys217), the position of this residue does not correspond to that of cysteine that forms a covalent bond with ligand in the PAS domain of PYP. Such comparative analysis demonstrates that IclR regulators form a new separate group in the diverse family of proteins containing PAS/GAF signaling domain and shows new functional capabilities of this domain as small-molecule-binding fold.

Based on C-AllIR and TM0065 structures analysis the minimal functional unit of this family of proteins appears to be a tetramer. The tetramerization of these proteins arises from the association of dimers. In TM0065 structure the stable dimer is formed by the hydrophobic interactions between α-helix 1 of the N-terminal DNA binding domains and α-helix 4 that links N- and C-terminal domains of each monomer. Tm0065 dimers then tetramerize through interactions between the ligand-binding domains, which involves the loop between the β4 and β4b strands. From C-AllIR structure analysis, it appears that AllIR use a similar oligomeric arrangement, except that the tetramerization interface contains an additional short β4a strand. The variation in the tetramerization interface may be important to minimize non-specific association among the various IclR paralogues (eight in *E. coli*) that are expressed in the same cell.
Superimposition of the AllR ligand-binding domain (PDB code 1T9L) with the structure of the photoactive yellow protein (PDB code 3PYP). AllR is depicted as a cyan worm, while the four segments of PYP are depicted with the following colors: the N-terminal cap, yellow; the PAS core, orange; the helical connector, violet; and the β-scaffold, green. Respective ligands (glyoxylate, 4-hydroxycinnamic acid) are shown in stick format. The amino and carboxyl termini are labeled, black for AllR, red for PYP. Figures 1, 2, 4 and 6 were prepared with a combination of SPOCK, Molscript, and Raster3D.

The functional relevance of tetramerization found both in TM0065 and AllR for their activity was defined by mutational analysis. The AllR dimers that were unable to tetramerize due to the mutation or alteration in β4a strand were functionally inactive in vivo. Thus the association of at least four wHTH domains was required for efficient binding of AllR to its operator region in the promoter regions of gcl and allA genes featuring the inverted repeat T₅G₆C₆G₅A₄A₃A₂A₁A₀T₁(A/T)₂T₃T₄C₅A₇A₈. Similarly most of the experimentally identified operator regions of the IclR family regulators contain palindromic or pseudo-palindromic sequences, underlining the symmetric nature of binding of these regulators to DNA. The founding member of the family – E. coli IclR – also is thought to bind its operator as a tetramer. Interestingly the
promoter regions under control of IclR family members such as PobR, PcaU and PcaR feature three fold repetition sequences\textsuperscript{34, 53}, this may imply a more complex oligomeric nature of functional units for these regulators. Further studies of AllR interactions with its operator will have a general impact on our understanding of IclR regulators interactions with DNA.

In \textit{in vitro} experiments, major part of the AllR-operator DNA complex was dissociated in the presence of 1 mM glyoxylate (Fig. 6B). Such significant AllR-operator DNA dissociation \textit{in vivo} would result in dramatic derepression of genes controlled by this regulator. However, only a 10-fold increase in the \textit{gel} gene transcription was detected during \textit{E. coli} strain growth in the presence of 25 mM glyoxylate (Fig. 4A). At the same time, the deletion of \textit{allR} gene prompted a 5000 fold increase in the transcription of this gene. This significant difference between \textit{in vivo} and \textit{in vitro} effects of glyoxylate on AllR function suggests that even in the presence of millimolar extracellular concentration of this compound, the intracellular concentration maybe too low for substantial inactivation of intracellular pool of AllR regulators. Although no experimental data are available on the levels of glyoxylate in \textit{E. coli} cells, a recent study\textsuperscript{54} on its close relative \textit{Salmonella typhimurium} determined that during growth in either 22 mM glucose or 135 mM acetate the intracellular concentration of glyoxylate was only 11 and 120 μM, respectively. As an important intermediate glyoxylate can be rapidly incorporated into the central metabolism by the combined action of malate synthase and the D-glycerate pathway\textsuperscript{3, 4}, or alternatively it can be reduced to glycolate by constitutive glyoxylate reductase activity\textsuperscript{4}. Low fluctuation of intracellular glyoxylate concentration would result in tight regulation of allantoin and glyoxylate degradation genes by AllR repressor.

Here and in our previous studies\textsuperscript{26} we have identified residue positions that are important for ligand binding. Extension of these positions to the rest of the IclR family members shows a wide variation of the amino acids involved in the ligand binding. While only five out of 12 AllR amino acids involved in glyoxylate binding are conserved in TM0065, none of these 12 amino acids involved in the interaction (Fig. 1 and 5B) including Cys217 is absolutely conserved across the
family. This feature indicates that dramatic changes in chemical environment of the ligand-binding pockets dictated by the variation among the ligands can be expected from structurally similar members of the IclR family. At the same time the IclR family members with conserved group of amino acids corresponding to the ones involved in the glyoxylate binding in AllR can be expected to bind chemically similar ligands.

According to the functional model, glyoxylate binding reduces the affinity of the regulator for its operator, thereby releasing the expression of the regulated genes. This effect must propagate from C-terminal ligand-binding domain to N-terminal DNA binding domain. Our analytical ultracentrifugation results suggest that signal is not propagated by destabilization of the tetramer required for DNA binding, suggesting that the signal may be directly propagated via the linker region. The TM0065 structure exhibits two alternate conformations of the linker region between the N- and C-terminal domains (residues 62-79), resulting in different relative orientations of the ligand-binding domains with respect to the wHTH dimer. This conformational flexibility may be important for signal propagation between the N- and C-terminal domains.

Another explanation of high conformational similarity between C-AllR and C-AllR-glyoxylate structures may be that the AllR apo- and glyoxylate-bound forms adapt partially overlapping ensembles of conformations. The crystal structures of each “freeze-out” the same, presumably highly sampled conformation that is shared by both forms. The analytical ultracentrifugation profile may reflect this ensemble. Glyoxylate binding also stabilizes AllR tetramer versus the two other oligomer states of this protein. The latter effect is probably due to the close proximity of the ligand-binding pocket and the AllR tetramerization interface and the increased number of inter-subunit contacts seen in the effector-bound form.

In summary the AllR studies presented in this paper reveal for the first time the details of the effector binding by IclR regulator. Nevertheless details of the promoter recognition by IclR regulators and the derepression mechanism upon the effector binding remain to be determined.
MATERIALS AND METHODS

DNA manipulations and cloning

Standard methods were used for site directed mutagenesis, chromosomal DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, ligation, and transformation\textsuperscript{55}. Plasmids were isolated using spin miniprep kits (Qiagen, USA), and PCR products were purified using Qiaquick purification kits (Qiagen, USA).

The \textit{allR} gene was amplified from wild type \textit{E. coli} BW25113 chromosomal DNA by PCR using the following primers (restriction sites \textit{Kpn}I and \textit{Hind}III are italicized and the Shine Dalgarno sequence added is underlined) all\textit{Re}fw:

\begin{verbatim}
GAGCTCGGTACCAGGAGGAAAACTATGACGGAAGTTAGACGGCGC
\end{verbatim}

and all\textit{Re}rv:

\begin{verbatim}
GCATGCAGCGCTTTATGGATGTGCTTTTCAGTCC
\end{verbatim}

The PCR products were purified, treated with \textit{Kpn}I and \textit{Hind}III, and then cloned into pBAD33 \textsuperscript{56}. For overexpression, full-length \textit{allR} and three C-terminal domain constructs (corresponding to Ala\textsubscript{91}, Asn\textsubscript{95} or Glu\textsubscript{97} - Pro\textsubscript{271}) were subcloned into the \textit{Nde}I and \textit{Bam}HI sites of a modified form of pET15b (EMD Biosciences, USA), in which a TEV protease\textsuperscript{1} cleavage site replaced the thrombin cleavage site and a double stop codon was introduced downstream from the \textit{Bam}HI site. This construct provides for an N-terminal hexahistidine tag separated from the protein by a TEV protease recognition site (ENLYFQ\textsubscript{↓}G). Site directed mutants were generated by QuikChange Site-Directed Mutagenesis Kit (Stratagene, USA) using primers listed in the Table 1.

Protein purification

The fusion proteins were overexpressed in \textit{E. coli} BL21-Star(DE3) (Stratagene, USA) harbouring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The cells were grown in LB at 37 °C to an OD\textsubscript{600} \textasciitilde 0.6 and expression induced with 0.4 mM IPTG. After
addition of IPTG, the cells were incubated with shaking at 15 °C overnight. The cells were harvested, resuspended in binding buffer (500 mM NaCl, 5% Glycerol, 20 mM Tris pH 9, 5 mM Imidazole), flash-frozen in liquid N\textsubscript{2} and stored at −70°C. The thawed cells were lysed by sonication after the addition of 0.5% NP-40 and 1 mM each of PMSF and benzamidine. The lysate was clarified by centrifugation (30 min @ 17000 rpm; Beckman Coulter Avanti J-25 centrifuge) and passed through a DE52 column pre-equilibrated in binding buffer. The flow-through fraction was then applied to a metal chelate affinity-column charged with Ni\textsuperscript{2+}. After the column was washed, the protein was eluted from the column in elution buffer (binding buffer with 500 mM Imidazole). The hexa-histidine tag was then cleaved from the protein by treatment with recombinant His-tagged TEV protease. The cleaved protein was then resolved from the cleaved His-tag and the His-tagged protease by flowing the mixture through a second Ni\textsuperscript{2+}-column.

The purified proteins were dialyzed against 10 mM Tris pH 9.0, 500 mM NaCl, and concentrated using a BioMax concentrator (Millipore, USA). Before crystallization, any particulate matter was removed from the sample by passage through a 0.2 μm Ultrafree-MC centrifugal filtration device (Millipore, USA).

Selenomethionine-labeled proteins were expressed using the same vector and host strain but in supplemented M9 media\textsuperscript{57}. The sample was prepared under the same conditions as the native protein except for the addition of 5 mM β-mercaptoethanol to the purification buffers.

Crystallization and data collection

The primary crystallization condition was determined by a sparse crystallization matrix (Hampton Research kits: Crystal Screen 1 and PEG/Ion Screen), at room temperature using the sitting drop vapor diffusion technique in 96-well plates. This condition was modified by varying the pH and the concentration of the solutes. The best condition was obtained using hanging drops (2 μl protein: 2 μl precipitant ratio) in crystallization condition: 0.1M Hepes pH 6.8, 1.9 M
Ammonium Sulfate, 4% MPD in 2-5 days at room temperature. For diffraction studies, the crystals were flash-frozen with the crystallization buffer plus 40% MPD as the cryoprotectant. For the complex structure, 1 mM glyoxylate was added to the protein solution prior to crystallization.

Diffraction data of crystals of C-AllR complexed with glyoxylate were collected at 100°K at the 19BM beam line of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The absorption peak and the rising inflection point were determined by calculating and plotting $f'$ and $f''$ values against energy from the fluorescence spectrum. The three-wavelength MAD data were collected from a Se-Met substituted protein crystal using an inverse-beam strategy. All crystallographic data were measured with the custom-built 3x3 tiled CCD (charge-coupled device) detector with a 210°—210 mm$^2$ active area and fast duty cycle (~1.7 sec). The experiment, data collection, and visualization were controlled with d*TREK$^{60}$ and all data were integrated and scaled with the program package HKL2000$^{61}$. Data collection and processing statistics are provided in Table 2.

Data corresponding to the peak wavelength was processed with HKL2000$^{61}$, and then input into the program SOLVE$^{62}$. SOLVE located 23 of the possible 32 selenium sites (including two in the N-terminal histidine tag), and gave a mean figure of merit of 0.39 (0.28 in the highest resolution bin). Density modification and automatic model building using RESOLVE resulted in excellent maps and partial models for each of the four subunits. The models were manually completed with the aid of the four-fold non-crystallographic symmetry and the graphics program O$^{65}$. Further sessions with O$^{65}$ and refinement of atomic positions and individual b-factors using the program CNS-1.1$^{66}$ resulted in a R-factor of 19.2% (R-free 23.2) for data from 33.7 to 1.7 Å. The final model comprises 5981 protein atoms (four C-AllR ligand binding domains), 20 ligand atoms (four glyoxylate molecules) and 504 water molecules. The model has excellent stereochemistry as judged by PROCHECK$^{67}$, with no Ramachandran violations. All residues of the ligand-binding domain, with the exception of the three C-terminal amino acids (two in the
case of subunit A), along with several residues of the N-terminal tag were located in the experiment.

