2006

The crystal structure of a DNA-binding protein, HucR, from Deinococcus radiodurans R1: a member of the MarR family of winged-helix transcriptional regulators

Hansen Bordelon
Louisiana State University and Agricultural and Mechanical College, hbordel@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
THE CRYSTAL STRUCTURE OF A DNA-BINDING PROTEIN, HUCR, FROM 
DEINOCOCCUS RADIJURANS R1: A MEMBER OF THE MARR FAMILY OF 
WINGED-HELIX TRANSCRIPTIONAL REGULATORS

A Dissertation
Submitted to the Graduate Faculty of the 
Louisiana State University and 
Agricultural and Mechanical College 
in partial fulfillment of the 
requirements for the degree of 
Doctor of Philosophy

In 
The Department of Biological Sciences

by 
Hansen D. Bordelon IV 
B.S., Louisiana State University, 1997 
M.S., Louisiana State University, 2000 
May 2006
DEDICATION

This work is in part dedicated to my parents, Chip and Susan Bordelon, who have in 30 years and counting never let me lose sight of my dreams and have always asked me to reach for the stars. They, along with the rest of my family, have been instrumental in keeping me motivated over this long and often-times tiresome pursuit for academic excellence.

Finally, above all, I dedicate this work to my beloved wife, Randa Carlisle Bordelon. If it wasn’t for her relentless support and encouragement throughout my Ph. D. candidacy, none of this would have been possible. Randa, you have been and will always be the wind beneath my wings.
ACKNOWLEDGEMENTS

It is with my deepest and sincere gratitude that I thank my advisor, Dr. Marcia Newcomer, for introducing me to the field of structural biology, specifically protein crystallography. I will be forever indebted to her unending generosity as she has supported me in every aspect of the word. Although Marcia has mentored several young scientists throughout her fruitful career in academics, I take great pleasure in the fact that I was her first graduate student here at LSU, as I am sure there will be many more to follow. I would also like to thank committee member Dr. Grover Waldrop for providing me the opportunity to work in his laboratory between my graduate school careers and for his persistent encouragement for me to pursue my aspirations of obtaining my Ph. D. Many thanks are also extended to committee members Dr. Anne Grove and Dr. Yong-Hwan Lee for their valuable insight and much needed advice throughout my graduate career. I would also like to thank Dr. Doug Gilman for participating as the Dean’s representative on my committee.

I am extremely thankful to Dr. Svetlana Pakhomova. She has taught me many aspects of x-ray crystallography and had it not been for her guidance, I would not have been able to complete my graduate studies as timely as I did. I would also like to thank Dr. Steven Wilkinson from the Grove lab for introducing me to HucR and providing me with the initial protein to start the HucR structure determination project. I also thank my current and former lab mates: Dr. Patrick Bilder, Dr. Michael Oldham, and Pratik Dhar for many valuable discussions that have aided me in the completion of this project, as well as a wonderful work environment.
I would like to thank Dr. Henry Bellamy and Dr. David Neau at CAMD for their help and assistance with MAD and native data collection.

Finally, I would like to thank those who have funded this work including the Louisiana Governors Biotechnology Initiative (MEN) and the LSU AgCenter (TB).
# TABLE OF CONTENTS

DEDICATION .................................................................................................................. ii

ACKNOWLEDGEMENTS .................................................................................................... iii

LIST OF TABLES .............................................................................................................. vi

LIST OF FIGURES .......................................................................................................... vii

ABSTRACT ....................................................................................................................... viii

CHAPTER 1. REVIEW OF LITERATURE ................................................................. 1
   Transcriptional Regulators and the HHelix-turn-helix DNA Binding Motif .......... 1
   The Winged Helix DNA Binding Motif ................................................................. 4
   Multidrug Resistance and the MarR Family of Transcriptional Regulators ....... 7
   Hypothetical uricase regulator (HucR) ................................................................. 9
   Structural Analysis of the MarR Family .............................................................. 11
     MarR ................................................................................................................ 14
     MexR .............................................................................................................. 16
     OhrR .............................................................................................................. 18

CHAPTER 2. THE CRYSTAL STRUCTURE OF HUCR: A MEMBER OF
   THE MARR FAMILY OF TRANSCRIPTIONAL REGULATORS .......... 23
   Materials and Methods ....................................................................................... 23
     Protein Expression and Purification .......................................................... 23
     Crystallization and Data Collection .......................................................... 25
     Structure Determination and Refinement .................................................. 27
     Data Deposition ......................................................................................... 32
   Results and Discussion ............................................................................... 32
     Overall Structure of HucR ......................................................................... 32
     The DNA Binding Domain ....................................................................... 34
     The Dimerization Domain ....................................................................... 37
     HucR-DNA Interactions ............................................................................ 40

CHAPTER 3. CONCLUDING REMARKS ................................................................ 45

REFERENCES .............................................................................................................. 48

VITA .............................................................................................................................. 53
LIST OF TABLES

1. Summary of the MarR family of transcriptional regulators...........................................12
2. Summary of data collection statistics............................................................................29
3. MAD phasing and structure refinement statistics..........................................................30
LIST OF FIGURES

1. Canonical helix-turn-helix DNA binding motif complexed with DNA………………..3
2. Winged helix DNA binding motif........................................................................6
3. The uric acid degradation reaction catalyzed by the enzyme uricase...............11
4. The crystal structure of MarR and its salicylate binding sites..............................17
5. Conformational flexibility observed between MexR dimer AB and dimer CD……20
6. Superposition of the reduced and DNA-bound OhrR dimers............................21
7. The crystal structure of OhrR complexed with DNA.........................................22
8. Purified native and SeMet HucR........................................................................25
9. Crystals of native and SeMet HucR....................................................................26
10. Native HucR diffraction image......................................................................28
11. X-ray fluorescence scan for SeMet crystals......................................................29
12. Patterson map.................................................................................................30
13. Electron density from initial phases...............................................................31
14. Representative 2F_o – F_c electron density map...........................................32
15. Overall structure of HucR................................................................................33
16. Structure based sequence alignment between HucR, MarR, MexR, and OhrR....36
17. Hydrophobic interactions of the DNA binding domain....................................37
18. Imidazole ring stacking at the HucR dimerization interface..............................40
19. Superposition of the HucR dimers..................................................................41
20. Superposition of HucR dimer on OhrR-DNA bound dimer and DNA...........43
21. Illustration of B-factors for HucR......................................................................44
ABSTRACT

The MarR family of transcriptional regulators is an important group of prokaryotic DNA binding proteins. As the family name implies, multiple antibiotic resistance, members of the MarR family often regulate the expression of resistance genes to multiple antibiotics, organic solvents, household disinfectants, detergents, and oxidative stress agents. Most MarR members act as transcriptional repressors and exist as homodimers in both free and DNA-bound states. DNA-binding is mediated via a winged-helix fold and is often relieved by anionic lipophilic ligands. *Deinococcus radiodurans* R1 was found to encode a 181 residue MarR homolog designated HucR (hypothetical uricase regulator). Biochemical evidence has shown that HucR negatively regulates expression of uricase and this repression is attenuated by the binding of uric acid, which is the natural substrate for uricase. In this study we present the crystal structure of HucR determined to 2.3 Å in the absence of ligand. In addition, a second crystal form of HucR was determined to 2.9 Å in which three dimers were observed in the asymmetric unit. Unlike the crystal structure of the MarR homolog, MexR, HucR does not display large conformational heterogeneity between dimers. Furthermore, superpositioning of the HucR dimer with the crystal structure of the OhrR dimer complexed with DNA suggests that HucR is in a “DNA ready” confirmation in which the lobes of the DNA binding domains are in a position compatible with DNA binding, with the exception of minor localized conformational changes needed at the amino termini of the recognition helices. This is in contrast to what is observed when comparing the crystal structures of the DNA-bound and unbound OhrR, in which there is a significant
displacement of the DNA binding domains as a result of conformational changes that originate at the dimerization interface. The crystal structure of HucR in the absence of either ligand or DNA suggests that HucR is likely to be fixed in a “DNA ready” conformation. Thus, the crystal structure of HucR has given new insight into the MarR family of transcriptional regulators, proving that although these family members share similar structural folds, their mechanisms for transcriptional regulation are likely very specialized.
CHAPTER 1

REVIEW OF LITERATURE

Transcriptional Regulators and the Helix-turn-helix DNA Binding Motif

Transcription is a very intricate cellular process whether it occurs in a prokaryotic or eukaryotic cell. Eukaryotic transcription is much more complex and far less understood compared to that of transcription in prokaryotes. Eukaryotes often require a multitude of trans-acting factors in addition to RNA polymerase in order to achieve successful transcript production. In either case, transcriptional regulators (i.e. transcription factors or trans-acting proteins) are DNA binding proteins that play a vital role in modulating gene expression within a cell. In fact, gene expression is primarily controlled at the level of transcription. Transcriptional regulators modulate gene expression by either being activators that up-regulate certain genes or repressors which down-regulate (i.e. repress) gene expression, or in some cases, they can do both. Interestingly, *E. coli* encodes for a higher proportion of transcriptional repressors than activators with 36% being repressors, 30% being activators, and 20% with dual functions (1). Moreover, transcription factors can be classified as either local or global regulators depending on the range of genes that they regulate.

