Title

The molecular basis of DNA-binding specificity of the yeast homeodomain proteins a1 and [alpha]2

Author

Stark, Martha R.

Publication Date

1997

Peer reviewed|Thesis/dissertation

THE MOLECULAR BASIS OF DNA-BINDING SPECIFICITY OF THE YEAST HOMEODOMAIN PROTEINS a1 AND α2

by

Martha R. Stark

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOCHEMISTRY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

ACKNOWLEDGMENTS

Many, many thanks go to my lab mates, current and ex-, without whose company graduate school would have been much less bearable, and without whose help graduate school might have taken me twice as long. I thank the older members of the lab, Caroline Goutte, Michael Redd, and Nancy Hollingsworth for enticing me to join the lab even though I didn't think transcriptional regulation was the most exciting thing in the world. I thank Caroline for her contagious enthusiasm in a1/α2 and for getting me started on the project that turned into my thesis work. I thank Michael for endless hours of entertainment, and the occasional scientific discussion which moved my work forward by leaps and bounds. I thank Kelly Komachi and Rebecca Smith for all of the favors they did for me late at night and on the weekends, and for all of the shared reagents, without which I would still undoubtedly be doing experiments. I thank Christina Hull and Rob Brazas, and Madhu Wahi and Danesh Moazed for their friendship and many hours of babysitting, which allowed me to regain some perspective in my attempt to juggle my family and my work.

My thanks go to my advisor Sandy Johnson for the good advice he has given me, and for maintaining an exciting lab where good projects flourish. I greatly appreciate his support of my decision to have a baby and his unhesitating confidence in my scientific abilities. Most of all I thank my husband, Stephen Rader, who has been here for me through all of graduate school's trials and triumphs, and who has been a constant source of encouragement and support.

Chapter 1 is reprinted with permission from Nature. Appendix A is reprinted with permission from Biochemistry. Appendix B is reprinted with permission from Science. Appendix C is reprinted with permission from Molecular and Cellular Biology.

18-DEC-1997 17:20 FROM NATURE TO 90014154760939 P.01

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO • SANTA BARBARA • SANTA CRUZ

SCHOOL OF MEDICINE
Department of Microbiology and Immunology

SAN FRANCISCO, CALIFORNIA 94143-0502
FAX: 415-476-0939

December 2, 1997

Nature
968 National Press Building
529 14th St NW
Washington DC 20045-1938
Fax: (202) 628-1609

To whom it may concern:

I would like permission to include in my dissertation a copy of the paper cited below, which was published in Nature. The dissertation will be microfilmed by University Microfilms Incorporated, and they request permission to supply single copies on demand.

Stark, M.R., and Johnson, A.D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. Nature *371*, 429-432.

Please respond by Fax to (415) 476-0939.

Thank you for your consideration.

Sincerely,
Martha Stark
Martha Stark

NATURE
Permission to reprint is granted by Macmillan Magazines Limited, Porters South, Crinan Street, London ... COPYRIGHT OWNER ... REQUIRED. ... Reprinted with permission from ... Copyright (year) Macmillan Magazines Limited
APPROVED BY L. Banbury DATE 18/12/97
[illegible]
☑ If this box is ticked ... permission is also required.
See ...

TOTAL P.01

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

SCHOOL OF MEDICINE

Department of Microbiology
and Immunology

SAN FRANCISCO, CALIFORNIA 94143-0502
FAX 415-476-0939

December 2, 1997

Biochemistry
ACS
Journal Publishing Operations
2540 Olentangy River Road
P.O. Box 3330
Columbus, OH 43210
Fax: (614) 447-3745

To whom it may concern:

I would like permission to include in my dissertation a copy of the paper cited below, which was published in Biochemistry. The dissertation will be microfilmed by University Microfilms Incorporated, and they request permission to supply single copies on demand.

Philips, C.L., Stark, M.R., Johnson, A.D., and Dahlquist, F.W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators α2 and a1 induces an interfacial helix in α2. Biochemistry *33*, 9294-9302.

Please respond by Fax to (415) 476-0939.

Thank you for your consideration.

Sincerely,

Martha Stark

American Chemical Society

PUBLICATIONS DIVISION
COPYRIGHT OFFICE
FAX NUMBER: (415) 476-0939

1155 SIXTEENTH STREET, N.W.
WASHINGTON, D.C. 20036
Phone (202) 872-4367 or -4368
Fax (202) 872-6060

DATE: December 8, 1997

MEMORANDUM

TO: Martha Stark, University of California, San Francisco
Department of Microbiology and Immunology
San Francisco CA 94143-0502

FROM: C. Arleen Courtney C. Arleen Courtney
Assistant Copyright Administrator

RE: Your letter dated December 2, 1997

Thank you for your recent letter, regarding your request for permission to include your paper(s) or portions of your paper(s), per your attached letter, in your thesis. Please note the following:

* If your paper has already been published by ACS, I would be happy to grant you this permission royalty free provided that you print the required ACS copyright credit line on the first page of your article: "Reprinted (or 'Reprinted in part') with permission from FULL REFERENCE CITATION. Copyright YEAR American Chemical Society."

  Note: If you plan to submit your thesis to UMI or to another dissertation publisher, please inform them that permission to include **your already published ACS article as part of your thesis** is granted for paper and microform copies only; the ACS copyright notice (see above) must appear on the first page of the ACS article.

* If your paper has not already been published by ACS, you may include it in your thesis provided that you print the following ACS copyright credit line on the first page of your article: "Reprinted (or 'Reprinted in part') with permission from JOURNAL NAME, in press (or 'submitted for publication'). Unpublished work copyright CURRENT YEAR American Chemical Society."

  Note: If you plan to submit your thesis to UMI or to another dissertation publisher, you may NOT include the ACS paper in the version that you submit to UMI or to another dissertation publisher until ACS has published your paper.

* Other: _____

Thank you for writing. If you have any questions, please call me at 202/872-4368.

4/17/97

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

SCHOOL OF MEDICINE

Department of Microbiology
and Immunology

SAN FRANCISCO, CALIFORNIA 94143-0502
FAX 415-476-0939

December 2, 1997

Science
AAAS
1200 New York Avenue NW
Washington DC 20005
Fax: (202) 682-0816

To whom it may concern:

I would like permission to include in my dissertation a copy of the paper cited below, which was published in Science. The dissertation will be microfilmed by University Microfilms Incorporated, and they request permission to supply single copies on demand.

Li, T., Stark, M., Johnson, A.D., and Wolberger, C. (1995). Crystal structure of the MAT a1/MAT alpha2 homeodomain heterodimer bound to DNA. Science *270*, 263-269.

Please respond by Fax to (415) 476-0939.

Thank you for your consideration.

Sincerely,

Martha Stark

THE GLOBAL WEEKLY OF RESEARCH
SCIENCE

Li:2139

DATE: December 9, 1997

VIA FAX: (# of pages)

TO: Martha Stark
Dept. of Microbiology & Immunology
University of California, San Francisco
513 Parnassus Ave., S447, Box 0414
San Francisco, CA 94143-0414

FROM: Emilie L. David, Permissions Assistant

RE: Your request for permission dated 12/2/97

Regarding your request (see attached copy), we are pleased to grant you non-exclusive permission, **but limited to print & microform formats only**, and provided that you meet the criteria checked below. Such permission is for **one-time use** and therefore does **not** include permission for future editions, revisions, additional printings, updates, ancillaries, customized forms, any electronic forms, braille editions, translations, or promotional pieces. We must be contacted for permission each time such use is planned. This permission does not apply to figures / artwork that are credited to non-AAAS sources. This permission does not include the right to modify AAAS material.

- ☒ Print the required copyright credit line on the first page that the material appears: "Reprinted (abstracted/excerpted) with permission from [**FULL REFERENCE CITATION**]. Copyright [**YEAR**] American Association for the Advancement of Science." Insert the appropriate information in place of the capitalized words.
- ☒ Permission is limited to the number of copies specified in your request or your first printing.
- ☐ Obtain the author's permission. *See original article for author's address.*
- ☐ Remit the following permission fee: $. This letter serves as your invoice. Please make your check payable to "The American Association for the Advancement of Science." A photocopy of this letter **must** be included with the check. Please remit to SCIENCE, P.O. Box 80144, Baltimore, MD 21280-0144, USA **Please note:** ***Permission is not valid unless payment is received within sixty (60) days of the date on this letter.***
- ☐ Ensure that no other material is attached to the AAAS article copy.
- ☐ AAAS **must** publish the full paper prior to use of any text.
- ☐ If you plan to use substantially more than 200 words, please let me know the full amount to be used. A royalty fee may be assessed depending on the text used.

Note: AAAS does not supply photos or artwork. Also, use of the AAAS material must not imply any endorsement by the American Association for the Advancement of Science.

Thank you for writing. If you have any questions please call me at (202) 326-6765 or write to me via FAX at (202) 682-0816. For international calls, +1 is the country code for the United States.

Headquarters
1200 New York Avenue, NW, Washington, DC 20005 USA • Telephone 202 326 7073 • Fax 202 682 0816

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

SCHOOL OF MEDICINE

Department of Microbiology
and Immunology

SAN FRANCISCO, CALIFORNIA 94143-0502
FAX: 415-476-0939

December 2, 1997

Molecular and Cellular Biology
ASM
1325 Massachusetts Avenue, NW
Washington DC 20005-4171
Fax: (202) 942-9346

To whom it may concern:

I would like permission to include in my dissertation a copy of the paper cited below, which was published in MCB. The dissertation will be microfilmed by University Microfilms Incorporated, and they request permission to supply single copies on demand.

Redd, M.J., Stark, M.R., and Johnson, A.D. (1996). Accessibility of alpha2-repressed promoters to the activator Gal4. MCB *16*, 2865-2869.

Please respond by Fax to (415) 476-0939.

Thank you for your consideration.

Sincerely,

Martha Stark

Martha Stark

PERMISSION GRANTED CONTINGENT ON AUTHOR PERMISSION
AND APPROPRIATE CREDIT
American Society for Microbiology
Journals Division

Date 12-16-97

THE MOLECULAR BASIS OF DNA-BINDING SPECIFICITY OF THE YEAST HOMEODOMAIN PROTEINS a1 AND α2

Martha R. Stark

ABSTRACT

Differential gene expression is the basis for many developmental processes, such as cell-type determination. One example of this, mating-type control in *Saccharomyces cerevisiae*, is regulated by a1 and α2, members of a family of proteins containing the homeodomain DNA-binding motif. This thesis describes the molecular details of the interactions that allow a1 and α2 to specifically recognize and bind their target DNA sequences. The mechanism of interaction between a1 and α2 may serve as a general model for other homeodomain protein-protein interactions, and thus provide some insight into how other homeodomain proteins achieve specificity.

Each chapter investigates a feature of the a1/α2 interaction that contributes to DNA-binding specificity. Chapter 1 demonstrates that the C-terminal tail of α2 is critical for heterodimer formation and even functions with a1 when attached to a heterologous homeodomain. The experiments in chapter 2 suggest that a conformational change is induced in a1 upon binding of the α2 tail, and chapter 3 describes an attempt to genetically isolate mutations in a1 that cause the conformational change in the absence of α2. Chapter 4 shows that the NH_2-terminal arm of a1 is not necessary for specific DNA binding in the context of the a1/α2 heterodimer. And chapter 5 provides evidence for the generality of the interactions seen in a1 and α2 by demonstrating that the α2 C-terminal tail can be replaced by heterologous sequences and still interact with a1. The ability of these α2 chimeras to substitute functionally in the a1/α2 heterodimer emphasizes the modular nature of homeodomain interactions.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Gene regulation is a fundamental level at which a cell can respond to environmental changes and influence its fate. Gene regulatory proteins bind to specific target sites in the DNA and regulate transcription either positively or negatively. A key question is how these DNA binding proteins recognize their target sequences amongst the vast excess of non-target DNA in the nucleus. One family of gene regulatory proteins, those containing the homeodomain DNA binding motif, appear to solve this problem by interacting with cofactors that increase DNA binding specificity. The experiments described in this thesis examine the molecular basis of this phenomenon for the *Saccharomyces cerevisiae* homeodomain proteins a1 and α2.

The life cycle of *S. cerevisiae* involves three specialized cell types; two haploid cell types, **a** and α, and a diploid cell type, **a**/α, generated by the mating of an **a** and an α cell (Herskowitz, 1988). In each of these cell types, different genes are expressed. The **a** and α cells express many genes whose products facilitate mating. In the α cell, **a**-specific genes are repressed, and in the **a** cell, α-specific genes are not expressed because of the absence of a specific activator. The **a**/α cell does not mate, but rather is specialized to undergo meiosis and sporulation, giving rise to **a** and α haploid spores. The **a**/α cell therefore expresses a different set of genes, as well as repressing the genes required for mating (the haploid-specific genes).

There are four DNA binding proteins that play a role in determining cell-type (Herskowitz, 1988; Johnson, 1992). These include a1 and α2, as well as α1 and MCM1 (a member of the MADS family of DNA binding proteins). a1, α2 and α1 are cell-type specific proteins, while MCM1 is present in all three cell types. a1 is only transcribed in **a** and **a**/α cells, α2 is only transcribed in α and **a**/α cells, and α1 is expressed only in α cells. Thus, a1 and α2 are only present together in the diploid cell, and it is here that they function cooperatively to repress the haploid-specific genes. α2 and MCM1 also interact and bind DNA cooperatively in the α cell type, where they repress the **a**-specific genes. The repression of entire sets of genes provides an organism with a powerful means of

developmental control. This relatively simple regulatory network has been studied as a model for the more complex developmental circuits of higher organisms.

a1 and α2 are of special interest because they are members of the homeodomain family of DNA binding proteins (Gehring et al., 1994; Laughon, 1991). The homeodomain is a 60 amino acid motif containing four highly conserved residues and a conserved secondary structure consisting of a helix-loop-helix-turn-helix motif (Burglin, 1994; Kappen et al., 1993; Scott et al., 1989). The helix-turn-helix portion of this structure is similar to that seen in many bacterial activator and repressor proteins and suggested initially that the homeodomain proteins also act as gene regulators. Homeodomain proteins have been found in all eukaryotic organisms investigated, including angiosperms, fungi, and metazoa. At the start of the work for this thesis about 300 homeodomain proteins had been identified. Now, less than six years later, this family of proteins contains almost 2000 members (Bharathan et al., 1997).

The majority of studies on the DNA-binding specificity of homeodomains have focused on elucidating the nature of the homeodomain-DNA complex interface. NMR and crystallographic structural studies have shown that most of the many molecular contacts between the homeodomain protein and its target site are with groups on the sugar-phosphate backbone of the DNA (Hirsch and Aggarwal, 1995; Kissinger et al., 1990; Klemm et al., 1994; Li et al., 1995; Otting et al., 1990; Philips et al., 1991; Qian et al., 1989; Wilson et al., 1995; Wolberger et al., 1991). These nonspecific interactions are important for providing an architectural framework that orients specificity-determining residues and their target base pairs. Bases in both the major and minor grooves of the DNA are contacted by a few residues which are displayed on the surface of the homeodomain. Based on experiments in several systems, much of the specificity of homeodomain binding is thought to arise from contacts between the flexible NH_2-terminal arm and the minor groove of the DNA (Chan and Mann, 1996; Furukubo-Tokunaga et al., 1993; Lu and Kamps, 1997; Phelan and Featherstone, 1997; Phelan et al., 1994; Zeng et

al., 1993). Bases in the major groove, however, are recognized by residues in helix three of the homeodomain. Some of these residues also seem to contribute to target specificity. For some homeodomain proteins changes of a single amino acid in the recognition helix can change the target site binding specificity (Treisman et al., 1989). Binding of the homeodomain to the DNA may also distort the DNA to increase the number of binding interactions (Li et al., 1995). Since the deformability of DNA is also highly sequence-specific, distortion provides an alternative way for the protein to detect DNA sequence information. Although these mechanisms for binding to specific targets are clearly important, in many cases they do not appear to be sufficient to achieve the tight and specific binding necessary for developmental control.

The abundance of DNA-binding features in homeodomains might suggest tight binding, but in fact, monomeric homeodomain proteins bind DNA with relatively modest affinity and specificity (Affolter et al., 1990; Goutte and Johnson, 1993; Hoey and Levine, 1988). For example, the α2 protein has an affinity for specific DNA of only 10^{-6} M and binds to specific DNA only 10-fold better than to nonspecific DNA (Goutte and Johnson, 1993). In contrast, the prokaryotic λ repressor binds to specific operator sequences with affinities as high as 10^{-13} M and binds to specific DNA 500,000-fold better than to nonspecific DNA (Sauer et al., 1990). It is possible for interactions to be very specific even if the affinity is low, and it is probably specificity, rather than affinity, that is the key ingredient for assembling functional gene regulatory complexes. So where does the increased specificity in homeodomain-DNA interactions come from?

Despite the fact that homeodomain proteins can bind DNA as monomers, it is now thought that, in most cases, the monomeric form is not the biologically functional unit. Cofactors have been shown to exist that enhance the DNA binding specificity of these proteins (Goutte and Johnson, 1988; Keleher et al., 1988; Mann and Chan, 1996; Xue et al., 1993). Cooperative, multiple interactions between the proteins and DNA can ensure that the overall specificity of the gene regulatory complex is high, even if some individual

interactions are of low specificity. Extremely tight or specific interactions might interfere with the combinatorial use of factors by many promoters. Larger, cooperatively interacting complexes lead to increased flexibility, and individual interactions with only modest specificity seem to be inherent to these designs. Homeodomain protein specificity appears to depend on cooperative protein-protein and protein-DNA interactions, and only when the entire complex is assembled is the true specificity achieved.

The best-studied example of the cooperative binding of homeodomain proteins involves the a1 and α2 proteins. a1 and α2 greatly increase their ability to recognize specific target sites in the DNA by forming a heterodimeric complex (Goutte and Johnson, 1993). The two proteins can interact in solution and then bind to the DNA, with the homeodomains of both proteins making contact with the DNA (Li et al., 1995; Philips et al., 1994). As described in this thesis, the a1 homeodomain and the α2 homeodomain plus a short region immediately C-terminal to it (the tail) are sufficient for cooperative DNA binding. The region of α2 involved in heterodimerization with a1 lies immediately C-terminal to the α2 homeodomain, and is referred to as the α2 tail. Hydrophobic residues in the α2 tail contact the exposed surface of the a1 homeodomain, made up of residues in helices one and two and the loop between them. Several of these residues form a hydrophobic patch with which the tail of α2 interacts. Conformational changes take place in both α2 and a1 upon heterodimer formation. The previously unstructured tail of α2 forms an α-helix when it contacts a1. The homeodomain of a1 undergoes an as yet undefined reciprocal conformational change upon interaction with the α2 tail, rendering it competent to bind DNA, whereas it was previously unable to do so.

Although the interaction with cofactors has been invoked as a means for gaining specificity for many other homeodomain proteins, specific cofactors have been identified, either genetically or biochemically, for only about ten of them. These include other mating type proteins: α2 and MCM1 from *S. cerevisiae* (Keleher et al., 1988), bE and bW from the plant pathogenic fungus *Ustilago maydis* (Kamper et al., 1995), and HD1 and HD2

from the mushroom *Coprinus cinereus* (Asante-Owusu et al., 1996). Interactions between developmental regulatory proteins in higher organisms include (1) the *Drosophila melanogaster* HOM class of proteins with the cofactor Exd, and their mammalian homologues, the HOX class of proteins with the cofactor Pbx (Mann and Chan, 1996), (2) Ftz and Ftz-F1 from *Drosophila* (Guichet et al., 1997; Yu et al., 1997), (3) Mec-3 and Unc-86 from *Caenorhabditis elegans* (Xue et al., 1993), and (4) the mammalian Oct-1 and viral protein VP16 (Herr and Cleary, 1995). The molecular details of a few of these interactions have been recently elucidated and some patterns of homeodomain-cofactor interactions are emerging, the most general of which is that homeodomain proteins are made up of modular structural units that permit the combinatorial use of several, or many, different factors.

The protein-protein interactions between Exd/Pbx (PBC) and the HOM/HOX (HOX) class of proteins contain some similarities to the a1/α2 heterodimer interaction. As is the case for α2 and a1, HOX and PBC proteins bind only very weakly or not at all, respectively, to the DNA (Phelan et al., 1995). Like a1, the Pbx homeodomain is sufficient for interaction with some Hox proteins (Chan et al., 1994; Lu and Kamps, 1997; Popperl et al., 1995). However, interaction with other members of the HOX class require an additional 15 amino acids C-terminal to the Pbx homeodomain (Chang et al., 1995; Lu and Kamps, 1997). Residues in the Pbx homeodomain that correspond to the residues in the a1 homeodomain that mediate interaction with α2 are also required for cooperative DNA binding involving Pbx and Hox (Lu and Kamps, 1997).

In contrast to α2, in which only the C-terminal tail is required for heterodimerization with a1, three regions of the Hox proteins appear to be important for heterodimerization with Pbx. These include the homeodomain itself (Chan et al., 1994; Lu and Kamps, 1997), a region N-terminal to the homeodomain (Chang et al., 1995; Johnson et al., 1995; Knoepfler and Kamps, 1995; Neuteboom et al., 1995; Phelan et al., 1995), and a region C-terminal to the homeodomain (Chan et al., 1994). The N-terminal domain

contains a short stretch of conserved amimo acids known as the hexapeptide motif (YPWM). This motif is not essential for the Exd/Ubx interaction, but it does enhance it (Chan et al., 1994; Johnson et al., 1995). Several residues in the Ubx homeodomain are required for interaction with Exd, and two of these lie on the exposed surface of the homeodomain, in the loop between helices one and two (Chan et al., 1994). The C-terminal tail of Ubx also does not seem to be essential, but enhances heterodimer formation. Neither the C-terminal tail of Pbx nor Hox contains sequence similarity to the α2 C-terminal tail, indicating that the hydrophobic interaction that is the basis of the a1/α2 interaction is probably not conserved in these heterodimers. It is possible that these tails interact in a different manner with the exposed surface of the partner protein. It has been suggested that heterodimerization of PBC/HOX results in conformational changes in both proteins (Chan and Mann, 1996). The identification of the PBC/HOX heterodimerization domains emphasizes the modular nature of homeodomain proteins, and, as is the case for α2, shows how different extensions of the homeodomain are used to make different protein-protein contacts.

The interaction between the mammalian Oct-1 homeodomain and the herpes simplex virus trans-activator protein VP16 has also been well characterized biochemically, and, surprisingly, shows similarities to the a1/α2 interaction (Chapter 5 and Appendix B). Complex formation is dependent on specific amino acids in the the Oct-1 homeodomain which are in positions analogous to those that form the hydrophobic patch of the a1 homeodomain (Lai et al., 1992; Pomerantz et al., 1992). Residues in this region of Oct-1 appear to form a hydrophobic patch with which hydrophobic residues from VP16 could interact. Intriguingly, the region of VP16 that has been mapped by mutagenesis studies to interact with Oct-1 contains some sequence similarity to the tail of α2, and is predicted to form an amphipathic helix (Baxter et al., 1994; Hayes and O'Hare, 1993; Lai and Herr, 1997; Li et al., 1995)

At the outset of the work described in this thesis, the idea that cofactors might play a role in enhancing the specificity of most homeodomain-DNA complexes was emerging. We knew that a1 and α2 formed a heterodimer which bound cooperatively to the DNA, and that the α2 C-terminal tail was necessary for heterodimerization. My goal was to understand how the interaction of these two proteins with low DNA binding affinity and specificity leads to a complex that binds DNA extremely tightly and specifically. Concurrent with this work, other studies identified the PBC cofactors and began to dissect the interactions between them and their partners. This work, and the work on Oct-1 and VP16, have demonstrated that the results presented in this thesis are generally relevant to other homeodomain proteins. At the culmination of this work it has become doubtful that homeodomain proteins function as monomers, and it has become clear that their interaction with other proteins largely determine their specificity.

The work in this thesis has been divided into five chapters. Chapters 1-3 (along with Appendices A and B) describe the molecular details of the interaction between the a1 homeodomain and the C-terminal tail of α2. These details give some insight into the mechanism of cooperative binding and the specificity of the protein-protein interaction. Chapter 4 describes the role of the NH_2-terminal arm of the a1 homeodomain in determining DNA binding specificity. Unlike the NH_2-terminal arms of other homeodomain proteins, the arm of a1 does not appear to bind in the minor groove of the DNA and does not contribute to the DNA binding specificity of the a1/α2 complex. These results suggest that the current dogma regarding homeodomain specificity determinants may need to be modified. The work described in Chapter 5 suggests that a mammalian homeodomain protein, Oct-1, and its cofactor VP16, interact in a manner similar to that of a1 and α2. This result provides evidence that the a1/α2 mode of interaction has been conserved throughout evolution, and is applicable to other homeodomain protein/cofactor pairs.

REFERNCES

Affolter, M., Percival-Smith, A., Muller, M., Leupin, W., and Gehring, W. J. (1990). DNA binding properties of the purified Antennapedia homeodomain. Proc Natl Acad Sci U S A *87*, 4093-7.

Asante-Owusu, R. N., Banham, A. H., Bohnert, H. U., Mellor, E. J., and Casselton, L. A. (1996). Heterodimerization between two classes of homeodomain proteins in the mushroom Coprinus cinereus brings together potential DNA-binding and activation domains. Gene *172*, 25-31.

Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F., and Dahlquist, F. W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and a1: secondary structure determination of the a1 homeodomain and changes produced by alpha 2 interactions. Biochemistry *33*, 15309-20.

Bharathan, G., Janssen, B., Kellogg, E. A., and Sinha, N. (1997). Did Homeodomain Proteins Duplicate Before the Origin of Angiosperms, Fungi, and Metazoa? Proceedings of the National Academy of Sciences, USA *94*, 13749-13753.

Burglin, T. (1994). . In Guidebook to the Homeobox Genes, D. Duboule, ed. (Oxford: Oxford University Press), pp. 27-71.

Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell *78*, 603-15.

Chan, S. K., and Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proc Natl Acad Sci U S A *93*, 5223-8.

Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., and Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes Dev *9*, 663-74.

Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. Proc Natl Acad Sci U S A *90*, 6360-4.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211-23.

Goutte, C., and Johnson, A. D. (1988). a1 Protein Alters the DNA Binding Specificity of α2 Repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and α2 Homeodomain Proteins Form a DNA-binding Activity with Properties Distinct from those of either Protein. J. Molec. Biol. *233*, 359-371.

Guichet, A., Copeland, J. W., Erdelyi, M., Hlousek, D., Zavorszky, P., Ho, J., Brown, S., Percival-Smith, A., Krause, H. M., and Ephrussi, A. (1997). The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. Nature *385*, 548-552.

Hayes, S., and O'Hare, P. (1993). Mapping of a major surface-exposed site in herpes simplex virus protein Vmw65 to a region of direct interaction in a transcription complex assembly. J Virol *67*, 852-62.

Herr, W., and Cleary, M. A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Genes Dev *9*, 1679-93.

Herskowitz, I. (1988). Life Cycle of the Budding Yeast *Saccharomyces cerevisiae*. Microbiol. Rev. *52*, 536-553.

Hirsch, J. A., and Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. Embo J *14*, 6280-91.

Hoey, T., and Levine, M. (1988). Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature *332*, 858-61.

Johnson, A. D. (1992). A Combinatorial Regulatory Circuit in Budding Yeast. In Transcriptional Regulation, S. L. McKnight and K. R. Yamamoto, eds. (Plainview: Cold Spring Harbor Laboratory Press), pp. 975-1005.

Johnson, F. B., Parker, E., and Krasnow, M. A. (1995). Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proc Natl Acad Sci U S A *92*, 739-43.

Kamper, J., Reichmann, M., Romeis, T., Bolker, M., and Kahmann, R. (1995). Multiallelic recognition: nonself-dependent dimerization of the bE and bW homeodomain proteins in Ustilago maydis. Cell *81*, 73-83.

Kappen, C., Schughart, K., and Ruddle, F. H. (1993). Early evolutionary origin of major homeodomain sequence classes. Genomics *18*, 54-70.

Keleher, C. A., Goutte, C., and Johnson, A. D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-36.

Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579-90.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Knoepfler, P. S., and Kamps, M. P. (1995). The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. Mol Cell Biol *15*, 5811-9.

Lai, J. S., Cleary, M. A., and Herr, W. (1992). A single amino acid exchange transfers VP16-induced positive control from the Oct-1 to the Oct-2 homeo domain [published erratum appears in Genes Dev 1992 Dec;6(12B):2663]. Genes Dev *6*, 2058-65.

Lai, J. S., and Herr, W. (1997). Interdigitated residues within a small region of VP16 interact with Oct-1, HCF, and DNA. Mol Cell Biol *17*, 3937-46.

Laughon, A. (1991). DNA binding specificity of homeodomains. Biochemistry *30*, 11357-67.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA [see comments]. Science *270*, 262-9.

Lu, Q., and Kamps, M. P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene *14*, 75-83.

Mann, R. S., and Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet *12*, 258-62.

Neuteboom, S. T., Peltenburg, L. T., van Dijk, M. A., and Murre, C. (1995). The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins. Proc Natl Acad Sci U S A *92*, 9166-70.

Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., and Wuthrich, K. (1990). Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. Embo J *9*, 3085-92.

Phelan, M. L., and Featherstone, M. S. (1997). Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers. J Biol Chem *272*, 8635-43.

Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Molecular and Cellular Biology *15*, 3989-3997.

Phelan, M. L., Sadoul, R., and Featherstone, M. S. (1994). Functional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm. Mol Cell Biol *14*, 5066-75.

Philips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994). Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2. Biochemistry *33*, 9294-9302.

Philips, C. L., Vershon, A. K., Johnson, A. D., and Dahlquist, F. W. (1991). Secondary structure of the homeo domain of yeast α2 repressor determined by NMR spectroscopy. Genes Dev. *5*, 764-772.

Pomerantz, J. L., Kristie, T. M., and Sharp, P. A. (1992). Recognition of the surface of a homeo domain protein. Genes Dev *6*, 2047-57.

Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S., and Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. Cell *81*, 1031-42.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors [published erratum appears in Cell 1990 May 4;61(3):548]. Cell *59*, 573-80.

Sauer, R. T., Jordan, S. R., and Pabo, C. O. (1990). Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. Adv Protein Chem *40*, 1-61.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W. d. (1989). The structure and function of the homeodomain. Biochim Biophys Acta *989*, 25-48.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell *59*, 553-62.

Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. Cell *82*, 709-19.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-528.

Xue, D., Tu, Y., and Chalfie, M. (1993). Cooperative Interactions Between the *Caenorhabditis elegans* Homeoproteins UNC-86 and MEC-3. Science *261*, 1324-1328.