To determine the ligand free structure of the AllR ligand-binding domain, a data set was collected on crystals formed in the absence of glyoxylate using an R-AxisIV++ detector and a MicroMax-007 generator (Rigaku MSC, USA). Improvement of the input model (1T9L, without glyoxylate or waters), was carried out by rigid-body refinement, energy minimization and b-factor refinement using CNS-1.1 \cite{CNS}, along with manual rebuilding in O \cite{O} (aided with omit maps). The final model comprises 5461 protein atoms and 595 water molecules, and has an R-factor of 21.4% (R-free 25.7%) for data from 19.7 to 1.8 Å. All residues of the ligand binding domain, with the exception of the three C-terminal amino acids (two in the case of subunit A), along with several residues of the N-terminal tag were located in the experiment, and the model has excellent stereochemistry as judged by PROCHECK \cite{PCHECK}, with no Ramachandran violations.

**Construction of allR deletion mutant**

All the *in vivo* studies were conducted in the genetic background of strain BW25113 \cite{strain}. Deletion mutants were generated by the methods described by Datsenko and Wanner (2000). To prepare competent cells for transformation, BW25113 containing pKD46 was cultured at 30°C in SOB broth \cite{SOB} containing 100 μg of ampicillin per ml. When the optical density at 600 nm (OD600) reached 0.5, the culture was centrifuged at 4000 rpm for 5 min, and the cells were washed three times with cold water before being resuspended in a minimal volume of water (1% of the original culture volume). The kanamycin resistance gene (*km*) was amplified by PCR from pKD4 by using the primers allRFw:

\[
\text{GCACAGGCGTTAGAGCGGGGAATTGCGATTCTGCAATATTTGGGTGTAGGCTGGAGCTGCTTC}
\]

and allRrv:
The PCR products were purified with a Qiagen kit, treated with DpnI, and repurified by electrophoresis. The \( km \) gene was transformed into BW25113-competent cells by electroporation (Gene Pulser; pulse controller at 200 \( \Omega \), capacitance at 250 \( \mu \)F, and voltage at 25 kV). After electroporation, the cells were grown with shaking in 1 ml of SOC medium at 37°C for 1 h, and the cultures were plated onto Luria-Bertani (LB) agar containing 25 \( \mu \)g of kanamycin per ml. The \( \text{Km}^+ \) transformants were purified on new kanamycin-LB plates. The mutants in which the target genes were replaced by the \( km \) gene were verified by PCR using the primers allRcfw: TTGCGATTCTGCAATATTTGG and allRcrv: CGATCTTCTGTCAGTCTTGAT. To delete the \( km \) gene from the chromosome, pKD46 was removed from the cells by growing the bacteria at 37°C, and then pCP20, expressing the FLP recombinase, was introduced by transformation. The transformants containing pCP20 were grown overnight with shaking at 42°C, and the cultures were plated on LB agar without antibiotics. Colonies were tested for sensitivity to kanamycin and ampicillin.

**qRT-PCR studies**

Bacterial cells were cultured in MOPS minimal media with 0.2% xylose as carbon source. When the OD600 reached 0.5, cells were collected by centrifugation at 4°C. Total RNA was subsequently isolated with the RNeasy Mini Kit (Qiagen, USA) in accordance with the manufacturer's protocol. Residual DNA present in the RNA preparations was removed by RNase-free DNase (Fermentas, Lithuania). cDNAs were synthesized with the superscript first-strand synthesis kit (Invitrogen, USA) in accordance with the manufacturer's instructions and stored at –20°C prior to use. Real-time quantitative PCR (qRT-PCR) was carried out on the Applied Biosystems 7300 apparatus (Applied Biosystems, USA) using Platinum® SYBR® Green qPCR SuperMix UDG (Invitrogen, USA) in accordance with the manufacturer's recommended protocol.
Primers used for the RT-PCR were as follows: for \textit{allA}, GTGGAGCGTTACCACGATTT and GGTCTGGTTTGTCGTCACCT; for \textit{gcl}, GCAAAATGCGGTTACAGTT and TCGGCTGGATTAACATTTC; for \textit{rrsC}, CAGCCACACTGGAACTGAGA and GTTAGCCGGTGTCTTTCTCTG. The relative expression values were normalized using the number of cycles obtained for the house-keeping gene \textit{rrsC} and then expressed in relation to the \textit{allR} strain under repressing conditions.

**Electrophoretic Mobility Shift Assays (EMSA)**

Electrophoretic Mobility Shift Assays for AllR and mutant proteins were performed in triplicates using proteins purified and concentrated according to the procedures described above. The oligonucleotides corresponding to AllR recognition region in \textit{gcl}-promoter region (5'-AAAGTTGGAAAAATTTTCCAATAAAT-3') were labeled and annealed by Biotin 3'End DNA labeling kit (Pierce, USA) according to the manufacturers instructions. Binding assays were performed at 37°C for 20 min. in the binding buffer (10 mM Tris-HCl, 50 mM KCl, 1mM DTT, pH7.5) supplied by LightShift Chemiluminescent EMSA Kit (Pierce, USA) in the presence of 50 ng/μl Poly (dl-dC) nonspecific competitor DNA. The binding reaction also contained 0.1 μM of purified protein, 5 mM MgCl₂, 5 nM Biotin-labeled oligonucleotide.

EMSA were performed in a BioRad Protein II apparatus using 6% polyacrylamide/0.5x TBE gels. 20 μl of the above binding reaction were loaded per lane. Electrophoresis was performed at 100V using ice-cold 0.5x TBE as a running buffer (5x TBE: 450 mM Tris, 450 mM Boric acid, 10 mM EDTA, pH 8.3). Biotin-labeled oligonucleotides and oligonucleotide-protein complexes were transferred from the polyacrylamide gel to the Biodyne B Positive Nylon Membrane (Pierce, USA) in 0.5xTBE immediately after electrophoresis by electroblotting at 380 milliampers for 40 minutes. Transfered DNA was crosslinked for 15 minutes using UV cross-linker equipped with 312 nm bulbs. The oligonucleotide-protein complexes were detected by horseradish
peroxidase/Super Signal Detection System. Membranes were exposed to Kodak X-ray film for 1 minute and quantified by image analysis.

Circular dichroism (CD) spectroscopy

CD spectra were collected at 25°C on a AVIV 62D CD Spectrophotometer from 200 to 260 nm using a 0.1 mm path length cell, with a scan rate of 100 nm/min, time constant 1.0 s, bandwidth of 1 nm, and sensitivity of 100 millidegrees. Each spectrum was averaged from 10 scans. After buffer subtraction, the spectra were calibrated in units of molar ellipticity.

Size-exclusion chromatography

FPLC size-exclusion chromatography was performed on a Superdex-200 10/30 column (GE Biosciences) pre-equilibrated with 10 mM Tris pH 9, 0.5 M NaCl, 0.5 mM TCEP. The column was calibrated with cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa), β-amylase (200 kDa). A 20-μl of protein sample at a 5 mg ml⁻¹ concentration or premixed with standard proteins was centrifuged at 14,000 rpm for 10 min before being injected into the column through a 20 μL injection loop. Filtration was carried out at 4°C at a flow rate of 0.5 ml min⁻¹. The eluted proteins were detected by measuring the absorbance at 280 nm.

Analytical ultracentrifugation

Sedimentation velocity (SV) experiments were performed with a Beckman XL-I analytical ultracentrifuge (Beckman-Coulter UK Ltd, High Wycombe, UK) using an 8-hole An-50 Ti rotor. Cells were equipped with 12 mm double sector, charcoal-filled Epon centerpieces and quartz windows. Measurements were carried out using the absorption optics of the instrument at suitable wavelengths for detection of the concentration gradient. Full-length AllR protein was extensively dialyzed against 20 mM Tris-HCl buffer (pH 9.0)
containing 0.5 M NaCl, 0.5 mM TCEP and 5% glycerol. SV experiments were recorded at two concentrations of AllR (25 and 50 μM) in the presence and absence of 100-fold molar excess of glyoxylate (Sigma). An aliquot of the sample buffer was used as a reference for all measurements. The partial specific volume (\(\nu\)) of AllR was calculated from the protein sequence using the program Sednterp v.1.08 by David Hayes and John Philo. The density and viscosity of the buffer were calculated from the buffer composition also using this software.

SV experiments were performed at 10° C and 50,000 rpm with 400 µl of protein solution in each cell. Scans were acquired every 180s at a single wavelength (280 nm) using a radial step size of 0.003 cm. Sedimentation coefficients were computed using the program Sedfit v.8.9g, which combines finite element solutions of the Lamm equation for a large number of discrete species, with maximum entropy regularization, to represent a continuous size-distribution. This method of processing the SV data yielded the highest resolution and sensitivity.

Protein Data Bank accession number

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 1TF1 and 1T9L), Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (www.rcsb.org).

REFERENCES


**ACKNOWLEDGEMENTS**

We thank all members of the SBC at ANL for their help in conducting experiments and Aled Edwards for critical reading of the manuscript. This work was supported by National Institutes of Health Grant GM62414-01, by the Ontario Research and Development Challenge Fund and by a grant from the Canadian Institutes of Health Research Grant.

GLYOXYLATE AND PYRUVATE ARE ANTAGONISTIC EFFECTORS OF THE ESCHERICHIA COLI ICLR TRANSCRIPTIONAL REGULATOR.


My Contribution to This Study:

I over-expressed, grew and purified the IclR protein and its C-terminal domain (C-IclR) for the structural and functional studies. I screened C-IclR against a library of small molecule compounds using a novel StarGazer method. I identified successful ligand candidates and tested their interaction with IclR by Isothermal Titration Calorimetry (ITC). I also produced C-IclR crystals for X-ray crystallography studies of C-IclR complexes with the newly characterized effectors.
GLYOXYLATE AND PYRUVATE ARE ANTAGONISTIC EFFECTORS OF THE ESCHERICHIA COLI ICLR TRANSCRIPTIONAL REGULATOR

Graciela L. Lorca1,†, Alexandra Ezersky1,2, Vladimir V. Lunin3, John R. Walker3,4, Svetlana Altamentova5, Elena Evdokimova6, Masoud Vedadi5, Alexey Bochkarev1, 4, and Alexei Savchenko1,3

From 1Banting and Best Department of Medical Research, 112 College St., Toronto, Ontario, Canada 2Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, 3Ontario Center for Structural Proteomics, Best Building, 112 College St., Toronto, Ontario, 4Canada Structural Genomics Consortium, Banting Building, 100 College St., Toronto, Ontario, Canada.

ABSTRACT

The Escherichia coli Isocitrate Lyase Regulator (IclR) regulates the expression of the glyoxylate bypass operon (aceBAK). Founding member of a large family of common fold transcriptional regulators - IclR comprises a DNA-binding domain that interacts with the operator sequence and a C-terminal domain (C-IclR) that binds a hitherto unknown small molecule. We screened a chemical library of over 150 metabolic scaffolds using a high-throughput protein stability assay to identify molecules that bind IclR, and then tested the active compounds in in vitro assays of operator binding. Glyoxylate and pyruvate, identified by this method bound C-IclR domain with KDs of 0.9 ± 0.2 μM and 156.2 ± 7.9 μM, as defined by isothermal titration calorimetry. Both compounds altered IclR interactions with operator DNA in EMSA assays but showed an antagonistic effect. Glyoxylate disrupted the formation of the IclR/operator complex in vitro by favoring the inactive dimeric state of the protein while pyruvate increased the binding of IclR to the aceBAK promoter by stabilizing the active tetrameric form of the protein. Structures of the C-IclR domain alone and in complex with each effector were determined at 2.3 Å, confirming the binding of both molecules in the effector recognition site previously
characterized for the other representative of the family, the E. coli AllR regulator. Site directed mutagenesis demonstrated the importance of hydrophobic patch formed by Met146, Leu154, Leu220, and Leu143 in interactions with effector molecules. In general, our strategy of combining chemical screens with functional assays and structural studies has uncovered two small molecules with antagonistic effects that regulate the IclR-dependent transcription of the aceBAK operon.

INTRODUCTION

A classical bacterial repressor, termed IclR, participates in the regulation of the aceBAK operon in E. coli (1, 2). The aceBAK operon encodes for the enzymes of the glyoxylate bypass (3) that are required during growth on acetate since it bypasses the two CO2-evolving steps of the Krebs cycle (4). The expression of these enzymes is induced during growth on minimal medium supplemented with acetate or fatty acids (3) as well as in rich medium as result of the acetate accumulation during exponential phase. According to early genetic studies (1, 2) IclR represses the expression of aceBAK operon as well as its own gene by binding to specific operator sequences in the promoter region. Two separate operator sequences, termed IclR-boxes, have been identified in the promoter region upstream of the aceB gene (5). The primary operator sequence (IclR box II) has been mapped between –52 and –19 bases of the aceB promoter, and binding of IclR to this site prevents the RNA polymerase-promoter interaction (5). Binding to the second site, IclR box I, which is located between –125 and –99 bases of the aceB promoter, disassembles the open complex through a protein-protein interaction between IclR and the RNA polymerase α-subunit (5). A third IclR-box was located between +14 and –21 bases of the iclR promoter (6). All three IclR-boxes identified in the promoters of aceB and iclR genes contain near-perfect palindromes. Although not identical, their sequences share a 15-mer consensus motif – 5’TGGAAATNATTTCCA, which was identified by in vitro random oligonucleotide selection experiments (7).
In growth media containing acetate IclR is believed to release the DNA as a result of binding a small molecule. Although this molecule is likely related to a metabolite within the glyoxylate bypass pathway, the precise nature of this IclR effector remains nevertheless in doubt. Phosphoenolpyruvate was reported to alter IclR binding to DNA in vitro (8), but recently was shown to have no effect on IclR activity by Yamamoto and Ishihama (5).