Transcriptional regulators have been placed into one of several groups primarily based on the structure of their DNA binding motifs. Some of these groups include the helix-turn-helix (HTH), zinc finger, leucine zipper, helix-loop-helix (HLH), basic region/leucine zipper (b/zip), basic region/helix-loop-helix/leucine zipper (b/HLH/zip),
homeodomain, β-sheet, and β-ribbon/hairpin proteins \(^2,3\). NMR and x-ray crystal structures of prokaryotic transcription factors have revealed three principal DNA binding motifs: the HTH, winged helix, and the β-ribbon motif \(^4\).

The helix-turn-helix DNA binding motif is the most common DNA binding structure in bacterial and archaeal transcription factors \(^1,4\). The classical HTH DNA binding motif was first structurally described in the crystal structures of bacteriophage \(\lambda\) Cro protein \(^5,6\) and \(E.\ coli\) catabolite activator protein (CAP) \(^7,8\). The canonical HTH motif consists of two consecutive helices (H2 and H3) that are packed at angles of 120º relative to each other and are connected by a tight four residue turn in which a glycine is usually found at the second position (Fig 1). Apparently the HTH motif alone is insufficient to undergo independent folding and requires an additional α-helix (H1) to render this motif as a compact, globular domain \(^4\), which is often collectively referred to as the DNA binding domain that typically ranges in size from 60 to 90 amino acids \(^9\). The second α-helix of the HTH motif, H3, is often referred to as the “recognition helix” and averages nine residues in length. The recognition helix, as its name implies, is important for making most of the sequence specific contacts with its cognate DNA \(^10\). This is typically done by insertion of the recognition helix into the major groove of the DNA, thus making critical base pair contacts that result in sequence specific DNA binding as illustrated in Fig 1. In fact, each of the four base pairs at the major groove provide a distinctive hydrogen bonding pattern (like a fingerprint) often referred to as direct readout, whereas the minor groove does not provide such a distinctive pattern \(^11\).

Nearly all prokaryotic proteins that bind DNA via the HTH motif function as homodimers and use their two-fold dimer related HTH motifs to bind their cognate
palindromic or pseudopalindromic DNA sequence which are also related by 2-fold symmetry. However, there is not an absolute requirement for dimerization amongst all HTH DNA binding proteins as demonstrated by the fact that some members of the AraC family, including *E. coli* MarA (12) and Rob (13), bind their cognate DNA as monomers. Both MarA and Rob have two HTH motifs per monomer. However, MarA uses both of its HTH motifs to make contacts with the major groove, whereas Rob only uses one HTH motif for major groove recognition and its other HTH motif contributes only DNA backbone contacts. Thus, it is difficult to speculate on the exact mechanism of protein-DNA interactions simply based on homology modeling in the absence of specific structural data.
The Winged Helix DNA Binding Motif

Since the original identification of the HTH DNA binding motif in the Cro and CAP proteins from bacteriophage λ and *E. coli*, respectively, structural studies on a variety of gene regulatory proteins have led to the identification of several variations of the canonical HTH motif. In fact, variations in this motif are quite common, with HTH variants occurring in both prokaryotic and eukaryotic gene regulatory proteins. Further structural disparity within this superfamily came when the co-crystal structure of the DNA binding domain of eukaryotic hepatocyte nuclear factor-3 (HNF-3) complexed with its target DNA (14) revealed a significant structural adaptation to the canonical HTH DNA binding motif. This new DNA binding element was called a “winged HTH motif” (also referred to as a winged helix motif) and was the basis for a new subclassification of the canonical HTH DNA binding proteins.

The winged helix fold has been identified in eukaryotes, prokaryotes, archaea, and viruses. Interestingly, sequence analysis of the DNA binding domains of archaeal proteins show that all archaea encode a large number of proteins that contain a HTH motif. The sequences of these motifs are more closely related to those of prokaryotic HTH domains than those of eukaryotes. Furthermore, the winged helix motif appears to be the predominant motif used for DNA binding proteins within the archaeal. (15).

The winged helix domain (Fig 2) is an adaptation of the canonical HTH motif with a compact α/β structure that has the following topology: H1-S1-H2-T-H3-S2-W1-S3-W2, where “H” denotes an α-helix, “S” denotes a β-strand, “T” denotes a turn, and “W” denotes a loop (16). As the name implies, the most striking feature of this class of DNA binding proteins is the presence of a structural element referred to as a “wing” that
is actually a β-hairpin that immediately follows the HTH motif with H2 and H3 of the winged helix motif corresponding to helices of the canonical HTH motif. The two wings, W1 and W2, flank the recognition helix, H3, similar to the wings of a butterfly (hence the name, winged helix motif) (16). Although W1 is always present in members of the winged-helix family of DNA binding proteins, W2 is sometimes absent, as observed in the structures of histone H5 (17), E2F, DP2 (18) and MarR (19). Besides the presence of the wing, winged helix proteins often deviate from the canonical HTH motif by the length of the turn region that separates H2 and H3. The turn varies among winged helix proteins but is typically longer than that of the 3-4 residues from the canonical HTH fold. For example, a 10 residue turn separates H2 and H3 in the crystal structure of OmpR (20). As a consequence of the extra residues in the turn, the steric hindrance between H2 and H3 is reduced, allowing for a greater variation in the angles between the two helices (100° to 150°) than that of the canonical HTH (120°) (16,4).

For canonical winged helix DNA binding proteins, the DNA recognition helix (H3) is primarily responsible for mediating most, but not all, of the sequence specific contacts. However, the wings seem to differ in their contributions to either sequence-specific DNA recognition or nonspecific protein/DNA stabilization via phosphate backbone contacts. The NMR structure of MuR from bacteriophage Mu in complex with its cognate DNA reveals that it uses its wing to make sequence specific contacts as its wing is bound to the minor groove where it makes direct hydrogen bonds to the nucleotide bases (22). Moreover, a comparison of the structures of unbound MuR (23) with that of DNA-bound MuR reveals that the wing undergoes a disorder-to-ordered transformation and becomes stabilized in the minor groove of the DNA (4). Strikingly,
Figure 2. **Winged helix DNA binding motif.** The winged helix DNA binding motif of MecI from *Staphylococcus aureus* in complex with a 24 bp dsDNA with a 1 bp overhang on either side. Incorporated within this oligonucleotide is the sequence of the *blaZ* dyad that repressor MecI binds. The recognition helix, H3, is embedded in the major groove of the DNA while the wing (yellow β-hairpin) makes only minor groove backbone contacts (21).
the co-crystal structure of the RFX1-DNA complex (24) revealed that the wing of RFX1 (W1) makes the majority of its contacts with its cognate DNA by inserting its wing into the major groove. Furthermore, the recognition helix of RFX1 sits over the minor groove where it only contributes a single DNA contact via a lysine sidechain and a cytosine. Thus, although the winged-helix HTH proteins share a common structural motif, there is significant variation in how this motif interacts with DNA.

**Multidrug Resistance and the MarR Family of Transcriptional Regulators**

Multidrug resistance (MDR) is a growing worldwide health issue, as various human pathogens are acquiring the ability to survive ordinarily lethal doses of structurally diverse drugs (25). The MDR phenotype can arise via several resistance mechanisms including the action of efflux transporters that actively transport a wide variety of structurally and chemically dissimilar compounds out of the cell (26). The crucial role these drug efflux transporter systems play in MDR has been well documented in eukaryotes as well as prokaryotes (27). Multidrug transporters are generally integral membrane proteins that are often regulated at the transcriptional level by repressors and/or activators. Multidrug pumps are usually upregulated in response to the presence of the natural substrates on which these pumps act, usually as a direct result of the regulators themselves binding to the same diverse drugs as the substrates for the pumps that they regulate (28,29,30). Due to the inherent difficulty of crystallizing integral membrane proteins, there is no structural data to help explain how these transporters recognize such a vast array of structurally diverse compounds. However, since the soluble cytosolic regulatory proteins of these transporters also recognize a similar range
of drugs, structural studies with transcription factors complexed with various drugs might help elucidate the mechanism by which these pumps recognize and bind multiple drugs.

Although the focus here is on prokaryotic MDR, it is important to mention that the development of MDR in tumor cells is also a serious concern and a major cause of failure in anti-tumor chemotherapy (31). Moreover, MDR in tumor cells is usually the result of an upregulation of the expression of a specific group of broad substrate spectrum membrane phosphoglycoproteins belonging to the ABC transporter superfamily (31). Thus, the phenomena of MDR must be addressed at both the prokaryotic and eukaryotic levels: to help combat emerging human pathogenic bacteria and to facilitate a better efficiency for tumor cell uptake of anticancer drugs, respectively.