Yu, Y., Li, W., Su, K., Yussa, M., Han, W., Perrimon, N., and Pick, L. (1997). The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. Nature *385*, 552-555.

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., and Scott, M. P. (1993). Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development *118*, 339-52.

CHAPTER 1

Interaction Between Two Homeodomain Proteins is Specified by a Short C-Terminal Tail

Martha R. Stark and Alexander D. Johnson

(Published in Nature *371*: 429-432, 1994)

Interaction between two homeodomain proteins is specified by a short C-terminal tail

Martha R. Stark* & Alexander D. Johnson*†

* Department of Biochemistry and Biophysics, and
† Department of Microbiology and Immunology,
School of Medicine, University of California, San Francisco,
California 94143–0502, USA

Two yeast homeodomain proteins, a1 and α2, interact and cooperatively bind the haploid-specific gene (*hsg*) operator, resulting in the repression of a set of genes involved in the determination of cell type[1–5]. The cooperative binding of a1 and α2 to DNA can be reconstituted *in vitro* using purified fragments of a1 and α2. Only the homeodomain is needed for a1, but for α2 a C-terminal 22-amino-acid tail is required as well[4,6–9]. As most of the specificity of DNA binding appears to derive from a1, we proposed[4] that α2 functions in the a1/α2 heterodimer to contact a1 with its tail. By construction and analysis of several chimaeric proteins, we investigate how two DNA-binding proteins, one with low intrinsic specificity (α2) and one with no apparent intrinsic DNA-binding ability (a1), can together create a highly specific DNA-binding activity[4]. We show that the 22-amino-acid region of α2 immediately C-terminal to the homeodomain, when grafted onto the a1 homeodomain, converts a1 to a strong DNA-binding protein. This α2 tail can also be attached to the *Drosophila engrailed* homeodomain, and the chimaeric protein now binds cooperatively to DNA with a1, showing how a simple change can create a new homeodomain combination that specifically recognizes a new DNA operator.

To test whether fusion of the $\alpha 2$ C-terminal 22-amino-acid tail onto the a1 homeodomain (a1::$\alpha 2$) can convert the a1 homeodomain into a form that binds DNA tightly and specifically in the absence of $\alpha 2$, we investigated the binding of the a1::$\alpha 2$ chimaeric protein to a synthetic *a1/a1* operator (Figs 1 and 2*a*) and to an *hsg* operator (not shown). We found that the chimaera bound both operators efficiently. On the basis of the migration of the a1::$\alpha 2$/DNA complex in gel mobility shift experiments[4] and DNase I footprinting results (Fig. 2*b*), we conclude that the a1::$\alpha 2$ chimaera binds to DNA as a homodimer. The simplest explanation of these results is that in each homodimer one molecule behaves as if it were a1 without an $\alpha 2$ tail attached, and the other molecule interacts with the first in the same way that $\alpha 2$ normally interacts with a1. To rule out the possibility that attachment of this tail to the end of the a1 homeodomain nonspecifically stabilizes the folding of the homeodomain and thereby allows a1 to bind to DNA, we constructed a similar chimaeric protein comprising the a1 homeodomain and an $\alpha 2$ tail that contains a single amino-acid substitution (leucine to serine at residue 196) known to abolish a1/$\alpha 2$ repression *in vivo*[7]. In contrast to the wild-type tail construct, this mutant construct bound only very weakly to the DNA (Fig. 2*a*), indicating that the interaction between the $\alpha 2$ tail and the a1 homeodomain is dependent on the amino-acid constitution of the tail, and that the results obtained with the chimaeric protein mirror the situation *in vivo*.

These results suggest that a novel combination of homeodomains could be constructed simply by grafting the tail of $\alpha 2$ onto a new homeodomain which should now bind DNA cooperatively with a1. To test this idea, the tail of $\alpha 2$ was grafted onto the *Drosophila engrailed* (*en*) homeodomain and the chimaera (En::$\alpha 2$) was tested for its ability to bind DNA cooperatively with the a1 homeodomain (Fig. 3). Under our conditions neither protein alone efficiently occupied a synthetic operator containing one binding site for a1 and one site for *engrailed* product (En). However, En::$\alpha 2$ and a1 together show strong cooperative binding, again suggesting that contacts between the tail of $\alpha 2$ and the homeodomain of a1 are all that are necessary for cooperative interaction. To rule out the possibility that the En homeodomain was responsible for the interaction with a1, we also tested for cooperative binding between a1 and the En homeodomain (not shown), and between a1 and the En homeodomain with its 40-amino-acid tail (Fig. 3). At high concentrations, both of these En proteins occupied the *a1/en* operator by themselves, indicating that the proteins are active, but neither of them showed any cooperative binding with a1. From these results we conclude that the 22-amino-acid tail of $\alpha 2$ constitutes a domain that is sufficient for cooperative interaction with a1. This domain is functional when attached to a1 itself or to a heterologous homeodomain, conferring on the chimaeric protein the ability to interact with a1.

These experiments suggest that a large component of the DNA-binding specificity is contributed by the a1 protein, which, under our set of biochemical conditions, does not bind DNA on its own. For example, the complex composed of a1 and En::$\alpha 2$, in addition to binding the *a1/en* operator, also recognizes the

hsg operator, but with reduced affinity (not shown). Similarly, the a1/α2 complex can efficiently bind to the *a1/en* operator (not shown) and to the *a1/a1* operator[10], despite the fact that the preferred binding site for the En or α2 homeodomains is not present in these operators. Binding is improved when the correct recognition site for the homeodomain is present, indicating that some En and α2 homeodomain/DNA contacts do contribute to DNA-binding affinity.

It remains unclear how interaction with the α2 tail induces DNA-binding activity in a1, but recent NMR data[9] show that the α2 tail undergoes a conformational change upon interaction with a1. The C-terminal tail of α2 is unstructured when α2 is in solution or when α2 is bound alone to the DNA[11,12]. But when a1 and α2 come into contact, the tail of α2 becomes structured, forming an α-helix[9]. It is plausible that a1 undergoes a reciprocal conformational change upon interaction with the α2 tail. This change would take place in the a1 homeodomain, which contains the receptor for the α2 tail as well as the DNA-binding domain, and would convert a1 to a form competent to bind DNA. The amino acids in the turn between helices two and three of the a1 homeodomain may be potential targets for α2 interaction. Mutations in two of these amino acids (108 and 110) lower the cooperative binding of a1/α2 by 10–100-fold, whereas mutations in residues 80 and 87 of helix one and residues 98, 101 and 105 of helix two have no observable effect on the cooperative binding of a1 and α2 to DNA (Fig. 4). All of these residues are predicted, on the basis of the X-ray structure of α2 complexed to DNA[12], to be solvent-exposed and away from the DNA-binding surface, although we cannot rule out the possibility that a change in residues 108 or 110 produces a structural change in a1 that diminishes its interaction with DNA.

In addition to its interaction with a1, α2 binds DNA cooperatively with a second DNA-binding protein, MCM1 (ref. 13). The α2/MCM1 interaction is similar in several respects to that between α2 and a1. In both, a short flexible region of α2 (for MCM1 this region is located immediately N-terminal to the α2 homeodomain[14]) interacts with the DNA-binding domain of MCM1 or a1. The C-terminal tail is not required for the α2/MCM1 interaction[6]; likewise, the flexible region N-terminal to the α2 homeodomain is not required for interaction with a1[4]. Thus we can think of the α2 homeodomain as having two flexible extenders, one specifying interaction with a1, the other specifying interaction with MCM1 (Fig. 1*c*).

The cooperative interaction of UNC-86 and MEC-3, two *C. elegans* homeodomain proteins, bears many similarities to the a1/α2 interaction[15]. The POU homeodomain of UNC-86 is sufficient for both DNA binding and heterodimerization, whereas for MEC-3 a region containing the homeodomain plus

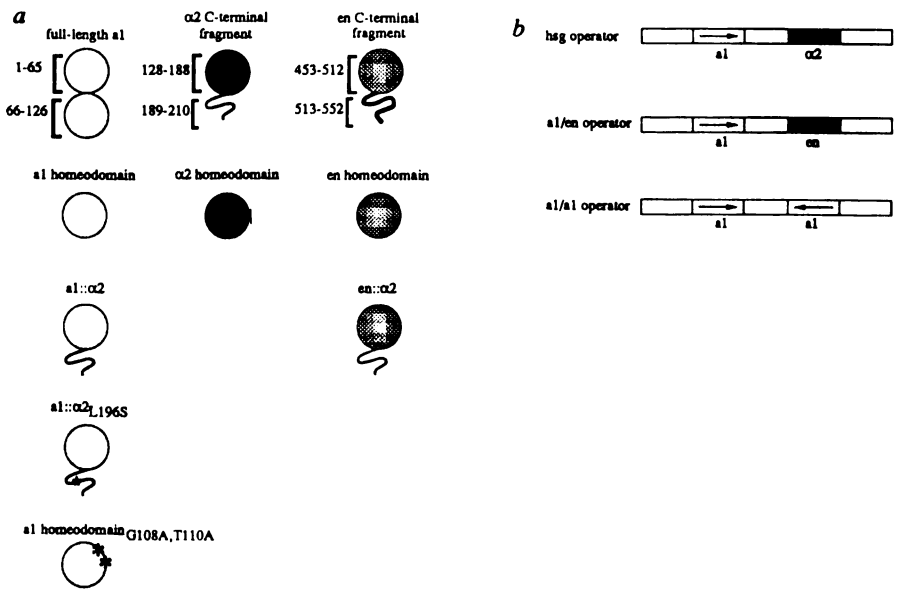


FIG. 1 Proteins (*a*) and operators (*b*) used in experiments. *hsg* (*a1/α2*): TCGACTATGATGTACTTTT-CTACATTGGGAAGC; *a1/en*: TCGACGACTATGATGTACT-TTTTAATTACCGGAAGC; *a1/a1*: TCGATGATGTAATT-AATTACATCA. Binding sites are underlined and arranged with dyad symmetry. The recognition sequences for the three proteins are: *a1*, TGATGTA; α2, ATGTA; *en*, TAATT. The *hsg* operator is that from upstream of the *MATα1* gene[4]. The sequence of the *en* binding site was that used for the crystallization of the En homeodomain on DNA[18]. *c*, Functional modules of α2. The short flexible region immediately N-terminal to the homeodomain (the hinge, residues 110–128) interacts with the core of MCM1[14], the homeodomain binds to the DNA operator[19], and the C-terminal tail interacts with the a1 homeodomain. In the absence of MCM1 and a1 the hinge and tail regions of α2 are unstructured[6,11,12,20]. It has been shown that upon interaction with a1 the tail of α2 assumes an α-helical structure[9]. Although we have no experimental evidence, we assume that upon interaction with MCM1 the hinge of α2 also becomes structured.

METHODS. All truncated and chimaeric versions of the proteins were generated by PCR mutagenesis[21], using the full-length genes as templates. An *Nde*I site was introduced at the desired N terminus, providing the ATG necessary for translation. The a1::α2 chimaeras were made by ligating the homeodomain and α2 tail products from two separate PCR reactions, using a naturally occurring *Bgl*II site at the C terminus of a1. The En::α2 chimaera was generated as follows. Two PCR products encoding the en homeodomain and the α2 tail were made to partially overlap in sequence at the desired junction between homeodomain and tail. These overlapping, primary products were denatured and allowed to reanneal, producing a heteroduplex product which was extended by *Taq* DNA polymerase to produce a fragment that is the sum of the two overlapping products. This fragment was then amplified using outside primers[22]. The a1 homeodomain with mutations G108A and T110A was generated in the same manner, using partially overlapping primers containing the two mutations. The α2 C-terminal fragment was a gift from A. Vershon. All other proteins were overexpressed and purified from *E. coli* strain BL21(DE3) or BL21(DE3)pLysS containing the relevant gene under the control of the T7 promoter in the plasmid pHB40P (ref. 23). Proteins expressed in BL21(DE3)pLysS were induced with 0.4 mM IPTG at $A_{600} = 0.5$ and grown for 5 h; those proteins expressed in BL21 (DE3) were grown without induction. All proteins were purified from cell lysates by adhesion to a cation-exchange resin (Sephadex SP-C50, Pharmacia) followed by elution with a NaCl gradient. The proteins were >90% pure by Coomassie staining following electrophoresis through SDS gels. All operators were designed based on the spacing and orientation of the binding sites in a *bona fide* hsg operator[8]. Complementary oligonucleotides were generated for each operator and then annealed to form duplexes with TCGA 5′-overhangs at each end. The oligonucleotides for the hsg and the a1/a1 operators were gifts from C. Goutte.

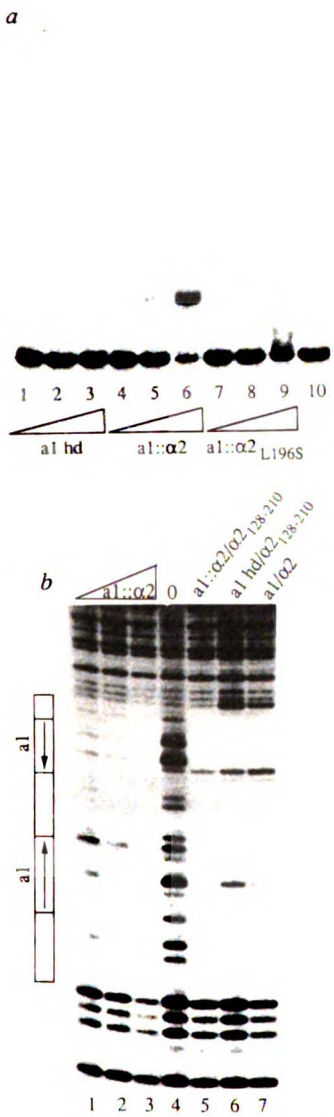


FIG. 2 The a1 homeodomain::α2 tail chimaera (a1::α2) binds a synthetic *a1/a1* operator. *a*, Electrophoretic mobility shift. A ^{32}P-labelled 80-bp DNA fragment containing two a1 binding sites (Fig. 1*b*) was incubated with purified protein. DNA concentration was approximately 10^{-10} M. Lanes: 1–3, successive 3-fold increases of a1 homeodomain protein beginning with a concentration of 30 nM; 4–6, successive 3-fold increases of a1::α2 protein beginning with a concentration of 30 nM; 7–9, successive 3-fold increases of a1:$\alpha 2_{L196S}$ protein beginning with a concentration of 30 nM; 10, labelled DNA alone. The size of the a1::α2 shift is the same as the shift when both half-sites of the hsg operator are filled by two molecules of the α2 C-terminal fragment[24]. *b*, Protection against DNase I. Labelled *a1/a1* operator was incubated with the proteins as indicated above the lanes of the gel, followed by treatment with DNase I. The two a1 binding sites of the *a1/a1* operator are indicated to the left of the figure. Lanes: 1–3, successive 2-fold increases of a1::α2 protein beginning with a concentration of 300 nM; 4, no protein; 5, 300 nM a1::α2 protein with 100 nM α2 C-terminal fragment; 6, 300 mM a1 homeodomain protein with 100 nM α2 C-terminal fragment; 7, 300 nM full-length a1 protein with 100 nM full-length α2 protein. The DNAse I footprint of full-length a1 and α2 proteins has been shown to represent the binding of a heterodimer to the DNA operator[8]. The fact that the a1::α2 chimaera footprint is nearly identical to the a1/α2 footprint indicates that there are two molecules of the a1::α2 protein binding to the operator. In addition, the protection seen with a1::α2 and the α2 C-terminal fragment together resembles that with a1 and α2, indicating that the tail of α2 which is attached to the a1 homeodomain does not interfere with the normal interaction between a1 and α2.

METHODS. The *a1/a1* operator duplex was cloned into the *Sal*I site of pUC18, excised as an 80-bp fragment, and ^{32}P-end-labelled using Klenow fragment. The DNA-binding experiments were performed as previously described[4,20]. Protein and DNA were incubated together at room temperature for 45 min in gel shift buffer (5% glycerol, 10 mM Tris-HCl, pH 8, 10 mg ml^{-1} BSA, 0.1 , EDTA, 10 μg ml^{-1} non-specific DNA, 5 mM $McCl_2$, 0.1% N-P40, 100 mM NaCl). *E. coli* genomic DNA digested with *Hae*III was used as nonspecific DNA. Samples were electrophoresed through a 5% native Tris–borate–EDTA (TBE) polyacrylamide gel for 1 h at 200 V. The DNase I protection experiment was performed as described[20], except the buffer used contained 10 mM Tris-HCl, pH 7.0, 5 mM $MgCl_2$, 10 mM $CaCl_2$, 0.5 mM EDTA, 50 μg ml^{-1} BSA, and 2.5 mg ml^{-1} calf thymus DNA. Reactions were incubated at 20 °C for 1 h before cleavage for 10 min with 1.5 μg DNase I (Worthington). Reactions were stopped and precipitated with 1.6 M ammonium acetate. Samples were electrophoresed through a 10% denaturing TBE gel.

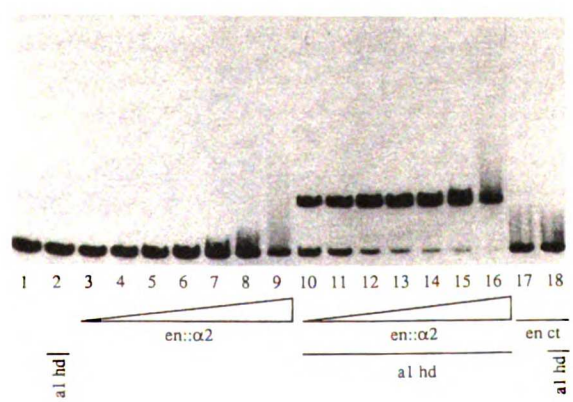


FIG. 3 Cooperative binding of the a1 homeodomain and the En homeodomain::α2 tail chimaera (En::α2) to a synthetic *a1/en* operator. A ^{32}P-labelled 92-bp DNA fragment containing the *a1/en* operator (Fig. 1*b*) was used in the electrophoretic mobility shift experiment. DNA concentration was ~10^{-10} M, and all proteins were >90% pure. Lanes: 1, labelled DNA alone; 2, 100 nM a1 homeodomain protein; 3–9, successive 3-fold increases of En::α2 chimaeric protein beginning with a concentration of 1 nM in lane 3, and ending with 1 μM in lane 9; 10–16, same titration of En::α2 as in lanes 3–9, but also containing 100 nM a1 homeodomain protein; 17, 300 nM En C-terminal fragment; 18, 300 nM En C-terminal fragment with 100 nM a1 homeodomain protein. The complex equilibrium dissociation constant describing the interaction of En::α2 and the a1 homeodomain with the *a1/en* operator is very similar to that observed for the cooperative binding of the C-terminal fragment of α2 and the a1 homeodomain to the *hsg* operator[4,9], ~10^{-16} M^2 as estimated from gel shift experiments.

METHODS. The synthetic *a1/en* operator duplex was cloned into pUC18 and was excised and radioactively labelled as a 92-bp fragment. DNA binding was assayed as for Fig. 2, except that proteins were incubated together at room temperature for 30 min in gel shift buffer containing 5 μg ml^{-1} nonspecific DNA before addition of the labelled DNA. Protein and DNA were incubated another 45 min at room temperature before being electrophoresed.

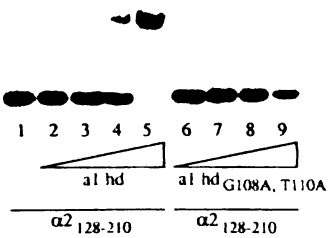


FIG. 4 Specific residues that comprise the turn between helix 2 and helix 3 of the a1 homeodomain are required for cooperative binding with $\alpha 2$. A ^{32}P-labelled 80-bp fragment containing the *hsg* operator (Fig. 1*b*) was incubated with purified proteins in the electrophoretic mobility shift experiment. Lanes: 1, 3 nM $\alpha 2$ C-terminal fragment; 2–5, 3 nM $\alpha 2$ C-terminal fragment with successive 3-fold increases of a1 homeodomain protein starting at 10 nM; 6–9, 3 nM $\alpha 2$ C-terminal fragment with successive 3-fold increases of the a1 homeodomain protein containing the two point mutations, G108A and T110A, beginning at 10 nM. The two forms of the a1 homeodomain, wild-type and mutant, were expressed at approximately the same levels, and their purification profile was identical. We therefore assume that these changes do not significantly destabilize the a1 homeodomain structure. The DNA-binding experiment is described in Fig. 2 legend.

the adjacent C-terminal 16 amino acids is required for these activities. Some HOM-C proteins from *Drosophila*, for example Antp, Ubx, and Dfd, may also use their C-terminal tails to alter target gene specificity by interacting with other proteins[16,17]. It is therefore plausible that the C-terminal tails of many homeodomain proteins could prove to be important protein interaction domains which determine protein partners and, as a result, the specific sets of genes to be regulated. □

Received 25 March; accepted 5 August 1994.

1. Strathern, J. N., Hicks, J. & Herskowitz, I. *J. molec. Biol.* **147,** 357–372 (1981).
2. Goutte, C. & Johnson, A. D. *Cell* **52,** 875–882 (1988).
3. Dranginis, A. M. *Nature* **347,** 682–685 (1990).
4. Goutte, C. & Johnson, A. D. *J. molec. Biol* **233,** 359–371 (1993).
5. Goutte, C. & Johnson, A. D. *EMBO J.* **13,** 1434–1442 (1994).
6. Mak, A. & Johnson, A. D. *Genes Dev.* **7,** 1862–1870 (1993).
7. Strathern, J., Shafer, B., Hicks, J. & McGill, C. *Genetics* **120,** 75–81 (1988).
8. Goutte, C. & Johnson, A. D. *EMBO J.* **13,** 1434–1442 (1994).
9. Philips, C. L., Stark, M. R., Johnson, A. D. & Dahlquist, F. W. *Biochemistry* **33,** 9294–9302 (1994).
10. Goutte, C. *Combinatorial Control of the Yeast Homeodomain Proteins a1 and α2* (UCSF Press, San Francisco, 1992).
11. Philips, C. L., Vershon, A. K., Johnson, A. D. & Dahlquist, F. W. *Genes Dev.* **5,** 764–772 (1991).
12. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. *Cell* **67,** 517–528 (1991).
13. Johnson, A. D. in *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K. R.) 975–1005 (Cold Spring Harbor Laboratory Press, Plainview, 1992).
14. Vershon, A. K. & Johnson, A. D. *Cell* **72,** 105–112 (1993).
15. Xue, D., Tu, Y. & Chalfie, M. *Science* **261,** 1324–1328 (1993).
16. Chan, S.-K. & Mann, R. S. *Genes Dev.* **7,** 796–811 (1993).
17. Lin, L. & McGinnis, W. *Genes Dev.* **6,** 1071–1081 (1992).
18. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. *Cell* **63,** 579–590 (1990).
19. Hall, M. N. & Johnson, A. D. *Science* **237,** 1007–1012 (1987).
20. Sauer, R. T., Smith, D. L. & Johnson, A. D. *Genes Dev.* **2,** 807–816 (1988).
21. Mullis, K. B. *et al. Cold Spring Harbor Symp. quant. Biol.* **51,** 263–273 (1986).
22. Higuchi, R. in *PCR Protocols: A Guide to Methods and Applications* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 177–183 (Academic, San Diego, 1990).
23. Studier, F. W. & Moffatt, B. A. *J. molec. Biol.* **189,** 113–130 (1986).
24. Smith, D. L. & Johnson, A. D. *Cell* **68,** 133–142 (1992).

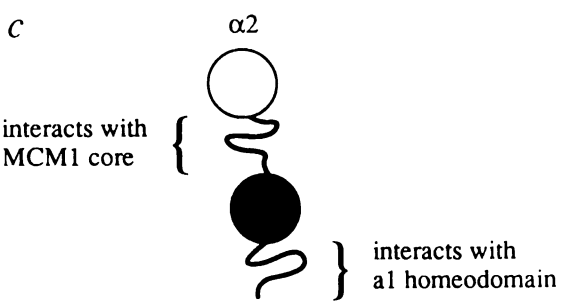
ACKNOWLEDGEMENTS. We thank C. Goutte for discussion and for the *hsg* and *a1/a1* operator constructs; A. Vershon for sharing unpublished results and for purified $\alpha 2$ C-terminal fragment; S. Rader, M. Wahi, D. Moazed and C. Kenyon for comments on the manuscript. This work was supported by a grant from the NIGMS.

ERRATUM

Interaction between two homeodomain proteins is specified by a short C-terminal tail

Martha R. Stark & Alexander D. Johnson

Nature **371,** 429–432 (1994)



PART *c* of Fig. 1 of this Letter was accidentally omitted and is now shown here. □

CHAPTER 2

A Conformational Change in a Homeodomain Protein Induced by a Peptide Ligand

Abstract

It has often been difficult to reconcile the highly selective biological activity of homeodomain proteins with their modest individual DNA binding specificities. Through the use of chimeric proteins and synthetic peptides we provide evidence that upon heterodimerization, one homeodomain protein (α2) contributes a ligand that induces a conformational change in a second homeodomain protein (a1) which enhances its DNA binding. This idea explains, in part, how high DNA-binding specificity is achieved only when the two homeodomains conjoin.

Introduction

In eucaryotes, proteins that regulate transcription typically act in combinations. A simple example of this principle is found in the specification of cell-types in the yeast *Saccharomyces cerevisiae*. In the **a**/α cell type two homeodomain proteins, a1 and α2, form a heterodimer that binds with high affinity and specificity to a DNA sequence called the haploid-specific gene (hsg) operator (Dranginis, 1990; Goutte and Johnson, 1988; Goutte and Johnson, 1994; Goutte and Johnson, 1993; Strathern et al., 1981) This operator is located upstream of many genes (collectively called the haploid-specific genes), and the binding of a1 and α2 to it recruits the SSN6/TUP1 repressor, which represses transcription of each haploid-specific gene (Keleher et al., 1992; Komachi et al., 1994; Mukai et al., 1991) These genes encode proteins required for **a** and α cells to mate as well as regulators of **a**/α cell-specific functions (Herskowitz et al., 1992; Johnson, 1995).

Although the a1/α2 heterodimer has been well studied, it has been difficult to rationalize the high DNA-binding specificity of the heterodimer in terms of the individual properties of its constituent proteins. Under experimental conditions in which the heterodimer specificity was determined to be at least 3,000-fold greater for the hsg operator than for non-specific DNA, α2 exhibited a DNA-binding specificity of approximately 10-fold and a1 showed no reproducible specific binding (Goutte and Johnson, 1993; Philips et al., 1994).

How is the DNA-binding specificity of the a1/α2 heterodimer generated? In this paper we test the hypothesis that, upon contact with α2 during heterodimer formation, a1 undergoes a conformational change which increases its affinity for the hsg operator. According to this hypothesis, the key protein-protein contact occurs between a short region of α2 (called the tail) immediately

C-terminal to its homeodomain and the homeodomain of a1. The α2 tail is unstructured in the α2 monomer (Philips et al., 1991; Wolberger et al., 1991) but folds into a short α helix upon contact with the a1 homeodomain (Li et al., 1995; Philips et al., 1991) We propose that the α2 tail acts as a small ligand to induce a conformational change in a1.

To test this hypothesis, we constructed three new chimeric molecules consisting of the homeodomain of a1 linked covalently to the tail of α2. These two elements were joined by linkers designed to be of sufficient length and flexibility to permit intramolecular interactions between a1 and the α2 tail. A prediction of the model described above is that such chimeric molecules should bind tightly and specifically to DNA as monomers, because they include both the ligand and its receptor, linked by a flexible tether. We show here that these a1::α2 chimeras containing flexible linkers are capable of binding DNA specifically as monomers. We further show that an α2 tail peptide supplied in *trans* can induce the a1 homeodomain to bind to DNA with increased affinity. We conclude that the α2 tail induces a conformational change in the homeodomain of a1 which is necessary for efficient DNA binding.

Results

Construction of the a1::α2 chimeric proteins

The design of the chimeric molecules (and in particular the length of the linkers) was based on inspection of the x-ray crystal structure of the a1/α2 heterodimer bound to DNA (Li et al., 1995). In two of the chimeric molecules, the α2 tail was attached via a linker to the C-terminus of the a1 homeodomain. In the x-ray structure the distance between the C-terminus of a1 and the N-terminus of the α2 tail is 32 Angstroms. A linker of 11 amino acids (present in the a1::11::α2 chimera, Figure 1A) should, in principle, span this distance if it is assumed that the linker is fully extended. A chimera with a linker of 16 amino acids (a1::16::α2) was also constructed to accommodate some degree of structure in the linker (Figure 1A). The linkers were composed of Glycine and Serine to provide both flexibility and solubility. A third chimeric molecule was constructed in which the α2 tail was attached via a linker to the N-terminus of the a1 homeodomain. The distance between the C-terminus of α2 and the N-terminus of the a1 homeodomain in the x-ray structure is only 13 Angstroms, and a Glycine/Serine linker of 6 amino acids was employed to span this distance (Figure 1A).

a1::α2 chimeric proteins bind as monomers to a modified hsg operator

The three chimeric proteins summarized in Fig. 1A were expressed in *E. coli,* purified to greater than 90% homogeneity, and tested for their binding to a synthetic operator composed of two a1 half-sites (a1/a1 in Fig. 1C). In contrast to the a1 homeodomain alone (Fig. 2 lanes 5-7), all three chimeras exhibited efficient DNA binding in the 30-100 nM range (lanes 11-13, 17-19, 23-25). Based on the electrophoretic migration of the chimera/DNA complexes (Goutte and Johnson, 1993; Smith and Johnson, 1992) it seemed likely that all three

Figure 2.1. Proteins (**A**), peptides (**B**) and operator (**C**) used in the experiments described in this paper. (**A**) The top two diagrams depict the portion of the wild-type a1 and α2 proteins sufficient for heterodimer formation and DNA binding. The additonal proteins are chimeras in which the tail of α2 has been fused to the homeodomain of a1. The linkers between the tail and the homeodomain are composed of alternating $(Glycine)_2$ and $(Serine)_2$, and are depicted as a string of filled black circles. (**B**) α2 tail peptides were synthesized by California Peptide Research, Inc. All three peptides are identical except for position 196; they are 19 amino acids in length, beginning at residue 189 of α2 and ending at residue 207. α2 ends with residue 210 and residues 208-210 are not required for a1/α2 heterodimer formation (Herskowitz, 1989). (**C**) The a1/a1 operator has the same spacing and binding site orientation as a naturally occurring hsg operator (which contains an a1 and α2 binding site), except the α2 binding site of the hsg operator has been replaced by a second a1 binding site (Keleher et al., 1992). The two a1 binding sites are separated by six base pairs and, for the experiments of Figs. 2.2-4, are contained within an 80-bp DNA fragment.

