

\( \beta \)-Sheet Interactions with DNA Revealed by the Arc Repressor-Operator Complex

by

Brigitte Elke Raumann

B.A., Biochemistry
University of California, Berkeley
May 1988

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology
February 1995

© 1994 by Brigitte Elke Raumann. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author

Certified by Dr. Robert T. Sauer, Thesis Supervisor

Accepted by Dr. Frank Solomon, Chairman, Biology Graduate Committee
β-Sheet Interactions with DNA Revealed by the Arc Repressor-Operator Complex

by

Brigitte Elke Raumann

submitted to the Department of Biology in December 1994 in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Bacteriophage P22 Arc repressor is a ribbon-helix-helix DNA-binding protein which specifically recognizes DNA using an antiparallel β-sheet. To gain a better understanding of how Arc recognizes its operator, I have solved the cocrystal structure of the Arc repressor-operator complex and characterized altered specificity mutants of the ribbon-helix-helix motif.

Chapter 1 gives an overview of mutations in DNA-binding proteins that alter the specificity of the protein-DNA interaction.

Chapter 2 describes the crystal structure at 2.6 Å resolution of two dimers of the P22 Arc repressor bound cooperatively to the wild-type 21 base pair operator site. Each Arc dimer uses an antiparallel β-sheet to recognize bases in the major groove, but the inherent symmetry of the β-sheet is broken as it makes an asymmetric set of contacts with a TAGA box subsite. Selected sections of this chapter have been previously published as "DNA Recognition by β-Sheets in the Arc Repressor-Operator Crystal Structure" (Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994) Nature 367:754-757).

Chapter 3 describes experiments altering the specificity of Arc and Mnt, two homologous, ribbon-helix-helix repressors which recognize very different operator sequences. A single amino acid change in the β-sheet of an Arc/Mnt hybrid protein switches the DNA-binding specificity from the arc operator to the mnt operator. In addition, a substitution in the N-terminal arm of the hybrid protein allows specific recognition of both the arc and mnt operators without increasing the affinity for non-specific DNA. How these dramatic specificity changes rely on three general features of the hybrid proteins is discussed. Dr. Kendall Knight collaborated on experiments described in this chapter.


Thesis Supervisor: Dr. Robert T. Sauer
Title: Whitehead Professor of Biochemistry
ACKNOWLEDGEMENTS

Above all, for their constant love and support, I thank my family and Kristen Fredricks.

For their very special friendship and caring over the years, I thank Asa Abeliovich, Robert Hayes, lab pal Bronwen Brown, and Alan Davidson.

For guidance and inspiration, I thank Herb Foerster, Susan Wente, Howard Schachman, and Carl Pabo.

For collaborating with me on experiments described in Chapter 2, I thank Mark Rould, and for collaborating with me on experiments described in Chapter 3, I thank Ken Knight.

For much needed advice and fun, I thank all of the Sauer lab members.

For being a wonderful advisor and setting an excellent example, I thank Bob Sauer.
**TABLE OF CONTENTS**

Abstract 2

Acknowledgements 3

Table of Contents 4

List of Figures 5

Chapter 1: Altered Specificity Mutants of DNA-Binding Proteins 7

Chapter 2: DNA Recognition by β-sheets in the Arc Repressor-Operator Complex 38

Chapter 3: Dramatic Changes in DNA-Binding Specificity Caused by Single Residue Substitutions in an Arc/Mnt Hybrid Repressor 72

Chapter 4: Major Groove DNA Recognition by β-Sheets: The Ribbon-Helix-Helix Family of Gene Regulatory Proteins 105
# LIST OF FIGURES

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Sequence of the DNA used for cocrystallization.</td>
<td>58</td>
</tr>
<tr>
<td>Figure 2a</td>
<td>The Arc-operator complex.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 2b</td>
<td>View of the β-sheet in the major groove of the right half-site.</td>
<td>61</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Schematic of hydrogen bonds between bases of the right operator half-site and the Arc dimer.</td>
<td>63</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Position of Phe10 in the Arc-operator complex.</td>
<td>65</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Details of the contacts between an Arc dimer and the right operator half-site.</td>
<td>67</td>
</tr>
<tr>
<td>Figure 6a</td>
<td>Electron density for β-strand residues 8 to 14 from an $</td>
<td>F_0</td>
</tr>
<tr>
<td>Figure 6b</td>
<td>Hydrogen bonds in the region of Gln9.</td>
<td>70</td>
</tr>
<tr>
<td>Figure 6c</td>
<td>Hydrogen bonds in the region of Gln9'.</td>
<td>71</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The arc and mnt operator sequences.</td>
<td>92</td>
</tr>
<tr>
<td>Figure 2</td>
<td>N-terminal sequences and DNA-binding affinities of Arc, Mnt, and the hybrid proteins.</td>
<td>94</td>
</tr>
<tr>
<td>Figure 3a</td>
<td>Gel mobility shift assay of the hybrid proteins.</td>
<td>96</td>
</tr>
<tr>
<td>Figure 3b</td>
<td>Methylation protection footprints of hybrid proteins.</td>
<td>97</td>
</tr>
</tbody>
</table>
Methylation protection footprint of the R5Q9 hybrid.

Hydroxy radical and methylation protection footprinting studies.

Contacts seen in the Arc-operator complex.

Plausible contacts between Mnt and the mnt operator based on genetic and biochemical studies.

Alternative DNA contact modes of the β4 position.

Figure 1. Ribbon representation of the Arc (βαα)2 motif.

Figure 2a. Ribbon representation of the MetJ dimer bound to an operator half-site.

Figure 2b. Ribbon representation of the Arc dimer bound to an operator half-site.

Figure 2c. View down the dyad axis of the MetJ tetramer bound to a DNA site containing the met operator.

Figure 2d. Corresponding view of the Arc tetramer bound to a DNA site containing the arc operator.

Figure 3. The helix B and tandem turn regions of the Arc and MetJ dimers contact six conserved phosphates in each operator half-site.

Figure 4. Schematic diagram of the contacts between β-sheet side chains and bases in an operator half-site for Arc and MetJ.
CHAPTER 1

Altered Specificity Mutants of DNA-Binding Proteins
Site specific DNA-binding proteins are essential to an array of cellular processes, including transcription, replication, and recombination. Key to the proper functioning of these proteins is the recognition of the proper DNA target site from the $10^6$ to $10^9$ potential non-specific sites in the cell. Understanding how this recognition is achieved is one of the major challenges in the field of molecular biology. One approach, illustrated in chapter 2, to elucidating the determinants of DNA-binding specificity is to solve the atomic structure of a specific protein-DNA complex. A complementary approach, taken in chapter 3, is to probe the protein-DNA interface with mutations that change specificity. This introductory chapter explores how these altered specificity mutations complement and expand available structural information.

Modifications of DNA-binding specificity can be grouped into several classes. Relaxed or broadened specificity occurs when a mutant protein recognizes new DNA sequences but has an equal or greater affinity for the wild-type sequence. Mutants with a reduction in affinity for alternate sites while retaining affinity for the wild-type site have sharpened specificity. Altered DNA-binding specificity requires the mutant protein not only to recognize a new site but to recognize this new site better than the wild-type site. In any case, the classification should depend only on the relative, not intrinsic, affinities of the protein for the DNA-binding sites. In fact, altered specificity mutants frequently have an overall reduction in DNA-binding activity compared to the wild-type protein.

When determining how a protein mutation modifies specificity, it is important to compare the affinities of various binding sites only for the given
mutant and not to compare DNA-binding values of the wild-type and mutant protein, particularly when using DNA-binding assays \textit{in vivo} as the intracellular levels of different proteins may not be the same. Another disadvantage of performing DNA-binding assays \textit{in vivo} is the difficulty of separating the effects of mutations on specificity from the effects on transcriptional activation, transcriptional repression, cofactor binding, or protein stability. For example, two mutants in the catabolite gene activator protein (CAP) were originally identified as altered specificity mutants based on their ability to activate the \textit{lacZ} gene fused to mutant CAP binding sites (Ebright et al., 1984). In later studies, the mutant proteins were purified and dissociation constants from a variety of mutated CAP binding sites were measured by gel retardation assays, revealing that these mutants had relaxed, but not altered, specificities (Ebright et al., 1987).

Altered specificity mutants have been successfully isolated by two methods. The first is by selecting or screening a pool of protein mutants with changes at sites believed to determine the specificity of the protein-DNA interaction. The second method takes advantage of homology between proteins with the same DNA-binding motif but different DNA-binding sites. In this case, amino acids thought to be critical for specificity of one protein are substituted into the homologous positions of another protein. No altered specificity mutant has been isolated that binds to a predetermined site without an example of a homologous interaction, although \textit{de novo} design of DNA-binding specificity may soon be realized. Although several altered specificity mutants have been obtained by exchanging entire DNA-binding domains of homologous proteins, the focus of this chapter is altered specificity achieved through the mutation of a very limited number of amino acids because this type of mutation is most useful in pinpointing individual amino acid-base interactions.
Since the first altered specificity mutant was isolated in 1983, many DNA-binding motifs have yielded altered specificity mutants of varying degrees. The most extensively studied motifs are the helix-turn-helix motif and, more recently, the Cys$_2$/His$_2$ zinc-finger motif. What follows is a summary of the altered specificity mutants found in several different classes of DNA-binding proteins.

**Ribbon-Helix-Helix Motif**

The first, and still one of the most dramatic, DNA specificity change was isolated in the phage P22 Mnt repressor (Youderian et al., 1983). Mutants were selected for the loss of affinity to the wild-type mnt operator and then screened for a gain in affinity to a mutant mnt operator containing symmetric mutations at two base pairs. The authors reasoned a single amino acid change could alter the specificity at two symmetrically related base pairs since symmetrically related subunits of Mnt probably recognize each operator half-site. A single mutation of His6 to Pro switched the specificity of Mnt to the mutant operator in vivo, and in vitro studies of the purified HP6 mutant showed the mutant mnt operator was recognized over the wild-type operator by a factor of 1000. Moreover, the HP6 mutant bound the mutant operator as tightly as Mnt bound the wild-type operator. Methylation at the N6 atom of the mutant adenine base was required for binding by the HP6 variant, indicating that, in the case of the mutant, specificity relied on a hydrophobic interaction between Pro6 and the N6-methyl group (Vershon et al., 1985).
This mutation identified a possible amino acid-base contact in the absence of any other biochemical or structural information. Later biochemical experiments supported the same direct His6-base contact (Knight & Sauer, 1989; Knight & Sauer, 1992), and in the cocrystal structures of the MetJ and Arc repressors, the analogous position in the β-sheet of both ribbon-helix-helix repressors makes direct base contacts (Somers & Phillips, 1992; Raumann et al., 1994).

Subsequently, the strategy of using symmetrical mutations in each half-site of DNA targets recognized by symmetric oligomers of DNA-binding proteins has been commonly used (see the helix-turn-helix motif and the basic region of bZIP and bHLH motif below).

*Helix-Turn-Helix Motif*

The specificities of the helix-turn-helix repressors from the lambdoid phages are among the earliest and most extensively studied. Switches of DNA-binding specificity by swapping residues in homologous positions were first carried out in this system. The solvent exposed residues of the DNA recognition α-helix of 434 repressor were exchanged with the homologous residues of P22 repressor, resulting in a protein with 5 mutations. These mutations almost completely switched the binding specificity of 434 repressor to that of P22 repressor, as judged by phage immunity *in vivo* and DNase I footprint analysis *in vitro* (Wharton & Ptashne, 1985). The mutant 434 repressor could discriminate
between the 434 and P22 operators to the same extent as wild type (100-fold), and the affinities for the cognate operators of mutant and wild-type repressors were approximately the same.

\[
\begin{array}{cccc}
X & X & A & C \\
X & X & T & G \\
\end{array} \quad \quad \begin{array}{cccc}
A & X & T & X \\
T & X & A & T \\
\end{array}
\]

Wild-Type 434 Repressor

Mutant 434 Repressor

X indicates a non-conserved base pair

Because several amino acid residues were mutated in this experiment and the P22 and 434 operators differ at several sites, it was not possible to define individual amino acid-base contacts. It was possible to conclude, however, that the recognition α-helix of 434 repressor is the sole determinant distinguishing between the 434 or P22 operators, and the recognition α-helices of 434 and P22 repressor approach the DNA in similar ways (Wharton & Ptashne, 1985).

The number of amino acid changes required for a specificity switch in the helix-turn-helix motif was narrowed when two amino acids in the λ repressor recognition α-helix were exchanged for the equivalent amino acids in λ Cro. The DNA-binding specificity was switched from OR1 to OR3 (which differ by 3 base pairs) by about 80-fold, as judged by DNase I footprinting studies, and the doubly mutant λ repressor bound OR3 as tightly as wild type bound OR1 (Hochschild et al., 1986).
This switch determined how λ repressor and λ Cro distinguish between $O_R1$ and $O_R3$ and suggested a contact between Ala at position 5 of the recognition helix and T5 later revealed by the cocrystal structure of the N-terminal domain of λ repressor (Jordan & Pabo, 1988).

Specificity switches in the helix-turn-helix motif can also result from single amino acid changes. Substitutions of the Gln in the first position of the recognition $\alpha$-helix of 434 repressor were screened for repression of the $lacZ$ gene controlled by mutant operators. The Gln to Ala mutation switched the specificity from the wild-type operator to a symmetric, doubly mutant operators; this switch in specificity was at least 150-fold when tested in vitro, and the interaction of the mutant protein for the mutant operator was as strong as wild-type repressor for the wild-type operator (Wharton & Ptashne, 1987).
This altered specificity mutant suggests a hydrophobic contact between Ala at position one of the recognition helix and T1 and again highlights the importance of van der Waals interactions in DNA-binding specificity. A contact between the wild-type Gln at position one and A1 was originally also suggested by this altered specificity mutant and later confirmed by the 434 repressor-DNA cocystal structure (Aggarwal et al., 1988).