The E. coli AcrAB-TolC multidrug efflux system functions as a tripartite complex that gives rise to a MDR phenotype which is also referred to as a mar phenotype (multiple antibiotic resistance) (32) The mar phenotype includes resistance to structurally diverse antibiotics, organic solvents, household disinfectants, and oxidative stresses (33,34,35). The mar phenotype in E. coli is directly associated with the expression of the marRAB operon, specifically expression of the activator protein, MarA, that activates gene expression of the AcrAB-TolC multidrug efflux pump (35,36,37,38). However, in the absence of the appropriate stimulus, MarR negatively regulates the marRAB operon (hence the multi drug efflux system), with repression being alleviated by the presence of a variety of mostly phenolic-like compounds, most notably salicylate (39,40).

Since the discovery of repressor MarR, several homologs have been identified and grouped into a common family, the MarR family of transcriptional regulators, named after the prototypical member MarR (Table 1). Proteins that belong to the MarR family
often act as environmental sensors by modulating gene expression by a rapid response to changes in environmental conditions. This is frequently achieved by binding to the same or oftentimes similar substrates as those for the gene products that they regulate. For example, the MarR homolog, HucR, is responsible for repression of the uricase gene in *Deinococcus radiodurans*. However, in the presence of uric acid which is the substrate for uricase, HucR-mediated repression of uricase is alleviated (41). Furthermore, most MarR homologs are repressors of the genes to which they regulate with the exception of BadR (42), NhhD (43), and ExpG (44) that play roles in activating gene expression, and SlyA that appears to have dual roles as both a repressor and an activator (45). Members of the MarR family mediate DNA binding via a winged helix DNA binding motif. Moreover, MarR homologs exist as homodimers in both free and DNA-bound states, with DNA binding occurring at either palindromic or pseudopalindromic sequence specific sites. Finally, several members of the MarR family have been shown to respond to various anionic lipophilic compounds (usually phenolic-like ligands), usually resulting in an attenuation of an ability to bind the cognate DNA sequence.

**Hypothetical uricase regulator (HucR)**

The eubacterial family *Deinococcaceae* contains some of earth’s most radiation-resistant organisms ever described. Of the several radiation-resistant species of *Deinococcus* identified to date (72), the most extensively studied to date is *Deinococcus radiodurans* R1, for which the complete genome has been sequenced (73). *D. radiodurans* is a nonpathogenic, gram positive, non-motile, aerobic mesophile best known for its remarkable ability to withstand both acute and chronic exposures to high levels of ionizing radiation that often leads to double-strand DNA breaks. *D.*
*radiodurans* is also highly resistant to other sources of DNA damage including UV radiation, desiccation, and oxidative stress (72). Moreover, *D. radiodurans* was found to encode orthologs to almost all known genes involved in stress responses in bacteria including pH, desiccation, temperature, phage, starvation, toxin, antibiotic, and oxidative stresses (72,73).

A gene designated *hucR* (hypothetical uricase regulator) encoding for a 181 amino acid protein from *D. radiodurans* was previously identified and characterized and is believed to play a critical role in the cellular response to oxidative stress (41,63). HucR shares 29% amino acid sequence identity with prototypical member MarR, thus placing HucR in the MarR family of transcriptional regulators. The biochemical characterization of HucR revealed that HucR is a transcriptional regulatory protein that binds as a homodimer to a single site located within its own promoter/operator region (*hucO*). The HucR binding site within *hucO* contains a pseudopalindromic sequence of 8-bp half sites separated by 2 bp, thus each HucR subunit of the dimer binds the same face of the DNA (41). Furthermore, *hucO* is also located within the promoter region of a separate gene encoding for uricase that is oriented in the opposite direction relative to *hucR*. Thus, this partial overlapping of promoter regions in which the transcriptional start site of each gene is located within the Pribnow box of the other provides for a genetically favorable mechanism for the simultaneous co-repression of both HucR and uricase. The biological substrate for uricase is uric acid, which is converted to allantoin during purine catabolism (Fig. 3) (74). Interestingly, *in vitro* studies have shown that uric acid appears to be the natural ligand for HucR as it binds HucR with an apparent *K_*$_d$ of 5.9 µM. A concomitant antagonistic effect on HucR-DNA
binding is observed upon HucR-uric acid binding. In fact, HucR-DNA complex formation is reduced by 50% at uric acid concentrations of 260 µM. Furthermore, *in vivo* studies have shown that when *D. radiodurans* was grown in the presence of excess uric acid, elevated levels of *hucR* and uricase transcripts were seen as well as increased uricase activity (41). Taken together, these data strongly support the identification of a novel regulatory mechanism for maintaining uric acid homeostasis in *D. radiodurans* in which HucR-mediated repression of uricase is relieved directly by uric acid.

![Figure 3. The uric acid degradation reaction catalyzed by the enzyme uricase.](image)

**Figure 3.** The uric acid degradation reaction catalyzed by the enzyme uricase.

**Structural Analysis of the MarR Family**

Currently there are four structures of members of the MarR family of transcriptional regulators including: MarR (19), MexR (68), SlyA-like protein (65), and OhrR (54). Although this family shares very little amino acid sequence identity, thus far, they appear to adopt very similar folds. Each structure has contributed its own unique information about this family of proteins while emphasizing similarities within the family. All structures have shown that they exist as homodimers with each subunit
<table>
<thead>
<tr>
<th>Organism</th>
<th>Repressor/Activator</th>
<th>Ligands</th>
<th>Footprint (bp)</th>
<th>(bp in ½ sites)</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FarR Neisseria gonorrhoeae</td>
<td>Repressor</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n</td>
<td>46</td>
</tr>
<tr>
<td>HpaR Escherichia coli</td>
<td>Repressor</td>
<td>4-hydroxyphenylacetic acid, 3-hydroxyphenylacetic acid, 3,4-hydroxyphenylacetic acid</td>
<td>27</td>
<td>9 (4)</td>
<td>n</td>
<td>47</td>
</tr>
<tr>
<td>HpcR Escherichia coli</td>
<td>Repressor</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n</td>
<td>48, 49</td>
</tr>
<tr>
<td>BadR Rhodopseudomonas palustris</td>
<td>Activator</td>
<td>Benzoate, 4-hydroxybenzoate</td>
<td>n/a</td>
<td>n/a</td>
<td>n</td>
<td>42</td>
</tr>
<tr>
<td>OhrR Xanthomonas campestris</td>
<td>Repressor</td>
<td>tert–butyl hydroperoxide, cumene hydroperoxide</td>
<td>44</td>
<td>8 (17)</td>
<td>n</td>
<td>50, 51</td>
</tr>
<tr>
<td>OhrR Bacillus subtilis</td>
<td>Repressor</td>
<td>tert–butyl hydroperoxide, cumene hydroperoxide</td>
<td>42b</td>
<td>7 (7)</td>
<td>y</td>
<td>52, 53, 54</td>
</tr>
<tr>
<td>CinR Butyrivibrio fibrisolvens</td>
<td>Repressor</td>
<td>Cinnamic acid sugar esters</td>
<td>8 (0)</td>
<td></td>
<td>n</td>
<td>55</td>
</tr>
<tr>
<td>MarR Escherichia coli</td>
<td>Repressor</td>
<td>salicylate, plumbagin, 2,4–dinitrophenol, menadione</td>
<td>21</td>
<td>5 (2)</td>
<td>y</td>
<td>39, 56, 19</td>
</tr>
<tr>
<td>EmrR Escherichia coli</td>
<td>Repressor</td>
<td>Nalidixic acid, salicylate caronyl cyanide m-chlorophenylhydrazone 2, 4–dinitrophenol, ethidium bromide</td>
<td>42</td>
<td>9 (3)</td>
<td>n</td>
<td>57, 58</td>
</tr>
</tbody>
</table>