Figure 2.1

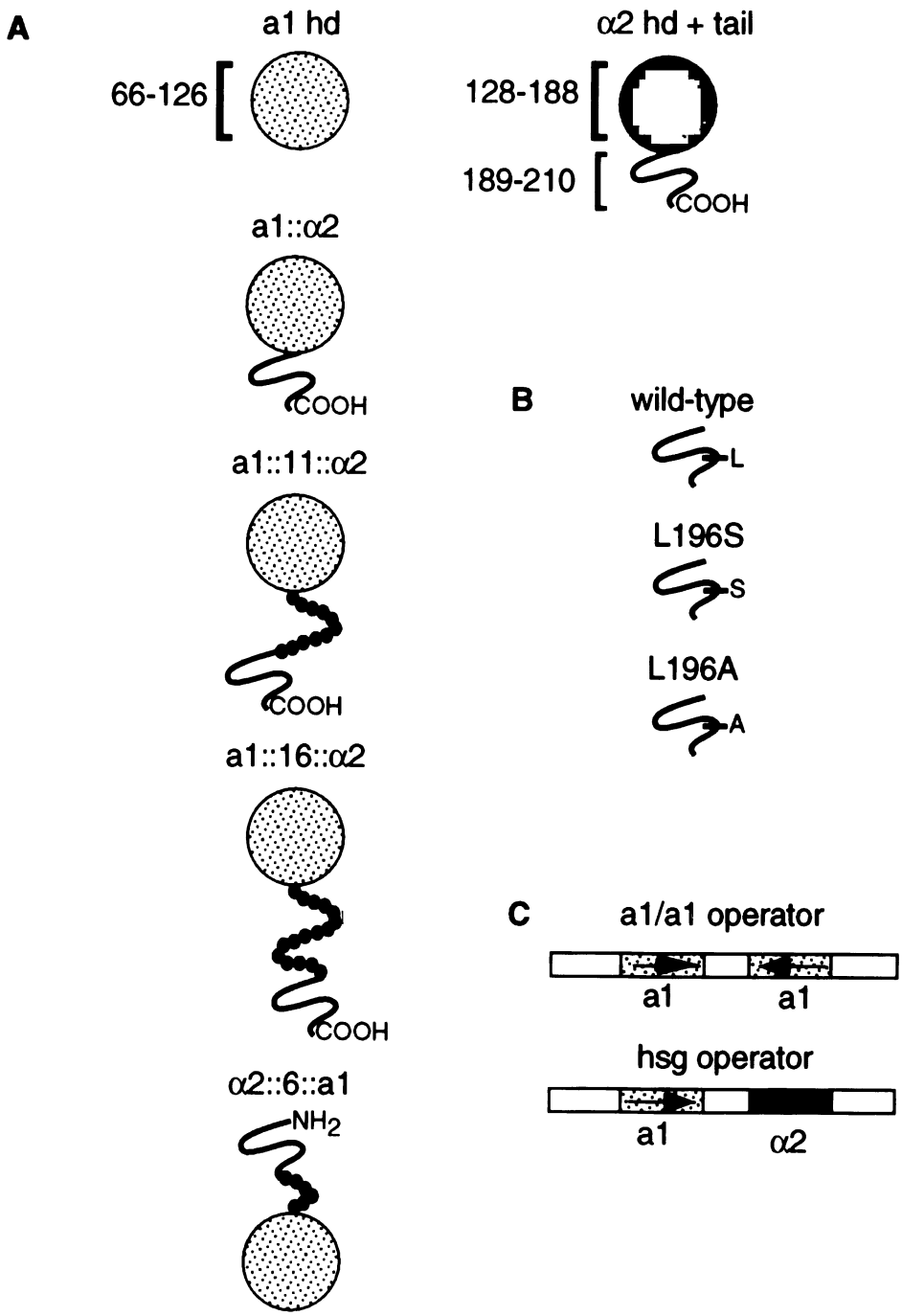
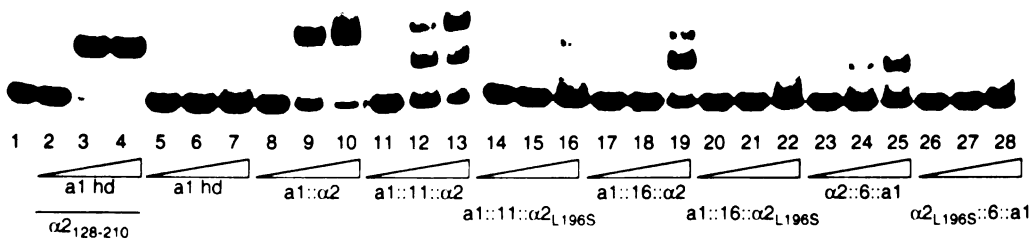

A
a1 hd
66-126
α2 hd + tail
128-188
189-210
COOH
a1::α2
COOH
a1::11::α2
COOH
a1::16::α2
COOH
α2::6::a1
NH2
B
wild-type
L
L196S
S
L196A
A
C
a1/a1 operator
a1
a1
hsg operator
a1
α2

Figure 2.2. Binding of the a1::α2 chimeras to the a1/a1 operator. The ^{32}P-labeled DNA fragment (80 nucleotide pairs) containing the a1/a1 operator was incubated with the indicated purified protein for 30 min at room temperature and electrophoresed through a 5% native Tris-borate-EDTA polyacrylamide gel . Lane 1 contains labeled DNA alone. Lanes 2-4 contain 3 nM α2 hd + tail ($\alpha2_{128\text{-}210}$) in addition to a1 hd. The a1 hd alone (lanes 5-7) and the a1::α2 chimera homodimer (lanes 8-10) were included along with the a1/α2 heterodimer in order to demonstrate the different mobility shifts expected for monomers and dimers bound to the DNA. a1::11::α2 (lanes 11-13) and a1::16::α2 (lanes 17-19) both give two shifts, consistent with monomeric and dimeric DNA binding, whereas α2::6::a1 (lanes 23-25) gives only one shift consistent with monomeric DNA binding. All three chimeras containing the α2 tail mutation have reduced DNA binding (lanes: 14-16, a1::11::$\alpha2_{L196S}$; 20-22, a1::16::$\alpha2_{L196S}$; 26-28, $\alpha2_{L196S}$::6::a1). The a1 homeodomain and chimera concentrations for each set of three reactions were 30 nM, 100 nM, and 300 nM.

Figure 2.2


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
a1 hd
a1 hd
a1::α2
a1::11::α2
a1::11::α2L196S
a1::16::α2
a1::16::α2L196S
α2::6::a1
α2L196S::6::a1
α2128-210

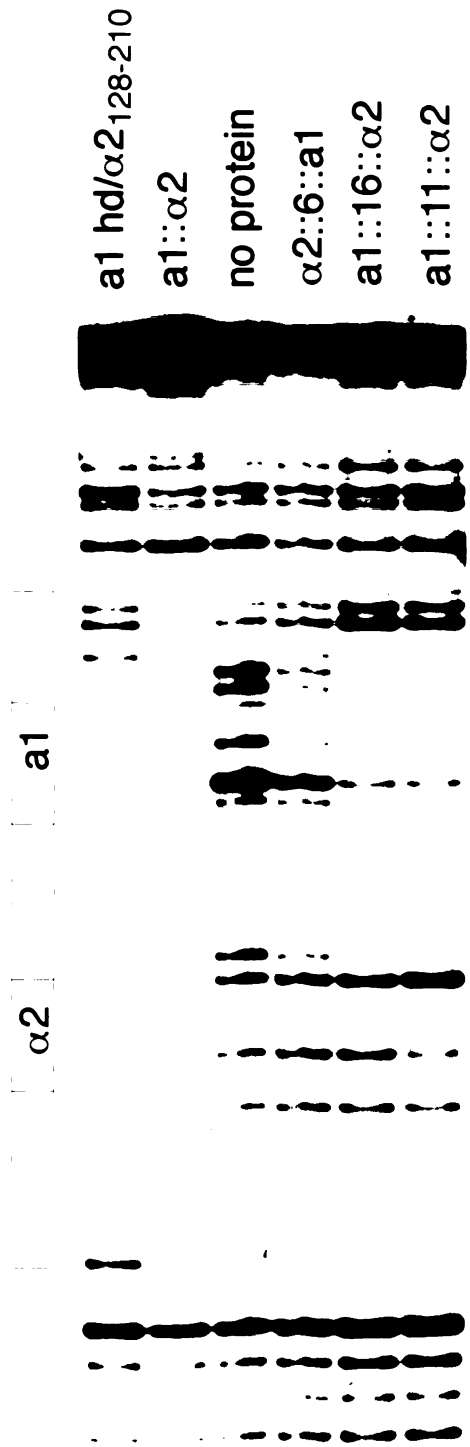
chimeric proteins were capable of binding DNA as monomers. DNase I footprinting of the chimeric proteins on an a1/α2 operator demonstrated that all three of the chimeric proteins with linkers bind only to the a1 half of the operator, whereas the a1::α2 chimera without a linker binds to both the a1 and α2 sites of the operator (Fig. 3). This result indicates that the linkers permit the α2 tail to interact with the a1 molecule to which it was attached and that this interaction stimulates DNA binding. This situation contrasts to that of the a1::α2 chimera that lacks a linker (see Fig. 1A) which can only bind DNA as a dimer (Fig. 2 lanes 8-10) (Stark and Johnson, 1994). a1::11::α2 and a1::16::α2 are capable of forming dimers on the DNA in addition to monomers, but the binding of α2::6::a1 appears solely monomeric. The simplest explanation for this difference is that the placement of the linker on the N-terminus of a1 in α2::6::a1 sterically blocks the binding of a second molecule to the inverted a1 half sites, while the C-terminal placements do not preclude this possibility.

Mutation of the tail reduces DNA binding by the a1::α2 chimeras

To rule out the possibility that the enhancement of DNA binding observed for the chimeric proteins was due to non-specific contributions of the linker, chimeras with mutant tails were constructed and tested for DNA binding. Residue 196 in the α2 tail of each chimera was changed from Leucine to Serine, a mutation known to disrupt a1/α2 function in vivo and in vitro (Stark and Johnson, 1994; Strathern et al., 1988) In the x-ray structure of the a1/α2 heterodimer, the Leucine at this position contacts the a1 homeodomain (Li et al., 1995) In all three mutants, a1::11::$\alpha 2_{L196S}$ (Figure 2 lanes 14-16), a1::16::$\alpha 2_{L196S}$ (lanes 20-22), and $\alpha 2_{L196S}$::6::a1 (lanes 26-28), DNA binding is significantly reduced, indicating that the leucine in this position of the

Figure 2.3. Monomeric DNA binding of the a1::α2 chimeras with linkers to an a1/α2 operator. DNase I protection of a ninety nucleotide pair ^{32}P-labeled fragment containing the hsg operator. Lanes: 1) 3μM a1 hd plus 100nM α2128-210 2) 50μM a1::α2 3) no protein 4) 50 μM α2::6::a1 5) 100μM a1::16::α2 6) 50μM a1::11::α2.

Figure 2.3



α2 tail is necessary for the maximal enhancement of a1 binding. Efficient DNA binding by the a1 homeodomain chimeras therefore specifically requires the α2 tail.

The α2::6::a1 chimeric protein binds the synthetic a1/a1 operator specifically

We verified that the DNA binding of one chimera, α2::6::a1, was specific, by comparing its affinity to two different DNA fragments, one with the synthetic a1/a1 binding site and the identical fragment that lacked these sites (Fig. 4). Both α2::6::a1 (lanes 9-11) and the a1/α2 heterodimer (lanes 3-5), used as a control, bind well to the a1/a1 operator but not to the operator that lacks a1 sites (lanes 6-8 and 12-14). These results indicate that the α2::6::a1 chimera shows a marked preference for known a1 binding sites over other DNA.

An α2 tail peptide is sufficient to induce the a1 homeodomain to bind DNA

It seemed plausible, based on the above results, that the tail of α2 could induce the a1 homeodomain to bind tightly to DNA even if it were not covalently attached to it. To test this idea, a wild-type and two mutant α2 tail peptides were synthesized. All three peptides are 19 amino acids in length and end at residue 207 of α2. One mutant contains a Serine at position 196 and the other an Alanine, changes known to disrupt the interaction of the tail with a1 in the a1/α2 heterodimer (Li et al., 1995) The experiment of Figure 5 shows that the affinity of the a1 homeodomain for DNA increases as the concentration of wild-type peptide is raised (lanes 2-6), with half-maximal stimulation reached at a peptide concentration of approximately 0.3 mM, a value in excellent agreement with the K_D of 0.2-0.3 mM for the interaction of the α2 tail with the a1 homeodomain as measured by NMR spectroscopy (Baxter et al., 1994; Philips et al., 1994) The two mutant peptides also stimulate a1 binding (lanes 7-11 and 12-16), but to a

significantly lesser extent. In other experiments, the difference between the wild-type and mutant peptides was less accentuated, suggesting that the a1/peptide complex is highly sensitive to the conditions used.

Figure 2.4. DNA binding specificity of the α2::6::a1 chimera. For this electrophoretic mobility shift experiment an 80 nucleotide pair ^{32}P-labeled fragment containing the a1/a1 operator and an identical labeled fragment with the operator deleted (consisting of 51 nucleotide pairs) were utilized. Lanes 1, 3-5 and 9-11 contain the a1/a1 fragment while 2, 6-8 and 12-14 contain the fragment that lacks specific a1-binding sites. Lanes 1 and 2 lack protein and the other lanes have the indicated protein or proteins added. The binding conditions and protein concentrations were the same as those used in Figure 2.2.

Figure 2.4

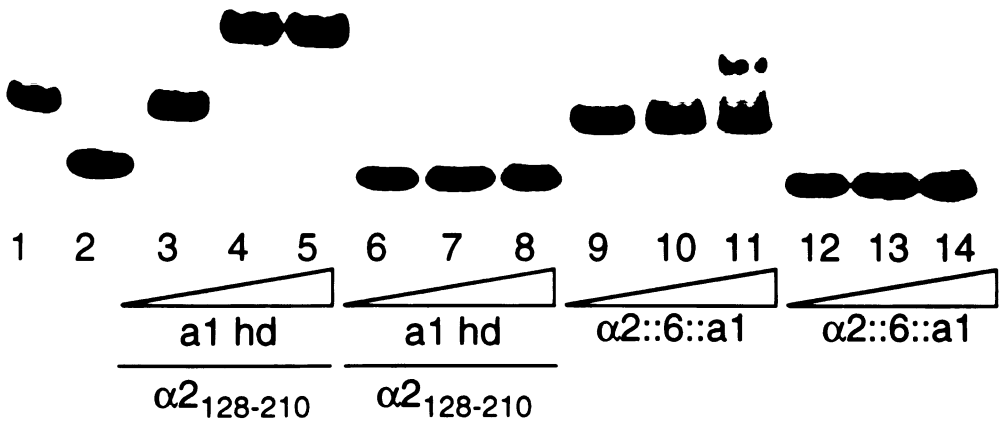
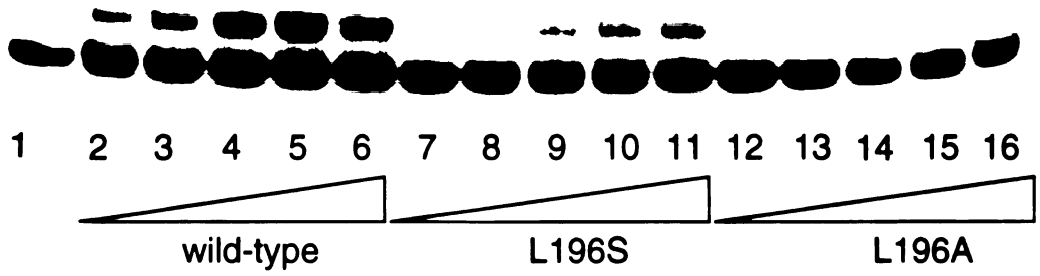

1 2 3 4 5 6 7 8 9 10 11 12 13 14
a1 hd
a1 hd
α2::6::a1
α2::6::a1
α2128-210
α2128-210

Figure 2.5. A 19 amino acid α2 tail peptide, supplied in *trans*, induces a1 to bind the a1/a1 operator. Each lane contains 30 nM a1 hd, which does not bind DNA on its own (lane 1). In addition to a1 hd, lanes 2-6 contain successive 2-fold increases of the wild-type α2 tail peptide beginning with a concentration of 0.15 mM in lane 2 and ending with 2.5 mM in lane 6. Both mutant peptides, present in the same concentrations as the wild-type peptide, are reduced in their ability to induce a1 DNA binding (L196S, lanes 7-11; L196A, lanes 12-16). The binding conditions are the same as those of Figure 2 except that the incubations of DNA and protein were carried out at 4°, as was the electrophoresis.

Figure 2.5



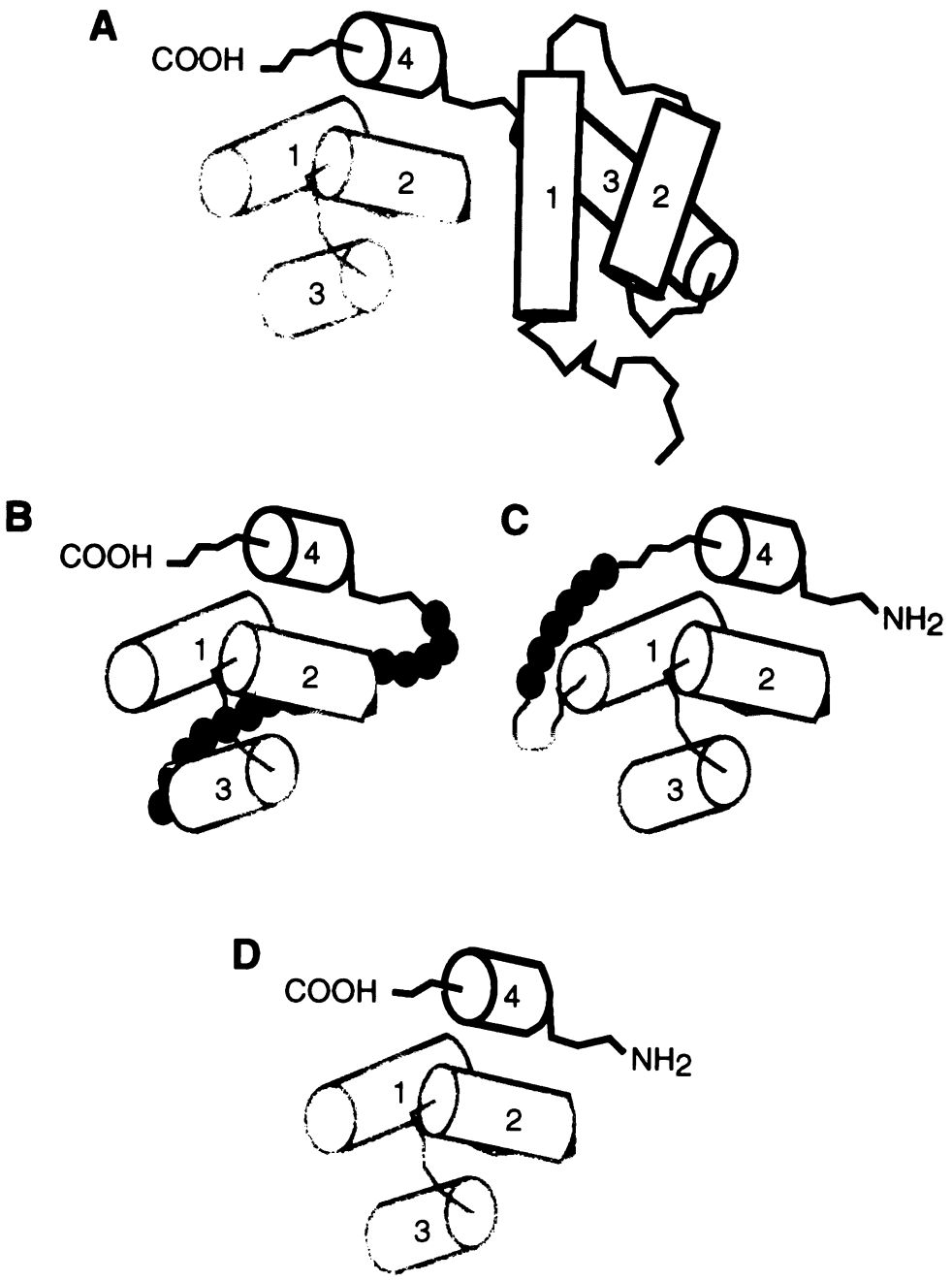
Discussion

a1 and α2 are homeodomain proteins that regulate cell identity in the budding yeast *S. cerevisiae*. These two proteins interact in solution and then bind to hsg operators, turning off the haploid-specific genes. Neither protein alone binds the DNA with the necessary affinity or specificity to efficiently select its target genes. A key protein-protein interaction interface in the heterodimer is comprised of the homeodomain of a1 and a 22 amino acid region of α2 C-terminal to the homeodomain. This α2 tail undergoes a conformational change upon interaction with a1, and in this paper we have now provided evidence (summarized in Fig. 6) that the homeodomain of a1 undergoes a reciprocal conformational change upon interaction with α2. We propose that this conformational change in a1 is required for tight and specific DNA binding by the heterodimer.

Upon a1/α2 heterodimerization, the tail of α2, which is unstructured in the monomer, assumes a distorted α helix upon interaction with a1 (Li et al., 1995; Philips et al., 1994) Since the tail is distant from the DNA in the bound heterodimer, this change in α2 has no consequence on the DNA-binding properties of α2 (Mak and Johnson, 1993) What could be the reciprocal change in a1 that increases its DNA-binding affinity? Baxter et al. (Baxter et al., 1994) showed, using ^{15}N labeled a1 and ^{14}N α2 and applying isotope-edited NMR spectroscopy, that the resonances of many positions of the a1 homeodomain changed upon addition of α2. Some of these changes can be accounted for by direct contact by the α2 tail; however, others lie in positions more distant from the sites of direct contact. A cluster of α2-induced changes lies in the loop between helix 1 and helix 2 of the a1 homeodomain. Since this loop makes contact with both the DNA (via a water molecule) and with the α2

Figure 2.6. Cartoon depicting interaction of the α2 tail with the a1 hd. a1 is shown in gray and α2 is shown in black. The filled black circles refer to the Glycine/Serine linkers (See Fig. 1). (**A**) The a1/α2 heterodimer as visualized in the x-ray structure of the heterodimer (Sauer et al., 1988). (**B**) Inferred structure of the two chimeras (a1::11::α2 and a1::16::α2) in which the α2 tail and linker were fused to the C-terminus of a1. (**C**) Inferred structure of the α2::6::a1 chimera, in which the α2 tail and linker were fused to the N-terminus of the a1 homeodomain. (**D**) Inferred structure of the a1 hd bound by wild-type α2 tail peptide. For each case, we have provided evidence that the interaction of the α2 tail with the homeodomain of a1 induces a conformational change which stimulates the binding of the a1 homeodomain to haploid-specific operator DNA. The contacts between a1 and DNA occur principally through helix 3 and the loop between helices 1 and 2 (Sauer et al., 1988). As described in the text, it has been proposed that the conformational change in a1 occurs in this loop.

Figure 2.6



tail in the heterodimer/DNA structure, it is plausible that contact with the α2 tail repositions the entire loop to maximize a1 binding (Li et al., 1995). A network of protein-DNA contacts involving the backbone of this loop is observed in the x-ray structure, and even a subtle change in the conformation of this loop (caused by dissociation of α2) could disrupt this network of contacts and significantly weaken the affinity of a1 for DNA.

We showed previously that an a1::α2 chimera containing the tail of α2 linked covalently to the C-terminal end of the a1 homeodomain binds DNA only as a dimer (Stark and Johnson, 1994). According to the a1/α2/DNA x-ray structure, this chimera is not capable of undergoing an intramolecular interaction to bring the α2 tail in contact with the proper surface of the a1 homeodomain because the linker is of insufficient length. Thus the only way for this chimera to efficiently bind the operator was through the interaction of the tail of one molecule with the homeodomain of a second. We believe this is the explanation for the dimer requirement of this chimera. In contrast, the chimeras described in this paper were capable of monomeric DNA binding, and therefore of intramolecular interactions between the α2 tail and the a1 homeodomain.

Other homeodomain proteins also appear to undergo partner-induced conformational changes which increase their DNA-binding affinities. The affinity of the mammalian homeodomain protein Pbx1 for DNA is enhanced by YPWM-containing peptides derived from several of the Hox proteins which are partners of Pbx1 (Knoepfler and Kamps, 1995; Peltenburg and Murre, 1996). Chan et al., (1996) have proposed that the homeodomain of the fly *labial* protein is masked by a domain located N-terminal to the homeodomain and that contact by a partner protein, extradenticle, relieves this inhibition and allows *labial* to bind DNA. Finally, the idea that α2 provides a ligand that increases the affinity of a1 for DNA is similar in principle to the many cases of small molecules

that activate the binding of proteins to DNA. Examples include such ligands as cAMP for *E. coli* CAP (Beckwith, 1987; Ebright, 1993; Reznikoff, 1992) and tryptophan for the *E. coli* trp repressor (Somerville, 1992; Yanofsky and Crawford, 1987). In the case of a1 and α2, the signal to bind DNA and to thereby turn off transcription of the haploid-specific genes is the presence of both proteins in the same cell. A ligand-induced conformational change seems an efficient way of ensuring that a1 is inactive in cells that lack α2 (**a** cells), but becomes activated only when α2 is also present, the combination that determines the **a**/α cell-type.

Materials and Methods

Construction of a1::α2 chimeras

The DNA encoding the Glycine/Serine linkers was synthesized as complimentary oligonucleotides, annealed, and ligated between DNA encoding the a1 hd and DNA encoding the tail of α2. The linkers are composed of alternating $(Glycine)_2$ and $(Serine)_2$. The oligonucleotides that made up the linkers in the a1::α2 chimeras are as follows: a1::11::α2 - GATCTAAAGGTGGTTCTTCTGGCGGCTCCTCCG; a1::16::α2 - GATCTAAAGGTGGTTCTTCTGGCGGCTCCTCCGGTGGCTCTTCCGGCG; α2::6::a1 - GGTGGTTCTTCTGGT. The first two oligonucleotides have an overhanging GATC at the 5' end of each oligonucleotide of the pair, and were cloned into the Bgl II site at the junction of the a1 homeodomain and the α2 tail in the a1::α2 chimera with no linker (Stark and Johnson, 1994). The third linker oligonucleotide was part of a larger oligonucleotide which contains the α2 tail sequence (aa 189-210) immediately upstream of the linker. This oligonucleotide pair has an overhanging TA at each 5' end and was cloned into the Nde I site at the 5' end of the a1 homeodomain (Philips et al., 1994).

Peptides

α2 tail peptides were synthesized by California Peptide Research, Inc. All three peptides are identical except for position 196; they are 19 amino acids in length, beginning at residue 189 of α2 and ending at residue 207. The amino acid sequence of the wild-type α2 tail peptide is TITIAPE**L**ADLLSGEPLAK. Residue 196 is shown in bold. α2 ends with residue 210 and residues 208-210 are not required for a1/α2 repression (Mak and Johnson, 1993). Peptides were

HPLC-purified, resuspended in H_2O, and concentrations were determined by the quantitative Ninhydrin assay (Sarin et al., 1981).

Operators

The a1/a1 operator has the same spacing and binding site orientation as a naturally occurring hsg operator (which contains an a1 and α2 binding site), except the α2 binding site of the hsg operator has been replaced by a second a1 binding site (Goutte and Johnson, 1994). The two a1 binding sites are separated by six base pairs and, for the experiments of Figs. 2-5, are contained within an 80-bp DNA fragment. A second DNA fragment, identical in sequence to the a1/a1 operator-containing fragment except that it contains no specific a1-binding sites, was used in the experiment of Figure 4. The removal of the a1-binding sites results in a 51 nucleotide pair DNA fragment.

Protein purification

The a1 hd protein and all a1::α2 chimeric proteins were overexpressed in *E. coli* strain BL21(DE3)pLysS. Protein purification from cell lysates was by adhesion to a cation-exchange resin (Sephadex SP-C50, Pharmacia) followed by elution with a NaCl gradient (Philips et al., 1994). The α2 hd + tail fragment was a gift from A. Vershon.

DNA-binding assays

For the electrophoretic mobility shifts purified proteins were incubated with a ^{32}P-labeled DNA fragment for 30 min at room temperature and electrophoresed through a 5% native Tris-borate-EDTA polyacrylamide gel as described in (Stark and Johnson, 1994). The binding conditions for the α2 tail peptides and labeled a1/a1 operator were the same as in the other experiments

except that the incubations of DNA and protein were carried out at 4^{o}C, as was the electrophoresis. The DNase I protection experiment was carried out under the same conditions as the mobility shift assay except that 10-50 times more DNA was used, and the binding buffer contained no glycerol or *E. coli* genomic DNA, but was supplemented with 10mM $CaCl_2$ and 2.5μg ml^{-1} calf thymus DNA. Reactions were cleaved for 10 min at room temperature with 1.5 μg DNase I (Worthington) and then stopped and precipitated with 1.6 M ammonium acetate. Samples were electrophoresed through a 10% denaturing TBE gel.

REFERENCES

Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F., and Dahlquist, F. W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and a1: secondary structure determination of the a1 homeodomain and changes produced by alpha 2 interactions. Biochemistry *33*, 15309-20.

Beckwith, J. (1987). The Lactose Operon. In Escherichia coli and Salmonella Typhinurium: Cellular and Molecular Biology, F. C. Neidhardt, ed. (Washington, D.C.: American Society for Microbiology), pp. 1444-1452.

Dranginis, A. M. (1990). Binding of yeast **a**1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. Nature *347*, 682-685.

Ebright, R. H. (1993). Transcription activation at Class I CAP-dependent promoters. Mol Microbiol *8*, 797-802.

Goutte, C., and Johnson, A. D. (1988). a1 Protein Alters the DNA Binding Specificity of α2 Repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1994). Recognition of a DNA Operator by a Dimer Composed of Two Different Homeodomain Proteins. EMBO *13*, 1434-1442.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and α2 Homeodomain Proteins Form a DNA-binding Activity with Properties Distinct from those of either Protein. J. Molec. Biol. *233*, 359-371.

Herskowitz, I. (1989). A regulatory hierarchy for cell specialization in yeast. Nature *342*, 749-757.

Herskowitz, I., Rine, J., and Strathern, J. (1992). Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*. In The molecular and cellular biology of the yeast Saccharomyces (Cold spring harbor: Cold spring harbor laboratory press), pp. 583-656.