Several altered specificity mutants have also been isolated in the bacterial helix-turn-helix repressors. The Lac repressor, in particular, has a long and extensive history of altered specificity mutants obtained using both homology with other helix-turn-helix repressors and selection or screening methods. These mutants allowed the identification of amino acid-base contacts and helped define the orientation of the recognition α-helix in the major groove (Lehming et al., 1987; Lehming et al., 1988; Sartorius et al., 1989; Sartorius et al., 1991), all before the NMR structure of the complex was solved (Chuprina et al., 1993). Because the Lac repressor mutants were tested in vivo by measuring the repression of the lacZ gene controlled by various mutant lac operators, it is difficult to quantify the extent of the specificity changes. Nevertheless, the repression assays in vivo allow the qualitative identification of altered specificity mutants. In 1987, Lehming and co-workers introduced changes in Lac repressor based on homology to the Gal repressor (Lehming et al., 1987). Individual mutations of the first or second residue of the recognition α-helix produced no change in specificity or broadened specificity, respectively. However, combining the mutations produced a mutant with altered specificity for a mutant operator with two, symmetric base changes.
In this case, it seems that residues at the first two positions of the recognition helix act cooperatively to determine binding specificity. Single site mutations, however, at position six of the recognition α-helix chosen by homology to phage repressors also showed specificity switches at two symmetric base pairs, and combining the altered specificity mutations at positions one, two, and six, produced the predicted specificity.

These mutants suggested that the orientation of the recognition α-helix of Lac repressor relative to the center of the operator was opposite to that of λ repressor (Lehming et al., 1988; Sartorius et al., 1989). Lastly, an altered specificity mutation was isolated by randomizing positions one, two, and five of the recognition α-helix of Lac repressor, resulting in the recognition of an operator with two symmetric nonconsensus base pairs.
Again, it was essential to randomize positions one and two in conjunction with position five to obtain mutants with specificity changes (Sartorius et al., 1991).

Mutations can also alter specificity in another bacterial helix-turn-helix DNA-binding protein, the Trp repressor. Using an screen in vivo based on P22 challenge phages, each of the possible 19 amino acid changes at positions one, two, and three of the Trp repressor recognition α-helix were tested against 28 operators with double, symmetric base pair mutations (Bass et al., 1988; Pfau et al., 1994a; Pfau et al., 1994b). A strong altered specificity phenotype was obtained by switching Ile to Lys at position 1 (Bass et al., 1988) and a weak phenotype by switching Thr to Lys at position 3 (Pfau et al., 1994a).
Since the specificity of these proteins was only measured using an assay \textit{in vivo}, it is difficult to quantify the extent of the specificity change. However, in the crystal structure of Trp repressor bound to the operator, these residues lie at the protein-DNA interface (Otwinowski et al., 1988; Lawson & Carey, 1993). Ile at position 1 is in contact with the edge of the bases in the major groove, but how these contacts contribute to specificity is unclear. Thr at position 3 makes a direct phosphate contact, and may contribute to specificity through indirect readout (Otwinowski et al., 1988).

Other examples of specificity changes in the helix-turn-helix motif have been found in the Tet repressor using \textit{lacZ} repression assays \textit{in vivo}. The binding specificity of the class A Tet repressor was switched to the specificity of the class B Tet repressor at two symmetrically related operator positions as a result of one amino acid exchange in position 3 of the recognition helix (Baumeister et al., 1988).
However, the reciprocal amino acid exchange in the class B Tet repressor showed only a very weak specificity switch, indicating perhaps position 3 of the class B recognition helix interacts more strongly with the class B operator than position 3 of the class A recognition helix interacts with the class A operator. Subsequent studies randomized and screened several single amino acid mutations all along the helix-turn-helix motif but found only one new altered specificity mutation, an Arg to Thr mutation in the helix preceding the recognition helix (Altschmied et al., 1988).

A direct contact between the Arg and a GC base pair was proposed. Although there is no structure of the Tet repressor-operator complex available to support this conclusion, Trp repressor also has an Arg at this position of the helix-turn helix motif which makes a direct contact to a GC base pair (Otwinowski et al., 1988). Like the Lac repressor system, these mutational studies placed the helix-turn-helix motif in the opposite orientation with respect to the center of the operator as the phage helix-turn-helix repressors. The crystal structure of the Tet repressor dimer later confirmed this orientation of the helix-turn-helix motif.

The σ\(^{70}\) subunit of *E. coli* RNA polymerase contains a region homologous to the helix-turn-helix motif, and mutations that alter the specificity of promoter recognition map to position 1 of the putative recognition α-helix. Selection and screening techniques identified one mutant of the σ\(^{70}\) subunit of *E. coli* RNA
polymerase that repressed a lac promoter differing from the consensus sequence by one base pair 10 fold better than the consensus lac promoter (Siegele et al., 1989).

This mutant and other modified specificity mutants isolated in this study identified possible amino acid-base contacts and oriented the σ$^{70}$ subunit with respect to the -35 and -10 regions of the promoter.

**Homeodomain Motif**

The homology among the eukaryotic homeodomain proteins has been used to isolate DNA-binding specificity switches caused by single amino acid mutations. Changing residue nine of the recognition α-helix of Biciod from Lys to Gln (found in the Antennapedia class of homeodomains) switched the specificity of transcriptional activation in yeast accordingly (Hanes & Brent, 1989). Subsequent studies in vivo dissected the Biciod binding site by mutation and showed the identity of primarily one base pair plays the decisive role in the specificity switch (Hanes & Brent, 1991).
These altered specificity mutants revealed that position nine of the recognition α-helix can play an important role in DNA-binding specificity, a prediction later borne out by NMR experiments on the Antennapedia complex and the cocrystal structure of the Engrailed complex (Qian et al., 1989; Kissinger et al., 1990). Later studies using purified Fushi Tarazu or Engrailed revealed Gln to Lys changes at position nine can influence the DNA-binding specificity at two base pairs by 30-60 fold (Percival-Smith et al., 1990; Ades & Sauer, 1994). In the case of Engrailed, the Lys mutant has a stronger interaction with the mutant operator than Gln has with the wild-type operator. Although the Gln to Lys mutant of Engrailed seems to indicated Gln in position 9 is an important determinant of specificity, a Gln to Ala mutation retains wild-type specificity (Ades & Sauer, 1994). This illustrates a limitation to the interpretation of altered specificity data. It cannot be assumed that because the mutant residue is responsible for the altered specificity, it has necessarily replaced a wild-type residue responsible for the wild-type specificity.

**Cys₂/His₂ Zinc-Finger Motif**

Recently, extensive studies have focused on the Cys₂/His₂ zinc-finger class of DNA-binding proteins in the hope that understanding altered specificity in this system will eventually allow the de novo design of a protein with a given DNA-binding specificity. The first mutations to change the specificity of a zinc-
finger protein were based on the homology between transcription factors Krox-20 and Sp1. Two residues in the helix of the second finger of Krox-20 were changed to the residues in the second finger of Sp1. Gel retardation assays using bacterial extracts showed the mutant Krox-20 now had the expected altered specificity at one base pair. The mutant bound this site with approximately the same affinity as wild-type Krox-20 recognizes the wild-type Krox-20 site and could discriminate it from wild type by 50-fold. The single mutations alone did not result in a specificity change, indicating that both residues cooperate in base-pair discrimination (Nardelli et al., 1991).

These experiments and similar substitutions in the first finger of Krox-20 identified two positions important for base selectivity and suggested each finger spanned a subsite of three nucleotides. Both conclusions were later supported by the cocrystal structures of Zif268 (Pavletich & Pabo, 1991), Gli (Pavletich & Pabo, 1993), and Tramtrack (Fairall et al., 1993). Subsequently, these results have been extended to different zinc-finger proteins (Desjarlais & Berg, 1992b; Desjarlais & Berg, 1992a; Thukral et al., 1992), including a synthetic consensus sequence zinc-finger (Desjarlais & Berg, 1993), showing in some, but not all cases, that the specificity of a given zinc-finger mutant can be predicted using the large data base of zinc-finger sequences and their cognate binding sites. Recently, altered
specificity mutants of Zif268 specific for an altered binding site by as much as 100-fold have been selected by phage affinity (Rebar & Pabo, 1994).

Another study also used phage display to select from a pool of mutants randomized at eight positions of the second Zn-finger of Zif268 (Choo & Klug, 1994b). The positions most important for specificity were -1, 3 and 6. A complementary experiment screened the selected mutants against libraries of DNA randomized at either the first, second, or third position of the recognition triplet (Choo & Klug, 1994a). This screen and subsequent equilibrium binding experiments with a subset of the mutants showed some mutants were specific for only one mutant triplet and could discriminate against closely related triplets by factors of about 10. These results lead the authors to propose a possible recognition code for a Zn-finger and a DNA triplet, however, all possible triplets cannot be recognized using this code and many of the possible Zn-finger-DNA triplet interactions have not yet been tested. As more and more altered specificity mutants of zinc-finger proteins accumulate, it seems the zinc-finger might be a useful motif for the de novo design of proteins with a predetermined DNA-binding specificity.
Basic Region of the bZIP and bHLH Motifs

Direct DNA contacts are made by the basic region of both the bZIP and bHLH proteins, while dimerization is mediated by either a leucine zipper, in the case of bZIP proteins, or a helix-loop-helix domain, as in the case of bHLH proteins (Murre et al., 1989; Davis et al., 1990; Talianian et al., 1990; O'Shea et al., 1991). Cocrystal structures of representatives of both families have been solved (Ellenberger et al., 1992; Ferré-D’Amaré et al., 1993; Ellenberger et al., 1994; Ma et al., 1994) and show in both families the basic region forms an α-helix and interacts with DNA in a similar manner. However, before the structures of either the bZIP or the bHLH complexes were solved, altered specificity mutants identified amino acid residues in the basic region important for DNA recognition.

One study used the homology between two yeast bHLH proteins, PHO4 and CPF-1, to identify a single amino acid change at position -14 of the basic region which switched the specificity from the PHO4 specificity to the CPF-1 specificity at two symmetrically related positions of the DNA site.

![Diagram of DNA with amino acid changes](image)

The gel retardation competition assays used to determine the specificity were not quantified, so the extent of the switch is unclear (Fisher & Goding, 1992). However, this altered specificity mutation provided strong evidence for one of the three models proposed at the time for the bHLH structure, and identified an
amino acid that plays a key role in DNA recognition in the basic regions of both the bHLH and bZIP proteins. This model and the DNA contact from position -14 was later confirmed by the cocrystal structures of three bHLH-DNA complexes, Max (Ferré-D’Amaré et al., 1993), MyoD (Ma et al., 1994), and E47 (Ellenberger et al., 1994) and the contact at position -14 was also seen in the GCN4-DNA complex (Ellenberger et al., 1992), a bZIP protein.

Positions -15 and -10 of the basic region of c-Myc, another bHLH protein, were changed to the analogous residues of another bHLH protein, E47. Gel retardation assays showed altered specificity at two symmetrically related base pairs only if both residues were mutated simultaneously. Since substitution of position -15 alone did not alter DNA-binding specificity, the authors concluded position -10 was the main determinant of sequence specificity for this particular nucleotide position.

![Diagram showing wild-type and mutant c-Myc sequences](image)

The extent of the switch was not determined, but it did not reach wild-type levels (Halazonetis & Kandil, 1992). Later studies of the E47-DNA complex structure revealed position -10 does not contact the DNA although it is in the protein-DNA interface (Ellenberger et al., 1994), however, the cocrystal structure of Max, another bHLH protein, shows position -10 in contact with the predicted bases (Ferré-D’Amaré et al., 1993). Thus, these mutations identified residues in the
protein-DNA interface, although, as in the Engrailed case, these residues may not determine specificity in the wild-type complex.

The specificity of N-Myc, a related bHLH protein, was also changed by homology swaps with other bHLH and bZIP proteins. Single amino acid substitutions at positions -10 or -14 in the basic region of N-Myc changed the DNA-binding specificity at one to two positions in the half-site.

Although many mutations in the basic region of GCN4, a bZIP protein, broaden specificity (Tzamarias et al., 1992; Kim et al., 1993; Suckow et al., 1993b;
Suckow et al., 1993a), only a few mutations alter the specificity. In the complex, alanine at position -14 contacts the methyl group of a thymine. Since this contact is not complicated by additional interactions at the protein-DNA interface (Ellenberger et al., 1992), it was judged to be a good candidate for isolating specificity changes. All 19 amino acid changes were introduced at position -14, and the resulting mutants were screened against all possible symmetric, two base pair changes of the GCN4 binding site. Crude bacterial extracts were used in gel retardation experiments to test the binding affinity (Suckow et al., 1993a).

One mutant altered the specificity at the expected base, and surprisingly, another mutant altered the specificity not at the base contacted in the wild-type complex, but at the adjacent base pair. This unexpected result highlights the danger of
relying solely on mutational data to define amino acid-base contacts in the wild-type structure.

Altered specificity mutants of GCN4 were also found at position -18. The codon at this position was randomized and mutants which could recognize symmetric, double mutations of the GCN4 binding site were selected. An Asp to Trp mutation slightly reduced the affinity to the wild-type site and strongly increased the affinity for a new site, as judged by \textit{in vitro} gel retardation assays (Tzamarias et al., 1992; Kim et al., 1993).

![Diagram of wild-type GCN4 and mutant GCN4](image)

In the cocrystal structure of GCN4, Asn at position -18 interacts with bases at position ±2 and ±3 of the symmetric binding site (Ellenberger et al., 1992), but the Trp mutation at position -18 affects the specificity only at position ±3, and it is unclear why this mutant protein retains wild-type specificity at position ±2. These studies demonstrate how mutations can test the importance of interactions seen in a protein-DNA complex.

\textit{Other Motifs}

In addition to the helix-turn-helix region of the \( \sigma^{70} \) subunit of \textit{E. coli} RNA polymerase, there is a region strongly conserved among the \( \sigma \) factors and predicted to form an \( \alpha \)-helix. An altered specificity mutation in this region, obtained by selection and assayed \textit{in vivo}, affects recognition at base pair -12 of
the $\text{P}_{\text{ant}}$ promoter, lending support to the model that this region is responsible for the recognition of the -10 promoter region (Waldburger et al., 1990).