Table cont
<table>
<thead>
<tr>
<th>Organism</th>
<th>Repressor/Activator</th>
<th>Ligands</th>
<th>Footprint (bp)</th>
<th>(bp in ½ sites)</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PecS Erwinia chrysanthemi</td>
<td>Repressor</td>
<td>n/a</td>
<td>45</td>
<td>5 (3)</td>
<td>n</td>
<td>59, 60, 61, 62</td>
</tr>
<tr>
<td>HucR Deinococcus radiodurans</td>
<td>Repressor</td>
<td>uric acid, salicylate</td>
<td>21&lt;sup&gt;e&lt;/sup&gt;</td>
<td>8 (2)</td>
<td>y</td>
<td>41, 63</td>
</tr>
<tr>
<td>SlyA Salmonella typhimurium</td>
<td>Repressor/Activator</td>
<td>n/a</td>
<td>25&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5 (2)</td>
<td>n</td>
<td>45, 64</td>
</tr>
<tr>
<td>SlyA Enterobacter faecalis</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>y</td>
<td>65</td>
</tr>
<tr>
<td>HcaR Acinetobacter sp. Strain ADP1</td>
<td>Repressor</td>
<td>hydroxycinnamoyl-CoA thioesters</td>
<td></td>
<td></td>
<td>n</td>
<td>66</td>
</tr>
<tr>
<td>MexR Pseudomonas aeruginosa</td>
<td>Repressor</td>
<td>n/a</td>
<td>28</td>
<td>5 (5)</td>
<td>y</td>
<td>67, 68</td>
</tr>
<tr>
<td>CbaR Comamonas testosterone BR60</td>
<td>Repressor</td>
<td>3-chlorobenzoate, protocatechuate</td>
<td>22</td>
<td>4 (6 or 9)</td>
<td>n</td>
<td>69</td>
</tr>
<tr>
<td>NhhD Rhodococcus rhodochrous J1</td>
<td>Activator</td>
<td>n/a</td>
<td></td>
<td>n/a</td>
<td>n</td>
<td>43</td>
</tr>
<tr>
<td>Hpr Bacillus subtilis</td>
<td>Repressor</td>
<td>n/a</td>
<td></td>
<td>n/a</td>
<td>n</td>
<td>70</td>
</tr>
<tr>
<td>ExpG Sinorhizobium meliloti</td>
<td>Activator</td>
<td>n/a</td>
<td>21&lt;sup&gt;e&lt;/sup&gt;</td>
<td>6 (3)</td>
<td>n</td>
<td>71, 44</td>
</tr>
</tbody>
</table>

<sup>a</sup> The smaller DNase I footprint size is indicated when footprint sizes are reported for both strands; footprint size is reported for sequences with lowest Kᵩ; <sup>b</sup> Values reported involve cooperative binding to DNA site containing adjacent perfect and imperfect repeats; <sup>c</sup> Methidiumpropyl-EDTA-Fe(II) footprint; <sup>d</sup> Footprint of a predicted high affinity site within a DNA sequence containing 5 predicted binding sites; <sup>e</sup> A conserved sequence found in three ExpG binding regions; <sup>f</sup> Number of base pairs for each half-site (number of base pairs between half sites in parentheses). This table was modified from Wilkinson et. al, 2006 (86).
consisting of two independent domains: 1) a globular DNA binding domain displaying a
winged helix motif and 2) a dimerization domain made up of the amino and carboxyl
termi of the protein. Extensive hydrophobic interactions appear to play a crucial role in
subunit association. Furthermore, dimerization requires that the equivalent region of
each dimerization domain intertwine to form the complete functional dimerization
domain with extensive surface area buried at the interface.

MarR

The first structure for the MarR family was the crystal structure of MarR that was
determined to 2.3 Å resolution in the presence of its ligand, salicylate (19). It revealed a
dimer having a pyramidal shape with overall dimensions of 50 X 55 X 45 Å³ (corresponding to the width, height, and depth relative to the orientation shown in Fig. 4a). The overall topological order of secondary structure in MarR is H1-H2-β1-H3-H4-
β2-W1-β3-H5-H6 with the winged helix DNA binding domain consisting of β1-H3-H4-
β2-W1-β3 (amino acids 55-100) and H1, H2, H5, and H6 making up the dimerization
domains. Furthermore, aside from being the first structure to represent this family of
proteins, the MarR structure revealed the binding sites of its ligand, salicylate.

In order to get well diffracting crystals suitable for structure determination, an
extremely high concentration of salicylate (250 mM) was used to co-crystallize MarR in
the presence of the ligand. Two salicylate binding sites per monomer were identified and
called Sal-A and Sal-B (Fig. 4bc). These sites are solvent exposed and are located on
each side of the recognition helix (H3) at the surface of the molecule. Site Sal-A is the
more conserved of the two sites with interactions taking place between residues from
both the wing and the recognition helix. Specifically, the guanidinium side chain of R86
from the wing and the hydroxyl side chain of T72 from the recognition helix make
hydrogen bonds with the salicylate carboxylate and hydroxyl groups, respectively (Fig. 4c). Further stabilization of site Sal-A comes from the pyrrolidone ring positioned within 3.5 Å of the salicylate ring. Site Sal-B differs in that all interactions with salicylate are mediated only from the recognition helix. The salicylate carboxylate makes a hydrogen bond to R77 while the backbone carbonyl of A70 makes a H-bond to the salicylate hydroxyl group (Fig. 4c). Furthermore, similar to what is seen for Sal-A, the hydrophobic side chain from M74 sits just above the salicylate ring, thus stabilizing the ligand binding pocket via hydrophobic interactions. The proximity of the ligand binding sites to the winged helix motif suggests a complete interlinkage of DNA and ligand binding regions.

Although the co-crystal structure of MarR reveals sites Sal-A and Sal-B as binding sites of salicylate, one must be reminded that MarR has an apparent $K_d$ for salicylate of 0.5-1 mM (56) and crystallization took place at a markedly high concentration of salicylate, 250 mM. Furthermore, site Sal-A is involved in the crystal packing of symmetry-related molecules throughout the crystal lattice. Thus, it is not possible to conclude whether these sites were an artifact of crystallization or true ligand binding sites of MarR. However, if these sites are the physiological binding sites for salicylate, it appears rather straightforward that the mechanism for attenuation of DNA binding as a result of MarR binding salicylate is steric hindrance. Binding two molecules of salicylate on each side of the recognition helix should inevitably prevent the recognition helix from recognizing and binding its cognate DNA sequence. Furthermore, the confirmation in which MarR crystallized was not in a confirmation suitable for DNA binding, as the spacing of the recognition helices were not compatible with both monomers making contacts within the major grooves of the DNA.
MexR

The second crystal structure for this family was that of MexR determined to 2.1 Å resolution (68). MexR is a repressor of the MexAB-OprM operon that encodes a tripartite multidrug efflux system in *Pseudomonas aeruginosa* (67). Furthermore, mutations in MexR result in the upregulation of the MexAB-OprM operon, thus conferring MDR, as evidenced by increased resistance to antimicrobials that include β-lactams, tetracycline, chloramphenicol, novobiocin, trimethoprim, and sulfonamides (75-88).

Repressor MexR was crystallized in the absence of an effector molecule (the physiological ligand for MexR has yet to be identified). However, fortuitously, in the absence of an effector molecule, four independent observations of the dimer were made within the asymmetric unit, thus revealing the plasticity of these proteins as indicated by the varied conformations for each dimer observed. This is markedly different than what was observed for MarR in which only a single monomer was observed in the asymmetric unit, with its dimer mate occurring on the crystallographic 2-fold, thus indicating relatively minimal flexibility, perhaps as a result of bound ligand. One particular conformation found in dimer CD indicates a dimer in an “open” state that can be docked onto B-form DNA with a “reasonable” good fit, with the largest spacing between the two recognition helices from each DNA binding domain with a Cα-Cα distance of 29.2 Å (measured from Arg-73 to Arg-73’, where prime denotes the other subunit). The spacing between major grooves in linear B-form DNA is approximately 34 Å. Another MexR conformation referred to as dimer AB was described as being in the “closed” conformation, unable to be docked onto DNA, with a short spacing between its recognition helices of 22.6 Å, far short of the 34 Å needed to reach successive major
Figure 4. **The crystal structure of MarR and its salicylate binding sites.** (A), Ribbon representation of the MarR dimer with each subunit represented by a different color. (B) The MarR monomer shown with its recognition helix (orange) and bound salicylate molecules (magenta) at sites Sal-A and Sal-B. (C) A stereo image of the salicylate binding sites Sal-A and Sal-B with salicylate molecules (magenta) and contact residues labeled from one monomer of MarR (19).
grooves on B-form DNA. Using the DNA binding domains from one monomer to superpose dimer AB onto dimer CD (Fig. 5), shows just how substantial the difference is in the “on” state (can bind DNA) compared to the “off” state (can not bind DNA). Interestingly, the structure of dimer AB shows the C-terminal tail (residues 140-147) of monomer C positioned between the lobes of the DNA binding domains of dimer AB resulting in a shortening of the distance between the recognition helices. Thus, it has been suggested by Lim et al. that the uncharacterized ligand of MexR might mimic that which is seen in dimer AB, where a ligand binds to MexR between the DNA binding domains resulting in a neutralization of the electrostatic repulsions of these domains that usually exist to keep the lobes at optimal spacing to facilitate DNA binding. Again, as in the MarR crystal structure, it is always possible that this observation is just an artifact of crystal packing and has no physiological relevance.