IclR is the founding member of a large family (Pfam 01614, COG1414) of sequence-related microbial transcriptional regulators with over 500 members identified in bacterial and archaeal genomes. Numerous bacterial genomes such as E. coli K12 (9) and Pseudomonas aeruginosa PA01 (10) contain multiple members of this family (8 and 9, respectively) with the largest number (40 members) identified so far in Bordetella bronchiseptica RB50 (11). The characterized members of this family are involved in the regulation of diverse catabolic pathways ranging from the degradation of plant cell polysaccharides in the plant pathogen Erwinia chrysanthemi (12) to the metabolism of aromatic acids in E. coli, Acinetobacter (13, 14) and Pseudomonas (15).

The characteristic structural features for members of this family, gained primarily from the crystal structure of TM0065 from Thermotoga maritime (PDB # code 1MKM), are fully consistent with the role of transcription regulators responding to chemical stimuli (16). The TM0065 structure confirmed the presence of N-terminal classic winged helix-turn-helix (wHTH) DNA-binding domain as well as demonstrated that the effector binding C-terminal domain has significant structural homology to the PAS/GAF domain, a known small molecule-binding motif (17, 18). The N- and C- terminal domains in the structure are connected by a linker helix, which together with N-terminal domain participates in protein dimerisation. In crystal lattice TM0065 dimers further oligomerise into tetramers, which appear to be the functional unit common for a number of IclR regulators (16). The C-terminal domain retains its structure and small molecule binding function when expressed without the DNA-binding domain and was structurally characterized for three E. coli representatives of IclR family – KdgR (PDB #1YSP), YiaJ (PDB #1YSQ) and AllR (20, PDB #1TF1). While sharing relatively low sequence identity between them and with
previously characterized TM0065, all four structures were very similar, confirming a common structural framework for most of the IclR family members.

AllR is involved in transcriptional regulation of the allantoin catabolism (19) and shares 42% sequence identity with IclR. The AllR C-terminal domain structure was recently determined in the presence and absence of its chemical effector glyoxylate (20), providing the first detailed view of the molecular interactions of an IclR family representative with a small molecule. According to the AllR-glyoxylate complex structure, the location of the effector binding pocket in IclR regulators corresponded to the co-factor accommodating cavity in PAS/GAF domains, however, the binding mechanisms were not conserved between these structurally similar domains (17, 18, 20).

Of the hundreds of IclR family members, only a few of the sensor chemicals have been identified, which apart from glyoxylate include 3-(3-hydroxyphenyl) propionic acid for E. coli MhpR (21); 2-keto-2-deoxyglyconate for Erwinia chrysanthemi KdgR (12); p-hydroxybenzoate and protocatechuate for Acinetobacter PobR and PcaU, respectively (13, 22), N-3-oxo-octanoyl homoserine lactone (3OC8HSL) for Agrobacterium tumefaciens AttJ (23); and 1-naphthol and 2,3-dihydroxynaphthalene for Pseudomonas putida Tt1gV (24). Thus, the effector specificity of hundreds IclR regulators, including the founding member of the family, remains uncertain.

To address these issues we used IclR as a test case and screened it against a large number of intermediates associated with glyoxylate bypass including phosphoenolpyruvate as part of a library of more than 150 metabolites by a high-throughput protein stability assay (25, 26). This methodology takes advantage of the well-documented phenomenon of enhanced thermal stability of the protein in the presence of a specifically bound ligand (27). The molecules that had a stabilizing effect were further assessed for the effector role for IclR using biophysical and functional assays.
Experimental Procedures

DNA manipulations and gene cloning

Standard methods were used for site directed mutagenesis, chromosomal DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, ligation, and transformation (28). Plasmids were isolated using spin miniprep kits (Qiagen, USA), and PCR products were purified using Qiaquick purification kits (Qiagen, USA). Mutagenesis was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, USA).

For protein expression and purification the iclR gene was amplified from wild-type E. coli BW25113 chromosomal DNA by PCR. To identify the boundaries of the ligand-binding domain of IclR, a multiple sequence alignment of IclR and its homologues was performed. The link between the DNA and ligand-binding domains in E. coli IclR was predicted to lie between amino acids Ser98 and Leu101. Two fragments corresponding to the 176 and 173 C-terminal amino acids were amplified and tested for expression of the soluble domain as previously described (20). The expression of a fragment of iclR corresponding to amino acids 98 to 274 yielded a soluble domain of IclR (C-IclR) that was used for further experiments. All amplified fragments were cloned into the NdeI and BamHI sites of a modified form of pET15b (EMD Biosciences, USA), in which a TEV protease cleavage site replaced the thrombin cleavage site and a double stop codon was introduced downstream from the BamHI site. This construct provides for an N-terminal hexahistidine tag separated from the protein by a TEV protease recognition site (ENLYFQ↓GS).

Construction of deletion mutants

An IclR deletion mutant was generated by the methods described by Datsenko and Wanner (29). To prepare competent cells for transformation, BW25113 (lacIq rrmBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78; 29) containing pKD46 was cultured at 30°C in SOB broth (28) containing 100 μg of ampicillin per ml. When the OD600 reached 0.5, the culture was
centrifuged at 4000 rpm for 5 min., and the cells were washed three times with cold water before being resuspended in 1% of the original culture volume of water. PCR methods were used to amplify the kanamycin resistance gene (km) from pKD4 by using the primers iclRFw: 5’-GTTCAACATTAACTCATCGGATCGTTCAGTAACTATTGCATTAGCTAAGCGTTGAGCTGCTTC-3’ and iclRrv: 5’-GCGATTAACAGACACCCTTATTCTATTGCCAUCTAGGTATGATCATATGATCATATGATATATCCTCCTTA-3’. The PCR products were purified with a Qiagen kit, treated with DpnI, and repurified by electrophoresis. The km gene was transformed into BW25113-competent cells by electroporation (Gene Pulser; pulse controller at 200 Ω, capacitance at 250 μF, and voltage at 25 kV). After electroporation, the cells were grown with shaking in 1 ml of SOC medium at 37°C for 1 h, and the cultures were plated onto Luria-Bertani (LB) agar containing 25 μg of kanamycin per ml. The Kmr transformants were purified on new kanamycin-LB plates. The mutants in which the target genes were replaced by the km gene were verified by PCR using the primers iclRcfw: 5’-CTCATCGGATCAGTTCAGTAACTATTGCATTAGCTAAGCGTTGAGCTGCTTC-3’ and iclRcrv: 5’-CTTATTCTATTGCCAUCTAGGTATGATCATATGATATATCCTCCTTA-3’.

To delete the km gene from the chromosome, pKD46 was removed from the cells by growing the bacteria at 37°C, and then pCP20, expressing the FLP recombinase, was introduced by transformation. The transformants containing pCP20 were grown overnight with shaking at 42°C, and the culture were plated on LB agar without antibiotics. Colonies were tested for sensitivity to kanamycin and ampicillin. Crp mutants were constructed by P1 transduction (30) using the strain JW3320 as donor (National BioResource Project, NIG, Japan, 31).

qRT-PCR studies

Bacterial cells were cultured in MOPS minimal media (32) with 10 mM pyruvate as carbon source and, when required, 10 mM glyoxylate was added. When the OD600 reached 0.5, cells were collected by centrifugation at 4°C. Total RNA was subsequently isolated with RiboPureTM-Bacteria (Ambion) in accordance with the manufacturer's protocol. cDNAs were synthesized with
the superscript first-strand synthesis kit (Invitrogen) in accordance with the manufacturer's instructions and stored at –80°C prior to use. Real-time quantitative PCR (qRT-PCR) was carried out on the Applied Biosystems 7300 apparatus (Applied Biosystems) using Platinum® SYBR® Green qPCR SuperMix UDG (Invitrogen) in accordance with the manufacturer's recommended protocol. Primers used for the RT-PCR were as follows: for rrsC, 5’-CAGCCACACTGGAACTGAGA-3’ and 5’-GTTAGCCGGTGCTTCTTCTG-3’; for aceB, 5’-CTGCGTGACCATTATTTGG-3’ and 5’-CAGGCGTGAGTAAGCATTCA-3’; for aceK, 5’-TTGTGCGCTGCTATCAACTG-3’ and 5’-AGATATTGAGCGGACCACC-3’.

**Protein purification**

The His-tagged fusion proteins were overexpressed in E. coli BL21-DE3 cells (Stratagene, USA) harbouring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The cells were grown in LB at 37°C to an OD600 ~0.6 and expression induced with 0.4 mM IPTG. After addition of IPTG, the cells were incubated with shaking at 15°C overnight. The cells were harvested, resuspended in binding buffer (500 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 5 mM Imidazole), flash-frozen in liquid N2 and stored at –70°C. The thawed cells were lysed by sonication after the addition of 0.5% NP-40 and 1 mM each of PMSF and benzamidine. The lysate was clarified by centrifugation (30 min at 17000 rpm) and passed through a DE52 column pre-equilibrated in binding buffer. The flow-through fraction was then applied to a metal chelate affinity-column charged with Ni2+. After the column was washed, the protein was eluted from the column in elution buffer (binding buffer with 500 mM Imidazole). The hexa-histidine tag was then cleaved from the protein by treatment with recombinant His-tagged TEV protease. The cleaved protein was then resolved from the cleaved His-tag and the His-tagged protease by flowing the mixture through a second Ni2+-column.

The purified proteins were dialyzed against 10 mM HEPES, pH 7.5, 500 mM NaCl, and concentrated using a BioMax concentrator (Millipore, USA). Before crystallization, any
particulate matter was removed from the sample by passage through a 0.2 μm Ultrafree-MC centrifugal filtration device (Millipore, USA). Selenomethionine-labeled proteins were expressed using the same vector and host strain but in supplemented M9 media (33). The sample was prepared under the same conditions as the native protein except for the addition of 5 mM β-mercaptoethanol to the purification buffers.

Size-exclusion chromatography

100 μl protein sample contained 10 mM HEPES, pH 7.5, 500 mM NaCl, 25 μM IclR wild type or its mutant derivatives and, when indicated, 1 mM glyoxylate or 10 mM pyruvate. After 20 min incubation on ice, samples were injected onto a Superose 12 10/300 GL gel filtration column (Amersham Biosciences) installed on an Äkta system (Amersham Biosciences) equilibrated with 10 mM HEPES pH 7.5, 500 mM NaCl. Filtration was performed at 4°C at a flow rate of 0.5 ml/min and the protein concentration was monitored by measuring the absorbance at 280 nm. Blue dextran 2000 was used to determine the void volume. A mixture of protein molecular mass standards, containing β-amylase (200 kDa), BSA (66 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa), was applied to the column under similar conditions. The elution volumes and molecular masses of the protein standards were used to generate a standard curve from which the apparent molecular mass was determined.

Crystallization and Data Collection

The primary crystallization condition was determined by a sparse crystallization matrix (Hampton Research kits: Crystal Screen 1 and PEG/Ion Screen), at room temperature using the sitting drop vapor diffusion technique in 96-well plates. This condition was optimized by varying the pH and the concentration of the solutes. For apo-IclR the best crystals were obtained using hanging drops in 0.2M Potassium Acetate pH 7.0, 20% PEG 3350 for 2-5 days at room temperature. For diffraction studies, the crystals were flash-frozen with the crystallization buffer plus 30% ethylene glycol as cryoprotectant. For the complex structures the crystals obtained in the
crystallization condition described above were soaked in 10 mM solution of either glyoxylate or pyruvate for 1 h. For diffraction studies, the crystals were flash-frozen with the crystallization buffer plus 25% ethylene glycol as the cryoprotectant.

Diffraction data of crystals of C-IclR were collected at 100K at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The absorption peak and rising inflection point were determined by calculating and plotting f' and f” values against energy (34) from the fluorescence spectrum. The three-wavelength MAD data were collected from a Se-Met substituted protein crystal using an inverse beam strategy. All crystallographic data were measured with the custom-built 3x3 tiled CCD (charge-coupled device) detector (35). The experiment, data collection, visualization and processing were controlled with the HKL2000 program package (36). Statistics of data collection and processing are provided in Table 3. Diffraction data for glyoxylate and pyruvate complexes were collected at home source (Rigaku FR-E generator with Raxis4++ detector and the data were integrated and scaled with *TREK (37).