Johnson, A. D. (1995). Molecular mechanisms of cell-type determination in budding yeast. Current Biology *5*, 552-558.

Keleher, C. A., Redd, M. J., Schultz, J., Carlson, M., and Johnson, A. D. (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. Cell *68*, 709-719.

Knoepfler, P. S., and Kamps, M. P. (1995). The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. Mol Cell Biol *15*, 5811-9.

Komachi, K., Redd, M. J., and Johnson, A. D. (1994). The WD repeats of Tup1 interact with the homeo domain protein alpha 2. Genes Dev *8*, 2857-2867.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA [see comments]. Science *270*, 262-9.

Mak, A., and Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein α2 is required for function with a second homeo domain protein. Genes Dev. *7*, 1862-1870.

Mukai, Y., Harashima, S., and Oshima, Y. (1991). AAR1/TUP1 protein, with a structure similar to that of the beta subunit of G proteins, is required for a1-alpha 2 and alpha 2 repression in cell type control of Saccharomyces cerevisiae. Mol Cell Biol *11*, 3773-3779.

Peltenburg, L. T., and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. Embo J *15*, 3385-93.

Philips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994). Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2. Biochemistry *33*, 9294-9302.

Philips, C. L., Vershon, A. K., Johnson, A. D., and Dahlquist, F. W. (1991). Secondary structure of the homeo domain of yeast α2 repressor determined by NMR spectroscopy. Genes Dev. *5*, 764-772.

Reznikoff, W. S. (1992). Catabolite gene activator protein activation of lac transcription. J Bacteriol *174*, 655-8.

Sarin, V. K., Kent, S. B., Tam, J. P., and Merrifield, R. B. (1981). Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. Anal Biochem *117*, 147-57.

Sauer, R. T., Smith, D. L., and Johnson, A. D. (1988). Flexibility of the yeast α2 repressor enables it to occupy the ends of its operator, leaving the center free. Genes Dev. *2*, 807-816.

Smith, D. L., and Johnson, A. D. (1992). A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α2 dimer. Cell *68*, 133-142.

Somerville, R. (1992). The Trp repressor, a ligand-activated regulatory protein. Prog Nucleic Acid Res Mol Biol *42*, 1-38.

Stark, M. R., and Johnson, A. D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. Nature *371*, 429-32.

Strathern, J., Shafer, B., Hicks, J., and McGill, C. (1988). a/α-Specific Repression by MATα2. Genetics *120*, 75-81.

Strathern, J. N., Hicks, J., and Herskowitz, I. (1981). Control of cell type in yeast by the mating type locus: the α1-α2 hypothesis. J. Molec. Biol. *147*, 357-372.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-528.

Yanofsky, C., and Crawford, I. P. (1987). The Tryptophan Operon. In Escherichia coli and Salmonella Typhinurium: Cellular and Molecular Biology, F. C. Neidhardt, ed. (Washington, D.C.: American Society for Microbiology), pp. 1453-1472.

CHAPTER 3

Isolation and Characterization of Mutations in the a1 Homeodomain that Enable DNA Binding in the Absence of α2

Abstract

The yeast homeodomain proteins a1 and α2 interact to form a heterodimer that binds DNA with high affinity and specificity. Neither protein alone is able to bind DNA with high affinity, but upon heterodimerization both proteins appear to undergo conformational changes, allowing tight and specific binding to the DNA. The details of the conformational change in α2 have been determined by direct NMR experiments as well as by comparison of the crystal structures of α2 by itself with that of the a1/α2 complex. However, the crystal structure of a1 by itself has not yet been determined, so the details of the conformational change in a1, which are proposed to increase its DNA-binding affinity, are unknown. We have isolated mutations in the a1 homeodomain that allow a1 to bind to the DNA in the absence of α2. Some of these mutants may assume, at least in part, the a1 conformation normally evoked by contact with the α2 protein. These mutants may provide insight into the sites responsible for, and the nature of, the changes in a1 conformation provoked by α2.

Introduction

In the yeast *Saccharomyces cerevisiae*, the a1 and α2 proteins, in conjunction with the global repressor proteins, Tup1 and Ssn6, modulate the transcription of a number of cell-type specific genes (Johnson, 1992). a1 and α2 both contact DNA through a region known as a homeodomain - a highly conserved, compact three-helix structure. a1 and α2 act together to recognize a set of genes known as the haploid-specific genes (hsg), the repression of which is essential for the **a/α** diploid cell type (Dranginis, 1990; Goutte and Johnson, 1988; Goutte and Johnson, 1994). a1 and α2 interact in solution and bind to an hsg operator as a heterodimer.

Neither the a1 nor the α2 protein binds the DNA with the necessary affinity or specificity to form a stable complex by itself (Goutte and Johnson, 1993). a1 on its own binds an hsg operator in vitro very poorly, if at all. α2 is able to bind to the hsg operator; however, this binding is only moderately specific, with a dissociation constant only 10-fold lower than that for random DNA. In contrast, the a1/α2 heterodimer binds the same DNA with a dissociation constant 3,000-fold lower than that for random DNA.

The primary a1/α2 protein-protein interaction interface is comprised of the homeodomain of a1 and a 22 amino acid region of α2 C-terminal to the homeodomain (the α2 C-terminal tail) (Mak and Johnson, 1993; Stark and Johnson, 1994). Solution NMR and Xray crystallographic studies have shown that the tail of α2, which is unstructured when α2 is free in solution or bound to DNA without a1, forms a fourth helix upon interaction with a1 (Li et al., 1995; Philips et al., 1994). The α2 tail contacts only the a1 homeodomain, making no DNA contacts, and the homeodomain of α2 is not involved in the interaction with a1 (Li et al., 1995; Mak and Johnson, 1993; Stark and Johnson, 1994). These studies also delineated the surface of a1 that is contacted by the α2 tail. The a1 interface is comprised of residues in helix one and helix two, as well as in the loop that connects the two, that face away from the DNA and are therefore accessible (Li et al.,

1995). Three of these residues form a hydrophobic patch on the surface of a1, (Val19, Leu26, and Val34), on which the tail of α2 sits. In addition to these hydrophobic interactions, the heterodimer is stabilized by several hydrogen bonds.

The a1/α2 heterodimer forms an extensive set of contacts with the DNA. Both a1 and α2 contact DNA bases and the sugar-phosphate backbone of the DNA by means of direct side chain contacts and water-mediated hydrogen bonds (Li et al., 1995). The two proteins are positioned on the DNA in a very similar manner, although the exact bases contacted differ.

It is surprising, given the extensive set of contacts formed by the a1 hd with DNA, and given the data suggesting that a1 provides most of the DNA-binding specificity of the a1/α2 heterodimer, that a1 binds DNA detectably only in the presence of α2 (Goutte, 1992; Goutte and Johnson, 1993; Li et al., 1995; Stark and Johnson, 1994; Vershon et al., 1995). Dimerization of a1 and α2 is clearly important for the generation of the highly specific a1/α2 repressor activity, as is a large conformational change in the C-terminal tail of α2 (Goutte and Johnson, 1993; Li et al., 1995; Philips et al., 1994). We have suggested that in the absence of α2, the a1 protein folds into a conformation that does not favor DNA binding (Stark and Johnson, 1994). According to this proposal, the α2 tail, upon contact with a1, induces a small conformational change in the a1 homeodomain, increasing its overall DNA-binding affinity and rendering it competent to bind to the DNA. The following lines of evidence, described in Chapter 2 and Appendices A and B (Li et al., 1995; Philips et al., 1994), support this model. First, a peptide corresponding to the tail of α2, either supplied in *trans* or covalently attached to the a1 homeodomain via a short Glycine-Serine linker, can induce the a1 homeodomain to bind to the DNA as a monomer (Chapter 2). Second, solution NMR studies show that residues in the a1 homeodomain that lie in portions of helices one and two and in the loop connecting them are perturbed upon addition of α2 (Philips et al., 1994). And, third, the crystal structure shows a complex set of interactions between helix three, the DNA backbone, a bound water

molecule, and the backbone of a1 in the loop between helices one and two (Li et al., 1995). Changes in the a1 loop or surrounding helices could easily affect this network of contacts, diminishing or increasing, the affinity of a1 for DNA depending on whether or not α2 contacted a1.

To address the question of how the conformational change in the a1 homeodomain is brought about by the tail of α2, we designed a screen to isolate mutations in the a1 homeodomain that would allow it to bind to DNA in the absence of α2. Three types of mutations could lead to the desired phenotype: 1) mutations that alter the conformation of a1, making it more like that of a1 in the presence of α2; 2) mutations that solely act to increase the DNA binding affinity of a1 through the formation of additional DNA contacts, not involving any changes in the overall conformation of the a1 protein, and; 3) mutations that create an interaction interface for a DNA-binding protein other than α2 (see Discussion). The first class of mutations could provide a means to define the sites responsible for, and the nature of, the change(s) in a1 conformation induced by α2. Mutations that fall into all three classes appear to have been isolated from this screen.

Results

In order to detect mutants of a1 capable of binding DNA in the absence of α2, the strong activation domain of Gal4 was attached to the amino-terminus of the a1 homeodomain (a1 hd). This allows the use of a reporter system which is not activated unless the Gal4::a1 fusion binds upstream of the reporter gene. The plasmid containing this Gal4::a1 fusion was transformed into a matΔ strain containing an integrated CYC1-LacZ reporter with four a1 binding sites located upstream of the CYC1 TATA (yMS1), in place of all the endogenous CYC1 upstream activating sequences (see Fig. 3.1). The default state of this reporter is OFF, meaning that it requires an activator in order for LacZ to be transcribed. Therefore, LacZ will only be produced if the a1 hd binds to the sites in the reporter, bringing the Gal4 activation domain to the DNA. When a1 does not bind the reporter construct, the colonies will be white, whereas mutants of a1 which can bind to the reporter construct will result in blue colonies in a β-Galactosidase filter assay.

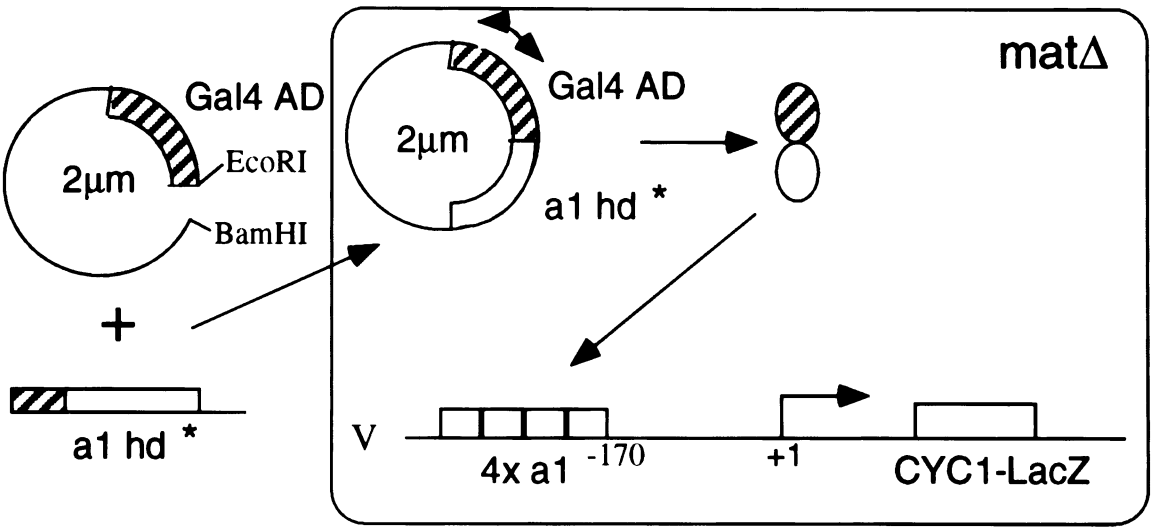
The a1 hd was mutagenized by PCR, and the mutagenized fragment was transformed into yMS1 along with a high copy plasmid containing the Gal4 activation domain under the control of the ADH1 promoter (pGAD424) that had been linearized with two restriction enzymes that cut in the polylinker immediately downstream of the activation domain (Fig. 3.1). The a1 hd PCR product was designed to contain a region homologous to the COOH-terminus of the Gal4 activation domain on its 5' end and a region homologous to pGAD424 vector sequences on the 3' end, so that upon cotransformation with the gapped pGAD424 plasmid, the plasmid and the PCR product will recombine in vivo, fusing the mutagenized a1 hd to the Gal4 activation domain.

Approximately 25,000 colonies were screened using a β-Galactosidase filter assay. Thirty five colonies out of the 25,000 were pale blue on the filters, and 12 of these remained blue upon retesting. The plasmids containing the a1 hd mutants were rescued from these 12 isolates and sequenced (Fig. 3.1). Nine out of 12 of the mutants contain a

Figure 3.1. a1 homeodomain mutations isolated from screen.

The schematic at the top of the figure illustrates the components involved in the screen for a1 hd mutants capable of binding DNA in the absence of α2. The PCR-mutagenized a1 hd was cotransformed into a matΔ strain containing a UASless CYC1-LacZ reporter with four a1 binding sites upstream of the CYC1 transcriptional start site along with a gapped plasmid containing the Gal4 activation domain. In vivo recombination of these two DNAs results in the Gal4::a1* fusions. Only cells in which the Gal4::a1* fusion binds to the reporter will be blue in a β-Galactosidase filter assay. The list below the schematic shows the positive mutations isolated from the screen by original mutant number and by the corresponding amino acid mutations in each. The numbering of amino acids is in accord with the numbering of the a1 hd in the crystal structure of the a1/α2 ternary complex. K-3E corresponds to a mutation 3 amino acids upstream of the NH_2-terminal end of the homeodomain (the PCR-mutagenized a1 hd fragment contains 4 amino acids upstream of the homeodomain proper).

Figure 3.1



Blue

mutant	a1 hd amino acid mutations
10-1	Q24R, N27D, K31E, R55G
12-1	K31E
72-1	K31R
74-1	S7L, K31E
76-2	K52R
79-1	K31E
89-1	K31E
115-1	V19A
134-1	R22K
146-1	K31E
146-2	K-3E,K31Q
162-1	K31E, E32G

mutation in Lys31 of helix two in the a1 hd. Seven of these mutations have the lysine changed to a glutamate, one has a lysine -> glutamine change, and one a lysine -> arginine change. Some of these contain the single Lys31 mutation, while others contain additional mutations. Of the three remaining mutants, one contains a mutation in helix one of the hd (valine 19 -> alanine), one contains a change in the loop between helices one and two (arginine 22 -> lysine). and one contains a change in a residue in helix three (lysine 52 -> arginine) (Fig. 3.2).

Five mutations were chosen for further characterization: V19A, K31R, K31E, K52R, and the double mutation K-3E, K31Q. These mutations were transferred into an *E. coli* expression plasmid using PCR, and the mutant proteins were overexpressed and purified from *E. coli*. The purified proteins were analyzed by gel mobility shift assay for their ability to bind DNA in the absence of α2 (Fig. 3.3). Mutant K31R binds to a synthetic operator with two a1 binding sites at least 10-fold more tightly than the wild-type a1 hd. Two other mutants, V19A and K52R, are capable of binding to the DNA in the absence of α2, but the strength of these interactions is questionable, and in subsequent gel shift assays the affinity of these mutants for DNA looked similar to that of wild-type a1. Mutants K31E and K-3E, K31Q show no detectable binding to the DNA in the absence of α2. All of the a1 hd mutants except K31E bind to an a1/α2 operator cooperatively with the C-terminal fragment of α2 (aa 128-210) with a binding affinity equal to or better than that of wild-type a1 hd (data not shown).

Figure 3.2. Positions of mutations on the a1 homeodomain.

The four panels show a stick model of the a1 hd bound to DNA based on the crystal structure coordinates (Li et al., 1995). The NH_2-terminus of helix one is highlighted in yellow. The side chains of the residues in which mutations were isolated are highlighted in green. The green side chains correspond to those of the original residues, not the mutated ones. (A) Val19. (B) Arg22. (C) Lys52. (D) Lys31.

Figure 3.2

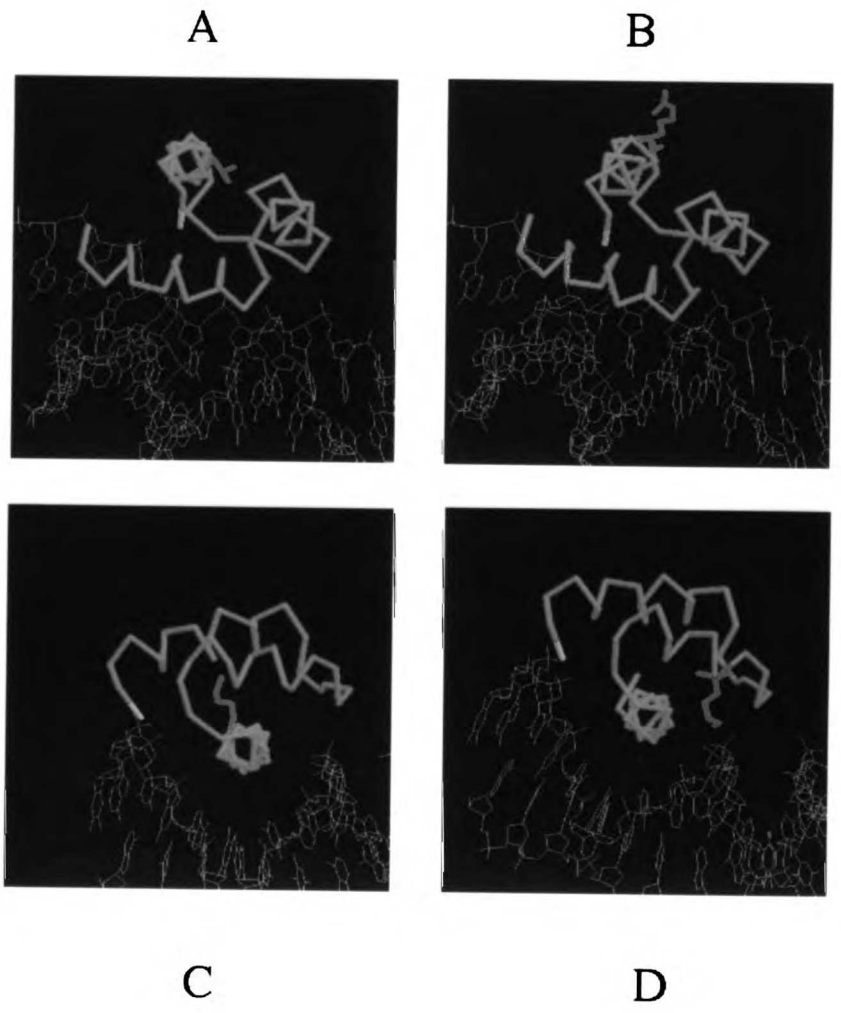


Figure 3.3. DNA binding analysis of a1 homeodomain mutants by gel shift assay. The ^{32}P-labelled DNA fragment containing the a1/a1 operator was incubated with the indicated purified protein for 30 min on ice and electrophoresed through a 5% native Tris-borate-EDTA polyacrylamide gel at 4°C. Lanes 1-2 contain 30nM and 100nM wild-type a1 hd, respectively. Lanes 3-4 contain 30 nM and 100nM a1 hd mutant K100R, respectively. Lanes 5-8 contain 100nM of the a1 hd mutant indicated.

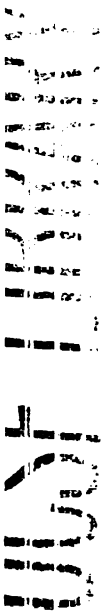
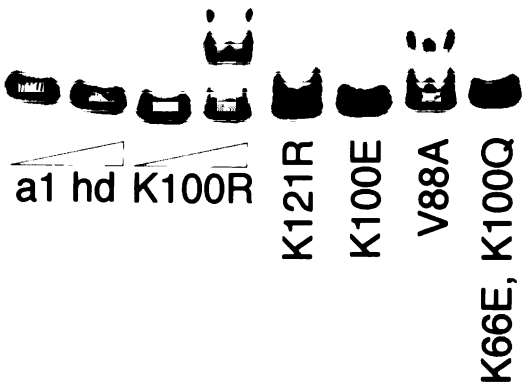


Figure 3.3



Discussion

The a1 protein is transcribed and translated in both **a** and **a**/α cells in the yeast *Saccharomyces cerevisiae* (Herskowitz, 1989). In the **a**/α diploid cell, a1 interacts with α2 to repress the haploid specific genes, but in the **a** cell a1 has no known function (Kassir and Simchen, 1976). This appears to be due to the fact that a1 is unable to bind to DNA in the absence of α2. Work presented in this thesis strongly supports a model in which the interaction with the tail of α2 induces a conformational change in the a1 hd, increasing its overall DNA-binding energy and allowing it to bind the DNA tightly and specifically as a heterodimer with α2.

The interaction between the short tail of α2 and the homeodomain of a1 is reminiscent of the interaction between the prokaryotic helix-turn-helix cyclic AMP (cAMP) receptor protein, CRP, and its ligand cAMP. cAMP acts as a small effector molecule, inducing a conformational change in CRP, and changing CRP's weak nonspecific binding activity into a tighter affinity for specific sites (Crothers and Steitz, 1992). Mutations have been isolated in CRP (CRP*) that enable CRP to bind to DNA and activate transcription without cAMP (Aiba et al., 1985; Garges and Adhya, 1985). These mutant proteins appear to have a conformation similar to cAMP-bound CRP. Some of the CRP* mutations map to a region outside of the cAMP-binding domain, analogous to the region of a1 shown by NMR and Xray crystallography to be involved in the interaction with α2.

The screen for a1 hd mutants capable of binding DNA in the absence of α2 was designed with the hope of isolating mutations in the a1 hd which alter the conformation of a1 to a form similar to that achieved in the presence of α2. Twelve mutations in the a1 hd were isolated from this screen. Two of these mutations encode residues in the region of a1 contacted by the α2 tail. These residues face away from the DNA, and are therefore incapable of making any DNA contacts. Theoretically, these two residues, Val19 and Arg22, are the best candidates for involvement in the a1 conformational change to come out

of the screen, as they are not likely to be simply increasing the DNA binding energy of the a1 hd. Val19 forms part of the hydrophobic patch on the surface of the a1 hd that interacts with the tail of α2. Arg22 is located in this same area, although it does not seem to be involved in the interaction with the α2 tail (see Appendix B, Fig.4A).

Several attempts were made to quantitate the levels of activation of the reporter construct in the mutant strains. Although there is a clear difference in color between the wild-type Gal4::a1 hd fusion control and the mutant Gal4::a1 hd fusions by the β-Galactosidase filter assay used to identify the mutants in the screen, the difference is small enough that it is unquantifiable by the available liquid β-Galactosidase quantitative assays. In addition to the relatively small differences in activation in vivo between wild-type and mutant a1, in vitro gel shift analysis on the V19A mutant shows some variability in its DNA binding affinity - in some instances it binds DNA more tightly than wild-type a1, and in other instances binding affinity of wild-type and mutant appear to be similar. No biochemical analysis was performed on the R22K mutant.

A mutation containing a change in helix three of the a1 hd, K52A, also came through the screen. The protein containing this mutation, like that containing the V19A mutation, shows some variablity in its DNA binding affinity; it sometimes binds DNA slightly more tightly than wild-type a1, and other times displays an equal binding affinity. Lys52 faces away from the DNA, and does not make any DNA contacts. It is therefore also a candidate for a mutation that induces a conformational change in a1. This lysine appears to pack up against residues in the end of helix one near the loop, in the region contacted by α2. The change from lysine to the slightly bulkier arginine may push helix three away from helices one and two, and into the major groove of the DNA.

The mutant showing the strongest binding to DNA in vitro in the absence of α2 contains the K31R mutation. The bacterially-produced mutant protein binds DNA at least 10-fold more tightly than does the wild-type a1 hd (Fig. 3.3). In the wild-type protein Lys31 indirectly contacts a phosphate in the DNA backbone through a water molecule. The

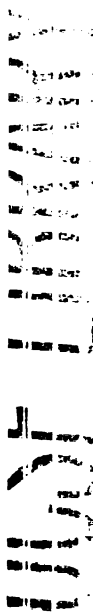
K31R mutant could be binding the DNA more tightly either because the mutation changes the conformation of the a1 hd into a more favorable DNA binding form, or because the mutation simply increases the overall DNA binding affinity of the a1 hd by forming an additional hydrogen bond with the water molecule through which it contacts the DNA. The available data are consistent with both models.

The most common mutation isolated was in Lys31 in helix 1 of the homeodomain. Seven mutations changed lysine 31 -> glutamate, and surprisingly, upon testing a bacterially-produced a1 hd protein containing this single mutation, I found that this mutant is unable to bind to DNA under the conditions used in the gel shift assay. In addition, this mutant does not bind the DNA cooperatively with α2. In the context of the full-length a1 protein, however, it does interact with α2, inhibiting repression of an hsg reporter in an **a**/α cell where repression of this reporter is normally strong (data not shown). The simplest explanation for why this mutant came out of a screen designed to isolate a1 hd mutants capable of binding DNA in the absence of α2 is that the K31E mutation creates a new interaction interface for another yeast DNA binding protein, (a homeodomain protein such as Pho2?), and that this new heterodimer (or heteromultimer) binds to the reporter construct.

The a1 hd mutant containing K31Q behaves similarly to the K31E mutant, in that it is unable to bind alone to the DNA in a gel shift assay. However, this mutant is capable of binding DNA cooperatively with α2 in the same assay. The creation of a new interaction interface for another DNA-binding protein remains a plausible explanation for why this mutant came through the screen.

Of the twelve mutations isolated from the screen, four appear to be worthy of further analysis, V19A, R22K, K52R, and K31R. NMR and/or crystallographic structures comparing the wild-type a1 hd alone to the a1/α2 complex, and to the various mutant a1 hds would provide the most information regarding the a1 conformational change. In the meantime, some information about the different conformations of a1 could come

from protease accessibility studies on a1 alone compared with a1 plus the α2 tail peptide and the four a1 hd mutants mentioned above.



Methods

Plasmids and strains

The synthetic reporter gene containing four a1 binding sites (4x a1) upstream of a UASless CYC1 promoter fused to the LacZ gene (pMS98) was constructed by inserting an oligonucleotide duplex containing four a1 binding sites (5'-TCGAGTGATGTAACTAATTACATCAGTTAAGTGATGTAATTAATTACATCAG -3'), each separated by six nucleotide pairs, into the XhoI site of a version of pLGΔSS (Johnson and Herskowitz, 1985) from which the 2μm sequences have been removed. pMS98 was linearized with StuI and integrated at the *URA3* locus of KT23αx8 (*matΔ trp1 leu2 ura3 his4*) (Siliciano and Tatchell, 1984; Tatchell et al., 1981). The resulting strain, yMS1, was the strain used in the screen. The high copy plasmid containing the Gal4 activation domain under the control of the ADH1 promoter is pGAD424 (Clontech). pGAD424 was cut with EcoRI and BamHI, and the wild-type a1 hd was cloned in frame into these sites to generate the Gal4::a1 fusion used as a negative control in the screen (pMS94). pGAD424 was cut with the same restriction enzymes before contransformation into yMS1 with the PCR-mutagenized a1 hd, to generate the Gal4::a1* fusions via in vivo recombination.

Several representative a1 hd mutations isolated from the screen were transferred by PCR into the bacterial expression vector pHB40P (Studier and Moffatt, 1986), cut with Nde I and Xba I, for expression in *E. coli* strain BL21 (DE3). The original mutant numbers corresponding to the bacterial expression plasmids are: 72-1 (pMS 157), 76-2 (pMS152), 89-1 (pMS153), 115-1 (pMS155), 146-2 (pMS156).

The synthetic a1/a1 operator fragment used in the gel shift assays was cut out of pCG46 with HindIII and EcoRI, resulting in an ~80 bp fragment (Goutte, 1992). The two a1 sites are separated by 6 bp and lie approximately in the middle of the fragment. Purified α2 C-terminal fragment protein ($\alpha 2_{128\text{-}210}$)was a gift of Andrew Vershon (Waksman Institute, Rutgers, Piscataway, NJ).

PCR mutagenesis

The a1 hd in pMS94 was mutagenized by PCR (Muhlrad et al., 1992). PCR primers for the mutagenic reaction corresponded to regions of pMS94 approximately 200 bp immediately upstream and downstream of the a1 hd gene. The upstream primer binds in the GAL4 gene (MRS113, 5'- CATGAATAATGAAATCACGGC) and the downstream primer binds to vector sequences (MRS114, 5'-AGATGGGCATTAATTCTAGTC). The primers were designed in this manner to provide suitable homology to allow homologous in vivo recombination with pGAD424 (same as pMS94 except contains only the GAL4 activation domain, and not the a1 hd) cut in the polylinker with EcoRI and BamHI, thus generating the library of Gal4::a1* fusions.

The mutagenic PCR reaction contained 10 ng pMS94 linearized with EcoRV in 25 µl, with 100 mM Tris-HCl, pH 8.3, 500 mM KCl, 0.1% (w/v) gelatin, 1 mM dGTP, dCTP, and dTTP, 200 µM dATP, 25 pmoles MRS113, 25 pmoles MRS114, 3mM $MgCl_2$, 0.6 mM $MnCl_2$, and 2.5 U Taq polymerase (Boehringer Mannheim). Each reaction was preheated to 94°C for an initial denaturation step followed by 30 cycles: 94°C 30 sec, 53°C 30 sec, 76°C 30 sec.

The frequency of point mutations is approximately 2% based on DNA sequence analysis of a small number of randomly picked clones.

Isolation of mutants

Yeast transformations were performed by the LiOAc method (Ito et al., 1983). LiOAc-treated yeast (yMS1) were added directly to a tube containing carrier DNA, 100 ng gel-purified gapped plasmid (pGAD424 cut with EcoRI and BamHI) and 100 ng mutagenized PCR fragment (0.5 µl of the PCR reaction). These conditions generate between 500 and 2000 recombinants, depending on the transformation efficiency. The transformed yeast were plated at a density of 50 -200 colonies per plate on plates lacking uracil and leucine (-

Ura-Leu plates). Transformants were replica plated onto Whatman 3MM filters on -Ura-Leu plates containing galactose as the sole carbon source, and grown for 24 hr at 30°C. (The presence of galactose ensures the activity of the Gal4 activation domain.) The colonies were scored for β-Galactosidase production by immersing the filter in liquid nitrogen for 10 sec, placing the filter on a disc of Whatman 3MM paper in a petri dish containing 2 ml of 0.3 μg/ml 5-bromo-4-chloroindoyl-β-D-galactopyranoside (Xgal) in Z buffer (Miller, 1972), and incubating the filter for 2-5 hr.