The family of single-subunit RNA polymerases, which includes T7 and T3 RNA polymerase, does not contain any of the well characterized DNA-binding motifs. The T7 and T3 RNA polymerases are 85% homologous, yet they are highly specific for their respective promoters. This homology was exploited to construct, purify, and assay by transcription run off, mutants in T7 RNA polymerase with substitutions of residues from T3 RNA polymerase (Raskin et al., 1992). A single mutation altered the recognition of one, possibly two, base pairs and switched promoter specificity to the T3 promoter by 20-fold, although the activity of the mutant enzyme was only about 15% of wild type.
The reciprocal mutation in T3 RNA polymerase also conferred T7 promoter specificity. These experiments identified interactions responsible for the promoter specificity of T7 and T3 RNA polymerase. This residue is located in the putative DNA-binding cleft in the crystal structure of T7 RNA polymerase (Sousa et al., 1993). Subsequently, all possible amino acids were substituted at this position of T7 RNA polymerase, and three other altered specificity mutations were found with similar phenotypes (Raskin et al., 1993).

\[
\begin{align*}
\text{Wild-Type T7 RNAP} & \quad -11 \quad G \\
& \quad C \\
& \quad \text{Asn} \\
& \quad 748 \\
\text{Mutant T7 RNAP} & \quad -11 \quad C \\
& \quad G \\
& \quad \text{Trp} \\
& \quad 748 \\
\text{Wild-Type T7 RNAP} & \quad -10 \quad A \\
& \quad T \\
& \quad \text{Asn} \\
& \quad 748 \\
\text{Mutant T7 RNAP} & \quad -10 \quad C \\
& \quad G \\
& \quad \text{Thr or Tyr} \\
& \quad 748
\end{align*}
\]

Concluding Remarks

Altered specificity mutants are rare, especially mutants with the same degree of specificity and affinity as wild type. Unlike the relatively abundant relaxed specificity mutants, altered specificity mutants require two phenotypes;
the mutant must gain the ability to recognize a new DNA site while losing the ability to recognize the wild-type site. This phenotype can occur by either a gain of a favorable interaction or loss of an unfavorable interaction with the new site in conjunction with either a loss of a favorable interaction or gain of an unfavorable interaction with the wild-type site. Another reason altered specificity mutants are rare is because the ability of a particular residue to recognize a particular base often depends on the identity of the neighboring residues and base pairs. Because of this dependence on context, an optimal screen for altered specificity would vary several residues at once while simultaneously varying several base pairs. Technical reasons, however, limit many screens to varying only one or two amino acid residues and base pairs at a time.

A limitation of altered specificity mutations is that they cannot distinguish between direct amino acid-base contacts and indirect effects on specificity. For example, specificity changes may depend on global conformational changes or a change in phosphate contacts used to sense the DNA sequence through the structure of the DNA. In practice, however, most cases of altered specificity perturb amino acids in direct contact with the DNA base pairs or at least residues directly at the protein-DNA interface, making altered specificity mutations more valuable than other types of modifications of specificity, such as relaxed specificity mutants. One exception is the Thr to Lys mutation at position 3 in the recognition helix of Trp repressor which alters a position in contact with the DNA backbone in the cocrystal structure (Otwinowski et al., 1988). However, this mutant was assayed using a relatively indirect, in vivo assay, and experiments in vitro should be done to determine the effect this mutation has on specificity. Another limitation is that a mutation does not always change the
specificity at the base that is contacted by the wild-type residue, for example some mutants at position -14 of the basic region of GCN4. In the absence of structural information this can potentially lead to false identification of an amino acid-base interaction in the wild-type complex. In rare cases, altered specificity mutations do not identify residues determining specificity in the wild-type complex, such as mutations isolated in Engrailed or c-Myc, although the residues do lie in the protein-DNA interface.

From the above discussion it is clear mutations resulting in altered specificity can suggest possible amino acid-base contacts, such as in the case of Mnt, 434, Bicoid, or Lac repressors, or test the role of an amino acid-base contact seen in the structure of a protein-DNA complex, such as in the case of Engrailed or GCN4. Altered specificity mutations can also be helpful in defining the orientation of the protein with respect to the DNA-binding site, as in the case of Lac repressor and the σ70 subunit of E. coli RNA polymerase, or suggest residues that cooperate to recognize a specific DNA sequence, such as the first and second positions in the recognition helix of Lac repressor and residues three and six in the helix of the PHO4 Zn-finger. Because of the limitations of altered specificity mutants noted above a better genetic and biochemical method of identifying amino acid-base pair contacts is the energetic analysis of changes in specificity resulting from alanine mutations (reviewed in Ebright, 1991). Instead, the study of altered specificity mutants can provide an understanding of how nature has satisfied the stereochemical constraints of recognition resulting from the limited number of building blocks in both DNA and protein. An understanding of how to alter DNA-binding specificity should eventually lead to the de novo design DNA-binding specificity.
REFERENCES


CHAPTER 2

DNA Recognition by β-sheets in the Arc Repressor-Operator Complex
INTRODUCTION

Structural studies have shown how α-helices can be used in DNA recognition (reviewed in Harrison & Aggarwal, 1990; Pabo & Sauer, 1992), but less is known about the role of β-sheets (reviewed in Phillips et al., 1993). The two-stranded, antiparallel β-sheet of the phage P22 Arc repressor has been implicated in operator DNA recognition by a variety of biochemical, genetic, and structural studies (Vershon et al., 1986; Bowie & Sauer, 1989; Knight & Sauer, 1989; Breg et al., 1990). Two Arc dimers bind cooperatively to the 21 base pair operator to form a stable protein-DNA complex (Brown et al., 1990). The cocrystal structure of this Arc tetramer-operator complex confirms the central role of the Arc β-sheet in recognition and provides new insights into β-sheet-DNA interactions.

Arc is a small (53 residue) repressor protein encoded by P22, a bacteriophage of Salmonella typhimurium (Sauer et al., 1983). By binding to a 21 base pair operator located between the -10 and -35 regions of two overlapping promoters, P\textsubscript{ant} and P\textsubscript{mnt}, Arc negatively regulates transcription from these promoters during lytic growth of the phage (Susskind & Youderian, 1983). The structure of the Arc dimer, the stable oligomeric form in solution, has been solved by NMR (Breg et al., 1990; Bonvin et al., 1994) and by x-ray crystallography (U. Obeyeskare, C. Kissenger, L. Keefe, M. Rould, B. Raumann, R. Sauer & C. Pabo, in preparation). Arc is a member of a family that includes the P22 Mnt repressor, and the MetJ and TraY proteins of Escherichia coli (Raumann et al., 1994). Arc and Mnt share 40% homology and the DNA-binding specificity of Mnt can be changed to that of Arc by swapping their N-terminal arm and β-sheet regions (Knight & Sauer, 1989).
Other family members share more limited sequence homology. The structure of the MetJ protein dimer (Rafferty et al., 1989) and the cocrystal structure of the DNA-bound MetJ tetramer (Somers & Phillips, 1992) are known. Chapter 4 contains a detailed comparison of the protein-DNA interactions of this family.

MATERIALS AND METHODS

**Crystal Growth**

Arc repressor was purified as previously described (Brown & Sauer, 1993), and full DNA-binding activity was confirmed by gel mobility shift assays. The DNA used for cocrystallization with Arc contains a double-stranded 21 base pair region corresponding to the wild-type *arc* operator site (Vershon et al., 1987) and 5' overhangs which were added to facilitate cocrystallization (Fig. 1). Synthetic DNA was purified by two steps of reverse phase chromatography. Crystals (≈ 0.4 mm x 0.4 mm x 1.5 mm) were grown using the hanging drop, vapor diffusion method with drops containing one μl of 0.58 mM Arc tetramer, 0.93 mM DNA duplex, 30 mM bis tris propane-Cl (pH 7.4), 1 mM NaN₃, and one μl of well solution containing 15% PEG400 and 100 mM MgCl₂. The reproducibility of crystal growth was vastly improved through macroseeding techniques. Crystals form in space group P2₁ with *a* = 65.7 Å, *b* = 56.7 Å, *c* = 53.8 Å, and β = 107°, and the asymmetric unit contains one tetramer-operator complex and 43% solvent. Derivative crystals were obtained by macroseeding with native crystals that were less than 5% of the volume of the final derivative crystals.
Data Collection

Data were collected at ambient temperature on a Rigaku R-AXIS IIC imaging plate area detector and reduced using the programs DENZO and SCALEPACK (Z. Otwinowski). Native data to 2.6 Å resolution were collected from two crystals, giving an Rsym of 11.4% for merging 161,619 observations to 11,253 unique reflections (95% complete, Rsym = Σ|I - <I>| / ΣI where I is the observed intensity of a reflection and <I> is the average intensity of the symmetry related reflections). Five heavy atom derivatives were prepared by cocrystallization using DNA duplexes with substitutions of 5-iodo-uridine for thymine at selected positions. Only one derivative, containing 5-iodo-uridine substituted for thymine at positions -9 and +9' of the operator site, showed sufficient phasing power to be useful. Data to 2.8 Å from this derivative were collected from one crystal, giving a phasing power of 1.45 (phasing power = \((\Sigma(F_{H(calc)}^2 - F_{H(obs)}^2))^1/2\) and an Rsym of 11.9% for merging 137,747 observations to 9,335 unique reflections (99% complete). The mean isomorphous difference in intensities was 25%. A strong signal in an anomalous difference Patterson map, allowed anomalous data to be included in phase calculations. Heavy atom parameters were refined with the CCP4 program REFINE (1979), and single isomorphous replacement/anomalous scattering phases were calculated with PHARE (1979). The phases had a figure of merit of 0.58 from 21 to 2.8 Å.

Model Building and Refinement

The initial map revealed several base pairs, most of the DNA phosphates, and about 80% of the protein main chain. B-form DNA and the crystal structure of the Arc dimer (U. Obeysekare, C. Kissenger, L. Keefe, M.
Rould, B. Raumann, R. Sauer, and C. Pabo, in preparation) were used as guides during initial model building. After rigid body refinement and positional refinement using the program X-PLOR (Brunger et al., 1987; Brunger, 1992), the initial model (which contained 76% of all atoms) gave a crystallographic R-factor of 32.7% to 3.0 Å. Refinement proceeded with further cycles of positional refinement and simulated annealing. The program FRODO (Jones, 1978) was used for rebuilding and adding new residues as they became clear in the electron density. Although the complex has an approximate two-fold symmetry axis relating the two dimers in the tetramer and the two halves of the operator (we refer to this as the tetramer symmetry axis), this restraint was not included during refinement. In later stages of refinement, the model was checked and rebuilt using simulated annealing omit maps (X-PLOR). Conventional least squares refinement with the TNT package of programs (Tronrud et al., 1987) was used to obtain the final model. At this stage, tightly restrained individual B-factors were used, and 45 water molecules were added.

The final structure includes all of the DNA and 98% of the protein tetramer (the last residue of one monomer and the last three residues of another monomer could not be reliably built). The crystallographic R-factor of the final model is 22.5% for all reflections from 21 to 2.6 Å. The 2,653 atoms in the final model have an rms deviation from ideal bond lengths of 0.010 Å, an rms deviation from ideal bond angles of 1.51°, and an rms difference in B values between bonded atoms of 1.99 Å². Only one of 199 Φ,Ψ torsion angle combinations lies in a non-favorable region of the Ramachandran plot. Helicoidal DNA parameters were calculated by the Dials and Windows program of the Molecular Dynamics Analysis Toolchest.
(Ravishanker et al., 1989) using a local helix axis calculated by the program CURVES (Lavery & Sklenar, 1988). The approximate two-fold symmetry of the Arc-DNA complex around the tetramer axis raises the possibility that the complex can exist in two orientations in the asymmetric unit, and some evidence for conformational averaging was observed in derivatives not used for phasing. Because the TAGA box is symmetric about the tetramer symmetry axis, however, conformational averaging in the native crystal would not affect the observed asymmetry of contacts to the TAGA box. As a further check of the asymmetry of dimer/half-site interactions, monomers were deliberately swapped within a dimer, and simulated annealing omit maps were calculated. These maps clearly regenerate the asymmetry described for the dimer/half-site complex.

RESULTS AND DISCUSSION

Overview of the Protein-DNA Complex

As shown in Figure 2a, Arc binds to the 21 base pair operator as a tetramer with a dimer recognizing each half-site. The operator DNA is basically B-form DNA, but it is bent and slightly unwound. Because the left and right halves of the complex are very similar, we focus primarily on a single dimer bound to a half-site in describing the structure.

The Arc dimer consists of identical, intertwined monomers, each containing a β-strand and two α-helices. The two-stranded β-sheet is formed by antiparallel pairing of residues 8-14 of each monomer, and the α-helices are formed by residues 15-30 (helix A) and by residues 32-48 (helix B). We will
refer to the dimer structural motif as a \( (\beta\alpha\alpha)_2 \) fold. The structure of the DNA bound dimer is similar to the solution structure (Breg et al., 1990; Bonvin et al., 1994) and to the crystal structure of the unbound dimer (0.7 Å rms deviation for the C\(\alpha\) atoms of residues 7-46). The C-terminal residues and some long surface side chains have larger rms deviations, but in most cases, these residues are poorly ordered or are influenced by crystal packing. There are, however, some important changes that occur upon complex formation. In particular, the six N-terminal residues adopt a compact structure in which residues one to four form a type II turn, residues three to six form a hydrogen bonded type III turn, and residues four to seven form a hydrogen bonded type I turn. This conformation is stabilized by numerous DNA contacts and by the packing of the Met4 side chain, and to a lesser extent the Met1 and Met7 side chains, against the body of the protein. By contrast, in the crystal structure of the uncomplexed dimer, the six N-terminal residues either appear disordered (residues one to three) or have different conformations and high thermal factors (residues four to six). The N-terminal residues of Arc are also disordered in the NMR structure (Breg et al., 1990; Bonvin et al., 1994), although the amides of Gly3 and Ser5 are observed to exchange slowly, suggesting these positions are hydrogen bonded. In the complex, these amides hydrogen bond only to the DNA phosphates. Hence, if the N-terminal arm does adopt some structure in solution, it is likely to be different from the structure of the protein-DNA complex.

The \( \beta \)-sheet of the dimer lies flat in the center of the major groove (Fig. 2b), and the side chains facing the DNA make an extensive set of base contacts. There are also numerous contacts between Arc and the DNA phosphate oxygens. These contacts straddle the major groove and position
the β-sheet for binding. Most of the phosphate contacts are made by main chain atoms in the N-terminal arm and at the N-terminus of helix B. Although there are no direct contacts between the polypeptide backbone of the β-sheet and the sugar-phosphate backbone of the DNA, water molecules link several of the peptide amides to phosphate oxygens.