Structure solution and refinement

Data from each of the three wavelengths was processed with the HKL2000 (36) and then input into the program SOLVE (38). SOLVE located 24 of a possible 28 selenium sites, and gave a mean figure of merit of 0.66 (0.51 in the highest resolution bin). Density modification and automatic model building using RESOLVE (39, 40) resulted in excellent maps and partial models for each of the four subunits. The models were completed with the aid of four-fold non-crystallographic symmetry and the graphics program O (41). Refinement of atomic positions and individual b-factors using the program CNS-1.1 resulted in an Rfactor of 23.0% (Rfree 29.9%) for data from 19.91 to 2.30 Å. The final model comprises 5499 protein atoms (four C-IclR domains) and 335 water molecules. The model has excellent stereochemistry as judged by PROCHECK (42), with no Ramachandran violations. All residues of the ligand binding domain except for three C-terminal amino acids (two in the case of subunit D) were located in the
experiment. Both glyoxylate and pyruvate complexes structures were solved by molecular replacement using apo-structure as starting model. Structures were refined using REFMAC5 (43, 44).

**Electrophoretic Mobility Shift Assays (EMSA)**

Electrophoretic Mobility Shift Assays for IclR and mutant proteins were performed using proteins purified and concentrated according the procedures described above. A fragment of the aceBAK promoter region, from position –117 to +17 (containing IclR box II) was generated by PCR using biotin prelabeled (5’-end) primers, and purified using QIAquick spin columns (Qiagen). Incubation mixtures for EMSA (20 μl) contained 2.5 nM of a 5’-labelled DNA fragment, 50 mM Tris-HCl pH 7.5, 150 mM KCl, 10 mM MgCl, 0.01% Triton X100, 50 ng/μl Poly(dI-dC) nonspecific competitor DNA, purified IclR protein (0-100 nM) and ligand (0-1 mM) where indicated.

After incubation for 20 min at 37°C, samples were separated on 5% acrylamide-bisacrylamide nondenaturing gels in 0.5x Tris borate-EDTA buffer, pH 8.3 (TBE). Electrophoresis was performed at 100V using ice-cold 0.5x TBE as a running buffer and DNA was transferred from the polyacrylamide gel to the Biodyne B Positive Nylon Membrane (Pierce,USA) by electroblotting at 380 mAmps for 40 minutes in 0.5x TBE. Transferred DNA was cross-linked for 15 minutes using UV cross-linker equipped with 312 nm bulbs. Biotin labeled DNA was detected by horseradish peroxidase/Super Signal Detection System (Pierce, USA). Membranes were exposed to Kodak X-ray film. For quantitative EMSA, the distribution of bound and unbound DNA was quantified using ImageJ (NIH, USA).

**In vitro transcription runoff assays**

For In vitro transcription runoff the complete aceB promoter was amplified from position –297 to +434 and cloned in a derivative of the pGFPuv vector (Clontech) in which the gfp gene was replaced with the aceB promotor in the SapI-EcoRI position. The expected specific transcript is
434 bases-long. The recombinant plasmids were amplified in E. coli DH5α, cut with EcoRI and gel purified. The reaction mixtures (20 μl) contained: template DNA (5 nM), purified IclR or its mutant forms (100 nM) and glyoxylate (1 to 1000 μM), in a buffer consisting of 40 mM Tris-HCl (pH 8.0), 50 mM KCl, 10 mM MgCl2, 0.01% triton X-100, and 1U Rnase inhibitor. The reaction mix was preincubated for 10 min at 37°C and 20 nM RNAP (Epicentre) was added. The reaction mix was further incubated for 5 min at 37°C and transcription was initiated by adding 2 μl of a cocktail containing 2 mM each ATP, GTP, CTP, 1.5 mM UTP and 0.5 mM biotin-11-UTP (Perkin Elmer). After 20 min at 37°C, the reactions were terminated with 10 mM EDTA. The ethanol-precipitated transcripts were analysed on 6% acrylamide-7 M urea gels. Electrophoresis was performed at 100V using ice-cold 0.5x TBE as a running buffer and RNA was transferred from the polyacrylamide gel to the Biodyne B Positive Nylon Membrane (Pierce, USA) by electroblotting at 380 mAmps for 40 minutes in 0.5x TBE. Transferred RNA was cross-linked for 15 minutes using UV cross-linker equipped with 312 nm bulbs. Detection was performed as described above. Pre-heated Biotinylated 2-log DNA ladder (BioLabs) was used as molecular weight marker.

**Screening using Static light scattering**

Purified C-IclR protein was screened against a library of 160 compounds using static light scattering (StarGazer) (45). Protein samples were diluted to a final concentration of 0.4 mg/ml in 100 mM Hepes pH 7, 150 mM NaCl (26). 50 μl aliquots of protein solution containing the chemical compounds at 1 mM were placed in duplicate into clear bottom 384 well plates (NuncTM, Nalgene Nunc International) and heated from 27°C to 80°C at the rate of 1°C per minute. The formation of protein aggregates was monitored by static light scattering. Images of scattered light were captured every 30 seconds and the light intensities were translated to arbitrary numbers using StarGazer™ proprietary software. Intensities were plotted against temperature for each sample well and transition curves were fitted using the Boltzmann equation. The midpoint of
each transition was calculated and compared with the one calculated for the reference sample. If the difference between them was greater than 1.5°C the corresponding compound was considered to be a “hit” and the experiment was repeated to confirm the effect.

**Circular dichroism (CD) spectroscopy**

CD spectra were collected at 25°C on an AVIV 62D CD Spectrophotometer from 200 to 260 nm using a 0.1-mm path length cell, with a scan rate of 100 nm/min, a time constant 1.0 s, a bandwidth of 1 nm, and a sensitivity of 100 millidegrees. Each spectrum was the average of 10 scans. After background subtraction, the spectrum was expressed in molar ellipticity. The proteins were analyzed at a final concentration of 1 mg/ml and the assays were performed in 10 mM HEPES, pH 7.5, 0.5 M NaCl. When necessary, glyoxylate or pyruvate was added at 1 mM.

**Isothermal Titration Calorimetry**

Measurements were performed on a VP-Microcalorimeter (MicroCal, Northampton, MA) at 25°C. The protein was thoroughly dialyzed against 10 mM HEPES, pH 7.5, and 500 mM NaCl. Solutions of glyoxylate (0.5 mM) and pyruvate (1.5 mM) were directly prepared in dialysis buffer. Each titration involved a series of 4 μl injections of effector molecules into the protein solution. The mean enthalpies measured from injection of the ligand in the buffer were subtracted from raw titration data before data analysis with ORIGIN software (MicroCal). Titration curves were fitted by a nonlinear least squares method to a function for the binding of a ligand to a macromolecule (46). From the curve thus fitted, the parameters ΔH (reaction enthalpy), KA (binding constant, KA = 1/KD), and n (reaction stoichiometry) were determined. From the values of KA and ΔH, the changes in free energy (ΔG) and in entropy (ΔS) were calculated with the equation: ΔG = -RT lnKA = ΔH - TΔS, where R is the universal molar gas constant and T is the absolute temperature.
Protein Data Bank accession

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 1TD5, 2O99 and 2O9A), Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (www.rcsb.org).

RESULTS

Screening of small molecule library for binding the IclR

Significant sequence similarity between IclR and previously characterized AllR regulators particularly among amino acids involved in glyoxylate binding (20) indicated that the IclR effector may have chemical features similar to glyoxylate rather than phosphoenolpyruvate, which was proposed as the IclR effector by Cortay et al. (8). To clarify the nature of the IclR effector molecule we decided to perform an unbiased screen of IclR against a variety of glyoxylate-associated compounds along with phosphoenolpyruvate as part of large set of metabolic scaffolds.

The expression of the IclR Ser97 to Arg287 fragment produced a soluble polypeptide (see Experimental Procedures for details) corresponding to the effector-binding domain. This protein, termed C-IclR, was tested for binding against 158 metabolic scaffolds (47) by high-throughput protein stability assay using static light scattering technology (25, 26). The library contained intermediates of the glyoxylate bypass and the three carboxylic acids cycle, including glyoxylate, malate, oxaloacetate, citrate, isocitrate, α-ketoglutarate, succinate and fumarate as well as phosphoenolpyruvate. The glyoxylate-binding domain of AllR regulator, C-AllR (20), was also tested against the same library as a control. The thermal melt conditions for C-IclR and C-AllR were established to generate interpretable unfolding data (described in the Materials and Methods). The compounds that induced a shift in the midpoint transition temperature (ΔTm) of proteins by more than 1.5°C are listed in Table 1.
TABLE 1. **Stabilization effect of ligand binding.** The thermal stabilization of each protein by 1 mM ligand concentration was evaluated using differential static light scattering.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Ligand</th>
<th>Delta temperature (ºC)(^1)</th>
<th>Chemical structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-IclR</td>
<td>Glyoxylate</td>
<td>11.1 ± 0.8</td>
<td><img src="structure1.png" alt="Glyoxylate structure" /></td>
</tr>
<tr>
<td></td>
<td>Allantoate</td>
<td>7.1 ± 0.3</td>
<td><img src="structure2.png" alt="Allantoate structure" /></td>
</tr>
<tr>
<td></td>
<td>DL-glyceraldehyde 3-phosphate</td>
<td>3.2 ± 0.1</td>
<td><img src="structure3.png" alt="DL-glyceraldehyde structure" /></td>
</tr>
<tr>
<td></td>
<td>Phosphoenolpyruvate</td>
<td>1.9 ± 0.2</td>
<td><img src="structure4.png" alt="Phosphoenolpyruvate structure" /></td>
</tr>
<tr>
<td></td>
<td>Pyruvate</td>
<td>1.6 ± 0.2</td>
<td><img src="structure5.png" alt="Pyruvate structure" /></td>
</tr>
<tr>
<td>C-AllR</td>
<td>Glyoxylate</td>
<td>9.9 ± 0.5</td>
<td><img src="structure1.png" alt="Glyoxylate structure" /></td>
</tr>
<tr>
<td></td>
<td>Allantoate</td>
<td>7.8 ± 0.1</td>
<td><img src="structure2.png" alt="Allantoate structure" /></td>
</tr>
<tr>
<td></td>
<td>Allantoin</td>
<td>3.0 ± 0.2</td>
<td><img src="structure6.png" alt="Allantoin structure" /></td>
</tr>
<tr>
<td></td>
<td>DL-glyceraldehyde 3-phosphate</td>
<td>3.7 ± 0.1</td>
<td><img src="structure3.png" alt="DL-glyceraldehyde structure" /></td>
</tr>
<tr>
<td></td>
<td>Phosphoenolpyruvate</td>
<td>2.4 ± 0.1</td>
<td><img src="structure4.png" alt="Phosphoenolpyruvate structure" /></td>
</tr>
<tr>
<td></td>
<td>Pyruvate</td>
<td>0.7 ± 0.1</td>
<td><img src="structure5.png" alt="Pyruvate structure" /></td>
</tr>
</tbody>
</table>

\(^1\) Delta temperature was calculated as the difference in the transition temperature between the protein in the absence and in the presence of a given ligand.
Glyoxylate was identified as the strongest thermostabilizing compound for C-IclR. The shift in C-IclR $\Delta$Tm in the presence of glyoxylate (11.1°C) was close to the effect of this compound (9.9°C) on C-AllR. Additional compounds inducing significant stabilization of C-IclR included allantoate, pyruvate, and phosphoenolpyruvate. Apart from glyoxylate, allantoate DL-glyceraldehyde-3-phosphate, and phosphoenolpyruvate C-AllR was stabilized by allantoin but not by pyruvate (Table 1). The thermal stabilizing effect of glyoxylate on C-IclR was confirmed by circular dichroism spectroscopy (Fig. 1). In agreement with the differential light scattering results, the circular dichroism analysis showed an increase in 12°C in the melting temperature of C-IclR in the presence of glyoxylate.

All five compounds identified for C-IclR contained a common chemical motif in form of adjacent carbonyl or hydroxyl groups (Table 1) with the strongest thermal stabilizing agent, glyoxylate, being the simplest representation of this chemical scaffold. The intermediate role played by glyoxylate in the pathway regulated by IclR also strengthens its case as a candidate for the native IclR effector molecule. At the same time, identification of several metabolites causing IclR thermal stabilization and having a common chemical signature indicates that this sensor protein might be reactive to more than one effector molecule.

![Figure 1](image)

**Figure 1 Circular dichroism of C-IclR plus (square) and minus glyoxylate 1 mM (circle).**

The proteins were used at a final concentration 1 mg/ml and the assays were performed in 10 mM HEPES, pH 7.5, 0.5 M NaCl.
Effect of small molecule modulation of IclR binding in vitro

The effect of glyoxylate, allantoate, DL-glyceraldehyde-3-phosphate, pyruvate, and phosphoenolpyruvate on IclR activity was tested in vitro by electrophoretic mobility shift assays (EMSA). The binding of IclR to the DNA fragment corresponding to the promoter region of the aceBAK operon and containing IclR box II (5) was analyzed in the presence and the absence of 100 μM or 1 mM of each compound. According to EMSA results glyoxylate caused dissociation of the IclR regulator from DNA thus increasing the fraction of unbound DNA (Fig. 2). On the contrary, pyruvate and, to a lesser extend, DL-glyceraldehyde 3-phosphate and phosphoenolpyruvate, stabilized the IclR-DNA complex and resulted in a decrease of unbound DNA (Fig. 2). Allantoate did not affect the IclR/aceB complex (data not shown).