Potential positives were picked from the original transformation plate, streaked for single colonies, and retested for blueness by the filter assay. Mutant plasmids were isolated from positive colonies (Schena et al., 1989), retransformed into the original yMS1 strain, and retested for blueness. Mutant plasmids passing these tests were transformed into *E. coli* strain BL21 (DE3) by the $CaCl_2$ method.

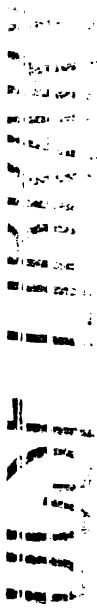
Protein purification

All a1 hd mutant proteins were overexpressed and purified without induction from *E. coli* strain BL21 (DE3) containing the relevant gene under the control of the T7 promoter in the plasmid pHB40P (Studier and Moffatt, 1986). Proteins were purified form cell lysates by adhesion to a cation-exchange resin (Sephadex SP-C50, Pharmacia) followed by elution with a NaCl gradient. The wild-type a1 hd was purified in the same manner, except that it was overexpressed in strain BL21 (DE3) pLysS with IPTG induction as described previously (Philips et al., 1994).

Gel shift assay

Purified proteins were quantitated by electrophoresis through an SDS gel, followed by Coomassie staining. Proteins were incubated with an 80 nucleotide pair ^{32}P-labelled DNA fragment containing two a1 binding sites (Goutte, 1992) for 30 min on ice in a buffer containing 5% glycerol, 10mM Tris-HCl, pH 8, 10 mg ml^{-1} BSA, 0.1 mM EDTA, 5 mM

$MgCl_2$, 0.1% N-P40, 100 mM NaCl, and 3 μg ml^{-1} nonspecific DNA (*E. coli* genomic DNA digested with Hae III). Samples were electrophoresed at 4°C through a 5% native Tris-borate-EDTA polyacrylamide gel for 70 min at 150 V.



REFERENCES

Aiba, H., Nakamura, T., Mitani, H., and Mori, H. (1985). Mutations that alter the allosteric nature of cAMP receptor protein of Escherichia coli. Embo J *4*, 3329-32.

Crothers, D. M., and Steitz, T. A. (1992). Transcriptional Activation by Escherichia coli CAP Protein. In Transcriptional Regulation, S. L. McKnight and K. Yamamoto, ed. (Plainview, New York: Cold Spring Harbor Laboratory Press), pp. 501-534.

Dranginis, A. M. (1990). Binding of yeast **a**1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. Nature *347*, 682-685.

Garges, S., and Adhya, S. (1985). Sites of allosteric shift in the structure of the cyclic AMP receptor protein. Cell *41*, 745-51.

Goutte, C. (1992). Combinatorial control of the yeast homeodomain proteins a1 and α2: UCSF).

Goutte, C., and Johnson, A. D. (1988). a1 Protein Alters the DNA Binding Specificity of α2 Repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1994). Recognition of a DNA Operator by a Dimer Composed of Two Different Homeodomain Proteins. EMBO *13*, 1434-1442.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and α2 Homeodomain Proteins Form a DNA-binding Activity with Properties Distinct from those of either Protein. J. Molec. Biol. *233*, 359-371.

Herskowitz, I. (1989). A regulatory hierarchy for cell specialization in yeast. Nature *342*, 749-757.

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. J Bacteriol *153*, 163-8.

Johnson, A. D. (1992). A Combinatorial Regulatory Circuit in Budding Yeast. In Transcriptional Regulation, S. L. McKnight and K. R. Yamamoto, eds. (Plainview: Cold Spring Harbor Laboratory Press), pp. 975-1005.

Johnson, A. D., and Herskowitz, I. (1985). A repressor (MAT alpha 2 Product) and its operator control expression of a set of cell type specific genes in yeast. Cell *42*, 237-47.

Kassir, Y., and Simchen, G. (1976). Regulation of mating and meiosis in yeast by the mating-type region. Genetics *82*, 187-206.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA [see comments]. Science *270*, 262-9.

Mak, A., and Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein α2 is required for function with a second homeo domain protein. Genes Dev. *7*, 1862-1870.

Miller, J. H. (1972). Experiments in Molecular Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Muhlrad, D., Hunter, R., and Parker, R. (1992). A rapid method for localized mutagenesis of yeast genes. Yeast *8*, 79-82.

Philips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994). Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2. Biochemistry *33*, 9294-9302.

Schena, M., Freedman, L. P., and Yamamoto, K. R. (1989). Mutations in the glucocorticoid receptor zinc finger region that distinguish interdigitated DNA binding and transcriptional enhancement activities. Genes Dev *3*, 1590-601.

Siliciano, P. G., and Tatchell, K. (1984). Transcription and regulatory signals at the mating type locus in yeast. Cell *37*, 969-978.

Stark, M. R., and Johnson, A. D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. Nature *371*, 429-32.

Studier, F. W., and Moffatt, B. A. (1986). Use of Bacteriophage T7 RNA Polymerase to direct selective high-level expression of cloned genes. J. Molec. Biol. *189*, 113-130.

Tatchell, K., Nasmyth, K. A., Hall, B. D., Astell, C., and Smith, M. (1981). In vitro mutation analysis of the mating-type locus in yeast. Cell *27*, 25-35.

Vershon, A. K., Jin, Y., and Johnson, A. D. (1995). A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. Genes Dev *9*, 182-92.

CHAPTER 4

The NH_2-Terminal Arm of the a1 Homeodomain is not Required for DNA Binding Specificity

Abstract

Homeodomain proteins regulate a wide variety of developmental pathways in eukaryote organisms ranging from yeast to man. In order to achieve tight developmental control, these proteins must bind with high specificity to target sites in the DNA. The determinants involved in DNA binding specificity of these proteins are not well understood in general, although several specific models have been suggested. One model suggests that interaction with cofactors raises the binding specificity of the homeodomain protein. Another model predicts that interaction of the highly variable and flexible NH_2-terminal arm of the homeodomain with bases in the minor groove of the DNA largely determines the DNA-binding specificity of some homedomain proteins. We show here, through deletion and mutational analysis, that the NH_2-terminal arm of at least one homeodomain protein (the yeast a1 protein) does not contribute to the DNA binding specificity and biological function of the a1 protein as measured both *in vivo* and *in vitro*.

Introduction

The assembly on DNA of specific combinations of proteins is a fundamental aspect of the regulation of gene expression. A class of gene regulatory proteins that contains a conserved sequence element known as the homeodomain specifies cell type in yeast, and also regulates cell fate, body plan, and development in multicellular organisms (Gehring et al., 1994; Laughon, 1991; Scott et al., 1989). The homeodomain is a DNA binding motif, consisting of a flexible NH_2-terminal arm followed by three α-helices. How do these homeodomain proteins, which generally have only modest specificity for their target sites, find these sites amongst the vast amount of DNA that makes up the genome?

a1 and α2 are homeodomain proteins from the yeast *Saccharomyces cerevisiae* that regulate cell identity (Herskowitz, 1989; Johnson, 1992). These two proteins act together to turn off a set of genes known as the haploid-specific genes (hsg), the repression of which is essential for the **a**/α diploid cell type. Neither protein alone binds the DNA with the necessary affinity or specificity to form a stable complex, but by heterodimerizing they boost their DNA binding specificity, increasing their ability to recognize specific target DNA over random DNA by a factor of at least 300 (Goutte and Johnson, 1993).

The interactions between the a1 and α2 homeodomain proteins have provided some insight into the problem of DNA binding specificity in higher organisms. Flies, for example, contain many homeodomain proteins that are responsible for highly specific effects during development (Gehring et al., 1994; Scott et al., 1989). However, *in vitro*, these same homeodomain proteins bind to similar or identical DNA sequences (Desplan et al., 1988; Ekker et al., 1994; Hoey and Levine, 1988; Kalionis and O'Farrell, 1993). While this apparent discrepancy was confusing at first, it is now thought that many homeodomain proteins typically interact with other proteins in order to achieve high DNA binding specificity which is necessary for tight and specific developmental control.

It has been proposed for many *Drosophila* homeodomain proteins that key DNA-binding specificity determinants of the protein lie in the NH_2-terminal arm of the homeodomain , a span of eight amino acids immediately upstream of the first helix of the homeodomain (Gehring et al., 1994). For example, mutations in NH_2-terminal arm residues known to make DNA contacts in the minor groove have been shown to change the specificity of a homeodomain from one DNA binding site to another, and deletion of the arm can result in severe loss of binding affinity (Furukubo-Tokunaga et al., 1993; Lin and McGinnis, 1992; Zeng et al., 1993).

For the yeast α2 protein, structural, biochemical, and genetic experiments all point to the importance of the NH_2-terminal arm and its interaction with the minor groove (Philips et al., 1991; Vershon et al., 1995; Wolberger et al., 1991). However, the situation for a1 appears quite different. Structural data from the a1/α2/DNA complex shows no electron density for the NH_2-terminal arm of a1 (Li et al., 1995). This observation was surprising given the structural and biochemical data from other homeodomain proteins (including α2) and given the chemical modification experiments that suggested the binding of the arm of a1 in the minor groove (Goutte and Johnson, 1994; Hirsch and Aggarwal, 1995; Kissinger et al., 1990; Klemm et al., 1994; Wilson et al., 1995; Wolberger et al., 1991) . One possibility is that the lack of structure in the a1 NH_2-terminal arm is an artifact of crystal packing forces (Li et al., 1995). The DNA oligonucleotide duplex in the a1/α2/DNA ternary complex is bent by 60^o, causing the DNA to stack end-to-end in the crystal, forming a pseudocontinuous helix with Watson-Crick base pairing between the overhanging 5' nucleotide at the end of one complex and the complementary unpaired bases at the end of the adjacent complex (see Appendix B, Fig. 3B). A 5' phosphate is missing at the junction of each pair of oligos, and it is possible that this missing phosphate causes the a1 arm to be displaced from the minor groove. An alternative possibility is that the a1 NH_2-terminal arm is unstructured because it does not acually lie in the minor groove of the DNA, and does not make any DNA contacts.

In order to investigate the role of the NH_2-terminal arm of a1 in DNA binding affinity and specificity, we isolated mutations (including deletions) in the arm of a1 and examined the ability of the mutated proteins to bind DNA in conjunction with α2, both *in vitro* and *in vivo*. The results show that complete removal of the NH_2-terminal arm does not affect the DNA binding specificity of the a1/α2 heterodimer, and only slightly reduces its DNA binding affinity. Of the three charged residues in the a1 arm that contact the DNA, mutation of two of them does not affect the ability of a1 to cooperate with α2 to repress a synthetic target gene. We conclude that the NH_2-terminal arm of a1 makes no significant DNA contacts in the minor groove and does not significantly contribute to the DNA binding specificity of the a1/α2 complex.

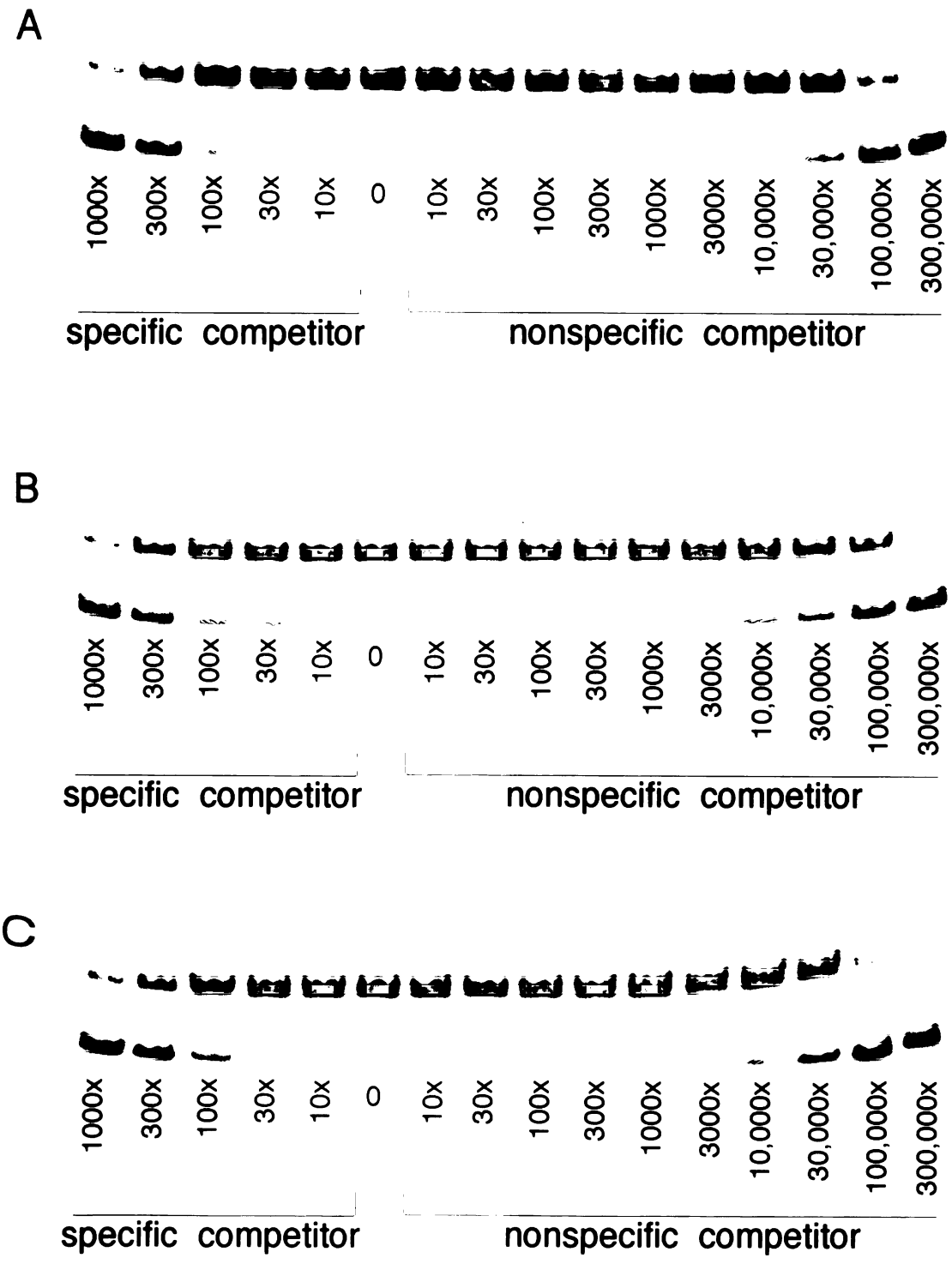
Results

To determine whether the NH_2-terminal arm (amino acids 0-7 using the standard homeodomain numbering, see Appendix B, Fig. 2A) of the a1 homeodomain (a1 hd) is important for DNA binding specificity and/or affinity, we deleted this region from the a1 hd (a1 hd_{9-57}) and examined the ability of this "armless" mutant to bind DNA with α2. In order to compare the specificity of the wild-type a1 hd with that of the armless a1 hd, a series of DNA competition experiments were performed in the presence of the C-terminal fragment of α2. The a1 hd used in these experiments is contained within a bacterial extract while the α2 protein fragment was purified from a bacterial extract. The ability of two different unlabeled DNA fragments to compete with a labeled hsg operator for the binding activity of either wild-type a1 hd/α2 or armless a1 hd/α2 was compared. One of the unlabeled DNA sequences is that of the hsg operator (80 bp) and the other corresponds to the same fragment of DNA with the hsg operator removed from it (51 bp).

Figure 4.1A shows the results of the wild-type a1 hd/α2 competition. The lane marked 0 contains no competitor DNA and represents the starting amount of a1/α2 activity bound to the labeled hsg operator. Increasing amounts of either the unlabeled specific (hsg - lanes to the left) or the nonspecific (no hsg - lanes to the right) competitor DNA were added to the reactions before addition of the proteins. Figure 4.1B shows the results of the same experiment carried out with the armless a1 mutant. For both the mutant and wild-type proteins a 100-fold molar excess of specific competitor DNA over labeled DNA fragment shows detectable competition. In contrast, a 30,000-fold molar excess of nonspecific competitor DNA is required before an approximately equivalent level of competition with the labeled DNA fragment is detected (phosphorimager quantitative data not shown). Therefore, both the wild-type a1 hd/α2 complex and the armless a1 hd/α2 complex prefer the hsg operator over random DNA by a factor of 300. The heterodimer composed of both full-length a1 and α2 proteins displays similar specificity under these

Figure 4.1. Specificity of the a1 hd and a1 NH_2-terminal arm mutants complexed with α2. Unlabeled competitor DNA was incubated with 10nM a1 hd and 10 nM α2 C-terminal fragment (panel A), or with 15 nM a1 hd_{9-57} and 10 nM α2 C-terminal fragment (panel B), or with 10 nM a1 $hd_{K3A, K5A}$ and 10 nM α2 C-terminal fragment (panel C) in a 20 μl reaction. All a1 hd proteins were present in a bacterial extract and the α2 C-terminal fragment was purified. After 30 min on ice, approximately 10^{-10} M ^{32}P-labeled hsg operator was added to each reaction, followed by another 30 min incubation on ice. The concentration of competitor DNA used in each reaction is indicated as the fold excess over the labeled DNA fragment. The first five lanes in each panel received a competitor DNA containing the hsg operator itself, and the last 10 lanes received a competitor DNA containing no a1/α2 binding sites. The lane marked 0 in each panel contains no competitor DNA.

Figure 4.1


A
1000x
300x
100x
30x
10x
0
10x
30x
100x
300x
1000x
3000x
10,000x
30,000x
100,000x
300,000x
specific competitor
nonspecific competitor
B
1000x
300x
100x
30x
10x
0
10x
30x
100x
300x
1000x
3000x
10,000x
30,000x
100,000x
300,000x
specific competitor
nonspecific competitor
C
1000x
300x
100x
30x
10x
0
10x
30x
100x
300x
1000x
3000x
10,000x
30,000x
100,000x
300,000x
specific competitor
nonspecific competitor

conditions (data not shown), suggesting that all of the components that determine DNA binding specificity are present in the truncated homeodomain-containing fragments.

The DNA binding affinity of the armless a1 hd/α2 complex appears to be only slightly lower than that of the wild-type a1 hd/α2 complex (data not shown, but compare the lanes marked 0 in Fig.4.1, panels A and B). This 2-3-fold difference may be due to a small difference in activity between the two protein extracts, to reduced folding of the mutant, or to a slightly lower affinity of the mutant a1 for either α2 or the DNA.

We next tested the effects of muations in the NH_2-terminal arm of a1 *in vivo*. It seemed unlikely that the full-length a1 protein would fold properly with the 9 amino acids of the NH_2-terminal arm of the homeodomain removed from the middle of the protein. We therefore mutated two lysines in the NH_2-terminal arm of the a1 hd, in the context of the full-length protein, to alanines (K3A, K5A). The NH_2-terminal arms of all homeodomains are highly positively charged, and in the crystal structures that have been solved, the arginines and lysines which contain these positive charges have been shown to be responsible for making most of the base pair contacts in the minor groove of the DNA.

A high copy plasmid contanining either a wild-type copy of a1 or the a1 NH_2-terminal arm mutant, $a1_{K3A, K5A}$, was transformed into a MATα strain containing an integrated reporter gene with a single hsg operator inserted in the promoter of the reporter. The a1 mutant was then assayed for its ability to repress transcription of the reporter gene in conjunction with α2 (Fig. 4.2). $a1_{K3A, K5A}$ represses with α2 at least as well as, and in this case, better than, wild-type a1/α2, suggesting that these two lysines do not contact DNA in the a1/α2/DNA ternary complex. Western blot analysis shows that the two strains contained roughly equivalent amounts of a1 protein (data not shown), ruling out the possiblity that the mutant represses α2 equally well due to its overexpression.

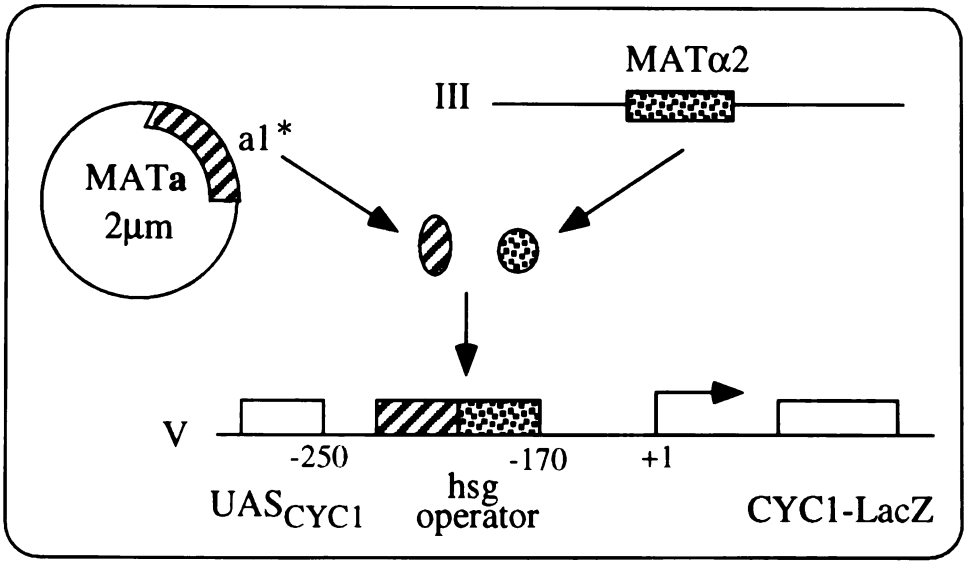
These two mutations also had no significant effect on the binding in vitro of the a1 hd and α2. When these same lysines are mutated to alanine in the context of the a1 hd (a1 $hd_{K3A, K5A}$), this mutant, like the armless a1 hd mutant and the wild-type a1 hd protein

complexed with α2, distinguished between the hsg operator and random DNA by a factor of 300 in a competition assay (Fig. 4.1C).

Figure 4.2. An a1 NH_2-terminal arm mutant can repress transcription of an hsg reporter gene with α2 in yeast.

The β–Galactosidase assay was used to quantitate the ability of wild-type a1 and the a1 NH_2-terminal arm mutant, $a1_{K3A, K5A}$, to repress a test promoter whose expression is controlled by a single hsg operator. A high copy plasmid containing either no a1, wild-type a1, or mutant a1, was transformed into a MATα cell containing the integrated reporter. Three independent transformants from each were selected and assayed in triplicate for β-Galactosidase enzyme activity.

Figure 4.2



	β-gal units
vector only	83
MAT a1	9.6
mat a1$_{K3A,K5A}$	6.4

Discussion

Gene regulatory proteins must bind with high affinity and specificity to the proper site on the DNA, while at the same time discriminating against closely related DNA sites. The structures of a number of homeodomain protein-DNA complexes have now been solved, providing a detailed molecular view of the interactions with the DNA bases and sugar-phosphate backbone that stabilize the correct protein-DNA complex (Gehring et al., 1994; Wolberger, 1996). In these complexes, the majority of interactions with the DNA bases are localized to the major groove and the phosphate backbone. However, interactions with bases in the minor groove have been suggested to play a large role in determining DNA-binding specificity.

The homeodomain-DNA cocrystal structures of engrailed (Kissinger et al., 1990), α2 (Wolberger et al., 1991), Oct-1 (Klemm et al., 1994), even-skipped (Hirsch and Aggarwal, 1995), and paired (Wilson et al., 1995), all show the flexible, NH_2-terminal arm of the homeodomain lying in the minor groove of the DNA and making specific base-pair contacts, as well as sugar-phosphate backbone contacts. The crystal structure of the a1/α2/DNA complex (Li et al., 1995) shows the NH_2-terminal arm of α2 in the minor groove of the DNA, as in the structure of α2 bound alone to the DNA (Wolberger et al., 1991), but the arm of a1 is unstructured in the ternary complex, and there is no electron density in the minor groove where one might have expected the NH_2-terminal arm. We have used biochemical and genetic techniques to address the lack of structure and DNA minor groove contacts of the NH_2-terminal arm of a1 in the a1/α2/DNA complex, and its importance to DNA binding affinity and specificity.

We have shown that complete removal of the NH_2-terminal arm of the a1 hd has no effect on the DNA binding specificity of the a1/α2 heterodimer, as determined in vitro, and only a slight effect on affinity. Furthermore, mutation of two out of the three charged residues in the a1 arm, which were predicted to make minor groove contacts based on

comparison with other homeodomain structures, has no affect on the ability of a1 to repress a reporter by an hsg operator construct in conjunction with α2. These results suggest that the NH_2-terminal arm of a1 is relatively unimportant for determining DNA binding specificity of the a1/α2 complex, and that in all likelihood, the a1 arm does not contact the minor groove of the DNA.

Based on the handful of homeodomain structures mentioned above, it was assumed that all homeodomain proteins would interact in similar ways with their target DNA sites. The lack of NH_2-terminal arm binding in the minor groove of the DNA by a1 may prove to be an exception to the rule, or it may be the case that other homeodomain proteins fall into this class. With the exception of the paired homodimer bound to DNA, all other structural data come from monomeric homeodomain proteins bound to DNA. It is now thought that most homeodomain proteins must interact with cofactors in order to attain the specificity necessary for developmental control. It is possible that in some of these more biologically relevant complexes the NH_2-terminal arms of the homeodomains may not be required for tight and specific binding. We suggest that although the helical core of different homeodomain proteins folds and interacts with DNA in a similar manner, there may be major variations in the positioning of the flexible NH_2-terminal arm.

The NH_2-terminal arm does appear to play a role in determining specificity for some fly homeodomain complexes, as suggested by genetic data. Functional studies of the Drosophila Antennapedia (Antp) and Sex combs reduced (Scr) homeodomain proteins suggest that residues in the NH_2-terminal arm of the homeodomains contribute to DNA binding specificity (Furukubo-Tokunaga et al., 1993; Zeng et al., 1993). The homeodomains of these two proteins differ at only five amino acids, four of which lie in the NH_2-terminal arm. Mutation of four of these specific residues in the arm of Scr to those of the Antp protein, changes the specificity of the Scr protein to that of the Antp protein in vivo. The mammalian homeodomain protein that is most closely related to the yeast a1 and α2 proteins, Pbx1, has been shown to modulate the DNA binding specificity

of a large subset of Hox proteins (Mann and Chan, 1996). Structural models suggest that the specificity of the Hox protein is altered by a conformational change involving residues in the NH_2-terminal arm of the Hox homeodomain, and mutational analysis supports the hypothesis that unique sequences in the Hox arm are at least partially responsible for mediating this specificity (Chan and Mann, 1996; Chan et al., 1996; Lu and Kamps, 1997; Phelan and Featherstone, 1997; Phelan et al., 1994).

Many different components appear to play roles in determining the biological specificity of homeodomain proteins. In the case of a1 and α2, it is well established that these proteins interact to bind cooperatively to DNA and that neither protein alone contains the necessary determinants for tight and specific DNA binding (Goutte and Johnson, 1993). α2 can also interact with another gene regulatory protein, MCM1, to bind a distinct DNA sequence (Keleher et al., 1988); two arginines in the NH_2-terminal arm of α2 have been suggested to play a crucial role in determining specificity for this heterotetramer (Vershon et al., 1995). Although mutation of these residues dramatically reduces α2/MCM1 affinity for their DNA site, the mutations have little effect on the ability of α2 to repress with a1. These results suggest that even though the arm of α2 contacts bases in the minor groove of the DNA, these contacts are not important for the specificity of the a1/α2 complex. Mutation of three residues in the recognition helix of α2 that contact bases in the major groove of the DNA display the same phenotype -- repression with MCM1 is greatly reduced but repression with a1 is only slightly affected (Vershon et al., 1995).

These results have been interpreted to mean that most of the DNA binding specificity of the a1/α2 heterodimer arises from the a1 protein. It is therefore even more surprising that the NH_2-terminal arm of the a1 homeodomain does not appear to contribute to specificity, suggesting that base pair contacts made by the a1 recognition helix and contacts with the DNA backbone must be supplying a large part of the specificity to the complex. This network of contacts can only be achieved once a1 interacts with α2, presumably through a conformational change in a1, which then allows it to bind to the

DNA. The conformational changes that are induced in a1 and α2 upon interaction with each other must act to greatly increase the overall binding energy of the complex, and at the same time increase the specificity. Although a1 and α2 can interact in solution, they do not bind tightly to the DNA unless the binding sites for the two proteins are appropriately spaced and the sequence is inherently bendable, both in the a1 binding site, and in the "spacer" between the a1 and α2 binding sites (Jin et al., 1995). The DNA must bend 60^0 in order for the interacting regions of the proteins to be in register, and if the two half-sites are too far apart the heterodimer cannot span them (Jin et al., 1995; Li et al., 1995). Thus, the bendability and configuration of the DNA also plays an important role in conferring specificity.

An understanding of how the a1/α2 heterodimer binds to its target sequences should aid in our understanding of how specificity is achieved by the multitude of known homeodomains from other organisms. Many of these homeodomain proteins are comprised of almost identical amino acid sequences, yet they regulate diverse developmental pathways. The available data suggests that specificity may arise from a combination of sources, including the NH_2-terminal arm of the homeodomain, the recognition helix of the homeodomain, the configuration of the DNA, the interaction with cofactors, and conformational changes in the homeodomain and/or the DNA.

Methods

Plasmids

The a1 NH_2-terminal arm deletion was generated by PCR, using a 5' primer that contained an NdeI site immediately upstream of codon 9 of the a1 hd (MRS96 5'-GGGAAAATCACATATGTCACCCCAAG), and a 3' primer that bound to the complementary strand of the vector, immediately downstream of the 3' end of the a1 hd (o11918 5'-GAAACTAAAAGAAAATCTAGACTATGC) . Wild-type a1 hd in the bacterial expression vector, pHB40P (pMS83), was used as a template for the PCR reaction. The PCR fragments were cut with NdeI and XbaI, ligated into these sites in the pHB40P vector (pMS71), and confirmed by DNA sequencing. The a1 NH_2-terminal arm double mutation, K3A and K5A (pMS68), was generated in the a1 hd by PCR using the same template and 3' primer as above, and a 5' primer containing the double mutation (MRS91 5'-GAAATATACCACCATATGAAAAAAGAGAAGAGCCCAGCGGGAGCATCATCAATATCACCC). The same mutations were generated in the full-length a1 gene using the Quickchange Mutagenesis Kit (Stratagene). The wild-type a1 gene contained within a 4 Kb HindII fragment (Astell et al., 1981) in pGEM (Promega), was used as the template (pMS11), and the mutagenic primers were: 5' - GAGAAGAGCCCAGCGGGAGCATCATCAATATCACCCCAAGC - 3', and its complement. Positive clones were confirmed by sequence analysis (pMS149). pMS159 was constructed for expressing the full-length double a1 arm mutant in yeast. The yeast 2μm plasmid, CV13 (Broach et al., 1979), was cut with HindIII and the HindII fragment from pMS149 was ligated into it. The corresponding wild-type a1 yeast expression plasmid is pAJ83.