OhrR

Very recently, the crystal structure of a MarR family repressor, OhrR, from Bacillus subtilis was solved in the absence and presence of DNA to resolutions of 2.5 and 2.64 Å, respectively (54). Upon exposure to oxidative stress, a conserved cysteine (C15) is oxidized to Cys-sulphenic acid which in turn causes a conformational change within the dimerization domain which is propagated to the DNA binding domain that ultimately results in the inability of OhrR to bind the ohrA operator, resulting in the derepression and upregulation of OhrA that encodes an organic hydroperoxide resistance protein (52,53). In fact, a comparison of the unbound (i.e. reduced OhrR) and DNA bound structures of OhrR reveals a root mean square deviation (rmsd) of 1.5 Å to 1.7 Å, with differences primarily confined to the DNA binding domain. Furthermore, in order for the unbound OhrR structure to assume the conformation of the DNA-bound form, the DNA
binding domain of one of the subunits must rotate ~25º relative to the other monomer and
the wing must translocate ~16 Å (as measured from the Ca of E91’) (Fig. 6).

A 29 bp oligonucleotide containing the ohrA operator sequence was used to
crystallize the OhrR-DNA complex. The recognition helices (H4 and H4’) of the winged
helix motif bind consecutive major grooves of the ohrA operator with the contacts
primarily mediated by the N-termini of these helices (Fig. 7). A total of 44 residues of
the OhrR dimer were found to make 60 contacts with the DNA spanning over 22
nucleotides. Moreover, each subunit of the dimer exhibits almost identical 2-fold related
interactions with the DNA. The crystal structure revealed that the bound DNA is bent ~
10º globally while the major grooves are substantially widened (from 11 Å to 15.5 Å) and
depened (from 4.0 Å to 6.0 Å) as a result of the insertion of the recognition helix (H4).
The wing from the winged helix DNA binding motif is inserted directly into the minor
groove resulting in localized conformational change of the DNA (overtwisting of a DNA
bp). Wing residues R86, D92, and R94, which are highly conserved throughout the
MarR family, perform the following functions in ohrA binding: R86 makes nonspecific
phosphate backbone contacts, D92 helps coordinate the correct positioning of the
guanidinium group of R94 through side chain electrostatic interactions, and R94 makes
specific contacts with the nucleobase thymine. Finally, DNA binding is further stabilized
by several H-bonds to the phosphate groups along the phosphate backbone of the minor
groove from residues on H2, and most strikingly, H1. Thus, it appears that helix H1 not
only plays a role in dimerization but also contributes to DNA binding.
Figure 5. **Conformational flexibility observed between MexR dimer AB and dimer CD.** Dimers AB (blue) and CD (red) were superposed onto one another relative to their DNA binding domain from one monomer. The recognition helices are shown in thicker rendering with cyan corresponding to dimer AB and brown corresponding to dimer CD. The distances between the recognition helices are measured from the Cα carbons of Arg-73 to Arg-73′. This figure is an adaptation of Fig. 2b from Lim et al, 2002.
Figure 6. **Superposition of the reduced and DNA-bound OhrR dimers.** A superposition of the dimers of reduced OhrR (cyan) and DNA-bound OhrR (red) was done using Cα carbons from one monomer. The curved and straight orange arrows indicate the rotation angle and wing translocation of the nonsuperposed subunit of the DNA-bound OhrR relative to the reduced OhrR, respectively. The bottom figure is identical to the top figure, except for its rotation 90º about the x-axis.
Figure 7. The crystal structure of OhrR complexed with DNA. OhrR is shown as a ribbon representation bound to the 29 bp ohrA operator depicted in stick format. The recognition helices H4 and H4’ are embedded in the major groove of the DNA with the wings making phosphate backbone contacts.
CHAPTER 2

THE CRYSTAL STRUCTURE OF HUCR: A MEMBER OF THE MARR FAMILY OF TRANSCRIPTIONAL REGULATORS

Materials and Methods

Protein Expression and Purification

For expression of native HucR protein, E. coli BL21(DE3) pLysS was transformed with pSPW1 (41) in which the gene encoding HucR was cloned into the NdeI/EcoRI sites of the pET-5a expression vector. A single colony was picked from a plate and incubated overnight in Terrific Broth (TB) containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. A 1:100 dilution of this culture was made using fresh TB containing ampicillin and chloramphenicol and allowed to incubate at 37°C with constant shaking (225 rpm) until the A_{600} reached 0.6 and isopropyl-ß-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM to induce HucR expression. Upon induction, the temperature was reduced to 20 °C and the cells were allowed to incubate for an additional 24 h with continuous shaking. Cells were pelleted by centrifugation at 5,500 X g for 10 min at 4 °C followed by immediate freezing at -80 °C. Cells were lysed with Bugbuster (Novagen) followed by the addition of DNase I and protease inhibitors pepstatin and leupeptin. Additionally, the cell lysate was sonicated twice in 5 min intervals to ensure complete cell lysis and to help solubilize HucR protein. All subsequent purification procedures were carried out at 4 °C. The crude lysate was centrifuged at 11,000 X g for 30 min with the resulting clarified lysate applied to a 10 ml Affi-Gel Blue Gel (Biorad) column equilibrated with Buffer A (10 mM Hepes, pH 7.0,
50 mM KCl, and 2 mM β-Mercaptoethanol) in which HucR was allowed to bind. The column was then washed with 10 column volumes of Buffer A followed by the elution of HucR from the column using 2 column volumes of Buffer A supplemented with 1 M KCl. The eluted HucR was dialyzed in 2L of Buffer A overnight and subsequently applied to a 15 ml DEAE cellulose column equilibrated with Buffer A and washed with an additional column volume of Buffer A in which both the flow-through and wash volumes were collected. HucR does not bind to this column. This flow-through was immediately added to a 15 ml CM-cellulose column also equilibrated with Buffer A and washed with one column volume of Buffer A. Again, HucR does not bind under these conditions and was collected in the flow-through volume. The flow-through was then applied to a 12 ml hydroxylapatite column equilibrated with Buffer A and the column was washed with 5 column volumes of Buffer A. HucR was eluted from this column in batch mode using 2 column volumes of Buffer A supplemented with 1M Potassium Phosphate. As a final purification step, HucR was then applied to a Superdex 200 size exclusion column in which the purest fractions of HucR (as determined by SDS-PAGE) were pooled together and concentrated to 15 mg/mL and flash-frozen in liquid nitrogen for storage at -80 °C (Fig. 8a). Protein concentration was determined spectrophotometrically using $A_{280}$ and a molar extinction coefficient for HucR ($\varepsilon = 13,940 \text{ M}^{-1}\text{cm}^{-1}$).

Selenomethionine-labeled (SeMet) HucR protein was expressed from pSPW1 using *E. coli* BL21(DE3) pLysS. An overnight culture grown in LB was used to make a 1:100 dilution with fresh M9 media (6g Na$_2$HPO$_4$, 3g KH$_2$PO$_4$, 1g NH$_4$Cl and 0.5 g NaCl) supplemented with 2 mM MgSO$_4$, 0.1 mM CaCl$_2$, and 4 g/L glycerol. Cells were
grown until $A_{600}$ reached 0.5 at which point 100 mg/L of amino acids Thr, Lys, Phe, and 50 mg/L of Leu, Val, and Ile were added to the media to inhibit the methionine biosynthetic pathway. Cells were grown for an additional 45 min to deplete residual cellular levels of L-methionine followed by a 0.2 mM IPTG induction, the addition of 50 mg/L selenomethionine, and the reduction in temperature to 24 ºC. Production of SeMet-labeled HucR was allowed to proceed for 24 h before the cells were harvested. Purification of SeMet HucR was carried out as described above for the native HucR protein with the exception that Buffer A was supplemented with 10 mM $\beta$-Mercaptoethanol (Fig. 8b).

![Image of SDS-PAGE gel with bands at 15 kDa and 25 kDa](image)

**Figure 8.** Purified native and SeMet HucR. Purified protein was analyzed on an SDS-PAGE gel to confirm purity. (A) Native HucR. (B) SeMet HucR

**Crystallization and Data Collection**

HucR crystals were grown by the hanging-drop vapor-diffusion method with 1:1 mixtures of 12 mg/mL of protein and 22 % PEG 3350, 500 mM MgCl$_2$, and 100 mM Bis-Tris, pH 7.0 at 4 ºC. Long rod-like hexagonal shaped crystals appeared after 4 days for the native HucR and after two weeks for the SeMet-labeled HucR (Fig. 9). Native and
SeMet crystals belong to the hexagonal space groups P6$_1$ (a=b=45.0, c=284.6) and P6$_1$22 (a=b=44.9, c=286.4), respectively, and were found to contain two and one molecule within their asymmetric unit, respectively. A third crystal form of space group P3$_1$21 (a=b=77.3, c=266.4) was produced at room temperature in the same conditions supplemented with 1 mM Pb(NO$_3$)$_2$. The room temperature crystal form, which has three dimers in the asymmetric unit, was merohedrally twinned with a twin fraction of 0.266 as determined in CNS (79).