In vitro transcription run-off (Fig. 3) and EMSA (Fig. 4) experiments were performed at different concentrations of glyoxylate. In the in vitro transcription run-off experiment the level of the aceB transcript was normalized to the concentration of the bla transcript used as internal control. The presence of IclR reduced the level of the aceB transcript to 30% of the level measured in the absence of this protein (Fig. 3, lanes 1 and 2). The addition of glyoxylate resulted in gradual increase of the aceB transcription reaching up to 74% of original value at 50 μM of glyoxylate (Fig. 3, lane 6). In EMSA experiments more than half of the bound operator DNA was released by IclR in the presence of 5 μM glyoxylate (Fig. 4A) while 50 μM of pyruvate were required to achieve 100% DNA-shift (Fig. 4B). While a substantial difference in the concentrations of glyoxylate needed to decrease IclR binding to DNA was observed between the two techniques tested they both confirmed the dissociation of the IclR-operator complex in a concentration dependent manner. This apparent discrepancy could be partially explained by the IclR interactions with the C-terminal domain of the RNA polymerase α subunit, as described earlier (5).
Figure 2 Effect of different ligands on the binding of IclR to the aceB promoter.

EMSA results using 2.5 nM of biotin-labeled aceB promoter and 25 nM (A) or 50 nM (B) of IclR with different inducer molecules. The full binding conditions are described in “Experimental procedures”. Lane 1 and 11 show the migration of the target DNA fragment; lanes 2-10, IclR 25 nM, lanes 12-14, IclR 50 nM. Glyoxylate (lanes 3, 4 and 13), pyruvate (lanes 5, 6 and 14), phosphoenolpyruvate (lanes 7 and 8) and glyceraldehyde-3-phosphate (lanes 9 and 10) were added at a concentration of 0.1 mM (3, 5, 7, 9 and 13) or 1 mM (4, 6, 8, 10 and 14).

Phosphoenolpyruvate and DL-glyceraldehyde-3-phosphate were able to stabilize the IclR-DNA complex (Fig. 2) however, the concentrations required to obtain this effect (over 10 mM, data not shown) were out of the physiological range for these compounds (47, 48, 49). Thus phosphoenolpyruvate and DL-glyceraldehyde-3-phosphate were not further considered for an IclR effector role.

Since glyoxylate and pyruvate had distinctive and antagonist effects on the formation of the IclR-DNA complex, in the next series of EMSA experiments we analyzed the combined effect of glyoxylate and pyruvate on the stability of the IclR-DNA complex. In this setting, one of compounds was added in increasing amounts while the concentration of the other was kept constant. The results obtained (Fig. 4A) demonstrated that the glyoxylate-induced dissociation of the IclR-operator complex was overturned by 50-fold excess of pyruvate (500 μM pyruvate with 10 μM glyoxylate). These results where corroborated by doing the reciprocal experiment (Fig. 4B). The common chemical features of these compounds suggested that this effect might be due
to the competition for the same binding site in IclR, with glyoxylate having a significantly higher
binding affinity than pyruvate.

Figure 3 In vitro transcription runoff experiments using the aceB promoter, IclR and
different concentrations of glyoxylate.

A. In vitro transcription assays were performed as described in “Experimental procedures”. The
assays were performed at 37 °C for 30 min in the absence (lane 1) or in the presence of 100 nM
IclR added prior to 20 nM of RNA polymerase (lanes 2-6). Glyoxylate was added at 1 μM (lane
3), 5 μM (lane 4), 10 μM (lane 5) and 50 μM (lane 6). The arrows denote the aceB (434
nucleotides) and the bla (860 nucleotides) transcripts. B. Quantitative analysis of the in vitro
transcription runoff experiment presented on (A). For each lane the amount of aceB transcript was
normalized (average of three experiments) to the expression level of the bla transcript. Expression
of aceB in the absence of IclR represented 10% of the expression of bla transcript (internal
control). Each bar represents the percentage of the aceB transcript relative to the transcription
level in the absence of IclR protein (lane 1).
Figure 4 Effect of glyoxylate and pyruvate on the binding of IclR to its operator site.

A, EMSA results using 2.5 nM of biotin-label aceB promoter, 50 nM of IclR and increasing concentrations of glyoxylate (0-500 μM) as shown on the top of the picture. Pyruvate 500 μM was added to a second set of samples that also contained increasing concentrations of glyoxylate as shown on the top of the panel. No protein was added to the first lane. B, EMSA results using 25 nM of IclR and increasing concentrations of pyruvate (0-500 μM) as shown on the top of the picture. Glyoxylate 500 μM was added to a second set of samples that also contained increasing concentrations of pyruvate as shown on the top of the panel. C and D, EMSA results from Fig. 4A and B, respectively, quantified by image analysis. Each graph represents the percentage (average of three experiments) of the bound DNA in the presence of the ligand relative to the amount of the DNA bound to the same protein in the absence of the ligand.

In order to further characterize IclR’s interactions with potential effector molecules, the thermodynamic properties of IclR interactions with glyoxylate and pyruvate were determined using isothermal titration calorimetry. The titration of C-IclR with each compound followed an exothermal heat change profile giving rise to a sigmoidal binding curve with glyoxylate or hyperbolic with pyruvate (Fig. 5). The calculated thermodynamic parameters are summarized in Table 2. The stoichiometry of the reaction, 0.5, is consistent with the binding of one ligand molecule per IclR dimer. In accordance with EMSA results, the C-IclR dissociation constant (KD) for glyoxylate (0.9 ± 0.2 μM) was significantly lower (~150 times) than that for pyruvate.
(156.2 ± 7.9 μM). While the KD values obtained for binding of glyoxylate and pyruvate to the full-length IclR were higher (14.2 ± 1.2 μM and 321.7 ± 20 μM, respectively) their relative ratio was within the same range as obtained for C-IclR.

In conclusion, our EMSA results show glyoxylate as the strongest candidate for the IclR effector molecule, which destabilized IclR interactions with DNA in vitro. By contrast, pyruvate affected IclR’s interactions by stabilizing the IclR-DNA complex.

Figure 5 Isothermal titration calorimetry data for the binding of glyoxylate (A) or pyruvate (B) to C-IclR.

Heat changes (upper panel) and integrated peak areas (lower panel) for the injection of a series of 4 μl aliquots of 500 μM glyoxylate in a solution of 50 μM of protein. Data was fitted with ORIGIN.
Table 2 Thermodynamic parameters derived from the calorimetric titration of C-IclR and C-AllR with effector molecules.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Ligand</th>
<th>$K_D$ (μM)</th>
<th>$n$</th>
<th>$K_A$ (M$^{-1}$)</th>
<th>$\Delta H$ (kcal/mol)</th>
<th>$\Delta S$ (cal/mol·K)</th>
<th>$\Delta G$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-IclR</td>
<td>glyoxylate</td>
<td>0.9 ± 0.2</td>
<td>0.5</td>
<td>(1.2 ± 0.2) x 10$^6$</td>
<td>-10.1</td>
<td>-1.96</td>
<td>-8.1</td>
</tr>
<tr>
<td>C-IclR</td>
<td>pyruvate</td>
<td>156.2 ± 7.9</td>
<td>0.5</td>
<td>(6.4 ± 0.3) x 10$^3$</td>
<td>-1.37</td>
<td>-8.58</td>
<td>7.21</td>
</tr>
<tr>
<td>C-AllR</td>
<td>glyoxylate</td>
<td>9.4 ± 2.2</td>
<td>0.5</td>
<td>(1.1 ± 0.3) x 10$^5$</td>
<td>-8.2</td>
<td>-7.21</td>
<td>-1.06</td>
</tr>
<tr>
<td>IclR</td>
<td>glyoxylate</td>
<td>14.2 ± 1.2</td>
<td>0.5</td>
<td>(7.6 ± 0.1) x 10$^4$</td>
<td>-5.8</td>
<td>0.8</td>
<td>-5.0</td>
</tr>
<tr>
<td>IclR</td>
<td>pyruvate</td>
<td>321.7 ± 20</td>
<td>ND$^2$</td>
<td>(3.1 ± 0.4) x 10$^3$</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

$^1$(kcal/mol)

$^2$ND, not determined

Protein solutions at 50 μM in 10 mM KH$_2$PO$_4$/K$_2$HPO$_4$, pH 7.4, and 500 mM NaCl were titrated with 0.5-1.5 mM solutions of effectors. Experiments were carried out at 25°C. Further assay conditions are given under “Experimental Procedures”.

**Table 2 Thermodynamic parameters derived from the calorimetric titration of C-IclR and C-AllR with effector molecules.**

**Modulation of IclR binding by glyoxylate in vivo**

To assess the effect of glyoxylate on IclR activity in vivo the expression level of aceBAK operon was measured during growth on minimal media in the presence and absence of this compound. Although early genetic studies identified IclR as the main regulator of aceBAK operon (51), recent E. coli transcriptome analysis indicated that this operon is also under control of the global catabolite regulation protein, Crp. The Crp potential binding sequence was found overlapping the IclR box II (52). In light of this new data the IclR activity was monitored in the presence and the absence of a functional Crp regulator.

Single ΔiclR and Δcrp as well as double (ΔiclRΔcrp) deletion strains were prepared in the E. coli BW25113 strain background. Deletion strains were grown to exponential phase on MOPS minimal medium with pyruvate (10 mM) as the carbon source in the presence and in the absence of 5 mM glyoxylate. The expression levels of aceB and aceK genes were measured using real time quantitative PCR (qRT-PCR). The expression profiles of both genes followed the same trend, thus only the results for aceB gene will be discussed further on.

The expression of aceB in the ΔiclR and Δcrp strains was increased by 3.3 and 24 fold respectively compared to the wild type strain. The double ΔiclRΔcrp deletion caused a 31-fold
increase in aceB expression corresponding to the combined effect of each gene deletion. The presence of glyoxylate in the media had no significant effect on the level of aceB expression in ΔiclR and ΔiclRΔcrp strains lacking the functional IclR regulator, while it triggered a 1.4 fold increase when the IclR regulator was present in the cells (Δcrp strain) (Fig. 6). The glyoxylate-mediated induction was dose dependent since in the presence of 10 mM glyoxylate a 2.7 fold induction in aceB expression was observed (data not shown). In the case of the wild type strain containing both IclR and Crp regulators the presence of glyoxylate in the media did not result in a detectable increase of aceBAK, expression probably due to effector independent tight repression by the functional Crp regulator.

The substantial derepression effect of the Crp deletion confirmed the major role played by this global regulator on the expression of the aceBAK operon. The effect of the IclR deletion was significantly smaller, yet consistent with its role as a specific regulator of this operon. The derepression effect in the presence of glyoxylate observed in the Δcrp strain appears to be specific to the presence of functional IclR indicating that glyoxylate can affect the activity of the IclR regulator in a manner expected from the native effector molecule. Unfortunately we were unable to study the corepressor effect of pyruvate in vivo since the crp deletion strain is unable to grow on the variety of carbon sources (52) required for these kinds of studies.
Figure 6 Effect of glyoxylate on the expression of aceB on different strains measured by quantitative real time PCR.

The strains BW25113 (wild type), ∆iclR, ∆crp or ∆iclR∆crp were grown under aerobic conditions in MOPS with pyruvate 10 mM as carbon source with (light grey bars) or without 5 mM glyoxylate (dark grey bars). RNA extractions and qRT-PCR was performed as described in “Experimental procedures”. The amplification values obtained were corrected for those obtained using rrsC as internal control. The values shown are relative to those observed for the same gene in the wild type strain grown with pyruvate.

Structure of the IclR effector-binding domain

To provide a structural framework for IclR interactions with the newly identified small molecules, the crystal structures of the IclR effector binding domain alone (apo-structure) and in complex with glyoxylate and pyruvate were determined at 2.3 Å, 1.8 Å and 2.3 Å, respectively. Final R/Rfree for the apo-structure was 0.23/0.299 (PDB #1TD5). Final R/Rfree for complex structures was 0.215/0.262 (PDB #2O99) for the glyoxylate and 0.177/0.233 (PDB #2O9A) for the pyruvate complexes. All three models possessed excellent stereochemistry as judged by PROCHECK (42), with no Ramachandran violations. Statistics for all three structures are shown in Table 3. In order to ease the description and comparison to other structures the assigned numbers for β sheets and α-helices are given in relation to the full-length homolog TM0065 (16). The C-IclR domain displays a similar architecture to the two other IclR family members whose structures have been determined, TM0065 (16) and AllR (20). The C-IclR structure comprises a
centrally located 6-stranded anti-parallel β-sheet, surrounded on one side by two long α-helices (α5 and α9), and on the other by three shorter α-helices (α6-α8). The β-sheet is strongly curved with the shortest helix (α6) fitting on the inside of the half-barrel (Fig. 7).