Rotation about the tetramer symmetry axis superimposes the phosphorous atoms of each half-site and the $C_\alpha$ atoms of residues 1-46 of each dimer with an rms deviation of 0.5 Å. Twelve of the twenty-one operator base pairs are symmetric around this central axis, including the TAGA boxes which have been identified as critical determinants of Arc binding (Vershon et al., 1987). In addition, there are also local symmetry axes associated with each dimer (ca. 0.7 Å rms deviation for superposition of the $C_\alpha$ atoms of residues 1-48 of each monomer). The backbone of the DNA half-site is less symmetric about the dimer axis (2.6 Å rms deviation for phosphorous atoms), and the base sequence of the half-site shows very little symmetry. As discussed below, many of the protein-DNA contacts in each half-site are also asymmetric.

**Base Contacts**

Side chains from the β-sheet make all of the hydrogen bonds with the DNA bases. As shown in Figure 3, contacts are made by Gln9, Asn11, and Arg13 from one strand and by Gln9', Asn11', and Arg13' from the other strand of the β-sheet. Five of these six side chains interact with first three base pairs of the TAGA box and also hydrogen bond with each other (Fig. 3) to form an intricate network at the protein-DNA interface. It seems likely that the observed interactions between the side chains stabilize the direct DNA
contacts. In addition, conformational adjustments in the β-sheet appear to play important roles in the recognition process. In the unbound protein, the Phe10 and Phe10' side chains are buried in the hydrophobic core, but each side chain swings out and packs between adjacent, unesterified, phosphate oxygens in the complex (Fig. 4). At the same time, the Cα atoms of residues 9, 9', 10, and 10' move 1.0-1.7 Å to partially fill the cavities left by the repositioning of the phenyl rings. The main chain movements of Gln9 and Gln9', in turn, help to position the side chains of these residues for base contacts.

Figures 5 and 6 show details of the side chain-base contacts. Hydrogen bonds to the TAGA box contacts in the right half-site include pairs of hydrogen bonds between Gln9/Gln9' and adenines 6 and 5, respectively, single hydrogen bonds between Asn11 and cytosine 4 and Asn11' and thymine 5, and one hydrogen bond between Arg13 and guanine 4 (all of the corresponding contacts are also observed in the left half-site, plus an additional, water-mediated hydrogen bond between Arg13 and guanine -4). The only hydrogen bonds to bases outside of the TAGA box are made by Arg13', which makes two hydrogen bonds with guanine 8 in the right half-site (Fig. 3a; Fig. 5) and one hydrogen bond with adenine -8 in the left half-site.

Contacts between Arc and the TAGA box also include van der Waals and hydrophobic interactions. The methyl groups of the three thymines in the TAGA box are all solvent inaccessible because of tight packing against the protein. In the most hydrophobic of these interactions, the methyl group of thymine 3 packs against the side chains of Met4' and the aliphatic region of Arg13. These are the sole contacts by the protein to the last base pair of the
TAGA box. The combination of hydrophobic interactions and hydrogen bonds between the protein and the TAGA sequence appears to be sufficiently specific to allow Arc to discriminate between TAGA and any other sequence.

Although Arc is an inherently symmetric protein dimer, it recognizes a largely asymmetric base sequence. The symmetry axis of the dimer lies between the first and second base pairs of the TAGA box and is roughly coincident with a symmetry axis that could be defined from the set of phosphate contacts (discussed below). The T:A and A:T base pairs at the first and second positions of the TAGA box are related by the dimer symmetry axis (Fig. 1; Fig. 3a), and indeed, Gln9 and Gln9' do make symmetric contacts with the adenines at these positions. However, there is no further symmetry in the base contacts. The breakdown of symmetry involves not only the sequence but even the position of the contacted bases (Fig. 3a), as illustrated by the contacts made by Asn11 and Asn11'. Asn11' contacts the thymine in the second base pair of the TAGA box, and therefore, one might expect Asn11 to contact the symmetric thymine in the first base pair, but instead, it contacts the cytosine in the third base pair of the TAGA box. The cross β-sheet hydrogen bond between Asn11 and Asn11' enforces asymmetry in the base contacts of these side chains, but even if this hydrogen bond was broken, the asparagine side chains would not be positioned to make symmetric contacts, since the symmetry axis defined by the dimer is not perfectly coincident with the symmetry axis defined by the DNA half-site. A similar positional asymmetry is seen with the contacts made by Arg13 and Arg13'.
Contacts with the DNA Backbone

In contrast to the asymmetric base contacts described above, each Arc dimer makes a symmetric set of phosphate contacts in each half-site. Eight phosphates, four consecutive positions on each DNA strand flanking the β-sheet, are contacted by each dimer. Since these contacts are symmetric, both within and between half-sites, we describe only one set of interactions. As indicated in Figure 5, phosphate 10' accepts hydrogen bonds from the main chain NH's of Gly3 and Ser5, and from the side chain of Ser5. Phosphate 9' is contacted by the α-NH₃⁺ group of Met1, and phosphate 8' accepts hydrogen bonds from the main chain NH's of Val33 and Asn34. Since Val33 and Asn34 are at the N-terminus of helix B, these interactions direct the helical dipole towards phosphate 8'. Phe10' makes van der Waals contacts with phosphates 7' and 8'. Finally, the guanidinium group of Arg23 from the first helix is close enough to phosphate 7' for favorable electrostatic interactions, however, any possible hydrogen bonds have generally poor geometry. Biochemical studies indicate that this interaction is not very important for binding (Vershon et al., 1987; Brown et al., 1994).

Arc does not rely on long, flexible side chains for phosphate contacts, instead most of the phosphate contacts in the Arc complex directly link the peptide backbone of the protein to the sugar-phosphate backbone of the DNA. These backbone-to-backbone contacts should help to rigidly fix the β-sheet against the DNA and thereby contribute to the specificity of base recognition. There also are a set of hydrogen bonding interactions that connect the phosphate contacts of helix B, the base contacts of the β-sheet, and the phosphate contacts of the N-terminal arm. For example, the peptide NH of Asn34' contacts a phosphate, while the side chain of this residue hydrogen
bonds to the backbone of Arg13. The side chain of Arg13 continues this network by contacting a base and also contacting the backbone of Ser5', a residue which makes both a main chain-phosphate and a side chain-phosphate contact.

**DNA Conformation**

The operator in the complex is generally B-form DNA but has several distinctive features: (i) The average twist is $32.3^\circ$ per base pair, resulting in 11.1 base pairs/turn. (ii) There is a bend at the center of the operator and smaller bends near the center of each half-site (giving an overall bend of about $50^\circ$). (iii) There are some localized distortions in the sugar phosphate backbone and significant variations in groove width (measured by the phosphorous to phosphorous distances less 5.8 Å). In the regions around the β-sheets, the major grooves are widened to an average of 14.3 Å (compared to 11.7 Å in B-DNA; Saenger, 1984). At the center of the operator site, the minor groove is compressed to a width of only 1.2 Å (compared to 5.7 Å in B-DNA). The electrostatic repulsion between the two phosphates that approach each other most closely across the minor groove (+2 and -2) is probably minimized because each of them is contacted by an α-NH$_3^+$ group of the protein.

The above mentioned distortions of the DNA are unlikely to be crystal packing artifacts because the two halves of the operator site are in different crystal environments, and yet each half-site shows similar variations in DNA structure. Moreover, the DNA distortions that lead to bending of the DNA appear to be required for tetramer formation, as modeling the dimers on linear DNA indicates that they could not interact cooperatively.
**Dimer-Dimer Interface**

The two Arc dimers in the complex contact each other above the center of the operator site. These cooperative contacts involve only a few residues at the junction of helices A and B and bury a total of 650 Å² surface area. The key interactions are made by the side chains of the two Arg31' residues, each of which reaches across the tetramer interface to make two hydrogen bonds with the backbone carbonyl of Asn29' in the other subunit. Biochemical studies have shown that the dimer-dimer contacts contribute roughly 5 kcal/mol of binding energy (Brown & Sauer, 1993), and mutational studies indicate that the contacts made by Arg31' account for most of this stabilization (Brown et al., 1994). Two factors are likely to contribute to the strength of these contacts: the hydrogen bonds are largely solvent inaccessible in the complex, and the Arg31' side chain is held in place by a buried hydrogen bond/salt bridge with Glu36'. Other contacts that span the tetramer interface include symmetric hydrogen bonds made by bridging water molecules between the backbone carbonyls of residues 30' and 32', and a single hydrogen bond between the Gln39' side chain and the peptide carbonyl oxygen of position 28'.

Biochemical and genetic studies have shown that changing the spacing between operator half-sites by even one base pair drastically reduces Arc binding (Vershon et al., 1989; Smith & Sauer, 1994). This observation can be readily explained by the crystal structure. Each dimer is relatively rigid, and the hydrogen bonding interactions across the tetramer interface are likely to be highly sensitive to even small changes in distance or orientation. As a result, the formation of a stably bound tetrameric complex requires the precise spacing of half-sites observed in the wild-type operator.
Comparisons with Biochemical and Genetic Data

The Arc-operator co-crystal structure is generally consistent with data from chemical, biochemical and genetic studies of the Arc-operator interactions (Vershon et al., 1986; Vershon et al., 1987; Bowie & Sauer, 1989; Knight & Sauer, 1989; Vershon et al., 1989; Brown et al., 1990; Brown & Sauer, 1993; Brown et al., 1994). For example, the pH dependence of binding implicates the $\alpha$-NH$_3^+$ group of Arc in the DNA-binding reaction (Vershon et al., 1987), and here we observe contacts of all four $\alpha$-NH$_3^+$ groups in the tetramer with the DNA backbone phosphates. Solution footprinting data provides another test of the relevance of the cocrystal structure. The four guanine N7 positions identified in methylation protection and interference assays are all sites of Arc contact or are solvent inaccessible in the cocrystal structure (Vershon et al., 1987). At twelve of sixteen phosphate positions contacted by Arc, ethylation of the phosphate interferes with Arc binding. Backbone positions at which Arc contacts both unesterified phosphate oxygens show strong ethylation interference, whereas positions contacted at only one unesterified oxygen show weak interference (Vershon et al., 1987). The phosphate contacts made by Arg23 are not detected in ethylation interference assays, but this may simply indicate these contacts do not contribute significantly to the free energy of binding. This conclusion is supported by the mild reduction in DNA-binding affinity of an Arg23 to Ala mutant (Brown et al., 1994).

Most mutations of amino acid residues to alanine at positions involved in the protein-DNA interface cause substantial decreases in operator affinity (Brown et al., 1994). The mutational sensitivity of bases in the operator is also largely consistent with the cocrystal structure (Vershon et al.,
1989). For example, base changes in the TAGA boxes are the most deleterious to DNA-binding (Vershon et al., 1989), and Arc makes contacts with all of these base pairs. Comparing the structural and biochemical data suggests that the effects of some base changes in the arc operator may be mediated by changes in DNA structure. Thus, positions (±7, ±9, ±10) show some sensitivity to base changes, but only the phosphates of these positions are contacted in the crystal structure. Base sequence is not important at each position of phosphate contact, however, as positions ±1 and ±2 are insensitive to sequence changes.

**Major Groove Recognition by β-Sheets and α-Helices**

An α-helix is complementary in shape to the major groove, and there are numerous instances in protein-DNA complexes where contacts to both the bases and phosphates are made from a single α-helix (for review, see Harrison & Aggarwal, 1990; Pabo & Sauer, 1992). In fact, all of the DNA contacts come from single, extended α-helices in the GCN4 and Max proteins (Ellenberger et al., 1992; Ferré-D’Amaré et al., 1993). By contrast, the width of a two-stranded β-sheet seems to allow room for wobbling or twisting in the major groove, and a more extensive and more rigid framework may be needed to secure the β-sheet. In Arc, this framework includes other parts of the (βαα)2 fold and employs a large number of main chain NH groups to directly couple the backbone of the protein to the backbone of the DNA. These phosphate contacts are then connected via hydrogen bonding or packing interactions to the β-sheet.

How does the relative rigidity of the protein main chain in α-helices and β-sheets influence their use in DNA recognition? Using a rigid scaffold,
like an α-helix, should be energetically preferred, as long as side chains from the helix can make optimal interactions with the DNA bases. If movements of the main chain are required to optimize the interactions between side chains and bases, then use of a more flexible scaffold, like a β-sheet, might be favored. Upon DNA-binding, main chain movements do occur in the β-sheet of Arc and seem to be important in positioning key side chains for base contacts. The flexibility of Arc’s β-sheet may also be useful in allowing the protein to optimize contacts with its asymmetric operator sequence. We note that the overwhelming prevalence of symmetry in known protein-DNA interactions must be biased to some degree by the symmetric sequences which are generally chosen for crystallization.