**Figure 9.** **Crystals of native and selenomethionine-containing HucR.** HucR crystals were obtained using the vapor diffusion method and photographed using a light polarizing filter. (A) Native HucR. (B) SeMet HucR.

Crystals were flash frozen in liquid nitrogen and diffraction data were collected with a MarCCD detector (165mm) at the protein crystallography beamline of the Center for Advanced Microstructures and Devices (CAMD, Louisiana State University) (Fig. 10). All data processing and scaling was carried out with HKL2000 software (80). An x-ray fluorescence scan was done to verify the presence of selenium in the SeMet crystals.
and to determine the exact wavelengths that correspond to the peak and inflection wavelengths for selenium (Fig. 11). Data was collected at three wavelengths (peak, inflection, and remote) for the SeMet HucR crystals to 2.6 Å resolution. All data collection statistics are listed in Table 2.

Structure Determination and Refinement

Initial phases (Table 3) were obtained with multiple wavelength anomalous dispersion (MAD) phasing to 2.8 Å resolution with SeMet data sets. Two of a possible four selenomethionine residues were apparent in a three-wavelength (peak, inflection, and remote) anomalous difference Patterson calculated with CNS (Fig. 12) After density modification which resulted in an overall figure of merit of 0.49, an electron density map was drawn revealing large solvent channels and continuous regions of electron density mostly resembling that of alpha helical secondary structure (Fig 13a). To more easily decipher the electron density, MapMan (81) was used to skeletonize the electron density (Fig 13bc) and a partial polyalanine model was built into the density using the program O (82). Side chains were placed by using the selenomethionine sites and large tryptophan residues to indicate the correct register within the coding sequence. Phase combination of model and experimental phases was performed to improve map quality until approximately two-thirds of the model was built. At this point 2Fo-Fc maps at 2.3 Å resolution were calculated.

Once an $R_{free}$ value of 0.43 was obtained, the model derived from the SeMet data was positioned in the wild-type P6$_1$ unit cell for further refinement with non crystallographic symmetry (NCS) restraint weights of 400. The final 2.3 Å resolution model ($R_{crys}/R_{free} = 23.5/29.0$) includes 157 of a total of 181 residues: The amino
Figure 10. **Native HucR diffraction image.** A diffraction image of the native HucR taken at the PX beamline of CAMD using a MarCCD detector.
Figure 11. **X-ray fluorescence scan for SeMet crystals.** The peak and inflection wavelengths for selenium in the SeMet crystal were found by doing a fluorescence scan at the PX beamline of CAMD and determined to be 12658 eV (0.9795 Å) and 12653 eV (0.9797 Å), respectively.

<table>
<thead>
<tr>
<th>Data set</th>
<th>SeMet-MAD</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peak</td>
<td>Inflection</td>
</tr>
<tr>
<td>Space group</td>
<td>P6,22</td>
<td></td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a,b,c (Å)</td>
<td>44.9, 44.9, 286.4</td>
<td>45.0, 45.0, 284.6</td>
</tr>
<tr>
<td>α,β,γ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9795</td>
<td>0.9798</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50-2.65</td>
<td>50-2.65</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>5659</td>
<td>5714</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.9 (99.8)</td>
<td>98.9 (100)</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>13.1 (9.2)</td>
<td>10.0 (9.3)</td>
</tr>
<tr>
<td>I/σ a</td>
<td>13.9 (4.2)</td>
<td>15.9 (5.0)</td>
</tr>
<tr>
<td>R_sym b (%)</td>
<td>13.3 (49.0)</td>
<td>11.0 (47.3)</td>
</tr>
</tbody>
</table>

*aHighest resolution shell in parentheses

b$R_{sym} = \sum |I_j - \langle I \rangle| / \Sigma \langle I \rangle$, where $I_j$ is the intensity for reflection $j$ and $\langle I \rangle$ is the mean intensity of the reflection
Table 3. **MAD phasing and structure refinement statistics.**

<table>
<thead>
<tr>
<th></th>
<th>SeMet</th>
<th>Native P6₁</th>
<th>Native P₃₁2₁</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phasing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>500-2.8</td>
<td>500-2.3</td>
<td>50-2.9</td>
</tr>
<tr>
<td>Figure of merit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrics/acentrics/all</td>
<td>0.68/0.44/0.49</td>
<td>23.5/29.0</td>
<td>22.1/30.4</td>
</tr>
<tr>
<td><strong>Phasing power</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrics/acentrics/all</td>
<td>1.55/1.41/1.45</td>
<td>2358/187/2</td>
<td>7088/88/na</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td></td>
<td>500-2.3</td>
<td>50-2.9</td>
</tr>
<tr>
<td>R&lt;sub&gt;cryst&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt; (%)</td>
<td>23.5/29.0</td>
<td>22.1/30.4</td>
<td></td>
</tr>
<tr>
<td>No. of nonhydrogen atoms</td>
<td></td>
<td>2358/187/2</td>
<td>7088/88/na</td>
</tr>
<tr>
<td>Protein/water/chloride</td>
<td>0.006</td>
<td>0.007</td>
<td>1.20</td>
</tr>
<tr>
<td>rms from ideal geometry</td>
<td></td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.20</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td></td>
<td>52/51/41</td>
<td>27.3/19.2/na</td>
</tr>
<tr>
<td>Protein/water/chloride</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Res. in most favored regions (%)</td>
<td>90.9</td>
<td>87.0</td>
<td></td>
</tr>
<tr>
<td>Res. in add. allowed regions (%)</td>
<td>8.3</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>Residues in gen. allowed regions (%)</td>
<td>0.8</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Residues in disallowed region (%)</td>
<td>0.0</td>
<td>1.4</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>R = Σ||F<sub>o</sub>|| - |F<sub>c</sub>|| / Σ|F<sub>o</sub>|, where F<sub>o</sub> and F<sub>c</sub> are the observed and calculated structure factor amplitudes.

<sup>b</sup>R<sub>free</sub> was calculated by withholding 10% of the total reflections from refinement.

Figure 12. **Patterson map.** The z = 0.5 Harker section of a 3-wavelength anomalous difference Patterson map from the SeMet datasets.
Figure 13. **Electron density from initial phases.** (A) Electron density map contoured to 1σ after density modification (solvent flipping) in CNS. (B) The skeletonized electron density from panel A was made using mapman. (C) The final Cα trace of HucR (yellow), correctly positioned within the skeletonized density seen in panel B.

terminus (residues 1-7), the last two residues on the carboxyl terminus, the wing residues 121-127, and loop residues 27-35 were not visible in the electron density map. In addition, amino acid residues 25, 36, 88, and 120 have no convincing electron density for side-chain placement. Final refinement statistics are shown in Table 3 and a representative 2Fo –Fc electron density map contoured at 1σ is depicted in Figure 14. All figures were prepared using PYMOL (83), GRASP (84), or O.

The dimer was positioned in the P3₁2₁ unit cell with molecular replacement as implemented in CNS. Six dimers were identified in the asymmetric unit giving a Matthews coefficient (V_m) of 1.93, within the typical ranges of 1.62-3.53. Refinement was done in CNS keeping each monomer restrained by NCS weights of 500. The structure was refined to a resolution of 2.9 Å to give a final R_{cryst}/R_{free} of 22.2/30.5 (Table 3).
Figure 14. **Representative 2F_o – F_c electron density map.** A stereo image of a representative 2F_o – F_c electron density map contoured to 1σ showing the HucR dimerization interface with each subunit colored in either magenta or green.

**Data Deposition**

The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the PDB ID **2FBK**.

**Results and Discussion**

**Overall Structure of HucR**

The crystal structure of the HucR dimer was determined to 2.3 Å and refined to a final R_cryst of 23.5% and R_free of 29.0%. HucR has a “saddle-like” shape with overall dimensions of 62 Å wide, 42 Å tall and 37 Å deep relative to the orientation shown in Fig 15a. The protein is largely α-helical (55% α-helix, 5% β-strand) with the topology α1-α2-α3-β1-α4-α5-β2-β3-α6-α7 (Fig. 15b). The first helix, α1, has no counterpart in any other MarR family members for which structures have been described. Helices from the amino and carboxyl termini (α2, α6, and α7) form the dimerization domain, while the DNA-binding domain is composed of α3, β1, α4, α5, β2, and β3 (Fig. 15b). As predicted,
the structure reveals that HucR adopts a similar winged-helix fold as that seen in the structures for MarR, MexR, and OhrR (19,68,54). Consistent with that of other winged-helix proteins (16), the electrostatic surface potentials of HucR surrounding the recognition helix (α5) and the wing regions are electropositive, and are likely to interact with the negatively charged DNA (Fig. 15c,d).