In complex structures both the glyoxylate and pyruvate molecules occupied the cavity corresponding to the effector-binding pocket in the AllR structure (20). The map quality was assessed by constructing an omit Fo-Fc electron density map. The map was calculated after 10 cycles of refinement without the ligand in the model (Fig. 8). Similar to the C-AllR-glyoxylate complex structure, no significant conformational changes were observed in C-IclR complexes with pyruvate or glyoxylate when compared to C-IclR apo-structure. The ligands were bound in the same orientation with almost identical positions for atoms C1, C2, O1, O2 and O3 (Fig. 8). Oxygen atoms of glyoxylate and pyruvate were coordinated by an extensive network of hydrogen bonds with the side chains and backbone of the protein. For each ligand the O1 atom is hydrogen-bound to the Asp212Oδ2 and Gly160N, the O2 atom to the Ser239Oγ and Ser241Oγ, and the O3 to the Ala161N. The C1-Sg distance varies from 1.86 to 2.09 Å for glyoxylate complex and from 1.90 to 1.97 Å for the pyruvate complex. Both complex models were refined with the same restraints with no covalent bond between ligand C1 and Cys222Sγ in the dictionary. In agreement with previous sequence analysis, all but two amino acids, Ile134 and Leu143, in the effector binding area are conserved between IclR and AllR (20). The corresponding positions in AllR are represented by leucine (Leu129) and methionine (Met138), respectively (Fig. 9).
<table>
<thead>
<tr>
<th>PDB code</th>
<th>1TD5</th>
<th>2O99</th>
<th>2O9A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sets</td>
<td>Inflection (Se1)</td>
<td>Peak (Se2)</td>
<td>Remote (Se3)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9794</td>
<td>0.9790</td>
<td>0.9611</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions, Å</td>
<td>53.5, 81.5, 154.7</td>
<td>53.5, 81.4, 154.3</td>
<td></td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>31311</td>
<td>31815</td>
<td>31820</td>
</tr>
<tr>
<td>Resolution, Å</td>
<td>20-2.30 (2.38-2.30)</td>
<td>20-2.30 (2.38-2.30)</td>
<td>20-2.30 (2.38-2.30)</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>99.1 (98.4)</td>
<td>99.1 (98.4)</td>
<td>99.2 (98.5)</td>
</tr>
<tr>
<td>(R_{merge})</td>
<td>0.077 (0.310)</td>
<td>0.095 (0.311)</td>
<td>0.092 (0.384)</td>
</tr>
<tr>
<td>I/G(I)</td>
<td>22.4 (4.9)</td>
<td>22.5 (5.0)</td>
<td>19.9 (4.0)</td>
</tr>
<tr>
<td>Mean figure of merit</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Refinement

| Resolution range | 19.9-2.30 | 50-1.8 | 50-1.8 |
| \(R/R_{free}\) | 0.230/0.299 | 0.201/0.257 | 0.173/0.232 |
| Root mean square deviation | 0.014 | 0.020 | 0.019 |
| Bonds (Å) | 1.50 | 1.6 | 1.8 |
| Angles (degrees) | | | |
| Total atoms | | | |
| Protein atoms | 5499 | 5660 | 5680 |
| Solvent | 335 | 628 | 835 |
| Average B for protein atoms | 37.82 | 26.5 | 19.9 |
| Average B for water molecules | 37.82 | 37.0 | 33.9 |
| Average B for ligands | 22.9 | 21.4 |

\(^a\) Data in the parentheses represent data in the highest resolution shell.

\(^b\) \(R_{merge} = \frac{\sum[I-(\bar{I})]}{\sum I}\), where \(I\) is the observed intensity of an individual reflection, and \(\bar{I}\) is the mean intensity of that reflection.

\(^c\) \(R_{free}\) is the cross-validated \(R_{factor}\) computed with the test set of reflections.

**Table 3 Data collection and refinement statistics.**
Figure 7 Overall structure of C-IclR.

(A) Cartoon diagram of a monomer of the IclR ligand-binding domain in complex with glyoxylate (PDB #2O99). (B) C-IclR interdomain interface between two monomers in the (PDB #2O9A) structure. Loops are colored green, a helices red and β-strands yellow. Glyoxylate is represented as a stick figure in cyan. The amino and carboxyl termini are labeled.
The presence of an isoleucine residue at position 134 in the IclR structure results in the shortening of the distances between Leu154 and glyoxylate atoms by 2 Å in the IclR-glyoxylate complex structure compared to the same region in C-AllR-glyoxylate structure (contact distance Cδ1 to C2: 3.68 Å vs 4.44 Å; Cδ2 to C1: 4.17 Å vs 6.16 Å). In addition the average distances between protein atoms are shorter by an average of 1 Å in C-IclR-glyoxylate versus C-IclR structures (10.8 Å and 11.6 Å, respectively) while there is no such difference between C-AllR and C-AllR-glyoxylate structures (11.96 Å and 12.1 Å, respectively) for the same parameters. These structural dissimilarities between binding cavities in C-IclR and C-AllR may contribute toward close to tenfold difference in KD for glyoxylate as determined by isothermal titration calorimetry (Table 2).

Taking into account the almost identical position of glyoxylate and pyruvate in the C-IclR structures, the interactions involving the methyl group of pyruvate, which is absent in glyoxylate become critical in understanding the different effects of these compounds on IclR activity. The methyl group is well defined in the pyruvate electron density map (Fig. 8B) and is oriented toward the side chain of the Leu143 residue belonging to the other molecule of the C-IclR dimer. The distances between C3 of the pyruvate and Cδ2 atom of Leu143 are on average 4.24 ± 0.4 Å for the four molecules in the C-IclR asymmetric unit. Leu143 makes part of the hydrophobic patch formed by side chains of Met146, Leu154, His216 and Leu220, which are in the range of van de Waals interactions with pyruvate’s methyl group. The presence of methionine instead of leucine in the corresponding position in AllR would explain why pyruvate had no stabilization effect (Table 1).

In C-IclR the inter-domain interactions between C-IclR monomers are similar to those observed in the C-AllR and TM0065 structures. The C-IclR and C-AllR dimers are superimposable to an rmsd of 1.75 Å, compared to an rmsd of 1.01 Å for individual subunits. This common arrangement of the C-terminal domains of three different IclR family members, in three different unit cells, supports the physiological relevance of these interactions for tetramerization of IclR
regulators (Fig. 7B). The additional intermolecular interactions between the pyruvate methyl group and hydrophobic patch observed in C-IclR-pyruvate complex structure would thus stabilize the tetramerization state of IclR protein.

Figure 8 Molecular details of C-IclR interactions with glyoxylate (A, C) or pyruvate (B, D).

In A and B, the side chains of amino acids within a distance of 8 Å are shown and labeled in the figure. Protein residues (yellow), glyoxylate (pink) and pyruvate (green) are represented as stick figures. The back-bone of the protein is shown as a thin yellow line. Leu143 from another monomer is represented in cyan. In C and D, a 3 sigma Fo-Fc omit map was built using the REFMAC5 program (43, 44) for the glyoxylate and pyruvate complex structures, respectively. The map was calculated after 10 cycles of refinement without the ligand in the model. The omit Fo-Fc electron density map covering the ligands is shown in stick representation (in blue for C or cyan for D). Neighbouring residues are also shown in stick representation (cyan in C or yellow in D).
Figure 9 Ligand binding pocket superimposition of AllR and IclR.

Superimposition of the active site of C-IclR with that of C-AllR. Protein residues and glyoxylate are represented as stick figures. Components of C-AllR are represented in green and light pink, while components of IclR are colored in yellow and magenta.

Effect of glyoxylate and pyruvate on the oligomerization of IclR

In order to establish the effect of glyoxylate or pyruvate on the oligomeric state of IclR protein, this protein was preincubated with 1 mM glyoxylate or 10 mM pyruvate and analyzed by gel filtration. The untreated IclR protein was used as control.

In the absence of effector molecules IclR was eluted in two peaks, corresponding to the dimeric (60 ± 5 kDa) or tetrameric (120 ± 12 kDa) states. The abundance of both fractions (54/46 % relative percentage, respectively) indicated that under our experimental conditions the IclR protein is in a dynamic equilibrium between these two oligomeric states (Fig. 10, Table 4). After preincubation with glyoxylate IclR was eluted predominantly in form of the lower (60 kDa) molecular weight fraction indicating dramatic shift of equilibrium toward the dimeric state. On the other hand, preincubation with 10 mM pyruvate resulted in elution of IclR in form of the
higher (120 kDa) molecular weight fraction corresponding to the tetramer. Similar results were obtained if glyoxylate was included in the elution buffer however; the tetrameric state of IclR was more favorable if pyruvate was included in the elution buffer. These results can be explained by the difference in affinity between the ligands (Table 2). Overall, size exclusion experiments demonstrated that glyoxylate and pyruvate have a dramatic effect on the oligomeric state of the IclR regulator. This suggests that the glyoxylate effect on IclR activity is due to the destabilization of tetramers, which is required for efficient binding to the operator DNA, while the co-repressor effect observed for pyruvate might be due to the stabilization of the IclR tetramer on the DNA.

**Figure 10** Gel filtration analysis of IclR oligomerization in presence of glyoxylate or pyruvate.

IclR (thick line), IclR preincubated with pyruvate (continuous line) or with glyoxylate (dotted line) was passed through a Superose 12 column, as described in “Experimental Procedures”. As inset is shown the calibration curve. The standards used were β-amylase (200 kDa), BSA (66 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa).
Functional analysis of key amino acids involved in interdomain interactions and effector binding

In light of the important effect of glyoxylate and pyruvate on IclR’s oligomeric state we investigated the specific role of amino acids Leu143, Met146, Leu154 and Leu220 in the interdomain interactions and effector binding. In course of this analysis each of these amino acids was replaced by alanine or aspartic acid. The mutant proteins were purified and submitted to gel filtration analysis and EMSA with the IclR box II fragment in the presence and absence of glyoxylate or pyruvate (Table 4, Fig. 11-12).

The alanine substitution of Leu143 did not significantly affect the gel filtration profile of IclR (Fig. 11). However, according to the EMSA experiments a higher concentration (100 nM) of the Leu143Ala mutant protein was required to achieve the DNA binding comparable to the wild type IclR (Table 4). The addition of pyruvate stabilized this mutant’s complex with operator DNA similarly to the wild type IclR, while the glyoxylate had significantly lower effect on the Leu143Ala mutant (Fig. 12). In the presence of the latter ligand an additional band was observed on the EMSA gels. This additional band migrates faster than the one corresponding to the wild type IclR-DNA complex but slower that the free DNA (Fig. 12). This band would correspond to an IclR-DNA complex with a lower molecular weight than in case of the wild type IclR. Taking into account that IclR might bind DNA as a tetramer (5), this lower molecular weight complex would correspond to the association of operator DNA with the dimers of Leu143Ala as an intermediate in the releasing of IclR from the Protein-DNA complex.

The gel filtration profiles of the alanine substitutions of Met146, Leu154 and Leu220 showed them predominantly as dimers in solution (70/30 relative percentage, respectively, Table 4) however, they were still able to tetramerize over DNA as evidenced by a high molecular weight migrating complex (Fig. 12). The mutant proteins showed a significant variation of responses to glyoxylate and pyruvate (Table 4). While the Met146Ala mutant’s response to these ligands was
similar to the wild type IclR protein, the Leu220Ala mutant showed no response to glyoxylate but was stabilized by pyruvate. The Leu154Ala mutant to the contrary responded to glyoxylate but not to pyruvate. These results put in evidence the different involvement of these amino acids in ligand binding.

<table>
<thead>
<tr>
<th>Protein</th>
<th>EMSA&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Gel Filtration&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA binding&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Glyoxylate</td>
</tr>
<tr>
<td>IclR (wt)</td>
<td>50 nM</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L154A</td>
<td>150 nM</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L154D</td>
<td>75 nM</td>
<td>-</td>
</tr>
<tr>
<td>L220A</td>
<td>150 nM</td>
<td>NE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L220D</td>
<td>100 nM</td>
<td>NE</td>
</tr>
<tr>
<td>L143A</td>
<td>100 nM</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L143D</td>
<td>100 nM</td>
<td>-</td>
</tr>
<tr>
<td>M146A</td>
<td>75 nM</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M146D</td>
<td>100 nM</td>
<td>--</td>
</tr>
</tbody>
</table>

<sup>1</sup> Protein necessary to obtain a 50% DNA shift; glyoxylate and pyruvate were added at 1 mM.