ACKNOWLEDGMENTS

We would like to thank Simon Phillips for providing the coordinates of the MetJ complex, and Bronwen Brown, Dan Chasman, Nikola Pavletich, Chuck Kissenger, and Tom Ellenberger for valuable discussions and advice.
REFERENCES


Figure 1: Sequence of the DNA used for cocrystallization. The wild-type *arc* operator sequence is shown in bold, bases added for crystallization are in outline, and the TAGA box of each half-site is shaded. The locations of the tetramer and dimer pseudo-symmetry axes are indicated by filled and open ovals, respectively.
T A T A G T A G A G T G C T T C T A T C A T
T A T C A T C T C A C G A A G A T A G T A A
-10' -5' 0' 5' 10'
**Figure 2**: a. The Arc-operator complex. DNA atoms and the ribbon trace through phosphorous atoms are in white. The Cα ribbon traces of Arc monomers in the right dimer are yellow and red, those in the left dimer are blue and purple. Residues in the yellow and blue monomers are marked by primes in the text. For reference, the positions of residues 1, 13, 29, 34, and 48 are marked in one dimer. b. View of the β-sheet in the major groove of the right half-site. The operator half-site, the backbone of residues 8-14 of the dimer, and the side chains of residues 8, 9, 11, and 13 are shown. Atoms from one monomer are red and from the other are yellow. The DNA is white, except for the TAGA box which is blue. Phosphate oxygens contacted by the Arc dimer are represented as spheres. Hydrogen bonds are shown as green, dashed lines. Pro8 of each monomer is labeled.
a.
b.
Figure 3: Schematic of hydrogen bonds between bases of the right operator half-site and the Arc dimer. Hydrogen bonds between β-sheet side chains and bases are indicated by arrows, and hydrogen bonds between side chains in the β-sheet are indicated by dashed lines. The dimer symmetry axis is indicated by an open oval, and contacted bases are shaded. Gln9 and Gln9' accept hydrogen bonds from Asn11' and Asn11, respectively. Asn11 accepts a hydrogen bond from the Ne atom of Arg13 and donates a hydrogen bond to Asn11'. Equivalent contacts are formed in the left half-site, except Arg13' contacts an adenine at position -8 rather than a guanine.
**Figure 4:** Position of Phe10 in the Arc-operator complex. Side chains and C$_\alpha$ trace for residues 8 through 14 are shown in blue. DNA atoms are shown in white. The van der Waal surface of the Phe10 side chain and surrounding DNA backbone atoms are shown in blue and white, respectively.
Figure 5: Details of the contacts between an Arc dimer and the right operator half-site. Contacted phosphate oxygens or bases are shaded. The location of the dimer symmetry axis is indicated by an open oval. Hydrogen bonds are indicated by solid lines, electrostatic contacts by long dashes, van der Waals contacts by medium dashes, and hydrophobic interactions by short dashes. Triangles represent the methyl group of thymine, squares represent the N7 atom of adenine or guanine, circles represent the O4 or O6 atom of thymine or guanine, and diamonds represent the N4 or N6 atom of cytosine or adenine.
Figure 6: 

a. Electron density for β-strand residues 8 to 14 from an $|F_0| - |F_c|$ simulated annealing map in which these residues were omitted. For reference, the model for base pairs -4 to -6 and the water bridging Arg13 and base -4 are also shown. 

b. Hydrogen bonds in the region of Gln9. The side chains of Gln9, Asn11, and Asn11', and base pairs -5 and -6 are shown. 

c. Hydrogen bonds in the region of Gln9'. The side chains of Gln9', Asn11, and Arg13, and base pairs -4 and -5 are shown. In b and c, bonds near the viewer are darkened. Short dashes indicate hydrogen bonds between bases while all other hydrogen bonds are shown by long dashes. Figures b and c were generated using the program PREMA.
a.
CHAPTER 3

Dramatic Changes in DNA-Binding Specificity
Caused by Single Residue Substitutions in an
Arc/Mnt Hybrid Repressor
INTRODUCTION

One of the central questions in studies of protein-DNA interactions is how the sequence specificity of binding is achieved. The cocrystal structures of more than 20 protein-DNA complexes have now been solved, revealing discrete sets of molecular contacts between each protein and its target DNA (reviewed in Harrison, 1991; Pabo & Sauer, 1992. However, despite this growing catalogue of interactions, it has been difficult to change DNA-binding specificity by replacing one residue-base contact with another based upon manual or computer assisted modeling. Apparently understanding more than the atomic structure of the wild-type protein-DNA complex is needed to predict how mutations in either the protein or the DNA will affect specificity. Numerous altered, or at least broadened, specificity DNA-binding proteins have been described, but, in almost all cases, changes in specificity have been realized either by selection or screening methods (e.g., Youderian et al., 1983; Ebright et al., 1984; Wharton & Ptashne, 1987; Tzamarias et al., 1992; Jamieson et al., 1994; Rebar & Pabo, 1994) or by changing residues in a given DNA-binding protein to those present at homologous positions of a related protein (e.g., Wharton et al., 1984; Wharton & Ptashne, 1985; Lehming et al., 1987; Altschmied et al., 1988; Hanes & Brent, 1989; Knight & Sauer, 1989a; Treisman et al., 1989).

In this paper, we probe the determinants of DNA-binding specificity for Arc and Mnt, two members of the ribbon-helix-helix family of gene regulatory proteins (Raumann et al., 1994a). Arc and Mnt are small repressors which share 40% sequence homology but bind to distinct operator DNA sites in the genome of bacteriophage P22 (Sauer et al., 1983; Vershon et al., 1987b; Vershon et al., 1987a; Knight & Sauer, 1989a). In the cocrystal structure of the protein-DNA
complex, Arc dimers bind to each half of the 21 base pair operator. Most base-specific contacts are made from a two-stranded, anti-parallel $\beta$-sheet by side chains which reach into the major groove of the DNA (Raumann et al., 1994b). Residues N-terminal to the $\beta$-sheet are also involved in operator binding and form an arm which folds along the sugar-phosphate DNA backbone. Although Arc and Mnt recognize different operators, several studies indicate these proteins have similar structures and dock with their operator DNA’s in similar ways: (i) NMR studies show the first 45 residues of Mnt contain a short N-terminal arm and ribbon-helix-helix fold similar to Arc (Burgering et al., 1994); (ii) both proteins bind to 21 base pair operators as tetramers (Vershon et al., 1987b; Vershon et al., 1987a; Brown et al., 1990); (iii) residues in the N-terminal arm and $\beta$-sheet regions of both proteins are important for operator recognition (Youderian et al., 1983; Knight & Sauer, 1989b; Knight & Sauer, 1989a; Knight & Sauer, 1992; Brown et al., 1994); and, (iv) a hybrid composed of the arm and $\beta$-sheet residues of Arc and the remaining residues of Mnt binds specifically to the $arc$ operator (Knight & Sauer, 1989a). Mnt and Arc differ in their oligomeric state in solution: when not bound to DNA, Arc is a dimer (Vershon et al., 1985a), whereas Mnt has a C-terminal domain of approximately 30 residues that stabilizes tetramer formation (C. Waldburger and R. Sauer, in preparation).

Figure 1 shows the sequences of the $arc$ and $mnt$ operators and highlights the bases most important for binding of the cognate repressor (Vershon et al., 1989; Knight & Sauer, 1992). None of the four critical base pairs in each $arc$ operator half-site (the TAGA box) is conserved in the $mnt$ half-site. Conversely, only one of the six critical base pairs in each $mnt$ half-site is conserved in each $arc$ half-site. Because the key bases are very different in the two operator sequences, numerous protein-DNA contacts would apparently need to be altered to switch
recognition from one to the other operator. Structural and genetic studies suggest that three side chains in Arc and four in Mnt are responsible for many, if not all, of the contacts that permit these proteins to recognize their respective operators (Knight & Sauer, 1989b; Knight & Sauer, 1989a; Knight & Sauer, 1992; Brown et al., 1994; Raumann et al., 1994b). As shown in Figure 2, these critical side chains comprise three residues in the β-sheets of both Arc and Mnt (termed β2, β4, and β6) and a residue in the N-terminal arm of Mnt. The amino acid residues at β4 and β6 are identical in Arc and Mnt. This leaves β2 and the arm position as the residues most likely to be responsible for discriminating between the arc and mnt operators.

To help understand the role of β2 and the arm position in operator specificity, we have constructed hybrid Arc/Mnt molecules in which all four combinations of the Arc and Mnt residues are present at these two positions. We find that recognition of the arc operator requires the Arc residue at β2, while recognition of the mnt operator requires the presence of the Mnt residue at either the β2 or the arm positions. Thus, in the context of these hybrid proteins, a single residue change can alter binding specificity between two operators which differ at many critical base pairs. When these results are interpreted in light of the Arc-cocrystal structure, certain master side chains appear to determine specificity both directly by contacting bases and indirectly by programming other protein-base contacts. Moreover, this effect is amplified significantly since these proteins bind their operators as tetramers, with each monomer contacting the DNA.
RESULTS

Single amino acid changes can alter the specificity of the Arc/Mnt hybrid proteins

Each of the four proteins used for this study contains 85 residues and is a hybrid of Arc and Mnt (shown schematically in Fig. 2). The first 13 residues are largely identical to those in Arc, and the last 72 residues are identical to those in Mnt, including the C-terminal tetramerization domain. The four hybrid proteins differ only at the arm and \( \beta_2 \) positions (residues 5 and 9, respectively) and are named using the one-letter code for the residues at these positions. Thus, hybrid \( R^5Q^9 \) has arginine at the arm position and glutamine at \( \beta_2 \). The arm position is serine in Arc and arginine in Mnt, while \( \beta_2 \) is glutamine in Arc and histidine in Mnt.

The construction, purification, and DNA-binding properties of hybrid \( S^5Q^9 \) have been described previously (Knight & Sauer, 1989a). Hybrids \( S^5H^9 \), \( R^5Q^9 \), and \( R^5H^9 \) were constructed by standard methods of cassette mutagenesis. Each protein was purified to greater than 95% homogeneity and was found, as expected, to behave as a tetramer in gel filtration experiments (data not shown).

The binding affinities of the four hybrid proteins to DNA fragments containing the arc operator, mnt operator, or non-specific DNA were determined by gel mobility shift experiments (Fried & Crothers, 1981; Garner & Revzin, 1981) and are listed in Fig. 2. As shown in Fig. 3a, hybrid \( S^5Q^9 \) binds specifically to the arc operator, hybrids \( R^5H^9 \) and \( S^5H^9 \) bind specifically to the mnt operator, and hybrid \( R^5Q^9 \) binds to both operators. None of the hybrids bind strongly to non-specific DNA. Hybrids \( S^5Q^9 \), \( S^5H^9 \), and \( R^5H^9 \) show at least 2,000-fold preferences for one operator over the other operator, and all four hybrids show at least 2,000-fold preferences for operator DNA over non-specific DNA. Taken
together, these experiments show that the β2 and arm positions are sites of specificity control. In the Arc/Mnt hybrid context, binding to the arc operator requires glutamine at the β2 position, whereas binding to the mnt operator requires either histidine at β2 or arginine at the arm position.

The Arc/Mnt hybrid proteins bind to DNA in a manner similar to Arc and Mnt

To determine whether the hybrid proteins recognize operator DNA in the same basic way as Arc and Mnt, we performed footprinting experiments using hydroxyl radical cleavage (Dixon et al., 1991) and methylation protection (Gilbert et al., 1976). Fig. 3b shows examples of footprinting experiments, and Fig. 3c summarizes the resulting protection patterns in schematic form. Depending on the specificity of the hybrid, hydroxyl radical footprinting gives backbone protection patterns identical to those produced by Arc binding or Mnt binding. Thus, the hybrids seem to approach the sugar-phosphate backbone of the DNA in the same fashion as Arc and Mnt. For the arc operator, the protected backbone positions are known to be points of close approach between Arc and the operator DNA in the cocrystal structure (Raumann et al., 1994b).

In methylation experiments, Mnt and the R5H9 and R5Q9 hybrids protect the N7 positions of the same five guanines on the bottom strand of the mnt operator (Fig. 3c). The details of major groove packing are therefore likely to be the same for each of these proteins. Hybrid S5H9 protects four of these guanines but fails to protect G10. This result would be expected if R5 is responsible for the protection of G10 and is in accord with previous studies showing that an Arg→Lys mutation at the arm position of Mnt also fails to protect G10 (Knight & Sauer, 1992). Methylation protection is a less useful probe of the contacts in the major groove of the arc operator where Arc protects only two bases on the
bottom strand of the operator (G18 strongly and G15 weakly; Vershon et al., 1987b). Nevertheless, the methylation protection patterns observed on the arc operator for both the S$^5$Q$^9$ and R$^5$Q$^9$ hybrids differ from the pattern seen with Arc alone, indicating some differences in the fine structures of the protein-DNA interfaces.

*Altered specificity depends on protein context*

An S$^2$Q$^6$ variant of Mnt was constructed with the expectation that it would bind strongly to the arc operator, but in fact, this protein, which contains the arm and β2 residues of Arc, bound to neither operator. Apparently, the first three residues of Arc (present in the Arc/Mnt hybrid but absent in Mnt) are required for binding to the arc operator. This is reasonable as Met1 and Gly3 of Arc make a total of eight hydrogen bonds to the DNA backbone in the tetramer-operator complex (Raumann et al., 1994b). An R$^5$H$^9$ variant of Arc was also constructed with the hope that it might bind strongly to the mnt operator, but again this was not observed. Because R$^5$H$^9$ Arc and the R$^5$H$^9$ Arc/Mnt hybrid are identical over the first 13 residues but only the latter protein binds the mnt operator, residues C-terminal to position 13 and unique to Mnt must also be needed for binding to the mnt operator.

**DISCUSSION**

In previous work, the S$^5$Q$^9$ Arc/Mnt hybrid was shown to have the binding specificity of Arc (Knight & Sauer, 1989a). In light of the recently determined Arc-operator cocrystal structure, this result can be easily understood. The S$^5$Q$^9$ hybrid contains the arm and β-sheet residues of Arc, which make all of the base contacts and most of the phosphate contacts in the protein-DNA complex (Raumann et al., 1994b), and because Arc and Mnt have the same fold,
these residues would be positioned in the protein-DNA complex in similar ways. Although some Arc residues outside of the arm and β-sheet regions are known to be important for operator binding (Brown et al., 1994), these amino acids are almost completely conserved in Mnt and thus are retained in the hybrid protein.

Three additional findings emerge from the experiments reported in this paper. First and most importantly, a single change at the second position of the β-sheet (the β2 position) switches the binding specificity of the Arc/Mnt hybrid protein between the arc and mnt operators which differ at 8 to 10 important bases. Second, in the hybrid context, binding to the mnt operator can be conferred either by arginine at the arm position (R5Q9, R5H9) or histidine at the β2 position (S5H9, R5H9). Third, a single hybrid, R5Q9, binds tightly to both the arc and mnt operators but shows no increase in binding to non-specific DNA.

In the following sections, we discuss a plausible molecular model to account for these results. Fig. 4a shows the DNA contacts made by the arm and β-sheet positions of Arc in the operator complex (Raumann et al., 1994b). The cocrystal structure of Mnt has not been solved, but Fig. 4b shows a predicted set of protein-DNA contacts which were proposed originally based upon the binding of wild-type and mutant Mnt proteins to mnt operators altered by base substitution or chemical modification (Youderian et al., 1983; Vershon et al., 1987a; Knight & Sauer, 1989b; Knight & Sauer, 1992). We have confirmed that these predicted contacts are structurally plausible by model-building studies based upon the homologous Arc-operator complex.