Bacterial extracts

Bacterial extracts were made from the protease deficient *E. coli* strain CAG597D overexpressing plasmids pMS83, pMS71, and pMS68. Cells were grown overnight to saturation at 30°C, harvested, resuspended in 5 ml/g lysis buffer (100mM Tris-HCl, pH 8.0, 1 mM EDTA, 10 mM β–mercaptoethanol, 500 mM NaCl, and 0,1 mM AEBSF), sonicated to lyse the cells, and centrifuged at 30,000 g for 40 min. The supernatant was used in the gel shift assays after quantitating the amount of a1 protein in each extract by SDS gel followed by Coomassie staining.

The C-terminal fragment of α2 ($\alpha2_{128-210}$), a gift of Andrew Vershon, was overexpressed and purified from *E. coli* cells containing the plasmid pAV105 (Vershon and Johnson, 1993).

Gel shift competition

Competition experiments were performed as described previously (Goutte and Johnson, 1993), except that 3 fm - 300 pm of unlabeled competitor DNA was added to each reaction mix before 1 fm of the ^{32}P-labeled DNA was added. All incubations were carried out on ice. The labelled DNA was the hsg operator, contained in an 80 bp fragment, cut out of pCG25 with HindII and EcoRI (Goutte and Johnson, 1988). The same DNA fragment, unlabeled, was used as the specific competitor in the assay. The nonspecific competitor was the same HindIII-EcoRI DNA fragment with the hsg operator removed, resulting in a 51 bp fragment.

β–Galactosidase assays

β–Galactosidase assays were performed as described by (Miller, 1972), except that the yeast cells were permeabilized by snap-freezing in liquid N_2, and then thawing at 37°C for 1 min. To avoid a derepression of the CYC1 promoter due to glucose deprivation, glucose was added to 2% to each culture 1 hr before it was assayed.

REFERENCES

Astell, C. R., Ahlstrom-Jonasson, L., Smith, M., Tatchell, K., Nasmyth, K. A., and Hall, B. D. (1981). The sequence of the DNAs coding for the mating-type loci of Saccharomyces cerevisiae. Cell *27*, 15-23.

Broach, J. R., Strathern, J. N., and Hicks, J. B. (1979). Transformation in yeast: development of a hybrid cloning vector and isolation of the CAN1 gene. Gene *8*, 121-33.

Chan, S. K., and Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proc Natl Acad Sci U S A *93*, 5223-8.

Chan, S. K., Popperl, H., Krumlauf, R., and Mann, R. S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. Embo J *15*, 2476-87.

Desplan, C., Theis, J., and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. Cell *54*, 1081-90.

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. Embo J *13*, 3551-60.

Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. Proc Natl Acad Sci U S A *90*, 6360-4.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211-23.

Goutte, C., and Johnson, A. D. (1988). a1 Protein Alters the DNA Binding Specificity of α2 Repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1994). Recognition of a DNA Operator by a Dimer Composed of Two Different Homeodomain Proteins. EMBO *13*, 1434-1442.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and α2 Homeodomain Proteins Form a DNA-binding Activity with Properties Distinct from those of either Protein. J. Molec. Biol. *233*, 359-371.

Herskowitz, I. (1989). A regulatory hierarchy for cell specialization in yeast. Nature *342*, 749-757.

Hirsch, J. A., and Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. Embo J *14*, 6280-91.

Hoey, T., and Levine, M. (1988). Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature *332*, 858-61.

Jin, Y., Mead, J., Li, T., Wolberger, C., and Vershon, A. K. (1995). Altered DNA recognition and bending by insertions in the alpha 2 tail of the yeast a1/alpha 2 homeodomain heterodimer [see comments]. Science *270*, 290-3.

Johnson, A. D. (1992). A Combinatorial Regulatory Circuit in Budding Yeast. In Transcriptional Regulation, S. L. McKnight and K. R. Yamamoto, eds. (Plainview: Cold Spring Harbor Laboratory Press), pp. 975-1005.

Kalionis, B., and O'Farrell, P. H. (1993). A universal target sequence is bound in vitro by diverse homeodomains. Mech Dev *43*, 57-70.

Keleher, C. A., Goutte, C., and Johnson, A. D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-36.

Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579-90.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Laughon, A. (1991). DNA binding specificity of homeodomains. Biochemistry *30*, 11357-67.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA [see comments]. Science *270*, 262-9.

Lin, L., and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeodomains. Genes Dev *6*, 1071-1081.

Lu, Q., and Kamps, M. P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene *14*, 75-83.

Mann, R. S., and Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet *12*, 258-62.

Miller, J. H. (1972). Experiments in Molecular Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Phelan, M. L., and Featherstone, M. S. (1997). Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers. J Biol Chem *272*, 8635-43.

Phelan, M. L., Sadoul, R., and Featherstone, M. S. (1994). Functional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm. Mol Cell Biol *14*, 5066-75.

Philips, C. L., Vershon, A. K., Johnson, A. D., and Dahlquist, F. W. (1991). Secondary structure of the homeo domain of yeast α2 repressor determined by NMR spectroscopy. Genes Dev. *5*, 764-772.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W. d. (1989). The structure and function of the homeodomain. Biochim Biophys Acta *989*, 25-48.

Vershon, A. K., Jin, Y., and Johnson, A. D. (1995). A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. Genes Dev *9*, 182-92.

Vershon, A. K., and Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell *72*, 105-12.

Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. Cell *82*, 709-19.

Wolberger, C. (1996). Homeodomain interactions. Curr Opin Struct Biol *6*, 62-8.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-528.

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., and Scott, M. P. (1993). Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development *118*, 339-52.

CHAPTER 5

A Heterologous "Tail" Can Function as a Ligand for a1 When Attached to the α2 Homeodomain

Abstract

Current models suggest that homeodomain proteins regulate their target genes in a combinatorial fashion, with different heterodimers forming depending on the stage of development, place in the cell cycle, or mating type. Different heterodimers would therefore have different DNA-binding specificities and regulate different sets of genes. One prediction of this model is that the homeodomain interactions will be modular in the sense that the interacting regions (the "ligand" and "receptor" portions of the proteins) are conserved among the various homeodomains. Previous work examining the heterodimerization interface of two yeast homeodomain proteins, a1 and α2, has indicated that a short region C-terminal to the α2 homeodomain (the tail) is the "ligand" while a hydrophobic patch on the surface of the a1 homeodomain constitutes the "receptor." In this paper we test the modularity of this interaction by substituting two other tails, from the *Kluyveromyces lactis* α2 protein, and a region from the viral VP16 protein, in place of the *Saccharomyces cerevisiae* α2 tail. Our results indicate that these heterologous tails bind a1 in an analogous manner to the α2 tail and result in cooperative binding to specific DNA sequences. This provides strong evidence for the generality of the homeodomain ligand-receptor interaction.

Introduction

Homeodomain proteins, gene regulatory proteins found in all eukaryotes, contain four highly conserved residues, and a conserved secondary structure consisting of a helix-loop-helix-turn-helix motif (Burglin, 1994). Approximately 1,850 proteins currently meet these criteria and they occur in angiosperms, fungi, and metazoa (Bharathan et al., 1997). Homeodomain proteins bind to specific target sites in the DNA, and play a vital role in the regulation of a wide range of biological phenomena, including mating-type regulation in fungi and morphological development in many multicellular organisms. These proteins bind DNA with relatively low specificity, and it is now widely thought that many homeodomain proteins must interact with cofactors in order to achieve their high biological specificity.

The best-studied example of homeodomain heterodimeric proteins involves the a1 and α2 proteins from *Saccharomyces cerevisiae*. The DNA binding properties of the a1/α2 heterodimer are distinct from those of a1 and α2 alone (Goutte and Johnson, 1993). These two homeodomain proteins greatly increase their ability to recognize specific target sites by forming a heterodimeric complex (Dranginis, 1990; Goutte and Johnson, 1988). The α2 protein contacts the a1 hd through a region that lies immediately C-terminal to its hd, the α2 C-terminal tail (Mak and Johnson, 1993; Stark and Johnson, 1994). The previously unstructured α2 tail forms an α-helix upon interaction with the hydrophobic face of the a1 hd, which is composed of helices one and two, and with the loop connecting them (Li et al., 1995; Philips et al., 1994). Interaction with the tail of α2 also induces a conformational change in the a1 hd (Chapter 2). These reciprocal conformational changes, along with specific DNA base contacts made by residues in both homeodomain proteins and the conformability of the target DNA, are predicted to provide DNA binding specificity to the heterodimeric complex.

We have examined the specificity of the interactions between the residues in the α2 tail and those on the surface of the a1 hd to learn more about the specificity of the interaction between these proteins. To this end, we generated several chimeric proteins containing the α2 hd fused to an altered C-terminal tail. The altered tail was taken from an α2 homologue or from a region of a heterologous viral protein (VP16) that contains some sequence similiarity to α2. We show that both of these chimeric proteins can bind DNA cooperatively with a1 in vitro. Based on these results, we propose that the mechanism of interaction that has been described for a1 and α2 is conserved in other homeodomain-protein interactions.

Results

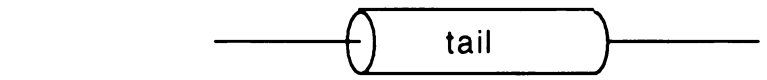
In an effort to understand the interaction between specific residues of the a1 hd and the C-terminal tail of α2, we generated several chimeric proteins containing the hd of α2 and the "tails" derived from two other proteins. As a starting point we put the C-terminal tail of the α2 protein from the related yeast, *Kluyveromyces lactis*, onto the *S. cerevisiae* α2 hd (α2 hd::*K.l.* tail). The *K. lactis* α2 tail is shorter than that of *S. cerevisiae* (Fig. 5.1), but the last nine residues of the *S. cerevisiae* α2 tail are known to be dispensable for its function. The hydrophobic residues that form the basis of the interaction between the *S. cerevisiae* α2 tail and the a1 hd are hydrophobic in the *K. lactis* α2 tail, but only one out of the four residues, Leu69, is identical.

We asked whether or not the α2 hd::*K.l.* tail protein was capable of interacting and binding DNA cooperatively with the *S. cerevisiae* a1 hd protein. Figure 5.2A shows that the chimeric protein is able to interact cooperatively with the a1 hd in an in vitro DNA binding assay. Lanes 2-5 show that the a1 hd does not bind detectably on its own at the concentrations used in this assay, but when the α2 hd::*K.l.* tail chimeric protein is added at high enough concentrations, a cooperative shift is seen. Approximately ten-fold more α2::*K.l.* tail/a1 hd protein than *S. cerevisiae* α2 hd/a1 hd protein is required to bind an equivalent amount of DNA (data not shown), indicating that the affinity of the interaction between the *K. lactis* α2 tail and the a1 hd is not as favorable as that between the two *S. cerevisiae* proteins.

To determine whether the *K. lactis* α2 tail was interacting with the a1 hd in a manner analogous to that of the *S. cerevisiae* α2 tail, we mutated one of the hydrophobic residues in the tail, isoleucine 65 -> serine, that we and others have previously shown to disrupt the interaction between a1 and α2 (Stark and Johnson, 1994; Strathern et al., 1988). This mutation in the chimeric protein, α2 hd::*K.l.* tail$_{I65S}$, also disrupts the interaction with the a1 hd (Fig. 5.2A). We conclude from these results that the C-terminal

Figure 5.1. Sequence allignment of the "tails" used in these experiments.
Amino acid sequence of the *S. cerevisiae* (*S.c.*) and *K. lactis* (*K.l.*) α2 C-terminal tails are numbered according to the conventional homeodomain numbering system. The VP16 sequence numbers correspond to residues 371-389 in the full-length VP16 protein (1-490). The residues shown in bold in the *S.c.* α2 tail are those that have been shown to interact with the surface of the a1 hd. The corresponding residues in the other two "tails" are also shown in bold for comparison.

Figure 5.1



S.c. α2: I T **I** A P E **L** A D **L L** S G E P L A K K K E stop

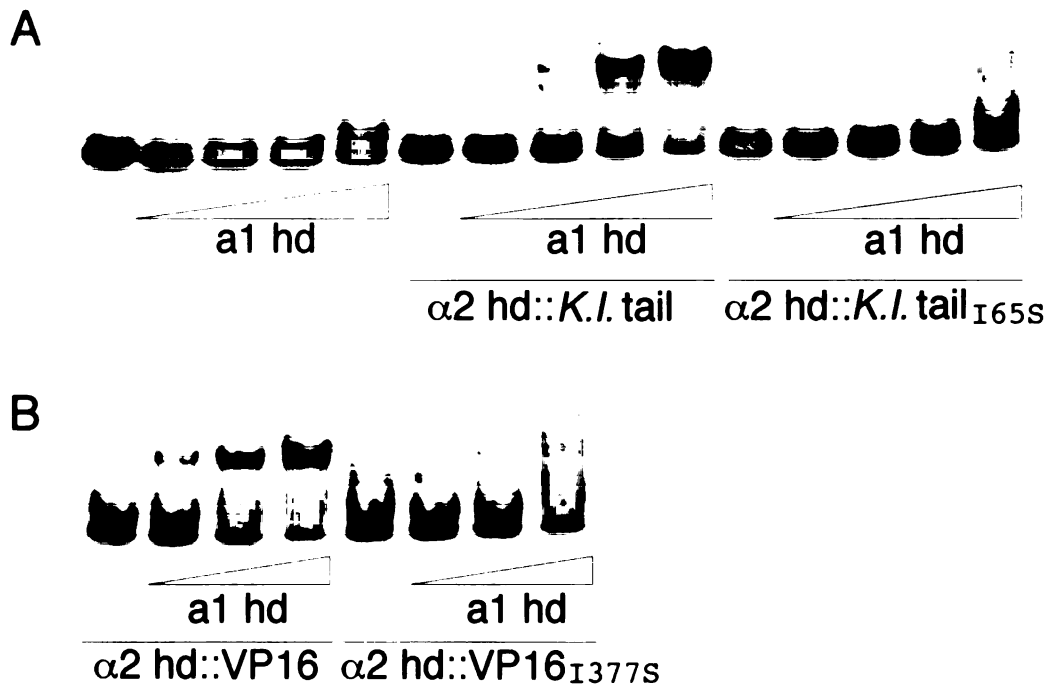
K.l. α2: T A **V** S S D **I** R N **I L** N stop

VP16: N N **Y** G S T **I** E G **L L** D L P D D D A P

Figure 5.2. The α2 chimeric proteins bind to an a1/α2 operator cooperatively with the a1 hd protein.

A ^{32}P-labelled DNA fragment containing an a1/α2 operator was incubated with the indicated proteins for 45 min on ice before being electrophoresed through a 5% native Tris-borate-EDTA polyacrylamide gel. The a1 hd protein was purified, and the α2 chimeric proteins were all present in bacterial extracts. (A) Lane 1 contains no protein. All lanes marked with the a1 hd contain three-fold increases of protein, beginning with 30nM and ending with 1μM. The α2 hd::*K.l.* tail and the α2 hd::*K.l.* tail$_{I65S}$ proteins are present in the lanes indicated at a concentration of approximately 2nM. (B) The lanes marked with the a1 hd protein contain three-fold increases of protein, beginning with 30nM and ending with 300nM. The α2 hd::VP16 and α2 hd::VP16$_{I377S}$ proteins are present in the lanes indicated at a concentration of approximately 80nM.

Figure 5.2


A
a1 hd
a1 hd
a1 hd
α2 hd::K.l. tail
α2 hd::K.l. tail I65S
B
a1 hd
a1 hd
α2 hd::VP16
α2 hd::VP16 I377S

tail of an α2 homologue from another yeast can be substituted for the *S. cerevisiae* α2 tail. The interaction with a1 is maintained even though the two tails are identical at only a single position.

To further test the generality of the a1 hd/ α2 tail interaction domain, we grafted a region of the heterologous herpes virus activator protein, VP16, that contains some sequence similarity to the α2 tail (Baxter et al., 1994; Li et al., 1995), onto the α2 hd, in place of α2's own tail (α2 hd::VP16). This region of VP16 is predicted to form an amphipathic helix and to interact with the exposed surface of the Oct-1 homeodomain protein (Hayes and O'Hare, 1993; Lai and Herr, 1997). In the VP16 "tail" there are four hydrophobic residues that correspond to the hydrophobic residues in the α2 tail that form the basis of the interaction with the a1 hd. Two out of four of these residues are identical in the VP16 "tail", one is conserved, and one is dissimilar (Fig. 5.1).

The gel shift in Figure 5.2B shows that this α2 hd::VP16 chimera is also capable of interacting with the a1 hd and cooperatively binding to the DNA, but the binding is approximately 50-fold weaker than with the wild-type α2 protein. The residue at position 61 in the α2 tail, isoleucine, is different from the corresponding residue in VP16, tyrosine 372, and this may account for the greater reduction in affinity seen with this chimera compared to the α2 hd::*K.l.* tail chimera. Although the α2 hd ::VP16 chimera can bind DNA cooperatively with the a1 hd in vitro, when the VP16 "tail" is fused to an α2 protein containing all of its N-terminal sequence (not just the hd), this interaction is apparently not strong enough to generate detectable repression with full-length a1 of a test promoter regulated by an a1/α2 operator (data not shown).

When the residue in VP16 that corresponds to Leu65 in the α2 tail is mutated to serine (I377S), the interaction of the mutant chimeric protein with the a1 hd is reduced by more than a factor of 10 (Fig. 5.2B), indicating that this residue plays a crucial role in the interaction with a1, and that the VP16 "tail" is interacting with the a1 hd in a manner similar to that of the α2 tail.

Discussion

The *Saccharomyces cerevisiae* a1 and α2 proteins constitute the first homeodomain heterodimer to be characterized, and it is now thought that many homeodomain proteins interact with other proteins in order to achieve the high DNA binding specificity necessary for tight and specific developmental control. Our work on a1 and α2 shows how the interaction between two homeodomain proteins can be mediated by a flexible tail that becomes ordered upon complex formation (Li et al., 1995; Philips et al., 1994; Stark and Johnson, 1994). The tail of α2 behaves like a ligand for a1, binding to the exposed surface of a1 and inducing a conformational change in a1 which renders a1 competent to bind DNA. The crystal structure provided the details of the a1/α2 interaction interface (Li et al., 1995). We have used this information to design chimeric α2 proteins with altered C-terminal tails, that are capable of forming cooperative heterodimers with the a1 hd.

Both "tails" used in this study, that of the *K. lactis* α2 protein and the short region of VP16 that interacts with Oct-1, contain some similarities with the *S. cerevisiae* α2 tail, and both are predicted to form amphipathic helices. The VP16 region contains two residues identical to those involved in the α2 tail interaction with a1, and the *K. lactis* tail contains only a single identical residue. However, the overall conservation of residue type across the entire tail region is much higher in the *K. lactis* tail than in VP16, and this difference may explain the noted difference in affinities for the a1 hd protein of the two chimeras. Nevertheless, this short region of VP16 is able to interact with the a1 hd when attached to the α2 hd, presumably in a manner similar to that of the α2 tail. a1 is induced to bind DNA in the presence of both chimeras, indicating that both tails induce the necessary conformational change in the a1 hd.

The *K. lactis* homologue of the a1 protein has not yet been identified, but we presume it exists based on the presence of other mating type genes in this yeast. Our data lead us to predict that the surface of the *K. lactis* a1 hd, encompassed by helices one and

two and the loop connecting them, will contain compensatory amino acid changes, relative to the *S. cerevisiae* a1 hd, that accomodate the changes in the *K. lactis* α2 tail residues.

The asymmetric nature of the a1/α2 interaction raises the possibility that a similar type of interaction could also occur between a homeodomain protein and a non-homeodomain protein containing a peptide similar to the tail of α2. The fact that the VP16 "tail" can interact with the a1 hd, suggests that this region of VP16 may interact with the Oct-1 hd in a similar manner. Mutations in this region of VP16 disrupt cooperative binding to the DNA with Oct-1, without affecting the ability of VP16 to interact with another factor, HCF (Lai and Herr, 1997; Shaw et al., 1995; Stern and Herr, 1991; Walker et al., 1994; Werstuck and Capone, 1989; Werstuck and Capone, 1989). Peptides corresponding to this region of VP16 can alter the DNA binding specificity of the Oct-1 hd, as does the intact VP16 protein (Stern and Herr, 1991). Similar peptides can inhibit formation of the Oct-1/VP16 complex, presumably by binding to the same surface of Oct-1 normally bound by the full-length VP16 protein (Haigh et al., 1990; Hayes and O'Hare, 1993; Wu et al., 1994).

The Oct-1/VP16 interaction domain has been mapped via mutagenesis studies to the exposed surface of the Oct-1 hd, which corresponds to the same region of the a1 hd contacted by α2 (Lai et al., 1992; Pomerantz et al., 1992). This interaction domain contains one important residue unique to the Oct-1 hd (residue 22) when compared to the Oct -2 hd. VP16 binds to the Oct-1 hd, which contains a glutamate at this position, but not the highly related Oct-2 hd, which contains an alanine at position 22 (Lai et al., 1992). Surprisingly, mutation of the corresponding arginine residue in the a1 hd to glutamate (R22E) did not increase the binding affinity of the α2 hd::VP16 chimera for the a1 hd (data not shown). However, despite the sequence identity between Oct-1 and Oct-2 , several of these residues have also been shown to be important for the interaction of Oct-1 with VP16 (Lai et al., 1992; Pomerantz et al., 1992).

The observations presented in this paper support the hypothesis that the principles underlying the a1/α2 interaction are conserved among other homeodomain proteins. There are now numerous examples of cooperative interactions involving homeodomain proteins from other organisms, some of which might be mediated by these same types of interactions. These include Exd/Ubx from flies, and the vertebrate homologues Pbx/Hox (Chan et al., 1994; Peifer and Wieschaus, 1990; van Dijk and Murre, 1994), and Mec-3/Unc-86 from worms (Xue et al., 1993). For Pbx/Exd and Mec-3 it has been shown that a region C-terminal to the homeodomain is necessary for interaction with the Hox/Ubx or Unc-86 homeodomains, respectively (Chang et al., 1995; Lu and Kamps, 1997; Xue et al., 1993). In addition, mutations in several residues in the Pbx hd that correspond to residues in the a1 hd that form the hydrophobic interaction interface with α2 severely affect the ability of Pbx to cooperatively interact with the Hox proteins (Peltenburg and Murre, 1997). The sequence similarity between a1, Pbx, and Exd (Rauskolb et al., 1993) also implies the existence of a common mechanism used by these proteins to recognize their homeodomain partners. Only a limited number of solvent exposed residues are available for interaction with other proteins on the homeodomain proper, so it is reasonable that the interaction domain that we have defined for a1 and α2 may have been conserved during the evolution and diversification of homeodomain proteins. We believe that details of the a1/α2 interaction should therefore serve as a general model.

If other homeodomain protein pairs prove to contain these interaction motifs, it may be possible to determine amino acid preferences at each of the important positions, and, in effect, come up with an amino acid recognition code for homeodomain protein-protein interactions. This type of analysis has been described for coiled-coil interactions (Hu et al., 1990). A code of this type might allow one to make predictions about as yet undiscovered protein partners strictly from amino acid sequence information. As the majority of homeodomain proteins have no known partners, and yet are predicted to need to interact with other factors in order to achieve the necessary specificity for developmental control,

Methods

Expression constructs and bacterial extracts

All α2 chimeric constructs were made by replacing the wild-type α2 tail with an oligo duplex consisting of either the *K. lactis* α2 tail (nucleotides corresponding to residues 212-223, or 59-70 using the homeodomain numbering scheme, MRS124 5'-TCGACGAAAAGAAAAAACAACTGCCGTTTCGTCAGATATAAGAAACATTCTTAATTAAG, MRS125 5'-GATCCTTAATTAAGAATGTTTCTTATATCTGACGAAACGGCAGTTGTTTTTTCTTTTCG), a mutant *K. lactic* α2 tail, I65S (MRS126 5'-TCGACGAAAAGAAAAAACAACTGCCGTTTCGTCAGATTCGTCAGATTCTAGAAACATTCTTAATTAAG, MRS127 5'-GATCCTTAATTAAGAATGTTTCTAGAATCTGACGAAACGGCAGTTGTTTTTTCTTTTCG), the VP16 "tail" (nucleotides corresponding to residues 371-389, MRS100 5'-TCGACGAAAAGAAAAAACAAACAATTACGGGTCTACCATCGAGGGCCTGCTCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCTAAG, MRS101 5'-GATCCTTAGGGGGCGTCGTCGTCCGGGAGATCGAGCAGGCCCTCGAGCAGGCCCTCGATGGTAGACCCGTAATTGTTTGTTTTTTCTTTTCG), or a mutant VP6 "tail", I377S (MRS118 5'-TCGACGAAAAGAAAAAACAAACAATTACGGGTCTACCTCTGAGGGCCTGCTCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCTAAG, MRS119 5'-GATCCTTAGGGGGCGTCGTCGTCCGGGAGATCGAGCAGGCCCTCGAGCAGGCCCTCAGAGGTAGACCCGTAATTGTTTGTTTTTTCTTTTCG). The tail oligo duplexes contain a SalI overhang on the 5' end and a BamHI overhang on the 3' end, with the first 19 nucleotides corresponding to α2 sequence upstream of the tail.

A SalI site was introduced into α2 19 bp upstream of the tail by site directed mutagenesis (pMS20). The SalI site was removed from the yeast CEN ARS vector

pAV115 containing the MATα locus (pMS21). pMS21 was cut with BglII and BamHI to remove the majority of the α2 gene. This was replaced with the corresponding BglII-BamHI fragment from pMS20 that contains α2 with the SalI site upstream of the α2 tail (pMS22). The SalI-BamHI fragment containing the α2 tail was removed from pMS22 and replaced with either the wild-type or mutant *K. lactis* α2 tail or VP16 "tail" (pMS104, pMS105, pMS85, pMS103, respectively). The α2 hd versions of these chimeras were generated by PCR using pMS104,105, 85, and 103 as templates. The 5' primer for all of these PCR reactions introduces an NdeI site at the beginning of the α2 hd (o14039 5'-GATAAACAAACATATGAAACCTTACAGAG). The 3' primers contain a BamHI site and are specific for each tail: *K. lactis* α2 tail - MRS128 5'-GCCGGATCCTTAATTAAGAATGTTTC; VP16 "tail" - MRS106 5'-GCCGGATCCTTAGGGGGCGTC. The resulting PCR fragments were cut with NdeI and BamHI and cloned into these sites in the bacterial expression vector, pHB40P, under the control of the T7 promoter (Studier and Moffatt, 1986), resulting in pMS109 - α2 hd::*K.l.* tail, pMS110 - α2 hd::*K.l.* tail$_{I65S}$, pMS89 - α2 hd::VP16, pMS108 - α2 hd::VP16$_{I377S}$.

Bacterial extracts were made from the protease deficient *E. coli* strain CAG597D overexpressing plasmids pMS109, pMS110, pMS89 and pMS108. Cells were grown overnight to saturation at 30°C, harvested, resuspended in 7 ml/g lysis buffer (100mM Tris-HCl, pH 8.0, 1 mM EDTA, 10 mM β–mercaptoethanol, 500 mM NaCl, and 0.1 mM AEBSF), sonicated to lyse the cells, and centrifuged at 30,000 g for 40 min. The supernatant was used in the gel shift assays after quantitating the amount of α2 hd chimeric protein in each extract by SDS gel followed by Coomassie staining.

The a1 hd protein was purified from pMS3 overexpressed in BL21 (DE3) pLysS, as described previously (Philips et al., 1994).

Gel shift assay

α2 hd chimeric proteins were incubated with an 80 nucleotide pair ^{32}P-labelled DNA fragment containing an hsg operator (Goutte and Johnson, 1988), either in the presence or absence of purified a1 hd protein, for 45 min on ice in a buffer containing 5% glycerol, 10mM Tris-HCl, pH 8, 10 mg ml^{-1} BSA, 0.1 mM EDTA, 5 mM $MgCl_2$, 0.1% N-P40, 100 mM NaCl, and 3 μg ml^{-1} nonspecific DNA (*E. coli* genomic DNA digested with Hae III). Samples were electrophoresed through a 5% native Tris-borate-EDTA polyacrylamide gel for 50 min at 150 V.

REFERENCES

Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F., and Dahlquist, F. W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and a1: secondary structure determination of the a1 homeodomain and changes produced by alpha 2 interactions. Biochemistry *33*, 15309-20.

Bharathan, G., Janssen, B., Kellogg, E. A., and Sinha, N. (1997). Did Homeodomain Proteins Duplicate Before the Origin of Angiosperms, Fungi, and Metazoa? Proceedings of the National Academy of Sciences, USA *94*, 13749-13753.

Burglin, T. (1994). . In Guidebook to the Homeobox Genes, D. Duboule, ed. (Oxford: Oxford University Press), pp. 27-71.

Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell *78*, 603-15.

Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., and Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes Dev *9*, 663-74.

Dranginis, A. M. (1990). Binding of yeast **a**1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. Nature *347*, 682-685.

Goutte, C., and Johnson, A. D. (1988). a1 Protein Alters the DNA Binding Specificity of α2 Repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and α2 Homeodomain Proteins Form a DNA-binding Activity with Properties Distinct from those of either Protein. J. Molec. Biol. *233*, 359-371.

Haigh, A., Greaves, R., and O'Hare, P. (1990). Interference with the assembly of a virus-host transcription complex by peptide competition. Nature *344*, 257-9.

Hayes, S., and O'Hare, P. (1993). Mapping of a major surface-exposed site in herpes simplex virus protein Vmw65 to a region of direct interaction in a transcription complex assembly. J Virol *67*, 852-62.

Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990). Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science *250*, 1400-3.

Lai, J. S., Cleary, M. A., and Herr, W. (1992). A single amino acid exchange transfers VP16-induced positive control from the Oct-1 to the Oct-2 homeo domain [published erratum appears in Genes Dev 1992 Dec;6(12B):2663]. Genes Dev *6*, 2058-65.

Lai, J. S., and Herr, W. (1997). Interdigitated residues within a small region of VP16 interact with Oct-1, HCF, and DNA. Mol Cell Biol *17*, 3937-46.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA [see comments]. Science *270*, 262-9.

Lu, Q., and Kamps, M. P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene *14*, 75-83.

Mak, A., and Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein α2 is required for function with a second homeo domain protein. Genes Dev. *7*, 1862-1870.