<sup>2</sup> Gel filtration was performed using a Superose 12 column. Protein 100 μg, glyoxylate and pyruvate were added at 1 or 10 mM, respectively. Relative percentages were calculated from the high of the peaks.

<sup>3</sup> Concentration of protein necessary to obtain a 50% DNA shift on EMSA.

<sup>4</sup> The effect of the ligand on the protein/DNA complex was classified as --- = 95%, -- = 50%, = 25% decrease or +++ = 95%, ++ = 50%, += 25% increase in binding. NE = no effect on the stability of the protein/DNA complex.

<sup>5</sup> ND, not determined

**Table 4 Summary of the effect of glyoxylate and pyruvate on DNA binding and on the oligomeric state of IclR or its mutant variants.**

In the next series of mutagenesis the hydrophobic region was altered by single replacement of Leu143, Met146, Leu154 or Leu220 residues to the charged residue, aspartate. The mutations had a dramatic effect on IclR oligomerization, with all mutant proteins being dimers as judged by gel filtration analysis (Table 4). According to EMSA results, Leu143Asp and Leu220Asp bound the operator DNA at a higher protein concentration (100-150 nM) if compared to the wild type
regulator (50 nM) and formed only low molecular weight complexes with DNA. Both mutants were not able to respond to either glyoxylate or to pyruvate (Fig. 12).

**Figure 11 Analysis of IclR L143A and M146A oligomerization in presence of glyoxylate and pyruvate.**

(A) L143A and (B) M146A (thick line) were preincubated with pyruvate (continuous line) or with glyoxylate (dotted line) and passed through a Superose 12 column, as described in “Experimental Procedures”.

EMSA experiments on Met146Asp showed low and high molecular weight protein-DNA complexes (Fig. 12). In agreement with the mutations to alanine, Met146Asp does not interfere in the binding of glyoxylate or pyruvate. Moreover, glyoxylate disrupted the formation of the high molecular weight Protein-DNA complex increasing the amount of the lower molecular weight complex (Fig. 12). The presence of pyruvate, on the contrary, increases the formation of the high
molecular weight complex. These results show that while the Met146 is important for the stabilization of tetramers in solution, it is not directly involved in the binding of the ligands although it might contribute to the overall hydrophobicity of the area.

In the case of Leu154Asp, the EMSA profile was similar to that of the wild type IclR, however, the response to glyoxylate was weaker (Fig. 12). Pyruvate had no stabilization effect on this mutant protein (Fig. 12) confirming the Leu154 role in the specific interactions with pyruvate.

![Figure 12](image)

**Figure 12 Functional analysis of key amino acids involved in interdomain interactions and effector binding.**

The glyoxylate and pyruvate binding on mutant IclR proteins was tested by EMSA. The name of the corresponding mutant protein and its concentration used for binding is indicated above each image. Glyoxylate and pyruvate were tested at 1 mM. The full binding conditions are described in “Experimental procedures”.
DISCUSSION

The IclR family of transcriptional regulators shares common structural features with their C-terminal domains representing small molecule binding modules adapted to recognize wide variety of cellular metabolites as effector molecules. The identity of the effector molecules remains unknown for a vast majority of these proteins. This reflects a general situation within the large transcriptional factor families such as LysR (53), LuxR (54), TetR (55) and GntR (56) accounting for thousands of transcriptional regulators that respond to unidentified chemical stimuli, which need to be addressed.

Using the founding member of the IclR family as a test case, we explored the possibility of screening a small molecule library using a high throughput protein stability assay as a general strategy to identify putative ligands for such families’ representatives. Although the phenomenon of protein thermostabilization by a bound small molecule was known for decades (57, 58, 59) only recently it became applicable in screening chemical libraries (25, 26, 60). The breakthrough came from the development of technologies enabling the thermo-denaturation of a large number of protein samples in parallel monitoring by static light scattering or fluorimetry adapted to microplate format (25, 26). This kind of technology enables the high-throughput screening of protein – ligand interactions, independent of the knowledge of the protein function. Similar technology named ThermoFluor was previously used to identify and elaborate upon the hypothetical function of a PLP-dependent amino transferase enzyme (61). In this study we screened a small molecule library containing representatives of the main chemical scaffolds identified in the E. coli metabolome (47) as well as a large number of chemicals associated with the three carboxylic acids cycle and the glyoxylate bypass. Glyoxylate was identified as the strongest thermo stabilizing compound for IclR. In EMSA and in vitro transcription runoff experiments the presence of glyoxylate disrupted IclR complex formation with operator DNA, an effect that would be expected from a physiologically relevant effector. Allantoate, DL-glyceraldehyde-3-phosphate, pyruvate and phosphoenolpyruvate also had a stabilizing effect on
IclR. Surprisingly, three of these compounds stabilize the binding of IclR to DNA. All five compounds demonstrated a common chemical moiety in form of vicinal carboxyl groups. Structural analysis of the C-IclR complexes with glyoxylate and pyruvate clearly showed that this chemical sub-structure is the key element in interactions with IclR. This observation indicates that the chemical screen based approach can provide the necessary information for identification of chemical moieties specifically recognized by IclR family members as well as for other effector-binding proteins in general.

Only two compounds, glyoxylate and pyruvate, altered the binding of IclR in vitro at a physiologically relevant concentration. The role of glyoxylate as IclR effector is supported by its role as an intermediate in the glyoxylate bypass, which is regulated by IclR. The importance of this pathway for cellular metabolism is reflected in the complex regulation of this bypass at transcriptional and post-transcriptional levels. At the transcriptional level it is tightly regulated by the binding of at least 6 different transcription factors and it is extremely dependent on the metabolic state of the cell. The aceBAK operon is under negative regulation by IclR, FadR (indirectly by regulation of the iclR expression, 62), Crp (52) and ArcA (63). Additionally, it is under positive regulation by IHF (64) and FruR (65, 66). Once induced at transcriptional level, the flow through the pathway is regulated, in part, by the inhibitory phosphorylation of isocitrate dehydrogenase (67, 68). Gluconeogenesis also plays an essential role during growth on acetate by synthesizing glucose from nonhexose precursors. The genes pckA and ppsA belong to two parallel pathways for the gluconeogenic conversions of the three carboxylic acids cycle intermediates to phosphoenolpyruvate with the malic enzymes (sfcA and maeB) and ppsA forming one path for the conversion of malate to phosphoenolpyruvate and pckA forming the other for the conversion of oxaloacetate to pyruvate (69, 70, 71).

Taking into account the physiological data we could hypothesize that the cell could maximize the efficiency of expression/repression of the aceBAK operon by using two antagonistic effectors on the same transcription factor. By EMSA and biophysical tests we were able to establish that IclR
binds glyoxylate and pyruvate. Pyruvate is the first non-phosphorylated intermediate in glycolysis and a key intermediate in catabolic and biosynthetic reaction pathways. There are several pathways that deliver this compound (48, 72, 73). Siddiquee et al. (48, 49) have shown that the intracellular concentration of pyruvate can be from 0.15 mM to 0.4 mM depending if samples were taken from batch or continuous cultures. These values are within the affinity constants defined for IclR binding. Glyoxylate, on the other hand, is an important intermediate of the central microbial metabolism in the glyoxylate bypass. Glyoxylate is also generated from glycolate or purine degradation (74, 75). The intracellular concentration for this metabolite in E. coli has not been reported, however, there are studies in the closely related bacterium Salmonella typhimurium (76) that suggest that the concentration of glyoxylate should be very low (~ 10 μM). Thus, the big difference in affinity between glyoxylate and pyruvate binding to IclR might be linked to their intracellular concentrations.

The posttranscriptional regulation of isocitrate dehydrogenase termed the branch point effect has been the focus of numerous studies (77, 78). This effect is a consequence of the difference in the affinities of isocitrate dehydrogenase and isocitrate lyase for isocitrate (Km of 8 and 600 mM, respectively) (77, 78). As a result, the flux of the glyoxylate bypass is strikingly sensitive to the phosphorylation state of isocitrate dehydrogenase. It has been proposed that the presence of pyruvate during the acetate metabolism alters the isocitrate dehydrogenase activity, resulting in inhibited growth due to reduced carbon flux through the glyoxylate shunt (79). Our results clearly show that the effect observed by El-Mansi et al. (79) could be explained by the binding of pyruvate to IclR, which would quickly repress the expression of aceK (encoding for isocitrate dehydrogenase kinase/phosphatase) and would be reflected as a decrease in isocitrate dehydrogenase activity.

According to the data on IclR presented in this study a mechanism of dual regulation by two effectors can be proposed. When cells grow on complex media or if glucose is added to the media, Crp is the primary transcription regulator (52). The DNA binding sequences for IclR (box
II) and CRP in the aceBAK operon have been predicted in silico to overlap (52). However, when the levels of cAMP increase and CRP is inactive (i.e., growth on acetate) the regulation due to IclR becomes predominant. Under these conditions the concentration of glyoxylate increases due to an increase in the flux through the shunt and is now able to compete with pyruvate for binding of IclR. We suggest that glyoxylate is the primary ligand for IclR, which will trigger IclR dissociation from the DNA. However, when the concentrations of pyruvate are increased due to gluconeogenesis the complex aceB-IclR-pyruvate might be formed as a safe mechanism to quickly repress the glyoxylate shunt. This mechanistic hypothesis is supported by the fact that both ligands bind to the same site in IclR.

According to the structural analysis and mutagenesis studies presented in this work the hydrophobic patch formed by Leu143, Met146, Leu154 and Leu220 plays a critical role in IclR interactions with pyruvate. Moreover, mutagenesis studies demonstrated that Leu154 substitution by alanine would abolish the pyruvate binding by IclR while leaving the binding of glyoxylate unaffected.

Glyoxylate interactions with IclR strongly resemble those of glyoxylate with another E. coli member of this protein family, the allantoin utilization regulatory protein (AllIR). The two proteins share 42% sequence identity and all but two amino acids, Ile134 and Leu143, are conserved in the effector binding area (20). The binding of glyoxylate causes conformational changes in both IclR and AllIR proteins resulting in their dissociation from corresponding DNA binding sites. Nevertheless in case of IclR the glyoxylate-bound conformation causes the shift from tetramer to dimer state, while in case of AllIR the binding of glyoxylate appeared not to affect its dimer to tetramer ratio (20). In the AllIR-glyoxylate complex structure Met138 (corresponding to the Leu143 in IclR) is involved in interactions with the glyoxylate bound to the adjacent AllIR molecule. This interaction would strengthen the tetramerisation interface in the glyoxylate-bound conformation of AllIR however, it is absent in the IclR-glyoxylate complex structure. Full understanding of the differences between IclR and AllIR interactions with
glyoxylate will have to wait for the structural information on the full-length IclR and AllR in complex with this ligand.

The presence of two similar regulators, AllR and IclR, responsive to the same effector in E. coli might be conditional upon different physiological roles played by these regulators. The pathways regulated by these transcriptional factors share glyoxylate as a common intermediate, however, the physiological conditions required for their expression are quite different. AllR regulates the utilization of allantoin as sole nitrogen source. The genes coding the enzymes involved in this metabolic process are organized in three transcriptional units (allA, gcl-hyi-glxR-o484-allB-o433-glxK and allD-f411-f261) that constitute the allantoin regulon. These three units are induced by growth on allantoin or glyoxylate and only seem to be regulated by the action of AllR and AllS (80). On the other hand, IclR represses the expression of the genes that code for the enzymes of the glyoxylate shunt and the regulation of this operon is extremely complex. Isothermal titration calorimetry experiments indicated that AllR has a 10-fold lower affinity for glyoxylate than does IclR. It may be possible that the regulators are tuned to launch their transcription programs as graded responses to the presence of glyoxylate, and that the intracellular concentrations of glyoxylate integrate different metabolic inputs, i.e. acetate metabolism for IclR and allantoin utilization for AllR. Oh et al. (81) characterized the transcript profile of E. coli in acetate cultures using microarrays. The authors observed that not only the genes involved in glyoxylate shunt pathway (ace and glc operons), but also those involved in other glyoxylate-related metabolic pathways, such as glycolate and allantoin metabolism, were all up-regulated. In a similar manner, Pellicer et al. (82) observed cross-induction of glcDEF and aceBAK operons attributable to pathway intersection at the glyoxylate level. However, in this case the signal molecule for each transcription factor is different.

In summary, we present the binding of two small molecule effectors to a single transcription factor (IclR) with antagonistic functional effects in E. coli. The two ligands recognized by the IclR transcriptional regulator act as corepressor (pyruvate) or activator (glyoxylate) on the
transcription of the aceBAK operon, likely in response to intracellular concentrations of these compounds in the bacterial cell.

ACKNOWLEDGEMENTS

We wish to thank all members of the Structural Biology Centre at Argonne National Laboratory and Greg Brown at the University of Toronto for their help in conducting experiments and to Aled Edwards, Alexander Iakounine, and Claudio Gonzalez for critical reading of the manuscript. This work was supported by National Institutes of Health Grants GM62414-01, the Ontario Research and Development Challenge Fund and by the Canadian Institutes of Health Research Grant.