Three key ideas are required to explain the affinity and specificity changes observed in the hybrid Arc/Mnt repressors: (i) Certain residues contact some bases directly and also influence other base contacts indirectly. For example, the
β2 side chains of Arc contact two adenines in the TAGA box of one half-site but also participate in a hydrogen bond network that helps to fix the conformations of the β4, β6, and arm residues, which themselves contact the DNA (Fig 4a). (ii) Some residues make degenerate sets of specific contacts depending upon structural context. For instance, in the cocrystal structure of Arc, the Asn11 and Asn11' side chains at the β4 positions contact two adjacent bases and also hydrogen bond to one another (Fig. 5) and to Gln9 and Gln9' at the β2 position. In the Mnt complex, the homologous asparagines at the β4 positions (Asn8 and Asn8') are predicted to have the same Cα positions but the side chains must rotate to make an alternative set of base contacts as shown in Figs. 4b and 5 (Knight & Sauer, 1992). Moreover, in the mnt-binding conformation, the hydrogen bond between the β4 side chains and the hydrogen bonds between the β4 and β2 side chains that are seen in the arc-complex could not form. (iii) These proteins bind operator DNA as tetramers (Vershon et al., 1987a; Brown et al., 1990), and thus the effects of a single mutation are amplified with the potential to alter as many as four sets of protein-DNA interactions.

The S5Q9 and S5H9 hybrids differ only by the Gln to His mutation at the β2 position (the β4, β6, and arm residues remain the same), and yet the S5Q9 hybrid recognizes the arc operator while the S5H9 hybrid recognizes the mnt operator. We assume that the change would alter several direct protein-DNA contacts (replacing four Gln-adenine contacts in the Arc complex with two His-guanine contacts in the Mnt complex, compare Fig. 4a and 4b) but would also cause the β4 and β6 residues to adopt conformations necessary for mnt operator recognition. This latter effect could be a result of forming a different set of hydrogen bonds and thereby stabilizing alternative conformations of these residues or a result of not participating in any hydrogen bonding network and
thereby allowing these residues to assume independent orientations. We favor
the latter for two reasons. First, Mnt variants bearing either Ala or Pro at the β2
position specifically recognize mnt operator variants altered at base pairs 5 and
17, indicating that operator recognition does not require a side chain capable of
hydrogen bonding at the β2 position (Youderian et al., 1983; Vershon et al.,
1985b; Knight & Sauer, 1989b). Additionally, the effects of operator mutations at
the two base pairs predicted to be contacted by the β2 and β4 positions (base
pairs 15 and 17) are additive, suggesting that Mnt interacts independently with
these two base pairs (Stormo et al., 1993).

The R5Q9 hybrid protein has the unusual property of being able to
recognize both the mnt and the arc operators. The binding to the arc operator is
easily rationalized since this hybrid has all of the Arc β-sheet residues, including
the critical glutamine at the β2 position. We presume the Ser5-DNA backbone
contact is replaced by a slightly stronger Arg5-backbone or base contact which
accounts for the increased affinity of the R5Q9 hybrid to the arc operator
compared to the S5Q9 hybrid. However, the increased affinity may also result
from some rearrangements at the protein-DNA interface, as the methylation
protection pattern of the R5Q9 hybrid bound to the arc operator is somewhat
different from wild-type Arc (Fig. 3c). The ability of the R5Q9 hybrid to bind the
mnt operator is somewhat surprising since it has the Arc β2 residue. The R5-mnt
operator contacts made by this hybrid must be sufficiently strong to replace the
lost contacts between the β2 histidines and the mnt operator and may also be
needed to reprogram the β4 and β6 residues to adopt the conformation necessary
for mnt operator recognition.

Both the β2 and arm residues appear to make direct base contacts in the
Mnt-operator complex (Fig. 4b) (Knight & Sauer, 1989a; Knight & Sauer, 1992),
but in the hybrid proteins only one of these positions needs to have the Mnt side chain for mnt operator recognition. Moreover, when both Mnt side chains are present at the β2 and arm positions in the R5H9 hybrid, only a very small (≈2-fold) increase in binding affinity is observed over hybrids S5H9 or R5Q9 which contain one or the other Mnt side chain (Fig. 2). This lack of additivity in the hybrid context suggests that both side chains are not capable of making optimal interactions with mnt operator DNA at the same time.

Single amino acid substitutions have been shown to change the specificity of other DNA-binding motifs such as the homeodomain (Hanes & Brent, 1989; Treisman et al., 1989), the helix-turn-helix motif (Wharton & Ptashne, 1987; Bass et al., 1988), and the bZIP family of proteins (Tzamarias et al., 1992; Suckow et al., 1993). In most of these cases, changes in specificity are between DNA sites differing at one or two base pairs. The Arc/Mnt hybrids provide a rare example in which a single amino acid substitution switches specificity at a much larger number of base pairs.

The altered specificity mutations described here provide a dramatic illustration of the cooperative nature of DNA recognition, in which the roles of some residues depend both on their ability to contact bases directly and their interactions with other residues. This type of cooperative DNA recognition is not limited to the ribbon-helix-helix family of DNA-binding proteins. Many protein-DNA complexes show intricate networks of side-chain interactions which seem to play important roles in DNA-binding (Aggarwal et al., 1988; Clarke et al., 1991; Luisi et al., 1991; Winkler et al., 1993; Benevides et al., 1994). The interdependence of individual protein-DNA contacts in such systems clearly makes the rules of DNA recognition more cryptic and makes the rational design of DNA-binding proteins more complicated. However, from an evolutionary
perspective, it may allow proteins to acquire dramatically new DNA-binding specificities by changing a minimal number of side chains.

**MATERIALS AND METHODS**

*Cloning and purification of proteins*

Plasmids encoding the R⁵Q⁹, S⁵H⁹, and R⁵H⁹ Arc/Mnt hybrids, R⁵H⁹ Arc, and S²Q⁶ Mnt were constructed with double stranded cassettes containing the appropriate base changes using a procedure described previously (Knight & Sauer, 1989a). Because of the low expression levels of R⁵H⁹ Arc, the R⁵H⁹ mutation of Arc was subcloned into an Arc gene containing the lt1 mutation. The lt1 mutation is a frame shift mutation which changes the last residue of Arc from Ala to Arg and adds 25 extra C-terminal amino acids and protects Arc from proteolytic degradation (Bowie & Sauer, 1989). *In vitro* biochemical experiments have shown that the lt1 mutation does not affect the structure, stability, or operator affinity of Arc. Hybrid proteins, S²Q⁶ Mnt, and R⁵H⁹ Arc-lt1 were purified by ion-exchange chromatography on a phosphocellulose column and by gel filtration chromatography on a Sephadex-G75 (fine) column (Knight & Sauer, 1989a) to greater than 95% homogeneity as judged by SDS-polyacrylamide gel electrophoresis.

*Determination of DNA-binding affinity*

DNA affinity was determined using a gel mobility shift assay. Operator DNA was the 240 base pair EcoRI/EcoRV fragment from pAO100 (arc operator) or pMO100 (mnt operator), and non-operator DNA was the 750 base pair PstI/EcoRI fragment from pBR322. Alternatively, DNA used in the gel mobility shift assay was a synthetic 30 base pair fragment which contained either the *arc*
operator, the mnt operator, or the lambda Oq1 operator (non-specific DNA). The DNA-binding affinity is the value obtained with the 30 base pair fragment for the R5Q9 hybrid and with the 240 base pair fragment for the rest of the proteins. However, control assays showed the size of the operator fragment did not affect DNA-binding affinity. DNA-binding reactions were performed at 22 °C in 50 mM Tris-Cl (pH 7.5), 10 mM MgCl2, 0.1 mM EDTA, 150 mM KCl, and 100 µg/ml bovine serum albumin. Reactions (50 µl) contained approximately 1000 cpm of end-labeled operator DNA and the appropriate concentration of Arc, Mnt, or Arc/Mnt hybrid. Following a 60 min incubation, 30 µl were electrophoresed on 5% TBE acrylamide gels. The amounts of bound and free operator were quantified by scanning laser densitometry of autoradiograms using an LKB 2202 Ultrosan laser densitometer with a 2202 recording integrator. Other details of the assay have been described (Vershon et al., 1987a; Knight & Sauer, 1988).

**DNA footprinting assays**

Hydroxyl radical protection assays were performed as described (Tullius & Dombroski, 1986; Knight & Sauer, 1988) at 22 °C in 10 mM Tris-Cl (pH 7.5), 10 mM MgCl2, 0.1 mM EDTA, 150 mM KCl, 1 mM CaCl2, 2.5 µg/ml sonicated salmon sperm DNA, and 100 µg/ml bovine serum albumin. Methylation protection reactions were performed as described (Knight & Sauer, 1988) at 22 °C in 50 mM sodium cacodylate (pH 7.5), 10 mM MgCl2, 0.1 mM EDTA, 1 mM DTT, 150 mM KCl, 1 mM CaCl2, 5 µg/ml sonicated salmon sperm DNA, and 250 µg/ml bovine serum albumin. Other details of the assay have been described (Knight & Sauer, 1988).
Molecular Modeling

The Mnt-operator interface was modeled based on the cocrystal structure of the Arc-operator complex. When modeling the *mnt* operator structure, the coordinates of the sugar-phosphate backbone and ribose atoms were not changed. The model was used to confirm that contacts previously predicted from biochemical and genetic studies of Mnt and its operator site are compatible with the framework of the Arc repressor-operator complex. The coordinates of the DNA were not changed except to replace the sequence of the *arc* operator with that of the *mnt* operator by superimposing the new base onto the old base using the C6, N1, and C2 atoms for pyrimidines and the C8, N9, and C4 atoms for purines. The amino acid residues of the Arc complex at the arm, β2, β4, and β6 positions were replaced with those of Mnt. Coordinates of the protein backbone and Cβ atom were not changed. Side chain dihedral angles were adjusted so that contacts determined from genetic and biochemical studies could be accounted for by the approach of a hydrogen bond donor and acceptor to within hydrogen bonding distance with no overlap of van der Waals surfaces. For one half-site, the protein/DNA contacts modeled were as follows: the Ne2 of His6 to the O6 of G5, the Nδ of Asn8 and Asn8' to the O6 of G7 and G4, respectively, the Nh2 of Arg10 to the O6 of G8, and the Nh2 of Arg2 to the N7 of G10.

ACKNOWLEDGMENTS

We thank Bronwen Brown, Tracy Smith, and Kevin Shoemaker for advice and discussions.
REFERENCES


Figure 1. The \textit{arc} and \textit{mnt} operator sequences. Base pairs critical for DNA-binding are boxed and shaded, and critical bases that are the same in the \textit{arc} and \textit{mnt} operators are marked by arrows. The vertical line indicates the pseudo-symmetry axes for both operators.
<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
</tbody>
</table>
**Figure 2.** N-terminal sequences and DNA-binding affinities of Arc, Mnt, and the hybrid proteins. The affinities shown are normalized protein concentrations (1 = 2 x 10^{-11} M) at which one-half of the DNA is bound as determined by a gel mobility shift assay. Values for Mnt and hybrid proteins are in terms of tetramer concentration, whereas values for Arc are in terms of dimer concentration. The values shown for Arc, Mnt and the S^5Q^9 hybrid are taken from Knight and Sauer (1989a). Filled circles under the sequences mark residues that make hydrogen bonds to bases in the Arc-operator cocrystal structure (Raumann et al., 1994b) or, in the case of Mnt, residues predicted to make direct base contacts from genetic and biochemical studies (Knight & Sauer, 1992).
Figure 3. DNA-binding of the hybrid proteins. a. Gel mobility shift assay of the hybrid proteins. End labeled DNA containing either the mnt operator (left panel), arc operator (middle panel), or non-specific DNA (right panel) was incubated in the absence of protein or in the presence of Mnt, Arc, $S^5Q^9$ hybrid, $S^5H^9$ hybrid, $R^5H^9$ hybrid, or $R^5Q^9$ hybrid at $5 \times 10^{-10}$ M. The products of each reaction were electrophoresed on a 5% acrylamide gel and visualized by autoradiography. b. Methylation protection footprints of hybrid proteins. End labeled DNA containing adjacent arc and mnt operators was incubated in the absence of protein (control) or in the presence of Mnt ($4 \times 10^{-8}$ M), Arc ($4 \times 10^{-7}$ M), $S^5Q^9$ hybrid ($2 \times 10^{-9}$ M), $S^5H^9$ hybrid ($2 \times 10^{-9}$ M), or $R^5H^9$ hybrid ($2 \times 10^{-9}$ M). Following methylation and DNA cleavage, products were electrophoresed on an 8% acrylamide gel and visualized by autoradiography. c. Methylation protection footprint of the $R^5Q^9$ hybrid. Footprinting was performed as above, using Mnt, Arc, and $R^5Q^9$ hybrid at $2 \times 10^{-7}$ M. d. Hydroxyl radical and methylation protection footprinting studies. A schematic summary of both the hydroxyl radical and methylation protection patterns. Footprinting patterns for Arc, Mnt and the $S^5Q^9$ hybrid are taken from Knight and Sauer, 1988 and 1989a. Bars above the operators indicate regions of hydroxyl radical protection, and filled ovals below the operators indicate bases protected from methylation by protein binding. Symbols are lightly shaded for weak protections. Arrows mark regions of missing protections when compared to Arc or Mnt.
R°Q° Hybrid Mnt

Arc Control

mnt operator
G-18
G-17
G-10
G-8
G-7

arc operator
G-18
G-15
G-13
G-11
Figure 4. Summary of DNA contacts made by the arm, β2, β4, and β6 positions. Solid lines with arrows represent hydrogen bonds between a base and a protein residue, and dashed lines represent hydrogen bonds between protein residues. Bases most important for DNA recognition, the arm position, and the β2 position are shaded. Filled ovals indicate the pseudo-symmetry axis of the operators. 

a. Contacts seen in the Arc-operator complex (Raumann et al., 1994b).

b. Plausible contacts between Mnt and the mnt operator based on genetic and biochemical studies (Vershon et al., 1987a; Knight & Sauer, 1989b; Knight & Sauer, 1992) and modeling based on the Arc-operator cocrystal structure.
Figure 5. Alternative DNA contact modes of the β4 position. Dashed green lines indicate hydrogen bonds. Relevant oxygen and nitrogen atoms are red and blue spheres, respectively. Coordinates for atoms in the Arc complex are taken from the cocrystal structure of the Arc-operator complex. Coordinates for atoms in the Mnt complex are a result of modeling the Mnt β-sheet residues and mnt operator sequence into the Arc-operator complex.
CHAPTER 4

Major Groove DNA Recognition by β-Sheets: The Ribbon-Helix-Helix Family of Gene Regulatory Proteins
Most well-characterized DNA-binding proteins have historically belonged to families employing α-helices for DNA recognition (for review, see Harrison, 1991; Pabo & Sauer, 1992). Recently, however, the structures and DNA recognition strategies of several proteins using β-sheets for DNA-binding have been elucidated. In this chapter, we will focus on the ribbon-helix-helix family which includes the bacterial MetJ repressor and TraY proteins and the phage P22 Arc and Mnt repressors (Breg et al., 1990; Phillips, 1991). Structural studies are most advanced for the MetJ and Arc repressors. The crystal structure of MetJ is known (Rafferty et al., 1989), as are the NMR structure (Breg et al., 1990; Bonvin et al., 1994) and crystal structure of Arc (U. Obeysekare, C. Kissenger, L. Keefe, M. Rould, B. Raumann, R. Sauer & C. Pabo, in preparation). More importantly, the cocrystal structures of both the MetJ-operator (Somers & Phillips, 1992) and Arc-operator complexes (Raumann et al., 1994) have been determined. These proteins use a two-stranded antiparallel β-sheet to recognize determinants in the major groove of their recognition sites. The eukaryotic TATA-binding proteins (TBP’s) comprise a second family that uses β-sheets to recognize DNA, although in this case, binding occurs in the minor groove (Kabsch & Sander, 1983; Kim et al., 1993a; Kim et al., 1993b).