Peifer, M., and Wieschaus, E. (1990). The segment polarity gene armadillo encodes a functionally modular protein that is the Drosophila homolog of human plakoglobin. Cell *63*, 1167-76.

Peltenburg, L. T., and Murre, C. (1997). Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. Development *124*, 1089-98.

Philips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994). Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2. Biochemistry *33*, 9294-9302.

Pomerantz, J. L., Kristie, T. M., and Sharp, P. A. (1992). Recognition of the surface of a homeo domain protein. Genes Dev *6*, 2047-57.

Rauskolb, C., Peifer, M., and Wieschaus, E. (1993). extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. Cell *74*, 1101-12.

Shaw, P., Knez, J., and Capone, J. P. (1995). Amino acid substitutions in the herpes simplex virus transactivator VP16 uncouple direct protein-protein interaction and DNA binding from complex assembly and transactivation. J Biol Chem *270*, 29030-7.

Stark, M. R., and Johnson, A. D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. Nature *371*, 429-32.

Stern, S., and Herr, W. (1991). The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain. Genes Dev *5*, 2555-66.

Strathern, J., Shafer, B., Hicks, J., and McGill, C. (1988). a/α-Specific Repression by MATα2. Genetics *120*, 75-81.

Studier, F. W., and Moffatt, B. A. (1986). Use of Bacteriophage T7 RNA Polymerase to direct selective high-level expression of cloned genes. J. Molec. Biol. *189*, 113-130.

van Dijk, M. A., and Murre, C. (1994). Extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell *78*, 617-24.

Walker, S., Hayes, S., and O'Hare, P. (1994). Site-specific conformational alteration of the Oct-1 POU domain-DNA complex as the basis for differential recognition by Vmw65 (VP16). Cell *79*, 841-52.

Werstuck, G., and Capone, J. P. (1989). Identification of a domain of the herpes simplex virus trans-activator Vmw65 required for protein-DNA complex formation through the use of protein A fusion proteins. J Virol *63*, 5509-13.

Werstuck, G., and Capone, J. P. (1989). Mutational analysis of the herpes simplex virus trans-inducing factor Vmw65. Gene *75*, 213-24.

Wu, T. J., Monokian, G., Mark, D. F., and Wobbe, C. R. (1994). Transcriptional activation by herpes simplex virus type 1 VP16 in vitro and its inhibition by oligopeptides. Mol Cell Biol *14*, 3484-93.

Xue, D., Tu, Y., and Chalfie, M. (1993). Cooperative Interactions Between the *Caenorhabditis elegans* Homeoproteins UNC-86 and MEC-3. Science *261*, 1324-1328.

CONCLUSIONS

The work presented in this thesis reveals many of the molecular mechanisms of DNA-binding specificity determination of the yeast homeodomain proteins a1 and α2. The combination of structural techniques along with biochemical and genetic approaches has been a powerful means of elucidating the details of the protein-protein and protein-DNA interactions involved in ternary complex formation.

The main question regarding the details of the a1/α2 interaction which remains unanswered by this work has to do with the proposed conformational change in the a1 homeodomain upon interaction with the tail of α2. The experiments in Chapter 2 indicate that the tail of α2 is sufficient to induce a1 to bind to the DNA as a monomer. Since the α2 tail does not contact the DNA, the simplest explanation of this result is that the tail of α2 induces a conformational change in the a1 homeodomain that stabilizes the DNA-binding form of the protein. NMR data reveal small changes in a1 homeodomain helices one and two and in the loop connecting them upon addition of α2 (Appendix A). In addition, I have isolated several mutations in the homeodomain of a1 that allow a1 to bind to DNA in the absence of α2 (Chapter 3). Some of these mutations are in positions that face away from the DNA, and thus are unlikely to affect the DNA-binding ability of a1 through direct DNA contacts. We have suggested that these mutations are in residues involved in the a1 conformational change.

The determination of the three dimensional structure of the a1 homeodomain alone, either by NMR or Xray crystallography, is necessary to definitively conclude that a1 undergoes a conformational change upon interaction with α2. The details of the change(s) may shed some light on the mutations in the a1 homeodomain which allow it to bind DNA on its own. If these mutations do indeed lie in regions affected by the conformational change, then structural studies on the mutants would also be informative.

The experiments of Chapter 4 challenge the homeodomain dogma that the NH_2-terminal arm of the homeodomain plays a key role in determining DNA-binding specificity. The observation that the NH_2-terminal arm of a1 does not contribute to the DNA-binding

specificity of the a1/α2 complex may be unique to this system, or may be true of some subset of homeodomain proteins, the specificity determinants of which have not yet been determined. Given the generality of the a1/α2 interaction (see below), it seems unlikely that this phenomenon singularly applies to a1/α2. It is possible that the discrepancy lies at least partly in the fact that DNA-binding specificity of homeodomains has been examined mostly in the context of the monomeric proteins. Now that several homeodomain cofactors have been identified, the role of the NH_2-terminal arms of these complexes in determining specificity should be re-examined.

The larger question raised by the work in this thesis pertains to the conservation of the mode of interaction between homeodomain proteins and their cofactors. I have shown in Chapter 5 that the region of the cofactor known to interact with the mammalian homeodomain-containing protein, Oct-1, can also interact with the a1 homeodomain when attached to α2, in place of α2's own tail. This rather surprising result suggests that Oct-1 and its cofactor, VP16, interact in a similar manner to a1 and α2, and raises the possibility that other homeodomain proteins also make use of this interaction interface. We believe that the details of the a1/α2 interaction will serve as a general model for homeodomain protein-protein interactions, and that if several more homeodomain partners are shown to use this mode of interaction it may become possible to predict other homeodomain protein partners based on sequence analysis.

APPENDIX A

Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in a2

Cynthia Phillips, Martha R. Stark, Alexander D. Johnson, and F. W. Dahlquist

(published in Biochemistry *33*: 9294-9302, 1994)

Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2†

Cynthia L. Phillips,‡,§ Martha R. Stark,‖ Alexander D. Johnson,‖ and F. W. Dahlquist*,‡

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, and Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Received March 16, 1994; Revised Manuscript Received May 18, 1994•

ABSTRACT: The homeodomain proteins a1 and α2 act cooperatively to regulate cell type specific genes in yeast. The basis of the cooperativity is a weak interaction between the two proteins which forms heterodimers that bind DNA tightly and specifically. In this paper, we examine the mechanism of heterodimerization. We show that two relatively small fragments of a1 and α2 are capable of heterodimerization and tight DNA binding. The α2 fragment contains the homeodomain followed by the natural 22 C-terminal amino acids of the protein; these 22 amino acids are unstructured in the α2 fragment. The a1 fragment contains only the homeodomain, indicating that the a1 homeodomain mediates both DNA binding and protein–protein interactions with α2. We used isotope-edited NMR spectroscopy to study the interaction in solution of these two fragments. Samples in which only the α2 fragment was uniformly labeled with ^{15}N allowed us to visualize changes in the NMR spectra of the α2 fragment produced by heterodimerization. We found that the a1 homeodomain perturbs the resonances of only the C-terminal tail of α2; moreover, contact with a1 converts a portion of this tail (residues 193–203) from its unstructured state to an α-helix, as determined by J coupling and NOE measurements. Thus the heterodimerization of two homeodomain proteins involves the specific interaction between a tail of one protein and the homeodomain of the other. This interaction is accompanied by the acquisition of secondary structure in the tail.

Transcriptional regulators often act in combination to increase their range of regulatory activity. Combinatorial control of mating-type gene expression in the yeast *Saccharomyces cerevisiae* involves the cell type specific transcriptional regulators a1 and α2 [for reviews, see Herskowitz (1989), Dolan and Fields (1991), Sprague (1990), and Johnson (1992)]. The DNA-binding specificity of α2 depends on which other transcriptional regulators are present in the cell. In both α and a/α diploid cells, an α2 homodimer acts in combination with the cell type nonspecific protein MCM1, to bind DNA target sequences upstream of a-specific genes, resulting in repression of these genes. The α2 protein has a second regulatory activity in the a/α cell, in which both α2 and a1 are present. In this case, α2 acts in combination with a1, forming a heterodimer that binds DNA target sequences upstream of haploid-specific genes, resulting in repression of these genes (see above reviews).

Structurally, a1 and α2 are related, as they both contain the homeodomain DNA-binding motif. This is a 61 amino acid segment found in many eukaryotic transcriptional regulators [for reviews, see Scott et al. (1989) and Qian et al. (1989)]. On the basis of crystallographic and NMR spectroscopic studies, the homeodomains from several transcription factors have been shown to adopt a common structure, despite great variation in their amino acid sequences (Qian et al., 1989; Kissinger et al., 1990; Phillips et al., 1991; Wolberger et al., 1991). A hydrophobic core is surrounded by three helices, one of which binds in the major groove of the target DNA. An N-terminal arm, unstructured in the free homeodomain, binds in the minor groove of target DNA.

Primarily on the basis of protease sensitivity (Sauer et al., 1988), deletion mapping (Hall & Johnson, 1987; Mak & Johnson, 1993), and an analysis of point mutations (Porter & Smith, 1986; Harashima et al., 1989; Strathern et al., 1988), α2 appears to have at least four regions, each associated with different functions. The homeodomain, located near the C-terminus, has the DNA-binding function. The region N-terminal to the homeodomain is believed to contain the homodimerization contact region, responsible for the α2 dimers that bind with MCM1 to a-specific gene target sites, as well as part of the a1 contact region (Goutte & Johnson, 1988, 1992; Harashima et al., 1989). The hinge that links the N-terminal domain to the homeodomain interacts with MCM1 (Vershon & Johnson, 1992). It is the tail of α2, C-terminal to the homeodomain, that is thought to contain the major a1 contact region required for stabilization of the a1/α2 complex with the haploid-specific gene target site (A. Mak and A. D. Johnson, submitted).

The a1 and α2 proteins heterodimerize in the absence of DNA (Mak & Johnson, 1993). The a1/α2 complex subsequently binds target DNA to coregulate haploid-specific genes (Goutte & Johnson, 1993; Dranginis, 1990). The work described in this paper provides a structural explanation for the a1/α2 interaction. To study the a1/α2 complex by NMR spectroscopy, we used the smallest available fragments of a1 and α2 that were capable of both heterodimerization and specific binding to haploid-specific gene operators. Intact a1 is 126 residues, and intact α2 is 210 residues. The α2 fragment ($\alpha 2_{128-210}$) contains the C-terminal 83 amino acid residues and has been previously well characterized both structurally and functionally (Sauer et al., 1988; Phillips et al., 1991; Wolberger et al., 1991). This C-terminal region includes the homeodomain (residues 128–189) and the adjacent 22 residues that comprise the C-terminal tail of the protein. NMR experiments on the α2 fragment free in solution have shown

† C.L.P. is a recipient of an NIH Training Grant in Molecular Biology. This work was supported by grants from NIH (GM37049) to A.D.J. and from the American Cancer Society (BE-74) to F.W.D.

‡ University of Oregon.

§ Present address: Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, 3181 SW Sam Jackson Park Rd., Portland, OR 97201.

‖ University of California, San Francisco.

• Abstract published in *Advance ACS Abstracts*, July 1, 1994.

that the tail is unstructured (Phillips et al., 1991). In the cocrystal with DNA, the tail of the $\alpha 2$ fragment cannot be seen, presumably because it is disordered (Wolberger et al., 1991). Furthermore, the tail of $\alpha 2$ is protease sensitive, while the adjacent homeodomain is relatively resistant to protease attack (Mak & Johnson, 1993). The a1 fragment ($a1_{66-126}$) used in these NMR experiments contains only the homeodomain. These two fragments, when mixed, are adequate for subsequent specific binding to haploid-specific gene operators (see Results). To date, no structural studies of the a1 protein have been reported.

In this paper we focus on the effects of heterodimerization on the $\alpha 2_{128-210}$ fragment. We have taken advantage of the previously established resonance assignments for the same $\alpha 2$ fragment (Phillips et al., 1991). We have also used ^{15}N-edited experiments to simplify otherwise exceedingly complex spectra. In this case, the $\alpha 2_{128-210}$ was uniformly ^{15}N-labeled, whereas the $a1_{66-126}$ remained unlabeled. The ^{15}N-labeled protein was studied with ^{15}N-edited NMR experiments designed to filter out most signals from the unlabeled a1. By comparing the spectra of $\alpha 2$ in the absence and presence of a1, we determined how the structure of $\alpha 2$ changed due to heterodimer formation. The binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ does not occur with pseudo-2-fold symmetry via complementary homeodomain contacts between a1 and $\alpha 2$. Rather, the homeodomain proper of $\alpha 2_{128-210}$ is unaffected by the binding of the a1 homeodomain, but dramatic structural changes are seen in the C-terminal tail of the $\alpha 2$ fragment. These changes in the tail are most probably a result of direct contact with the a1 homeodomain. These data also show that heterodimer formation induces an interfacial helix in the previously unstructured C-terminal tail of $\alpha 2_{128-210}$.

EXPERIMENTAL PROCEDURES

Plasmids. The $a1_{66-126}$ protein was purified from *Escherichia coli* cells containing the plasmid pCW/K66. This plasmid encodes residues 66–126 of a1, under the control of tandem P_{tac} promoters. The original $a1_{66-126}$ plasmid was under the control of the T7 promoter in the T7 expression system of Studier and Moffat (1986) and is described in the Electrophoretic Mobility Shift Assay section below. To achieve higher levels of protein expression, the a1 fragment was subsequently cloned into the *Nde*I and *Xba*I sites of pCWori+ (Muchmore et al., 1989; Gegner & Dahlquist, 1991) and subsequently transformed into *E. coli* TB1 cells. The construction of the resulting plasmid was confirmed by restriction digestion and DNA sequencing. The $\alpha 2_{128-210}$ plasmid was constructed as previously described (Phillips et al., 1991).

Protein Purification. The $a1_{66-126}$ protein was purified to >95% homogeneity from *E. coli* TB1 cells as follows. Each liter of LB media with 100 mg/L ampicillin was inoculated to an initial density of 1×10^7 cells/mL. These were grown at 37 °C to a density of 5×10^8 cells/mL, at which time the cells were induced to a concentration of 0.4 mM IPTG. The cells were harvested 6 h after induction and resuspended in 500 mM NaCl, 25 mM sodium phosphate, pH 6.5, 1 mM EDTA, 10 mM β-mercaptoethanol, 0.01% NaN_3, and 1 mM phenylmethanesulfonyl fluoride (PMSF) using 25 mL of lysis buffer per liter of harvested media. The cells were sonicated and spun at 20000*g* for 20 min. The supernatant was dialyzed overnight against 25 mM sodium phosphate, pH 6.5, 10 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.2 mM PMSF, and 0.01% NaN_3 and loaded on a column of CM Sepharose CL 6B (Sigma). The $a1_{66-126}$ protein was eluted with a gradient from 10 to 600 mM NaCl and found to be >95% pure as judged by SDS–PAGE. The $a1_{66-126}$ protein was concentrated to approximately 0.1 mM using an Amicon concentrator with a 3000 molecular weight cutoff filter and further concentrated to 1–5 mM using 3000 molecular weight cutoff concentrators from Filtron. The molecular mass of $a1_{66-126}$ is 7229 Da, and we used the calculated ϵ_{280} of 5700 cm^{-1} M^{-1} to determine the protein concentration. The yield of $a1_{66-126}$ was approximately 7 mg/L.

The $\alpha 2_{128-210}$ protein was purified to >95% homogeneity and concentrated as described previously (Phillips et al., 1991). Uniform ^{15}N labeling of the $\alpha 2_{128-210}$ was carried out as previously described (Phillips et al., 1991). Prior to use, both the $a1_{66-126}$ and $\alpha 2_{128-210}$ proteins were dialyzed against 25 mM potassium phosphate buffer, pH 4.5, with 100 mM KCl, 0.01% NaN_3, and D_2O added to 5% for the spectrometer lock.

Electrophoretic Mobility Shift Assay. DNA binding assays were performed using purified $a1_{66-126}$ and $\alpha 2_{128-210}$ proteins as previously described for full-length a1 and $\alpha 2$ (Goutte & Johnson, 1993). A DNA fragment coding for residues 66–126 of a1 was generated by polymerase chain reaction using oligonucleotides that introduced an *Nde*I site at the 5′ end and an *Xho*I site at the 3′ end of the fragment. pMSK66 was constructed by cloning the *Nde*I–*Xho*I fragment into pHB40P, a derivative of pET-3a, under the control of the T7 promoter (Studier & Moffat, 1986). The $a1_{66-126}$ protein was overexpressed and purified from BL21(DE3)-pLysS *E. coli* cells containing the plasmid pMSK66. Cells were grown to an OD_{600} of 0.5 and induced with a final concentration of 0.4 mM IPTG. The cells were harvested after 5 h, frozen in liquid nitrogen, and then resuspended in cold 100 mM Tris-HCl, pH 8.0, 1 mM EDTA, 10 mM β-mercaptoethanol, 500 mM NaCl, and 1 mM PMSF. The cells were sonicated on ice and spun at 30000*g* for 40 min. The supernatant was dialyzed overnight at 4 °C against 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.28 mM β-mercaptoethanol, and 100 mM NaCl, and then dialyzed for another 3 h against the same buffer but with 50 mM NaCl instead of 100 mM NaCl. The supernatant was cleared by spinning at 30000*g* for 20 min and then loaded onto an SP-C50 Sephadex column (Pharmacia) at 4 °C. A salt gradient from 0.5 to 1.0 M NaCl was used to elute the protein from the column. $a1_{66-126}$ came off the column between 325 and 425 mM NaCl and was >90% pure.

$\alpha 2_{128-210}$, a gift of Andrew Vershon, was overexpressed and purified from *E. coli* cells containing the plasmid pAV105 (Vershon & Johnson, 1992).

NMR Spectroscopy. All spectra were acquired at 25 °C on a General Electric Omega 500-MHz spectrometer operating at 11.9 T. The chemical shifts were set relative to an external proton reference of sodium 2,2-dimethyl-2-silapentane-5-sulfonate at 0.0 ppm and an external nitrogen standard of $^{15}NH_4Cl$ at 24.93 ppm relative to NH_3 (Levy & Lichter, 1979). The spectra were recorded with a spectral width of 6410 Hz in the ^1H dimensions and 3333 Hz in the ^{15}N dimensions (except for the 3D-HSMQC-NOESY, where it was 1300 Hz). The recycle times, including acquisition, were usually 1 s. In the ^{15}N experiments a delay of 4.8 instead of 5.4 ms was used as the nominal $(2J_{NH})^{-1}$ time period to reduce the loss of signal due to relaxation. Spectra were analyzed using FELIX software from Hare Research.

To observe line width and chemical shift changes during a titration of ^{15}N-labeled $\alpha 2_{128-210}$ with $a1_{66-126}$, single-bond ^1H–^{15}N correlation spectra were measured by heteronuclear single–multiple quantum coherence (HSMQC) (Zuiderweg

1990). Spectra were collected as described by McIntosh et al. (1990) with minor differences. The experiments were taken with 1024 complex data points in the t_2 domain and 128–256 complex increments in t_1, and they typically required 2–6 h to acquire. Spectra were processed with a 60° shifted sine-bell apodization and zero-filled to 2048 and 1024 data points in the ^1H and ^{15}N dimensions, respectively. Spectra of dilute protein where the H_2O signal was not well suppressed were processed with a baseline correction (Burg smoothing of the FID or a polynomial baseline correction) as well. ^{15}N-ω_2-edited two-dimensional COSY and NOESY spectra of the $a1_{66-126}/\alpha2_{128-210}$ complex were recorded and processed as described for T4 lysozyme by McIntosh et al. (1990). NOESY spectra were collected with 100- and 150-ms mixing times.

The 3D-HSMQC-NOESY experiment is a variation of the 3D-HMQC-NOESY experiment (Kay et al., 1989b; Zuiderweg & Fesik, 1989), in which an HSMQC ^1H–^{15}N pulse sequence is used to generate ^1H–^{15}N coherence. The mixing time was 150 ms. The spectral widths used were 6410 (^1H, ω_1), 1300 (^{15}N, ω_2), and 6410 Hz (^1H, ω_3). The 3D data matrix contained 256 (real) × 64 (real) × 512 (complex) points. Quadrature in the ω_1 and ω_2 dimensions was obtained using time-proportional phase incrementation (TPPI). Sixteen scans were collected for each increment. The recycle time was 750 ms not including the acquisition time. The spectrum was processed with a 60° shifted sine-bell apodization in ω_3 and ω_2 and a 70° shifted sine-bell apodization in ω_1. A polynomial baseline correction was applied in the direct dimension. The matrix was zero-filled to 1024 × 512 × 128 real points.

The $^3J_{HN-H\alpha}$ coupling constants of the $a1_{66-126}/\alpha2_{128-210}$ complex were measured as described previously (Phillips et al., 1991) using an HMQC-J experiment (Kay et al., 1989a; Forman-Kay et al., 1990; Kay & Bax, 1990). The measured $^3J_{HN-H\alpha}$ coupling constants have errors of up to approximately 0.5 Hz, based on the ratio $J_{measured}/J_{actual}$ calculated by Kay and Bax (1990) for a typical resonance with an 11-Hz line width in both the free and bound forms of $\alpha2_{128-210}$.

The semiselective and nonselective T_1's for the amide resonances of the complexed $\alpha2_{128-210}$ were measured using variations on the HSMQC experiment, using an inversion recovery sequence in which the final read pulse was replaced by the 2D-HSMQC sequence (Fraenkel et al., 1990; Valensin et al., 1982).

RESULTS

Cooperative DNA Binding by a1 and α2 Homeodomain Fragments. Full-length a1 and $\alpha2$ bind the haploid-specific gene (hsg) operator in a strongly cooperative fashion. The two proteins form a heterodimer in solution with a K_d of approximately 10^{-6} M (A. Mak and A. D. Johnson, submitted). This dimer then binds the hsg operator with a K_d of approximately 10^{-10} M with both homeodomains making contact with the DNA (Goutte & Johnson, 1993). The protein $\alpha2$, in the absence of a1, binds weakly to the hsg operator ($K_d \sim 10^{-6}$ M), while a1 alone shows no detectable specific binding under the same conditions (Goutte & Johnson, 1993).

We found that the cooperative binding of a1 and $\alpha2$ to the hsg operator could be reconstituted *in vitro* using short fragments of both proteins. For a1, the homeodomain alone (residues 66–126 of the full-length protein) was sufficient for this cooperative binding. For $\alpha2$, both the homeodomain and the 20 amino acid C-terminal tail (residues 128–210) were required. For example, deletion of the C-terminal tail destroyed the cooperative binding of the protein fragments to DNA [not shown; see also Mak and Johnson (1993)].

The experiment in Figure 1 illustrates the cooperative DNA binding by these minimal fragments. Lanes 13–18 and lanes

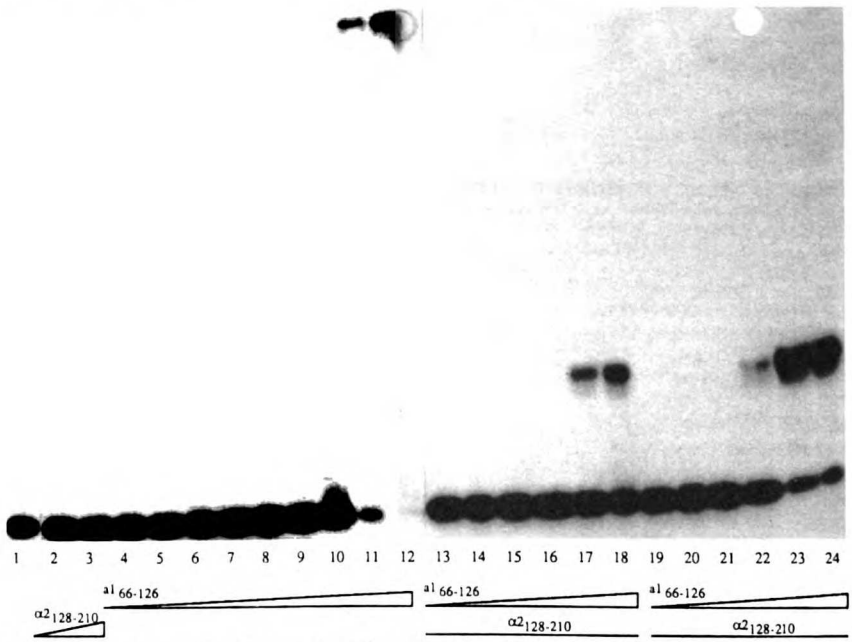


FIGURE 1: Cooperative binding of $a1_{66-126}$ and $\alpha2_{128-210}$ to the hsg operator. A 37-bp oligonucleotide duplex containing an hsg operator was used in the electrophoretic mobility shift experiment. Lane 1: labeled operator alone. Lanes 2 and 3: 3×10^{-9} M and 3×10^{-8} M $\alpha2_{128-210}$ protein, respectively. Lanes 4–12: successive 3-fold increases of $a1_{66-126}$ protein beginning with a concentration of 10^{-8} M in lane 4. Lanes 13–24 show titrations of $a1_{66-126}$ protein with a constant concentration of $\alpha2_{128-210}$ protein. Lanes 13–18: 3×10^{-9} M $\alpha2_{128-210}$ with successive 3-fold increases of $a1_{66-126}$ beginning with a concentration of 10^{-8} M. Lanes 19–24: 3×10^{-8} M $\alpha2_{128-210}$ with the same $a1_{66-126}$ titration as in previous lanes 13–18.

19–24 show the effect on DNA binding of increasing concentrations of the a1 homeodomain in a constant concentration of the α2 fragment. Both titrations show efficient operator binding at concentrations at which neither fragment alone binds to DNA (lanes 2–12). As can be seen in lanes 14–18, there is significant formation of a slower migrating form of the DNA as a result of increasing concentrations of $a1_{66-126}$ at a constant concentration of $\alpha2_{128-210}$ (3×10^{-9} M). At a 3-fold higher concentration of $\alpha2_{128-210}$, formation of the complex is observed at a corespondingly lower concentration of $a1_{66-126}$ (lanes 19–24). From these data we estimate the overall K_d for the reaction

$$a1_{66-126} + \alpha2_{128-210} + \text{hsg operator} \rightarrow a1/\alpha2/\text{operator}$$

to be approximately 10^{-14}–10^{-15} M[2]. This value is approximately 10–100-fold weaker than that observed for the cooperative DNA binding of full-length a1 and α2 [see Goutte and Johnson (1993)]. We attribute this difference to protein–protein interactions made between a1 and the amino terminus of α2, which has been deleted in these experiments (Goutte & Johnson, 1988).

We conclude from these results that the a1 homeodomain is sufficient to bind the hsg operator cooperatively with α2. We know that the a1 homeodomain contacts the operator in the $a1/\alpha2$/operator complex (Goutte & Johnson, 1993); these results suggest that the a1 homeodomain is also contacted by α2. For α2, both the homeodomain and the C-terminal 20 residue tail are required for cooperative DNA binding. Since this tail does not appear to contact DNA (Mak & Johnson, 1993), the simplest model is one where the tail makes direct contact with the a1 homeodomain. In order to test these ideas experimentally, we examined the behavior of the two homeodomain fragments in solution by NMR methods.

The a1 Homeodomain, $a1_{66-126}$, Forms a Heterodimer with $\alpha2_{128-210}$ in Solution. Figure 2 shows a region of the ^{15}N–^{1}H correlated spectrum (HSMQC) of 0.5 mM uniformly ^{15}N-labeled $\alpha2_{128-210}$ as it is titrated with unlabeled $a1_{66-126}$. Since only the $\alpha2_{128-210}$ fragment is ^{15}N labeled, we observe only the $\alpha2_{128-210}$ resonances and how they are affected by addition of $a1_{66-126}$. Over the course of the titration, some $\alpha2_{128-210}$ resonances shift to new positions and experience line broadening, while others do not. For example, resonances arising from Asn 178, Thr 159, and Ser 181 in the homeodomain remain in the same position. Other resonances, such as those arising from Asp 198 and Ser 201 (shaded in the figure), shift steadily across the spectrum from a "free" position to a "bound" position. On the basis of the chemical shift and line shape differences between the spectrum of the free protein and those observed in the presence of various amounts of $a1_{66-126}$, we can estimate the exchange lifetimes for the protein–protein association. Line shapes and resonance positions were simulated using well-known relationships for a two-site exchange process (Sandstrom, 1982), suggesting that $\alpha2_{128-210}$ interconverts between the free and the bound forms approximately 300 times per second.

From the resonances that do shift during the $a1_{66-126}$ titration, we can calculate a binding constant for $a1_{66-126}$ complexing with $\alpha2_{128-210}$. Figure 3 shows the chemical shift change of the representative resonance Asp 198 as a function of the concentration of $a1_{66-126}$ added to a constant concentration of $\alpha2_{128-210}$. The smooth line drawn through the data points is a theoretical curve based on a least-squares fit of the data to a single binding site with a dissociation constant (K_d) of 2×10^{-4} M and a total shift of 196 Hz (0.39 ppm) in a 1:1 complex. Similar least-squares analysis of other resonances

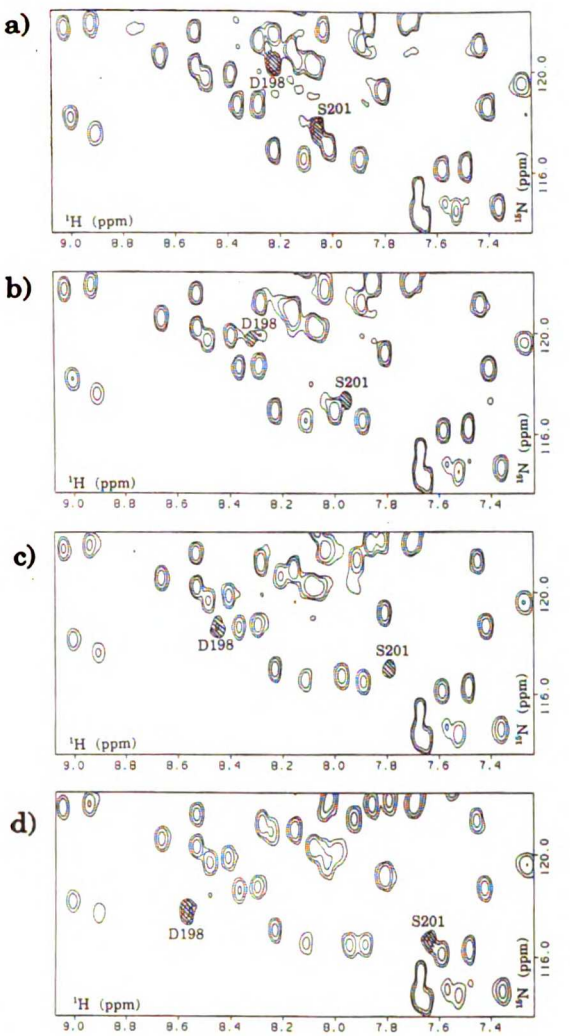


FIGURE 2: A small region of the HSMQC spectra during titration of $\alpha2_{128-210}$ by $a1_{66-126}$. The $\alpha2_{128-210}$ concentration was held constant at 0.5 mM. The resonance positions of Asp 198 and Ser 201 (shown shaded) are labeled as they shift during $a1_{66-126}$ addition. The spectra were taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. The spectra were apodized identically and are drawn at the same contour level. (a) $a1_{66-126}/\alpha2_{128-210} = 0$ (0 mM $a1_{66-126}$). (b) $a1_{66-126}/\alpha2_{128-210} = 0.25$ (0.125 mM $a1_{66-126}$). (c) $a1_{66-126}/\alpha2_{128-210} = 1.0$ (0.5 mM $a1_{66-126}$). (d) $a1_{66-126}/\alpha2_{128-210} = 4.0$ (2 mM $a1_{66-126}$).

observed to shift through the titration gave the same binding constant and 1:1 stoichiometry. Subsequent experiments on the structure of the bound form of $\alpha2_{128-210}$ were performed at an $a1_{66-126}$ concentration of 5 mM and an $\alpha2_{128-210}$ concentration of 3 mM, conditions where approximately 95% of the $\alpha2_{128-210}$ is bound to $a1_{66-126}$.