Figure 15. **Overall structure of HucR.** (A) The HucR dimer is shown in a ribbon drawing with one monomer colored by spectrum from N to C while the other monomer is depicted in gray. (B) The HucR monomer is shown with each secondary structural element colored and labeled. The DNA binding domain is encased in a black box (α3, β1, α4, α5, β2, and β3). The dimerization domain consists of α2, α6, and α7. (C,D) The electrostatic potential surface map of HucR with positively and negatively charged regions shown in blue and red, respectively. Figure C is shown in the same orientation relative to A above, while D is rotated 90° around the x-axis relative to C with the bottom of C facing outwards in D.
The DNA Binding Domain

The DNA binding domain (residues 65 – 133) corresponds to a contiguous stretch of polypeptide flanked by regions that form the dimerization domain and is comprised of α3, β1, α4, α5, β2, and β3 (Fig. 15b). The winged-helix motif characteristic of the MarR family is an adaptation of the classical helix-turn-helix DNA binding motif and in HucR corresponds to short α-helices, α4 and α5, that pack at approximately 110°, followed by two anti-parallel β-strands (β2 and β3) that form a β-hairpin referred to as the “wing.” Helix α3 appears to serve as a scaffold-like structure for this domain as it mediates the majority of interactions within its own domain and those with the dimerization domain as well. The crystal structure reveals that the DNA binding domains of HucR are completely independent of one another, in contrast to what is observed for MarR in which a salt bridge links the two domains together.

The 17 residue β-hairpin “wings” of HucR jut out from the globular portion of the DNA binding domains to provide an exceptionally electropositive surface (Fig. 15c,d): six of the 17 amino acids are arginines (118, 119, 123, 125, 126, and 130). Previous biochemical evidence has shown that R118A HucR binds DNA with an affinity reduced 6-fold compared to that for wild type HucR (63). Moreover, the R118A mutant had no effect on sequence specificity of DNA binding, thus it is tempting to speculate that the wing is important for maintaining non-specific HucR/DNA interactions similar to what is observed in other canonical winged helix proteins (16).

Most winged helix proteins make sequence specific contacts to their respective DNA targets via the recognition helix that immediately precedes the wing motif (16). In the crystal structure of HucR, helix α5 (residues 101-113) is designated the recognition
helix, and is unusual in the fact that its amino terminus is a very short 3-residue 3-10 helix which turns into a typical 8-residue right-handed α-helix. The electrostatic potential map shows that HucR’s recognition helix is electropositive on both sides, a common feature for many winged-helix DNA binding proteins (Fig. 15c,d). In the crystal structure, the recognition helix is anchored to the central scaffold helix of the DNA binding domain (α3) via a salt bridge between Arg-106 and Asp-73 and side chain hydrogen bonding between Arg-109 and Asn-68. In addition, Lys-113 makes a hydrogen bond to the main chain carbonyl group of Gly-66, located on a turn between helices α2 and α3. Furthermore, this Lys-113/Gly-66 interaction is functionally identical to side chain- main chain H-bonds in MarR (Lys-81/Asp-37) and MexR (Arg-78/Asp-34) and OhrR (Gln-81/Asn-37). Such an interaction may be necessary to anchor the C-terminal end of the recognition helix in a position suitable for DNA binding.

A structure-based sequence alignment of HucR, MarR, MexR, and OhrR is presented in Figure 16. Hydrophobic residues conserved throughout the DNA binding domain appear to be important in stabilization of the overall fold of the DNA binding domain. In fact, the DNA binding domain has about twice as many invariant and/or highly conserved residues compared to the dimerization domain (Fig. 16), while it represents just one third of the protein. The bulk of the conserved amino acids are located in the hydrophobic core and include Leu-67, -74, -75, -78, -87, -92, -110, -115 -132 and Ile-107, -116,-130 (Fig. 17). Another invariant residue from the MarR family of transcriptional regulators is Thr-133 which is located at the proximal end of the wing between β3 and helix α6. This residue appears to clamp the wing in place by making two hydrogen bonding interactions: its side chain hydroxyl group makes an H-bond (2.7 Å)
with the backbone carbonyl of Gly-114 and the main chain nitrogen participates in an H-bond (2.9 Å) with the backbone carbonyl of Leu-115. The length of the wing is further stabilized by typical β-sheet interactions and an additional salt bridge between Arg-131 and Glu-117. There is insufficient electron density in the map to model the residues that comprise the distal wing (121-127), most likely as a consequence of a high degree of flexibility in this region. Such mobility is common in winged helix proteins, especially when the wing is not in a complex with its target DNA (68,85).

Figure 16. **Structure based sequence alignment between HucR, MarR, MexR, and OhrR.** Secondary structural elements and residue numbering corresponds to that of HucR. Residues shown in red font have no interpretable electron density in the structure of HucR and were not used in the superpositioning of the structures. Regions of amino acid conservation corresponding to invariant and similar are shaded in cyan and yellow, respectively. The DNA binding domain is encased within the black box.
Figure 17. **Hydrophobic interactions of the DNA binding domain.** The DNA binding domain of HucR forms a compact globular domain with a hydrophobic core; many of the core residues are highly conserved within the MarR family, including Leu-67, -74, -75, -78, -87, -92, -110, -115 -132 and Ile-107, -116, -130.

**The Dimerization Domain**

Helices α2, α6, α7, α2’, α6’, and α7’ (where prime denotes the other subunit) form the intertwined helical bundle that constitutes the dimerization domain. Helices α2 and α2’ are oriented in an anti-parallel fashion to form the scaffold of the dimerization interface, while helices α6 and α7 straddle helix α2’ and α6’ and α7’ straddle α2. As a consequence of interdigitation of the helices, approximately 6300 Å² of surface area is buried at the dimerization interface. At the center of this packing located along the two-fold rotation axis of the dimer is a π-stacking interaction between the imidazole rings of
His-51 and His-51’ in which the rings are parallel and separated by a distance of 3.6 Å (Fig. 14). Helices α2 and α2’ also interact through hydrophobic interactions that include Leu-54/Leu-44’ and Leu-50/Leu-47’, and there two well-ordered pockets of water molecules located on either side of the stacked His-51/His-51’ rings. The C-terminal helix α7 forms a coiled coil leucine zipper-like structure with extensive hydrophobic interactions with α2’, while α6 and α2’ make only limited hydrophobic contacts. Additional interactions that contribute to dimer stabilization are the intermolecular salt bridges Arg-172/Glu-162’ (α7/α7’) and Arg-17/Glu-167’ (α1/α7’). An intramolecular salt bridge between Arg-153 and Glu-57 stabilizes an interaction between the amino and carboxyl terminal helices that flank the DNA binding domain.

Interestingly, the crystal structure of HucR and the structure-based alignment (Fig. 16) reveal two conserved residues that might play a critical role in dimerization for members of the MarR family. Residues Leu-166, which is strictly conserved in the MarR superfamily, and Leu-166’ of HucR make contact at the two-fold rotation axis. They, along with other hydrophobic residues, collectively participate in the formation of the compact core of the dimerization domain, thus stabilizing helices α7, α7’, α2, and α2’ at this site. The MarR, MexR, and OhrR equivalents (Leu-139, Leu-131, and Leu-132, respectively) of Leu-166 in those structures participate in similar van der Waals interactions. Leu-158 from HucR, also conserved in MexR and MarR, serves to anchor a turn region located between helices α6 and α7 to the carboxyl terminal helix α7’ of the opposing monomer. Residues corresponding to Leu-158 in both MarR and MexR (Leu-127 and Leu-123, respectively) perform a similar structural role in those proteins and are
likely to have a similar function in all MarR family members given that this residue is highly conserved throughout the superfamily.

As referred to above, near the midpoint of the 40 Å long α2 and α2’ helices that provide the framework for the dimerization domain, the imidazole rings of His-51 and His-51’ are stacked and separated by a distance of 3.6 Å (Figs. 14 and 18) At pH 7.0, the pH at which the crystals were obtained, the imidazoles would be expected to be deprotonated and thus such an interaction is permissible. However, this configuration of His-side chains suggests that the dimer interface can serve as a pH sensor, and that a transient drop in pH could lead to a conformational change induced by the repulsion of like charges. Conformational changes at the dimer interface of MarR family members result in repositioning of the DNA binding domains. Thus DNA binding is abrogated if the domains are repositioned such that spacing between the recognition helices of winged-helix motifs is not compatible with the distance between the major grooves of the cognate operator. In the MexR structure, determined in the absence of ligand or DNA, four independent dimers were observed and the relative orientations of the monomers within each dimer differ as a result of deviations at the dimerization interfaces, and structural heterogeneity in the monomer itself (68). The structure of the reduced form of the oxidative stress sensor OhrR in the absence and presence of DNA similarly revealed differences at the dimer interface (54). Furthermore, the authors propose that oxidation of the conserved Cys, located at the amino terminus of the α-helix of OhrR that corresponds to α2 in HucR, induces a conformational change at the dimer interface and thus abrogates DNA binding.
Figure 18. **Imidazole ring stacking at the HucR dimerization interface.** (A) A global view of the HucR dimer shown as a Ca trace with helices α2 and α2′ shown as a ribbon drawing (gold) and His-51 and His-51′ imidazoles shown in stick format (green and blue). (B) A close-up look at the imidazole ring stacking interactions.