Protein Structure

Arc, at 53 residues, is the smallest member of the ribbon-helix-helix family and can be used to illustrate the basic fold. As shown in Fig. 1, the Arc dimer is formed by the intertwining of identical monomers, each of which contains a β-strand and two α-helices (designated helix A and helix B). Dimerization is required for stable folding of Arc (Bowie & Sauer, 1989; Silva
et al., 1992; Peng et al., 1993), presumably because both monomers are needed to form the hydrophobic core, and because the β-strands of the monomers pair to form an antiparallel β-sheet at the dimer interface. Because the dimer contains two ribbon-helix-helix motifs, we will refer to the basic structure as the (βαα)2 fold.

All other family members appear to adopt basic structures similar to Arc. For example, although MetJ is larger (104 residues) and contains additional structural elements, it is stable only as a dimer (Johnson et al., 1992) and has a homologous (βαα)2 fold (Rafferty et al., 1989). There are, however, small variations in the (βαα)2 fold between Arc and MetJ. For example, the β-strand in MetJ is two residues longer, helix A is one turn longer at its C-terminus, and there are differences in the turn between the two helices. Despite these variations, the overall alignment of the (βαα)2 units is quite good (1.2 Å rms deviation of Cα positions for Arc residues 8-28 and 32-45 and MetJ residues 22-42 and 52-65 (Somers & Phillips, 1992)). The complete structure of Mnt (82 residues) is not yet known, but NMR studies confirm the presence of a (βαα)2 core (Burgering et al., 1994). TraY appears to use an interesting variation on the (βαα)2 theme. Sequence homology indicates that each monomer (131 residues) contains two repeats of the βαα motif (Nelson et al., 1993), and as a result, both repeats in the same polypeptide chain could fold together and allow TraY to adopt the (βαα)2 fold as a monomer. An exciting recent discovery is the presence of sequence patterns indicative of the ribbon-helix-helix fold in auxin-induced plant proteins (Abel et al., 1994). These proteins may represent the first eukaryotic members of the ribbon-helix-helix family.
Organization of Operator Sites

The naturally occurring operator DNA sites for Arc, Mnt, and TraY have adjacent binding sites for two protein dimers or \((\beta\alpha\alpha)_2\) units, and met operators have two to five tandem binding sites (met boxes) for MetJ dimers (Phillips, 1991). Biochemical studies show that Arc dimers can bind to operator half-sites with reasonable affinity (ca. 1 nM), but they also demonstrate that cooperative binding of a second dimer yields an additional 5 kcal/mol of interaction energy which results in a kinetically stable, tetrameric operator complex (Brown & Sauer, 1993). These observations suggest that the biologically active, DNA-bound forms of the \((\beta\alpha\alpha)_2\) proteins are likely to be tetramers or higher-order oligomers. Whether this will always be true for members of this family or whether a dimer or \((\beta\alpha\alpha)_2\) unit bound to a single site could, in some instances, regulate gene expression remains an open question.

Overview of MetJ and Arc Protein-DNA Interactions

The cocrystal structure of MetJ bound to a 16 bp operator site containing two consensus met boxes has been determined at 2.8 Å resolution (Somers & Phillips, 1992) and the structure of Arc bound to its wild-type 21 bp operator site has been determined at 2.6 Å resolution (Raumann et al., 1994). The DNA-binding properties of both wild-type proteins and numerous mutants have also been studied biochemically (Knight & Sauer, 1989; Phillips et al., 1989; Vershon et al., 1989; Brown et al., 1990; He et al., 1992; Brown & Sauer, 1993; Brown et al., 1994). These structural and biochemical studies reveal several similarities between the two proteins and their DNA complexes: (i) each protein binds to DNA as a tetramer, with dimers binding to each half of
the operator and making cooperative interactions with the neighboring dimer (see Fig. 2c and 2d); (ii) the left and right half of each protein-DNA complex is generally symmetric; (iii) the general position and orientation of the (βαα)_2 motif relative to each DNA half-site is conserved between the two complexes (1.5 Å rms deviation for common C_α and phosphorous atoms; compare Fig. 2a and 2b); (iv) all of the critical base contacts are made by residues from the antiparallel β-sheet which lies flat in the major groove and parallel to the DNA backbone; (v) homologous DNA phosphate contacts are made by main chain -NH groups from helix B of each protein subunit and from a tandem-turn region which precedes the β-strand; and (vi) linkage between the phosphate contacts and the β-sheet contacts appears to be crucial for binding. Notable differences between the MetJ-DNA and Arc-DNA complexes include: (vii) the spacing of half-sites in the operators and the nature of cooperative interactions between dimers; (viii) variations in DNA structure for the two operators; and (ix) the roles of β-sheet residues in terms of the symmetry and spacing of DNA contacts, interactions between side chains, and conformational adjustments accompanying DNA-binding. In the sections that follow, we discuss many of these issues in greater detail.

**Conserved Contacts with the Phosphate Backbone**

In both the Arc and MetJ protein-DNA complexes, phosphate backbone contacts seem to play critical roles in fixing the (βαα)_2 motif to the DNA and in orienting the β-sheet for DNA-binding. The pattern of phosphate backbone contacts made by each Arc or MetJ dimer is very similar (Somers & Phillips, 1992; Raumann et al., 1994). If we take the β-sheet in the major groove as a point of reference, then both Arc and MetJ contact three homologous phosphates on each DNA strand flanking the β-sheet.
Moreover, corresponding regions of the protein dimers contact the same phosphates. Thus, phosphates $a$, $a'$, $b$, and $b'$ are contacted by the main-chain and side chains of residues in the tandem-turn region of Arc and MetJ, and phosphates $c$ and $c'$ are contacted by the main-chain amide hydrogens from the N-terminus of helix B in both proteins (Fig. 3). There are also non-conserved backbone contacts, although biochemical and mutant analyses indicate that some of these are relatively unimportant for binding (He et al., 1992; Brown et al., 1994). In the Arc complex, phosphates $d$ and $d'$ are contacted by Arg23 (in helix A), while in the MetJ complex, Arg40 (in helix A) makes additional contacts to phosphates $c$ and $c'$ and Lys22 (in the $\beta$-sheet) contacts phosphates $e$ and $e'$.

A surprisingly large number of phosphate contacts in the arc and met operators (70% for Arc; 50% for MetJ) involve hydrogen bonds donated by --NH groups of the protein main-chain, as opposed to contacts mediated by flexible side chains such as Arg or Lys. These contacts between the protein backbone and the DNA backbone may provide an especially effective system for rigidly attaching the $(\beta\alpha)_{2}$ motif to the DNA. For both Arc and MetJ, the $\beta$-sheets lie in the center of the major groove, but the fit is not sufficiently snug to prevent twisting or wobbling of the sheet in the absence of additional stabilizing contacts. Both proteins solve this problem in the same way. They attach other parts of the protein (the tandem-turn region, or the N-terminus of helix B) firmly to the DNA backbone and then link these parts to the $\beta$-sheet. In Arc, for example, the main-chain of Asn34 in helix B makes a phosphate backbone contact while the side chain of the same residue forms hydrogen bonds to the main-chain of Arg13 in the $\beta$-sheet. Although the Asn34 side chain serves only to connect the DNA phosphate contacts to the $\beta$-
sheet contacts, mutagenesis studies show that this side chain is as energetically important as the majority of the β-sheet base contact residues (Brown et al., 1994). The side chain at the corresponding position in MetJ (Ser54) plays an analogous role in linking the helix B phosphate contacts to the β-sheet base contacts (Somers & Phillips, 1992).

**DNA Contacts from Flexible Regions**

In both the Arc and MetJ complexes (Somers & Phillips, 1992; Raumann et al., 1994), the seven residues before the β-strand adopt compact conformations consisting of two reverse turns, a type II turn (Arc residues 1-4; MetJ residues 13-16) followed by a type I turn (Arc residues 4-7; MetJ residues 16-19). In the unbound proteins, these residues assume different conformations. In MetJ they form a loop with high thermal factors and different conformations depending on the crystal environment (Rafferty et al., 1989), and in Arc they are largely disordered in solution (Breg et al., 1990; Bonvin et al., 1994) and in the crystal. The tandem-turn regions mediate four critical phosphate backbone interactions in each half-site of the protein-DNA complexes (Fig. 3). In Arc, the contacts made by the tandem-turn region are linked to the body of Arc, and thus to the β-sheet contacts, by a set of hydrophobic packing interactions mediated by the methionine side chains at positions 1, 4, and 7. Mutations at each of these residues reduce DNA-binding, suggesting that this linkage is important (Brown et al., 1994). In MetJ there seem to be fewer contacts linking the tandem-turn region to the β-sheet.

Why do both proteins use regions of the protein that are flexible in solution (or at least able to assume alternate conformations) to make
important backbone contacts in the complex? Some flexibility in these regions might allow the bound proteins to adapt to changes in DNA conformation in solution as the operator bends or writhes. Such flexibility could also be useful if dimers were to bind initially to straight DNA and then to adjust as the DNA assumed the bent conformation seen in the tetrameric complexes (see below). (The tandem-turn regions appear to retain some flexibility even when folded against the DNA backbone, as the main chain atoms of all four tandem-turns in MetJ and the outer two motifs in Arc have relatively high thermal factors; Somers & Phillips, 1992; Raumann et al., 1994.) Another possibility is that folding of the tandem-turn region against the phosphate backbone is coupled in some fashion to the formation of specific contacts by the β-sheet in the major groove. Such a mechanism could reduce binding to non-operator sites, while still allowing stable complex formation on the correct operator sequences.

DNA Structure

The met and arc operators both show an overall bend of about 50° in the protein-DNA complexes (Somers & Phillips, 1992; Raumann et al., 1994). These bends are probably required to allow the bound dimers to interact cooperatively, as dimers modeled on linear DNA do not contact each other. Nevertheless, the remaining details of the DNA structures are rather different in the two complexes, and hence are unlikely to be of significance for the (βαα)2 family as a whole. For example, there are differences in the average helical repeat in the met (10.6 bp/turn) and arc (11.1 bp/turn) operators, and most of the DNA bending in the met operator occurs near the center of each half-site, whereas the majority of the bending in the arc operator occurs near the center of the whole operator. Moreover, the groove
widths vary between the two complexes. In the met operator, the major grooves are compressed around the β-sheets, whereas in the arc operator, the major groove is widened around the β-sheets and the minor groove is severely compressed near the center of the site.

In both operators, some base mutations diminish operator binding even though the proteins contact only phosphates at these positions (Vershon et al., 1989; Phillips et al., 1993). Although such effects could be mediated by global conformational changes, it is also possible that the base sequence is important in determining local phosphate conformation. This appears especially likely for MetJ where the phosphates contacted by helix B are some 2 Å from their expected positions in B-DNA and the base sequence at these positions is sensitive to mutations (Somers & Phillips, 1992). Many of the phosphates contacted in the arc operator are also at positions where the backbone is distorted, however while base substitutions reduce binding at some of these positions, at others (including the most severely distorted near the center of the operator) DNA-binding is insensitive to sequence changes (Vershon et al., 1989).

Dimer-Dimer Contacts and Cooperativity

Cooperativity is extremely common in the binding of gene regulatory proteins to multivalent operators, and it is observed in the binding of both Arc and MetJ to their operators (Phillips et al., 1989; Brown et al., 1990; Brown & Sauer, 1993) (YY He, T McNally, I Manfield, I Parsons, SEV. Phillips, PG. Stockley, in preparation). In the MetJ-operator complex and the Arc-operator complex, we are afforded two molecular views of the types of protein-protein interactions that can stabilize tandemly-bound regulatory molecules.
The quaternary relationships between the dimers in the DNA-bound tetramers are quite different for Arc and MetJ, as are the spacings of the half-sites in the two operators (Somers & Phillips, 1992; Raumann et al., 1994). In the 21 bp arc operator, the centers of the half-sites are 11 bp or one turn of the DNA helix apart. This spacing allows both Arc dimers to bind to the same face of the DNA helix. By contrast, in the 16 bp met operator, the centers of the half-sites are only eight bp apart. Thus, the bound MetJ dimers are closer together than the Arc dimers by about 10 Å (3 x 3.4 Å) and are related by a rotation of approximately 100° (3 x 34°). The observed dimer-dimer interactions in both the MetJ and Arc complexes appear to require the precise wild-type spacing of half-sites. Indeed, changing these spacings by even a single base pair in either operator leads to substantial reductions in DNA-binding (Phillips et al., 1989; Vershon et al., 1989), presumably because the tetramer interface cannot form properly. As a consequence of the different half-site spacings, MetJ and Arc are forced to use different tetramer interfaces (cf. Fig. 2c and 2d). Arc dimers interact with each other over a limited region (ca. 325 Å²/dimer) using residues in the loop between helix A and helix B; MetJ dimers pack against each other over a larger region (ca. 550 Å²/dimer) along the length of helix A. In both cases, some of the stabilizing quaternary interactions involve hydrogen bonds. In Arc the primary interactions between dimers are made by two Arg31 side chains, which span the interface and make two charge-stabilized hydrogen bonds with the backbone carbonyl oxygen of Asn29 in the neighboring dimer. In MetJ, interactions at the dimer-dimer interface include extensive hydrophobic contacts as well as hydrogen bonds mediated by water molecules (e.g. Thr37-water-Gln44'). Mutation of Arg31 in Arc or Thr37 in MetJ results in severely reduced binding, confirming
the importance of these cooperative interactions in stabilizing the overall complexes (He et al., 1992; Brown et al., 1994).