REFERENCES


104

105

CONCLUSIONS AND FUTURE DIRECTIONS

Oligomerization and mechanism of action

Change in oligomerization state is one of the possible mechanisms of function of prokaryotic transcription regulators (Chen and Calvo, 2002; Schleif 2004; Vilar and Saiz, 2005). Operator DNA regions of many IclR-type regulators contain inverted repeats, palindromic and pseudo-palindromic sequences, as well as more complicated repeated patterns, which suggest that the regulators bind DNA as oligomers. The importance of formation of a tetramer (“dimer of dimers”) for AllR’s DNA-binding function was demonstrated by our mutagenesis experiments showing that mutation of residues at tetramerization interface or deletion of portion of it impairs the capability of AllR to bind DNA.

We used size-exclusion chromatography and analytical ultracentrifugation (AUC) to assess the oligomerization state of AllR and to study the effect of the physiological ligand glyoxylate on oligomerization. Our AUC results reveal that AllR is in equilibrium between tetrameric, dimeric and monomeric states, and addition of glyoxylate results in a shift toward the tetrameric state, suggesting stabilization of the tetrameric state when glyoxylate is bound. Previous experiments by Donald et al., who used mass spectrometry, has also shown that AllR, as well as IclR, appear mostly as tetramers, however a smaller fraction appears as dimers, suggesting dimer-tetramer equilibrium (Donald et al., 2001).

Our results are also supported by Guazzaroni and colleagues, who used AUC to study a multidrug efflux regulator TtgV from P.putida that also belongs to IclR family. TtgV binds DNA as a tetramer and dissociate from DNA in the presence of effectors (1-naphthol or 4-nitrotoluene), still as a tetramer (Guazzaroni et al., 2007(a)). The authors
proposed a mechanism for TtgV-DNA interaction, which involves binding of two sets of inverted repeats. They suggested that TtgV first binds one of the sets (IR1) and induces DNA distortion, which allows binding of the second set (IR2) that will stabilize the complex. Using Atomic Force Microscopy they demonstrated that binding of TtgV produces a $57^\circ$ convex bending of DNA, similarly to convex bending that was previously demonstrated for some LacI-GalR family members (Guazzaroni et al., 2007(a)).

Future studies of AllR and IclR interaction with operator DNA, ideally crystal structures of AllR and IclR complexes with DNA, might shed light on mechanism of interaction and the role oligomerization plays in it.

**Ligand binding pocket characterization by site-directed mutagenesis**

I used site directed mutagenesis to characterize the ligand binding pocket of the AllR regulator. The residues that comprise the ligand-binding pocket were suggested based on crystal structure of C-AllR-glyoxylate complex. *In vitro* and *in vivo* studies of AllR mutants identified the residues that play important roles in AllR-glyoxylate interaction, mutation of which has weakened or completely abolished the derepressing effect of glyoxylate on AllR interaction with its DNA operator.

My results are supported by recent findings of the TtgV regulator from *P. putida*. A model of TtgV ligand binding domain was created based on our structures of TM0065 and of the C-terminal domains of IclR family members from *E. coli*. TtgV has 21-29% sequence identity with the proteins we used. Based on this model the authors proposed a location for the ligand binding pocket in TtgV. By mutating some of the amino acids in this pocket they could alter the binding parameters of the protein to its various effectors. Four out of six residues that were mutated in TtgV correspond to residues that are
involved in glyoxylate binding in AllR (Leu-149, His-211, Leu-215, Ser-234), as was shown by our results (Guazzaroni et al., 2007(b)).

These findings speak in favour of our suggestion regarding specificity in ligand recognition. While the ligand binding domains of IclR family members demonstrate very high degree of sequence and structural conservation, the proteins selectively recognize a wide range of structurally and chemically diverse small molecules. The residues that we showed to be involved in glyoxylate binding in AllR are not conserved across the family, suggesting that variation of these residues creates the specific chemical environment in the ligand-binding pocket that adapts it for interaction with the specific ligand.

Mutation of the residues that are involved in interaction of TtgV with its ligands allowed altering the binding parameters (Guazzaroni et al., 2007(b)). Further research is needed for better understanding of the mechanism of protein-ligand interaction specificity in the IclR family. For majority of IclR family members, as well as for other transcriptional regulators, the ligands are not known, thus limiting the possibility of such research. Discovering of specific small molecule effectors that these regulators bind is therefore a priority.

**Search for Physiological Effectors**

Differential light scattering (DLS) utilizes static light scattering for high-throughput screening of proteins for stabilization against thermal denaturation and aggregation (Vedadi et al., 2006). The physiological effectors for most of the IclR family members, as well as for other families of transcription regulators, are unknown. I used DLS in order to search for putative ligands for IclR from *E. coli*. IclR was scanned against a library that included intermediates from the glyoxylate shunt, the Krebs cycle
and other major metabolic pathways. Five compounds stabilized the protein by at least 1.5°C (from the strongest to the weakest): glyoxylate, allantoate, DL-glyceraldehyde-3-phosphate, phosphoenolpyruvate and pyruvate. Interestingly, these molecules have a common element of two alcohol or carboxyl oxygen atoms at vicinal positions.

Out of these candidates two molecules affected DNA operator binding by IclR at physiologically relevant concentrations: glyoxylate (inhibited binding) and pyruvate (stabilized binding). The physiological relevance of the two candidates was confirmed by \textit{in-vitro} and \textit{in-vivo} methods.

Our results show that high throughput screen using static light scattering or fluorescence (Senisterra \textit{et al.}, 2006; Vedadi \textit{et al.}, 2006) can be employed for identification of ligands for other regulators in the IclR family and for regulators from other families which effectors are unknown.

While only for minority of transcription regulators the physiological effectors were identified, identification of new effectors might dramatically change our understanding of transcription regulation in prokaryotes. Not only it will allow elucidation of details of interaction between the proteins and the small molecules, but it also will assist in characterization of proteins and pathways under control of these regulators, as expected in case of MarR-like regulator MexR, which ligand is unknown (Lim \textit{et al.}, 2002), as well as many others. It might as well facilitate characterization of transcription regulators with unknown function – their function and pathways under regulation might be predicted with the help of newly characterized effectors. This in turn will facilitate new genomes annotation (Molina-Henares \textit{et al.}, 2006).
Binding of ligands is known to stabilize proteins and to favor crystallization (Vedadi et al., 2006). In this respect discovery of new ligands with also allow obtaining more 3D crystal structures of transcription regulator-ligand complexes. These structures will expand our understanding of the molecular mechanisms of protein-ligand interaction and of signal transduction and transcription regulation in general. This understanding is of special value in drug-discovery field there the majority of targets are proteins, especially these involved in signal transduction networks (Araujo et al., 2007).

One of the important applications of structural proteomics is discovery of new targets for antimicrobial therapy. As more and more pathogenic strains develop resistance to existing antibiotic treatments, there is increasing need for new therapeutic targets. Such targets have not been subject to natural selection pressure of antibiotics, and therefore are expected not to harbour resistance mutations (Schmid, 2004). Transcription is one of prokaryotic functions currently targeted by antimicrobial therapy, with an anti-tuberculosis drug rifampin that targets RNA-polymerase being the most successful (Villain-Guillot et al., 2007). Research of transcription regulators that unleash pathogenic factors or control multidrug resistance in bacteria might provide promising targets for the development of new anti-invectives. 3D structures of regulators’ complexes with their ligands will assist in structure-guided drug discovery, as potential inhibitors can be synthesised based on chemical scaffolds of the physiological small molecule ligands.

Another interesting application of transcription regulators and their effectors research is for engineering of regulators with novel specificities through mutagenesis. Such “à la carte” regulators might have various uses in biotechnology. For instance, they might be used as environmental biosensors for pollutants. Environmental pollutants could
be detected by fusing the promoters of biodegradative routes for the pollutant of interest to easily measurable reporter genes. Transcriptional regulators activated by the target contaminant interact with the promoter triggering production of the quantifiable marker. Mutants with new specificities have been generated from regulators that belong to XylR/NtrC, LysR, AraC/XylS and TetR families (Galvão and de Lorenzo, 2006). Identification of small molecules that selectively bind to transcription regulators and obtaining 3D structures of regulator-small molecule complexes will provide the basis for structure-guided development of narrower and/or novel specificities.

Another application for optimized/modified transcription regulators can be found in the revolutionary field of bacterial therapy, i.e., using bacteria as therapeutic vectors. The main aim of this field is treating malignant tumours through colonization with genetically engineered bacteria. Organisms such as Clostridium (Wei et al., 2007) or Salmonella typhimurium (Zhao et al., 2005) that are made apathogenic through genetic modifications are being tried. These organisms naturally target and colonize tumour and necrotic tissue. The bacterial are engineered to release active substance inside the tumour, e.g. cytotoxins, enzymatically activated pro-drugs or physiologically active biomolecules (St Jean et al., 2008). Another trend in bacterial therapy is study of lactic acid bacteria as vectors for antigens or drugs transfer to intestinal mucosa for the purpose of immunization or treatment of gastrointestinal diseases (Wells and Mercenier, 2008). Transcription regulators and their effectors can be very useful in bacterial therapy of any type, for construction of new regulatory networks in vector organisms that will allow controlling activation of the therapeutic/cytotoxic systems.
Glyoxylate and pyruvate as the effectors of IclR from E.coli.

In this study I have identified two effector molecules that participate in control of the glyoxylate bypass in *E.coli* by IclR: glyoxylate, which is an intermediate of the pathway under control, and pyruvate, one of glycolysis pathway intermediates. We characterized their interactions with IclR both structurally, by X-ray crystallography, and functionally, *in vitro* and *in vivo*. We suggested that two antagonistic effectors imply regulation on different levels, which reflects the importance of glyoxylate bypass and tight control on it, yet sensitive to environmental and metabolic changes.

Identification of glyoxylate as an effector of IclR answers a question asked earlier by Pellicer and her group (Pellicer *et al.*, 1999). They researched a cross-induction phenomenon between the *aceBAK* operon, regulated by IclR, and the *glc* operon which includes glycerate pathway genes, regulated by glcC. These two pathways involve malate synthase isoenzymes (known as aceB and glcB) that convert glyoxylate to malate. Inactivation of either of the isoenzymes produces glyoxylate accumulation and cross induction. The authors supposed that *aceBAK* is induced by phosphoenol pyruvate, based on a study by Cortay *et al.* (Cortay *et al.*, 1991), which was later disproved by Yamamoto and Ishihama (Yamamoto and Ishihama, 2003). Therefore, they suggested that cross-induction of *aceBAK* following inactivation of glcB could be activated through generation of phosphoenol pyruvate via the D-glycerate pathway from glycolate that in its turn was formed from glyoxylate by reduction (Pellicer *et al.*, 1999). However, our results can provide a simpler explanation to their observation: since glyoxylate was shown to reverse inhibition of *aceBAK* operon by IclR, accumulation of glyoxylate will directly activate *aceBAK*. 
Our results are also supported by recent work by Hasegawa et al. They found that AllR is also affected by two antagonistic regulators: in addition to inactivation of repression by glyoxylate they propose a co-repression function performed by allantoin. Allantoin, which is a product of purine degradation and an intermediate of the pathway controlled by AllR, enhances AllR binding to DNA operators. This effect was overcome by addition of glyoxylate, thus suggesting competition between two effectors for same binding site (similarly to competition between glyoxylate and pyruvate that we found in IclR). The authors also demonstrated that AllR controls the allS gene that encodes AllS, activator of genes for nitrogen assimilation from allantoin. They proposed a model of regulation by two effectors, in which accumulation of co-repressor allantoin redirects allantoin to nitrogen assimilation pathway, however accumulation of inducer glyoxylate switches to utilization of allantoin for energy production (Hasegawa et al., 2008).

Notably, while IclRs effectors, glyoxylate and pyruvate, are structurally very similar, the structure of allantoin is different. Yet, similarly to glyoxylate and pyruvate, it involves two closely spaced carboxyl groups. To characterize the binding of effectors to AllR and to confirm the competitive nature of their binding additional studies, such as Isothermal Titration Calorimetry are needed. A crystal structure of AllR-allantoin complex would be of much interest, as it will demonstrate how the ligand-binding pocket of AllR accommodates this ligand, what residues are involved in binding, and what allosteric changes this binding might induce that possibly contribute to stabilization of AllR-DNA complex.

To summarize, our and other groups’ results together present a picture of two representative repressors from the IclR family, IclR and AllR that control two major
metabolic junctions in \textit{E.coli}. Each regulator’s activity is controlled by two antagonistic small molecule effectors, while one of the effectors, glyoxylate is common for the two proteins and it is also a key intermediate in both pathways. More studies are needed to characterize in detail this interplay of regulatory mechanisms that control on a number of levels the balance of carbon, nitrogen and energy sources in a bacterial cell.
REFERENCES