**HucR-DNA Interactions**

In addition to the 2.3 Å structure of HucR, a second crystal form (space group P3₁2₁) was solved in the absence of ligand to 2.9 Å with three dimers found in the asymmetric unit. Superpositioning of all three dimers from the second crystal form with the original P6₁ dimer using 314 Ca carbons revealed that there is very little flexibility between each of the observed HucR dimers as indicated by a similar positioning of the DNA binding lobes from each monomer relative to the other (Fig. 19). In fact the variation in r.m.s. deviation between all four dimers was between 0.6 Å and 1.4 Å. This is in stark contrast to what is seen in the crystal structure of MexR in which there are significant differences in the positioning of the DNA binding domains, specifically for dimer AB compared to dimer CD (Fig. 5). Furthermore, the authors suggest that it is this flexibility that is likely responsible for determining whether MexR can or cannot bind its cognate DNA as a direct result of the positioning of the two DNA binding domains relative to one another (68). Additionally, the observation that MarR was unable to
crystallize in the absence of its ligand supports the possibility that it too has inherent conformational flexibility. Thus, the HucR dimer appears to lack the conformational flexibility that is observed in MexR and MarR and likely stays in a more fixed confirmation even in the absence of its ligand (uric acid) or its cognate DNA.

Figure 19. **Superposition of HucR dimers.** The three dimers from the P3_121 space group crystal form and the one dimer from the P6_1 crystal form were superposed using 314 Cα carbons to illustrate the lack of conformational heterogeneity between the four HucR dimers. The Cα traces for the dimers are shown (the P6_1 dimer is green).
The crystal structure of HucR also reveals that the lobes of the DNA binding domains are positioned with a spacing compatible with DNA binding. Superposition of the HucR dimer on the structure of the OhrR DNA-bound dimer puts the recognition helices from each subunit directly on top of one another, suggesting that no major reorientation of the domain interface is necessary to position the DNA binding domains to make contacts in the major grooves of its operator, *hucO* (Fig. 20a,b). Thus, due to the lack of conformational flexibility as noted above, it is our belief that HucR remains fixed in a “DNA ready” confirmation in which the lobes of the DNA binding domains are in a position ready to bind DNA. However, a localized conformational change is needed in residues 99-104 consisting of the loop between helices α4 and α5 (i.e. the turn of the HTH motif) and the N-terminal end of α5 (the recognition helix) in order to become fully compatible with DNA binding, as that defined by the OhrR-DNA structure (fig 20a,b). Furthermore, the N-termini of the recognition helices adopt a 3-10 helix secondary structure at the position of the largest degree of steric clashing between HucR and DNA (most notably Pro-101 and Ser-102), it is possible that DNA binding facilitates the reorientation of this region of the recognition helix to allow for proper major groove binding interactions. To further support this hypothesis, figure 21 shows a distribution of the thermal B-factors for HucR in which relatively high values (as compared to the other regions of HucR) are found at the N-termini of the recognition helices. High B-factors generally indicate increased thermal motion or regions of disorder as a consequence of increased motion. Thus in contrast to what is observed for OhrR in the absence of DNA, the dimer interface of HucR in the absence of DNA is compatible with DNA binding.
Figure 20. **Superposition of the HucR dimer on the OhrR-DNA bound dimer and DNA.** (A) Superposition of HucR (red) on the OhrR-DNA bound structure (blue) with heavy black circles indicating similar positioning of the recognition helices and heavy green circles representing the N-terminal regions of the recognition helices of HucR that require minor localized conformational changes to facilitate DNA binding. (B) The DNA binding domain of HucR (red, recognition helix yellow) modeled onto DNA according to the superposition in (A). Positions at which HucR clashes with the DNA are highlighted with white circles.
Figure 21. **Illustration of B-factors for HucR.** The B-factor distribution of the Cα carbons in HucR is shown so that B-factors are color coded from highest values to lowest values (i.e. red → orange → yellow → green → cyan → blue). Heavy black circles located at the N-terminal regions of the recognition helices correspond to regions that have high B-factors and would have to undergo minor localized conformational change to facilitate DNA binding.
CHAPTER 3

CONCLUDING REMARKS

The crystal structure of HucR has given new insight into the MarR family of transcriptional regulators as well as emphasized reoccurring themes throughout this family. The overall structure of HucR is largely alpha helical having structural topology very similar to the structures of other MarR family members, with the exception of an additional $\alpha$-helix at the amino terminus. The function of this helix is unknown and awaits future investigation, however, the crystal structure suggests that it might function to further stabilize the primary dimerization helix, $\alpha_2$, as demonstrated by several hydrophobic interactions. The HucR dimer has a “saddle-like” shape in which the subunits are extensively intertwined which results in approximately 6300 Å$^2$ of surface area being buried at the dimerization interface with most of these contacts mediated primarily by van der Waals interactions.

HucR has two domains: 1) a dimerization domain comprised of $\alpha$-helices from both the amino and carboxyl termini and 2) a globular DNA binding domain composed of residues encoded within the interior of the polypeptide. As predicted, HucR adopts a similar winged helix motif as seen for other MarR homologs for which structures have been described. Furthermore, the electrostatic surface potentials of HucR indicate that the regions around the recognition helices and the wing are highly electropositive, and are likely to provide a favorable mechanism for electrostatic interaction with the negatively charged DNA.
A second crystal form (P3₁21) of HucR was solved containing three dimers in the asymmetric unit. Comparison of these three dimers with the original dimer of the first crystal form (P6₁) reveals that there is very little conformational heterogeneity among the four dimers: the r.m.s. deviation range is from 0.6 Å to 1.4 Å. This is in contrast to what was described for the structure of MexR, in which significant variations in the MexR dimers were observed. Furthermore, a superposition of the HucR dimer with the structure of the DNA-bound OhrR, reveals that HucR is in a “DNA ready” confirmation in which the lobes of the DNA binding domains are in a position ready to bind DNA with the exception of some localized conformational changes needed at the amino termini of the recognition helices. Interestingly, a comparison of the OhrR DNA-bound structure with the nonbound OhrR reveals that significant reorientation of the DNA binding domains occurs upon DNA binding. Since HucR is in a DNA ready conformation in the absence of DNA, we suggest that HucR might remain fixed in this state at all times, even in the absence of DNA or ligand. The fact that four independent observations of the HucR dimer show that HucR maintains this state is further evidence that supports that HucR remains fixed in a DNA ready confirmation. If this is true, how does HucR lose its ability to bind DNA, as demonstrated by the presence of uric acid. One possibility is that the ligand binds to a region of the winged helix DNA binding motif and actively blocks DNA interactions via steric hindrance, similar to what is believed to occur with MarR and its ligand, salicylate. A second possibility, is that urate binding induces a conformational change resulting in a displacement of the DNA binding domain in a conformation not compatible with DNA binding. The fact that HucR crystals immediately crack when soaked with uric acid indicate that significant conformational
change likely occurs upon ligand binding. The site for which uric acid binds has yet to be determined. Interestingly, a third possibility is that the stacked imidazole rings from His-51 and His-51′ might act as a pH sensor and should there be a transient drop in pH so that the rings become protonated, the resulting repulsion of like charges could lead to a localized conformational change that would be propagated out to the lobes of the DNA binding domains placing them in a position not compatible with DNA binding. This might be advantageous to the cell if uric acid solubility is compromised during transient drops in pH which would result in an upregulation of cellular uricase levels to ensure uric acid is degraded before it precipitates within the cell. Interestingly, a substitution of phenylalanine for histidine at this position would eliminate the possibility of protonation while still allow the pi-stacking interactions to take place, thus providing a nice system to test this hypothesis of pH-mediated abrogation of DNA binding.
REFERENCES


Hansen Dudley Bordelon IV, who is better known by his family, friends, and colleagues as “Tee”, was born on March 13, 1975, in Baton Rouge, Louisiana. He has spent his entire life in Baton Rouge, where he graduated from Broadmoor High School in 1993 and went on to receive his bachelor and master degrees in microbiology from Louisiana State University in 1997 and 2000, respectively. He then worked as a Research Associate in the Department of Biological Sciences at Louisiana State University for two years in the laboratory of Dr. Grover Waldrop. On June 1, 2002, Tee joined Dr. Marcia Newcomer’s laboratory at LSU and began research in the field of protein crystallography where he officially enrolled into the graduate program in the spring semester of 2004. Tee will fulfill his requirements for the Doctor of Philosophy degree in biochemistry and will be graduating on May 19, 2006. On January 1, 2006, Tee plans to join a start-up biotechnology company in Baton Rouge where he will undertake various scientific endeavors.