Neither Arc nor MetJ forms tetramers in solution, indicating that the protein dimers need to be brought together and oriented, at least approximately, by DNA-binding before they can interact productively. In each DNA complex, one protein dimer can be superimposed on the other by a rotation about the two-fold axis passing through the center of the operator site, or by translation along and rotation about the DNA axis. The latter method of superposition provides a useful perspective because a second translation/rotation operation would generate a third cooperatively-bound dimer, another event would generate a fourth cooperatively-bound dimer, and so on. On a DNA template, such sequential addition of further dimers is presumably limited by the number of tandem binding sites (e.g., five in the metF operator).

β-Sheet/Base Contacts and DNA Recognition

So far, we have described many of the general architectural features of the MetJ-operator and Arc-operator complexes. However, the ability of these proteins to recognize their operator sites with high affinity requires specific readout of the sequence information in the major groove which is mediated by residues from the β-sheet (He et al., 1992; Brown et al., 1994). Side chains from the exposed face of the β-sheet make all of the base-specific hydrogen bonds in both MetJ and Arc, and not surprisingly, the general way in which the β-sheets of the two proteins approach the DNA is very similar (the Ca positions of corresponding β-sheet positions have an rms deviation of 1.15 Å in the aligned half-site complexes; Somers & Phillips, 1992; Raumann et al.,
1994). There are, however, some very surprising differences in the way MetJ and Arc use their β-sheet residues for base contacts. First, the side chains at analogous positions in the two proteins make radically different base contacts, both with respect to the spacing of these contacts and in terms of the symmetry of the contacts. Second, the β-sheet DNA-binding side chains in Arc are linked by an intricate hydrogen bonding network, which is not present in MetJ. Third, the β-sheet region of Arc, but not that of MetJ, undergoes significant conformational changes upon DNA-binding.

The base contacts of MetJ and Arc can be compared most easily by describing the interactions made by residues at three corresponding, solvent exposed positions in the β-sheets. For each of these corresponding positions, there are two symmetric side chains, one from each strand of the β-sheet. Reference to Fig. 4 will be helpful in comparing these contacts.

Position 1: In MetJ, Lys23 and Lys23' make symmetric contacts with guanines on opposite DNA strands. As shown in Fig. 4, these contacts are separated by four bp. The corresponding residues in Arc, Gln9 and Gln9', make symmetric contacts with adenines on opposite DNA strands, but these bases are in adjacent bp (Fig. 4). The dramatically different spacings of the contacts made by these analogous MetJ and Arc positions arise largely because of differences in side chain rotation. The Arc side chains are rotated towards the center of the half-site, while the MetJ side chains point outward.

Position 2: In MetJ, Thr25 and Thr25' make symmetric contacts with adenines on opposite DNA strands (these contacts are separated by two base pairs). Thr25 and Asn11 (the analogous residue in Arc) actually contact the same position in the aligned half-sites, but this is the only instance in which
corresponding residues contact corresponding DNA positions (see Fig. 4).
Asn11 and Asn11’ contact adjacent bases (cytosine and thymine) on the same DNA strand. Thus, the Asn11 and Asn11’ contacts are asymmetric in terms of DNA strands, in terms of the type of base contact, and in terms of their positions relative to the two-fold axis of the half-site.

Position 3: In Arc, Arg13 and Arg13’ make asymmetric contacts with guanines on the same DNA strand, separated by three bases (Fig. 4). At the corresponding β-sheet positions in MetJ, the Ser27 and Ser27’ side chains are too short to reach the bases. Although there are no DNA contacts from this position in MetJ, substituting alanine for Arg13 in Arc reduces operator binding by more than 15 kcal/mol (Brown et al., 1994). By contrast, Arc mutants with alanine substitutions at the other key β-sheet positions (Gln9 and Asn11), show smaller reductions in binding free energy (5-7 kcal/mol).

From the foregoing discussion, it can be seen that MetJ makes symmetric contacts with its operator, while Arc makes largely asymmetric contacts. Some of these differences undoubtedly arise because the consensus met box sequence is symmetric while the wild-type arc half-sites show only limited symmetry. In binding to its asymmetric operator, it is important to note that Arc does not maintain the inherent symmetry of the antiparallel β-sheet and then make the best DNA contacts possible within this constraint. Rather, the inherent symmetry is broken to allow Asn11, Asn11’, Arg13, and Arg13’ to make the best available DNA contacts. It is also worth noting that only two of 43 naturally occurring met boxes actually contain the symmetric 8 bp consensus sequence used for cocrystallization with MetJ (Somers & Phillips, 1992). It would be interesting to know how MetJ interacts with some of these natural but less symmetric operator sites.
In Arc, most of the exposed β-sheet side chains hydrogen bond with each other as well with the DNA bases (see Fig. 4). It is likely that these interactions contribute to the overall stabilization of the Arc-DNA complex both because they form only in the complex (only one of these hydrogen bonds is observed in the unbound Arc dimer) and because they reduce the unfavorable entropy associated with DNA-binding by stabilizing the conformations of the otherwise flexible β-sheet side chains. There are, however, no interactions between the β-sheet DNA contact residues of MetJ, suggesting that such stabilizing interactions are not a general family trait.

Excluding the tandem-turn region, only very minor conformational changes are observed between the unbound and DNA-bound MetJ dimers (Rafferty et al., 1989; Somers & Phillips, 1992). Although this is also generally true of Arc, there are some intriguing local changes in the β-sheet that accompany DNA-binding. In the complex, the Phe10 side chain of Arc swings out of the hydrophobic core and part of the phenyl ring becomes buried by DNA instead of protein. At the same time, there are adjustments of up to 1.7 Å in the Cα positions of residues 9, 10, and 11 in the β-sheet. These peptide backbone movements may be needed to optimize interactions between side chains and bases, and thereby to allow Arc to make asymmetric and energetically significant interactions with its operator DNA.

**DNA Recognition by Mnt**

Although neither the structure of P22 Mnt repressor nor its operator complex is known, the results of biochemical and genetic experiments allow some comparisons with Arc (Mnt's closest relative) and MetJ. Although the *mnt* and *arc* operators have very different DNA sequences, both are 21 bp in
length and have the same 11 bp spacing of half-sites (Knight & Sauer, 1992). Moreover, Mnt and Arc must bind operator DNA in a fundamentally similar manner since a hybrid protein containing the first nine residues of Arc and the last 75 residues of Mnt binds strongly and specifically to the arc operator (Knight & Sauer, 1989).

Because the operator binding of Mnt and Arc is similar, one might assume that Mnt’s β-sheet residues should interact with the mnt operator in the same basic manner observed for Arc. Indeed, for the first exposed β-sheet position in Mnt, genetic and biochemical studies (Youderian et al., 1983; Knight & Sauer, 1992) suggest it makes DNA contacts analogous to those seen for the corresponding position of Arc. However, the second and third exposed β-sheet residues are identical in Mnt and Arc (Asn8 and Arg10 in Mnt; Asn11 and Arg13 in Arc), and yet genetic and biochemical studies indicate that the Mnt side chains at these positions make different base contacts than the homologous Arc side chains (Knight & Sauer, 1992). In fact, the contacts made by the second position in the Mnt β-sheet are more similar to those made by the corresponding positions of MetJ. Since the Arc and Mnt β-sheets differ only at the first exposed position, it may be that the identity of this residue determines how the remaining β-sheet residues contact the DNA. This could be accomplished through hydrogen bonding interactions such as those seen between the β-sheet side chains in the Arc cocrystal.

Interestingly, the Mnt protein forms a tetramer in solution (Vershon et al., 1985) stabilized by a C-terminal domain of approximately 30 residues (C. Waldburger and R. Sauer, in preparation), which are absent in Arc. The solution tetramers seem to have D2 symmetry instead of the same symmetry as the DNA-bound tetramers of MetJ or Arc. This difference in symmetry can
explain why Mnt does not indefinitely polymerize in solution, however it raises the problem of mixed symmetry between the C-terminal tetramerization domain and the tetramer-operator complex. Perhaps the N-terminal, DNA-binding domain of Mnt is flexibly linked to the C-terminal domain.

_Closing Remarks_

From the available structural, genetic, and biochemical studies of ribbon-helix-helix proteins, it is possible to draw several conclusions:

(1) In MetJ and Arc, the structures and DNA-binding functions of the core \((\beta\alpha\alpha)_2\) fold and the tandem-turn regions are strongly conserved. Interactions that appear to be critical for the DNA recognition process are the major groove contacts from the \(\beta\)-sheet, the phosphate backbone contacts from the N-terminus of helix B and the tandem-turn region, and the structural linkage between the base-contacts and the phosphate-contacts.

(2) Although the biological activity of both MetJ and Arc depends upon cooperative interactions between DNA-bound dimers, the tetramer interfaces used by each repressor for cooperativity are quite different. This allows recognition of operators with different lengths and half-site spacings.

(3) The major groove DNA contacts made by the analogous \(\beta\)-sheet positions of Arc, MetJ, and Mnt are sufficiently different to preclude any simple rules that match particular \(\beta\)-sheet positions with particular base contacts.

(4) The antiparallel \(\beta\)-sheet provides a versatile scaffold from which to recognize bases in the major groove of DNA. The known ribbon-helix-helix
proteins recognize a variety of DNA sequences, and despite the inherent symmetry of the motif can recognize asymmetric DNA sequences.

(5) The \((\beta\alpha\alpha)_2\) motif is a minimal fold and members of the family often contain additional structural elements. The MetJ dimer contains a C-helix involved in binding the corepressor, S-adenosylmethionine (Rafferty et al., 1989). In Mnt, residues C-terminal to helix B permit stable tetramer formation in solution (Knight & Sauer, 1988).

Flexibility, the ability of the protein or DNA to change conformation, appears to play several roles in the formation of the \((\beta\alpha\alpha)_2\) motif-DNA complex. While some regions of the \((\beta\alpha\alpha)_2\) motif are completely disordered in the absence of DNA (e.g., the tandem turn region of Arc), other regions can exist in a limited number of energetically similar conformations within the confines of the folded structure, (e.g., the protein backbone surrounding Phe10 in the \(\beta\)-sheet of Arc). In principle, the highest degree of affinity and specificity is obtained by recognition between two completely rigid and complementary surfaces. If the surfaces, however, are not perfectly complementary, then rigidity can be a disadvantage. In this case a somewhat malleable molecule is preferred because it can adapt to maximize favorable contacts between the surfaces. For example, the rearrangement of the \(\beta\)-sheet atoms around Phe10 optimizes the fit between Arc and operator DNA. Regions of disorder in the unbound protein, such as the tandem turn region of Arc, can also aid in DNA recognition by contributing the energy gained from folding the disordered region to the overall energy of DNA binding. If, however, an excessive amount of the recognition surface of either the protein or the DNA is disordered, then DNA-binding specificity and affinity is lost.
It will be important to obtain additional structural and biochemical information about MetJ, Arc, and other ribbon-helix-helix family members. Are Arc and MetJ more similar in structure and function than other family members? How does MetJ interact with naturally occurring asymmetric sites? Could Arc bind to a more symmetric operator sequence? How can Mnt form a stable tetramer in solution? Does the first exposed β-sheet position in Arc and Mnt really control the contacts made by the other β-sheet residues? Is the (βαα)₂ motif in TraY actually formed by folding of tandem motifs in the same polypeptide? If so, does this permit greater versatility in DNA recognition by allowing a different set of amino acids on each strand of the β-sheet?

ACKNOWLEDGMENTS

We thank all of our colleagues who provided preprints and information about work in progress. We would also like to acknowledge Mark Rould, Carl Pabo, Tracy Smith, and Marcos Milla whose help and insights have been invaluable, and Simon Phillips for generously providing the coordinates of the MetJ-operator complex.
REFERENCES


Figure 1: Ribbon representation of the Arc (βαα)₂ motif. Each monomer is shaded differently. The locations of the N-terminus, C-terminus, and α-helices are indicated. Symmetric elements in different monomers are marked by primes. This figure and figure 2 were prepared using the program MOLSCRIPT (Kraulis, 1991).
Figure 2: a. Ribbon representation of the MetJ dimer bound to an operator half-site. b. Ribbon representation of the Arc dimer bound to an operator half-site. c. View down the dyad axis of the MetJ tetramer bound to a DNA site containing the met operator. d. Corresponding view of the Arc tetramer bound to a DNA site containing the arc operator. In panels a and b, each monomer is shaded differently; in panels c and d, each dimer is shaded differently. Although the MetJ and Arc dimers interact with half-sites in a similar fashion (panels a and b), the tetramer interfaces used in binding the whole operator are very different (panels c and d). In panels a and c, only the (βαα)2 and tandem turn regions of MetJ are shown. The remaining parts of the structure have been removed for clarity.
Figure 3: The helix B and tandem turn regions of the Arc and MetJ dimers contact six conserved phosphates in each operator half-site. Phosphates are shown as lettered circles. Arc contacts phosphates $a-d$ and $a'-d'$. MetJ contacts phosphate $a-c, a'-c', e$ and $e'$. 
Figure 4: Schematic diagram of the contacts between β-sheet side chains and bases in an operator half-site for Arc and MetJ. Contacted bases are shaded, and direct hydrogen bonds between side chains and bases are shown by an arrow (thin arrows are from one strand and thick arrows from the other strand of the β-sheet). Dashed lines represent hydrogen bonds between side chains. The closed oval represents the location of the two-fold axis within the half-site.
DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Some pages in the original document contain color pictures or graphics that will not scan or reproduce well.