The spectral changes that occur when $a1_{66-126}$ is added to $\alpha2_{128-210}$ appear to be due to interaction between $a1_{66-126}$ and a particular region of $\alpha2_{128-210}$. Some of the resonances are perturbed, while most are not, indicating that only a small region of the protein is affected by $a1_{66-126}$ binding. Also, the total number of $\alpha2_{128-210}$ amide resonances is conserved through the titration, reflecting a fast exchange average of the free and bound forms of $\alpha2_{128-210}$. The moderate 0.2 mM

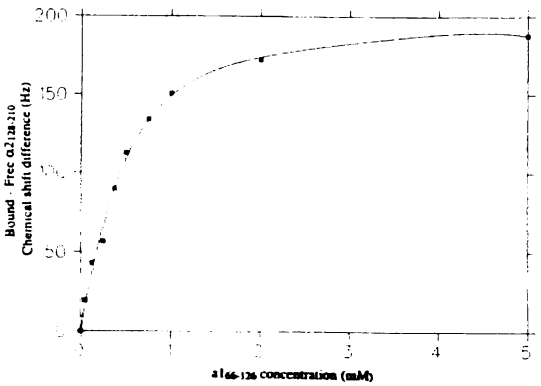


FIGURE 3: Chemical shift difference (Hz) between the bound and free forms of $\alpha 2_{128-210}$ for the representative resonance Asp 198, as a function of the concentration of added $a1_{66-126}$. The concentration of $\alpha 2_{128-210}$ was held constant at 0.5 mM. The smooth line drawn through the data points is a theoretical curve based on a least-squares fit of the data to a single binding site of $K_d = 2 \times 10^{-4}$ M and a total shift of 196 Hz (0.39 ppm).

K_d for heterodimer dissociation suggests that interactions distinct from simple electrostatics are important for heterodimer formation, especially since the two proteins are both highly positively charged. The $a1_{66-126}$ protein has a net charge of +13, and $\alpha 2_{128-210}$ has a net charge of +9, at neutral pH.

The spectra of $\alpha 2_{128-210}$ alone do not change over a concentration range of 0.2–6 mM (Phillips et al., 1991). Thus the changes we observe in the $\alpha 2_{128-210}$ spectra upon addition of $a1_{66-126}$ can only be attributed to the presence of $a1_{66-126}$ and are not a nonspecific effect of increasing total protein concentration. Finally, the HSMQC spectrum of 0.5 mM $\alpha 2_{128-210}$ with 5 mM $a1_{66-126}$ and the HSMQC spectrum of 3 mM $\alpha 2_{128-210}$ and 5 mM $a1_{66-126}$ are virtually indistinguishable in spite of the differing protein ratios (compare Figures 2d and 4). If a complex between $\alpha 2_{128-210}$ and $a1_{66-126}$ were formed with more than a single $a1_{66-126}$ per $\alpha 2_{128-210}$, this behavior would not be observed. This observation reinforces our conclusion above, based on least-squares fitting of the chemical shift data (see Figure 3), that $a1_{66-126}$ and $\alpha 2_{128-210}$ bind in a specific 1:1 complex.

Binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ Affects Only the C-Terminal Region of $\alpha 2_{128-210}$. With specific binding of the $a1_{66-126}$ and $\alpha 2_{128-210}$ fragments established, we turned to analyzing which $\alpha 2_{128-210}$ backbone resonances were affected by heterodimer formation. The 17-kDa $a1_{66-126}/\alpha 2_{128-210}$ complex is large for NMR characterization. Also, both $a1_{66-126}$ and $\alpha 2_{128-210}$ have limited chemical shift dispersion in their H^N and H^α protons, worsening the problem of spectral overlap (data for $a1_{66-126}$ not shown). Use of the added ^{15}N dimension in 2D-HSMQC spectra affords an advantage in dispersion over H^1-only spectra. In addition, the use of ^{15}N-edited experiments to observe only the uniformly ^{15}N-labeled $\alpha 2_{128-210}$ in the presence of unlabeled $a1_{66-126}$ further decreases the complexity of the $a1_{66-126}/\alpha 2_{128-210}$ heterodimer spectra. This ability to filter out signal from $a1_{66-126}$ in the $a1_{66-126}/\alpha 2_{128-210}$ complex was crucial to the assignment of the bound $\alpha 2_{128-210}$ backbone resonances.

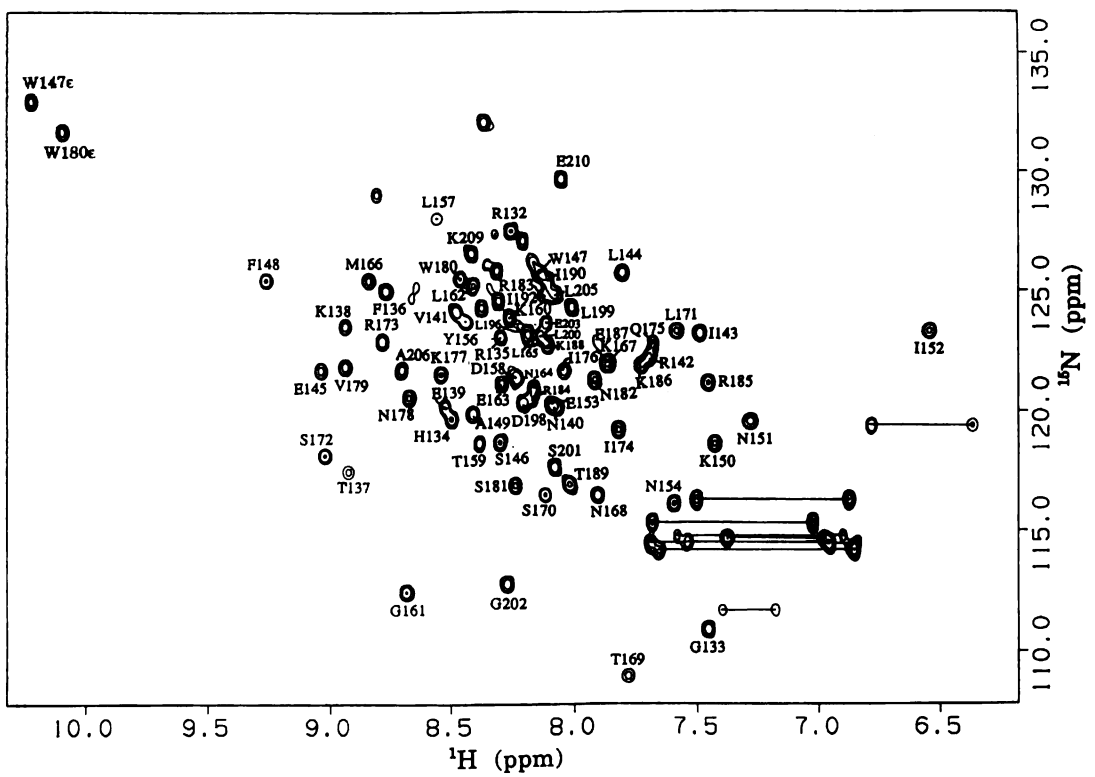


FIGURE 4: Amide region of the HSMQC of uniformly ^{15}N-labeled free $\alpha 2_{128-210}$ at 4 mM. Each cross peak arises from a proton directly bonded to a ^{15}N. In the lower right corner of the spectrum, each pair of side-chain $^{15}NH_2$'s is connected by a line. The assigned cross peaks are labeled. The spectrum was taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock.

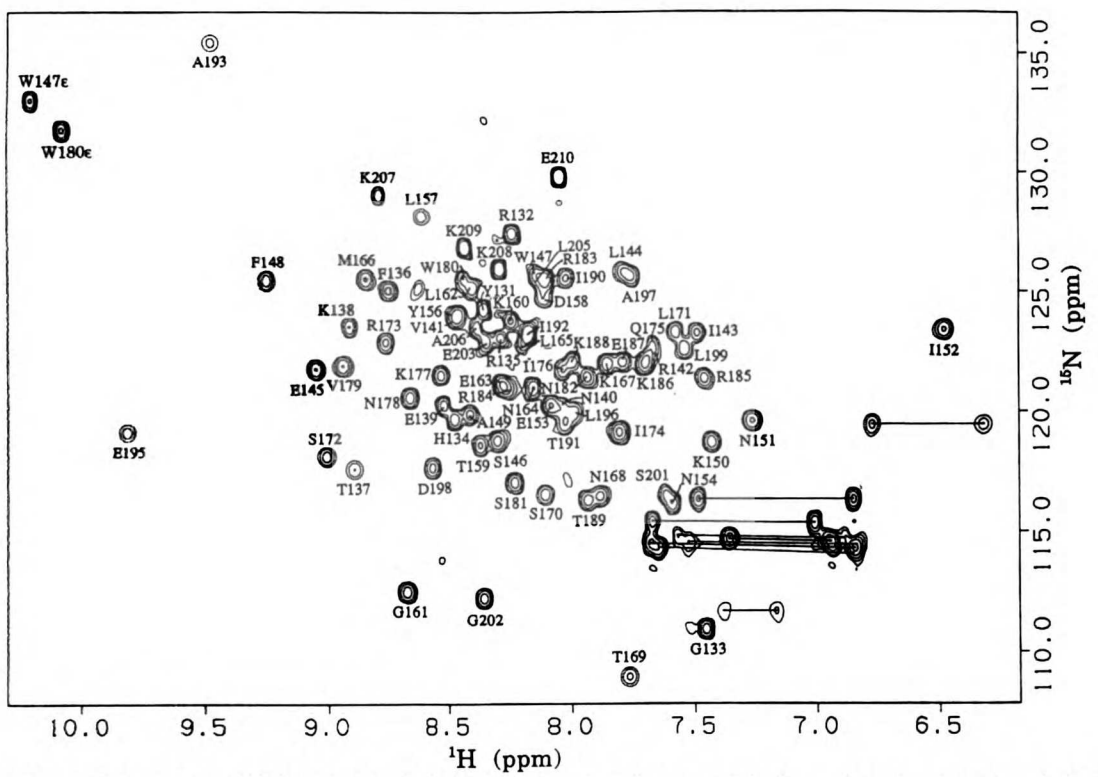


FIGURE 5: Amide region of the HSMQC of uniformly ^{15}N-labeled $\alpha 2_{128-210}$ bound to $a1_{66-126}$. The $\alpha 2_{128-210}$ is 3 mM, and the $a1_{66-126}$ is 5 mM. Greater than 95% of the $\alpha 2_{128-210}$ is in the bound form. In the lower right corner of the spectrum, each pair of side-chain $^{15}NH_2$'s is connected by a line. The assigned cross peaks are labeled. The assignments for K208 and K209 are ambiguous and may be reversed. The spectrum was taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. The resonances are broader in this figure than those in Figure 4 primarily because the contour levels shown are lower.

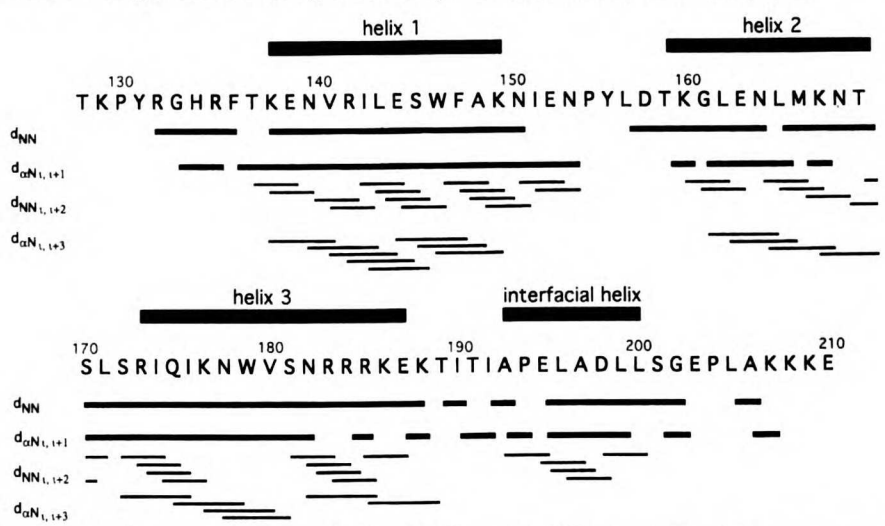


FIGURE 6: $\alpha 2_{128-210}$ amino acid sequence with sequential and intermediate-range NOE connectivities. Below the sequence, lines of varying widths indicate NOE connectivities. The uppermost two sets of broader lines indicate the d_{NN} and $d_{\alpha N}$ sequential connectivities. Below, the narrower sets of lines indicate the $d_{NN}(i, i+2)$ and the $d_{\alpha N}(i, i+3)$ connectivities. All NOE connectivities were observed in 3D-HSMQC-NOESY.

Figures 4 and 5show the full amide regions of the HSMQC spectra for the free and bound forms, respectively, of $\alpha 2_{128-210}$. The labeled resonances in Figure 5 were assigned by taking advantage of the previously determined assignments of free $\alpha 2_{128-210}$ (Figure 4). Due to spectral overlap, we were unable to simply track and reassign all of the shifting resonances in the bound form of $\alpha 2_{128-210}$ over the course of the titration. 3D-HSMQC-NOESY (not shown) of the $a1_{66-126}/\alpha 2_{128-210}$ complex, in which $\alpha 2_{128-210}$ (but not $a1_{66-126}$) was uniformly ^{15}N labeled, helped resolve resonances and allowed assignment

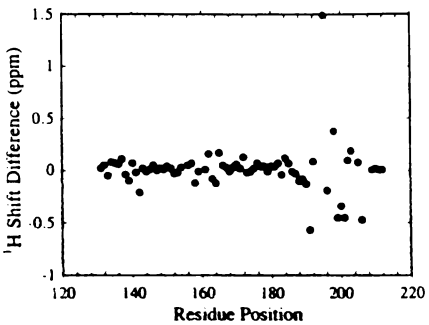


FIGURE 7: Amide ^{1}H chemical shift difference between bound and the free $\alpha 2_{128-210}$ as a function of residue position. Large chemical shift perturbations are seen in the C-terminal tail. Data is from the HSMQC spectra.

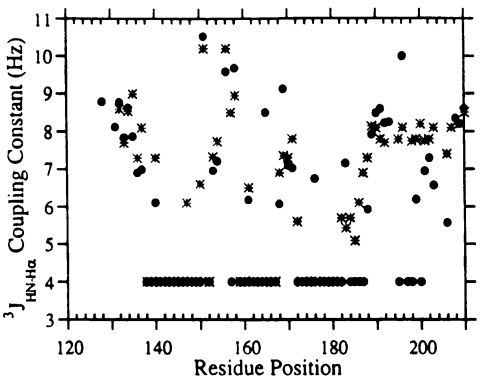


FIGURE 8: $^3J_{HN-H\alpha}$ coupling constants (Hz) as a function of residue position for free $\alpha 2_{128-210}$ (asterisks) and a 1_{66-126}-bound $\alpha 2_{128-210}$ (filled circles). No splittings could be measured for the $^3J_{HN-H\alpha}$ coupling constants shown as 4 Hz. These $^3J_{HN-H\alpha}$ couplings are smaller than the 5.5–6.0-Hz minimum for detection and were given the value of 4 Hz for presentation purposes. $^3J_{HN-H\alpha}$ coupling constants for overlapping or unassigned resonances are not shown.

of the spectrum. Sequential amide resonances were traced using H^N_i–H^N_{i+1} (d_{NN}) connectivities in the 3D-HSMQC-NOESY spectrum. Most of these d_{NN} sequential connectivities were corroborated with sequential H^N–H^α ($d_{\alpha N}$) connectivities of the (i, i+1) and (i, i+3) types. A ^{15}N-edited 2D-DQF-COSY experiment of the a $1_{66-126}/\alpha 2_{128-210}$ complex was used to identify the shifted H^N–H^α cross peaks in the 3D-HSMQC-NOESY. The backbone NOE connectivities are summarized in Figure 6. All of the amide and most of the H^α resonances (not shown) in the bound form of $\alpha 2_{128-210}$ were assigned by this procedure (Figure 5).

Remarkably, the amide resonances from residues 132–189 that comprise the homeodomain of $\alpha 2_{128-210}$ remain virtually unperturbed by the binding of a 1_{66-126}. In contrast, most of the resonances in the C-terminal tail region, following the homeodomain of $\alpha 2_{128-210}$, are affected by a 1_{66-126} binding. Figure 7is a plot of the change in $\alpha 2_{128-210}$ $^1H^N$ chemical shift as a function of residue position. The greatest chemical shift changes are seen for amide resonances belonging to residues 193–206. Residues 189–192, immediately following the C-terminal end of the homeodomain, show less dramatic perturbations. Amide chemical shift is extremely sensitive to environment and secondary structure. We conclude that the secondary structure of the $\alpha 2_{128-210}$ homeodomain proper is not affected by a 1_{66-126} binding. Absence of chemical shift change in the homeodomain argues strongly that the a 1_{66-126} homeodomain does not contact the $\alpha 2$ homeodomain. Conversely, a 1_{66-126} probably contacts the C-terminal tail of $\alpha 2_{128-210}$, causing the large chemical shift changes.

In the HSMQC spectrum of the free $\alpha 2_{128-210}$ (Figure 4), the C-terminal tail amide resonances are clustered in the center, where they are sharp and have chemical shifts consistent with a random coil structure. NOE connectivities and hydrogen exchange and $^3J_{HN-H\alpha}$ coupling constant data on the free $\alpha 2_{128-210}$ also indicate that this region is essentially unstructured (Phillips et al., 1991). In contrast, the HSMQC spectrum of the bound $\alpha 2_{128-210}$ (Figure 5) shows that the C-terminal tail amide resonances are considerably more disperse, and the resonances from amides 193 and 195 become the most extreme downfield shifted amide resonances in the spectrum. Such downfield chemical shifts in the proton dimension are very likely due to amide hydrogen bond formation.

In spite of a near doubling of molecular weight in the a $1_{66-126}/\alpha 2_{128-210}$ complex, the apparent line widths of the unperturbed resonances remain basically unchanged (see Figure 2, in which each spectrum in the titration is apodized identically and is drawn at the same contour level). The amide resonance line widths do not appear to be greatly influenced by the increase in correlation time upon complex formation. This may reflect relatively independent motion of the two monomers in the heterodimer. However, resonances arising from the contact region are somewhat broader than those arising from free $\alpha 2_{128-210}$. This broadening of the resonances in the contact region is probably due to chemical exchange with a 1_{66-126}.

Structure of the $\alpha 2_{128-210}$ C-Terminal Tail in the a $1_{66-126}/\alpha 2_{128-210}$ Complex. (A) $^3J_{HN-H\alpha}$ Coupling Constants. The scalar coupling constant, $^3J_{HN-H\alpha}$, is dependent on the Ramachandran dihedral angle ϕ of the protein backbone (Pardi et al., 1983; Wüthrich, 1986), and thus is sensitive to secondary structure. Figure 8 shows the $^3J_{HN-H\alpha}$ coupling constants for the free and bound forms of $\alpha 2_{128-210}$ as a function of residue number. These couplings were obtained using an HMQC-J experiment (Kay & Bax, 1990). Due to nonlinear effects of line width (10–11 Hz for most of the amide resonances) in this experiment, observed $^3J_{HN-H\alpha}$ coupling constants may be larger or smaller by about 0.5 Hz than the actual values (Kay & Bax, 1990).

A contiguous series of residues with $^3J_{HN-H\alpha}$ coupling constants of less than 6.5 Hz is strongly indicative of helical structure (Pardi et al., 1983; Wüthrich, 1986). No other regular secondary structure features such a series of low $^3J_{HN-H\alpha}$ coupling constants. The amides of the three helices of the homeodomain in both the free and bound $\alpha 2_{128-210}$ feature these low $^3J_{HN-H\alpha}$ coupling constants. This further confirms that the helical structure of the $\alpha 2_{128-210}$ homeodomain is unperturbed by a 1_{66-126} binding. Differences in the $^3J_{HN-H\alpha}$ coupling constants between the free and bound forms of $\alpha 2_{128-210}$ only become apparent in the tail region of $\alpha 2_{128-210}$. The $^3J_{HN-H\alpha}$ couplings were greater than 6.5 Hz for all of the C-terminal tail in the free $\alpha 2_{128-210}$. The $^3J_{HN-H\alpha}$ coupling constants of residues 193–203 of the bound form of $\alpha 2_{128-210}$ decrease to values consistent with a helical structure.

(B) Short-Range NOE Connectivities. In helices, the sequential d_{NN} connectivites are very strong, while in β-sheets and other extended structures, where sequential amide protons are not as close in space, the connectivities are weaker (Wüthrich, 1986). We wanted to compare the strength of the sequential d_{NN} NOEs in the tail region of bound $\alpha 2_{128-210}$ with the NOEs of the amide protons in helices in the

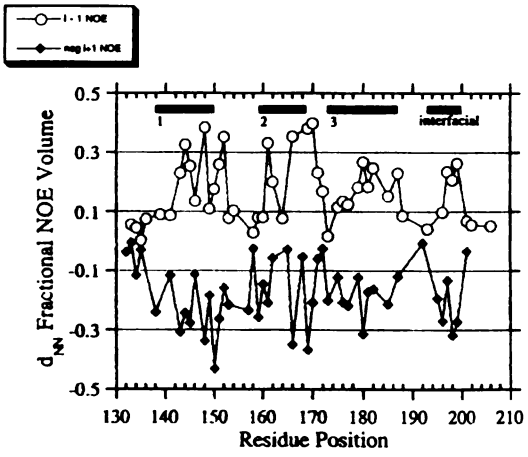


FIGURE 9: Ratio of the $d_{NN}(i, i+1)$ and $-(i, i-1)$ NOE volume to the diagonal peak volume as a function of residue position in $\alpha 2_{128-210}$ bound to $a1_{66-126}$. The $(i, i+1)$ fractional NOE volumes are shown as negative. The $(i, i+1)$ NOE values are shown as filled diamonds; the $(i, i-1)$ NOE values are shown as open circles. All data was from the 3D-HSMQC-NOESY. The positions of the helices are marked by filled bars at the top of the graph.

homeodomain that we observed to be unperturbed by $a1_{66-126}$ binding. Figure 9 shows the relative volumes of the sequential d_{NN} NOE cross peaks for the 3D-HSMQC-NOESY spectrum of bound $\alpha 2_{128-210}$. Each d_{NN} volume was divided by the volume of its own resolved "diagonal" peak for normalization. To ensure that we were measuring comparable NOEs at the NOESY mixing time of 150 ms, we measured the nonselective T_1's for the amides to be uniformly 900 ± 100 msec, and the semi-selective T_1's for the amides (also fairly uniform throughout the bound $\alpha 2_{128-210}$ form) at 250 ± 50 ms (see Experimental Procedures for a description of these experiments). Resonances 194–200 in the tail region of bound $\alpha 2_{128-210}$ have strong sequential d_{NN} NOEs, just as in the three homeodomain helices. The sequential d_{NN} NOEs of the N-terminal arm region and of the turns between the helices remain weak, as they are in free $\alpha 2_{128-210}$ (Phillips et al., 1991). Therefore, spin diffusion is not a complicating factor in this experiment on the larger complex, even at the 150-ms mixing time. These NOE cross peak volume data suggest that residues 194–200 are helical.

(C) Medium-Range NOE Connectivities. Very weak d_{NN} NOEs between an amide and its neighbor two amides away in the sequence are also commonly observed in helical structures [$(i, i+2)$ connectivities]. Figure 6 shows that these $d_{NN}(i, i+2)$ connectivities are present in part of the tail of $\alpha 2_{128-210}$ when it is bound to $a1_{66-126}$, as well as in the three helical regions of the homeodomain. The positions of these $d_{NN}(i, i+2)$ connectivities in the tail of bound $\alpha 2_{128-210}$ coincide with the position of the other helical types of NOE and the helical $^3J_{HN-H\alpha}$ coupling constants.

An NOE connectivity between an amide proton and the H^α three or four N-terminal to it in the sequence [$d^{\alpha N}(i, i+4)$ and $-(i, i+3)$] is also a marker for helical structure. These are observed in the three homeodomain helices [Figure 6; $(i, i+4)$ connectivities not shown]. Unfortunately, even with the increased dispersion of the 3D-HSMQC-NOESY, these connectivities are degenerate for most of the tail of bound $\alpha 2_{128-210}$. There are NOE cross peaks consistent with $d_{\alpha N}(i, i+3)$ and $-(i, i+4)$ connectivities, but they are degenerate with $(i, i+1)$ or other types of connectivities in this region of this experiment. We cannot be certain that the $d_{NN}(i, i+3)$ and $-(i, i+4)$ connectivities are present. However, the $^3J_{HN-H\alpha}$ coupling and the sequential d_{NN} and $d_{NN}(i, i+2)$ connectivities strongly suggest that $\alpha 2_{128-210}$ residues 194–206 of the C-terminal tail are helical in the presence of $a1_{66-126}$.

DISCUSSION

The $\alpha 2$ protein binds to two different classes of DNA target sites. In the case of a-specific gene target sites, $\alpha 2$ binds cooperatively with the cell type nonspecific MCM1 (Keleher et al., 1988; Passmore et al., 1989; Ammerer, 1990). To bind with high specificity to haploid-specific gene target sites, $\alpha 2$ protein heterodimerizes with the a1 protein. We have shown that, upon heterodimer formation with a1, a helix forms in the C-terminal tail of $\alpha 2_{128-210}$. At the $a1/\alpha 2$ interface, this helix may directly or indirectly create a new surface that is complementary to the DNA target site recognized by the heterodimer.

DNA-binding experiments and residue-specific chemical shift changes in $\alpha 2_{128-210}$ during titration with $a1_{66-126}$ show that the specific heterodimerization function is preserved in the a1 and $\alpha 2$ fragments studied in this work. Since the a1 fragment contains only the homeodomain, the homeodomain itself must serve as a protein interaction surface as well as a DNA-binding domain. We determined a moderate K_d of 2 × 10^{-4} M and a ratio of 1:1 for heterodimer dissociation. Since the a1 and $\alpha 2$ fragments retain this ability to specifically dimerize at a ratio of 1:1, and since the $a1_{66-126}/\alpha 2_{128-210}$ complex can bind target DNA with an affinity similar to that of intact $a1/\alpha 2$ (see Results), the $a1_{66-126}/\alpha 2_{128-210}$ complex retains the essential features of the intact $a1/\alpha 2$ complex. Given the supporting genetic evidence, the interactions we observe between the a1 and $\alpha 2$ fragments are also likely to exist between the intact a1 and $\alpha 2$.

Dividing our estimate of the overall dissociation constant of the ternary $a1_{66-126}/\alpha 2_{128-210}$ /operator complex of 10^{-14}–10^{-15} M^2 by the dissociation constant of the heterodimer of the a1 and $\alpha 2$ fragments of 2 × 10^{-4} M gives an estimate of the dissociation constant of the heterodimer and the hsg operator. This value of ~10^{-11} M agrees well with estimates of the affinity of the heterodimer formed from intact a1 and $\alpha 2$ (Goutte & Johnson, 1993) and suggests that most if not all of the energetically important interactions for specific DNA binding are retained in the heterodimer formed by the fragments.

Specific Binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ Induces an Interfacial Helix in $\alpha 2_{128-210}$. We assigned nearly all of the backbone ^{15}N, H^N, and H^α resonances of $\alpha 2_{128-210}$ bound to $a1_{66-126}$ and compared them with the resonances of uncomplexed $\alpha 2_{128-210}$. Only a few resonances are perturbed upon complex formation; these belong exclusively to the C-terminal tail of $\alpha 2_{128-210}$. This result suggests both that the C-terminal tail is the only region affected by $a1_{66-126}$ binding and that the homeodomain region of $\alpha 2_{128-210}$ is neither environmentally nor structurally perturbed by $a1_{66-126}$ binding.

Secondary structure determination of the bound $\alpha 2_{128-210}$ confirms this conclusion. We identified the sequential and intermediate-range NOE connectivities involving the H^N and H^α protons, roughly defining the secondary structure of the a1-bound form of $\alpha 2_{128-210}$ (Phillips et al., 1991). In addition, $^3J_{HN-H\alpha}$ coupling constants of less than 6.5 Hz helped to define the boundaries of the helices. The three helices of the homeodomain remain of the same length and position as in the free $\alpha 2_{128-210}$. Helix 1 spans residues 138–150, helix 2 spans residues 159–169, and helix 3 spans residues 173–187. The region C-terminal to the homeodomain was largely

unstructured in the free $\alpha 2_{128-210}$, whereas when bound to a1, this tail adopts a helical structure that extends from residue 193 to at least 200. This region shows comparable helical-type sequential d_{NN} NOE volumes and $d_{NN}(i, i+2)$ connectivities also seen in the homeodomain helices. Helical-type NOE connectivities are not observed beyond residue 200, but the $^3J_{HN-H\alpha}$ coupling constants are less than 6.5 Hz through residue 203, suggesting that the helix may extend this far. Absence of chemical shift change elsewhere in $\alpha 2_{128-210}$ during the titration with $a1_{66-126}$ argues strongly that direct contact with the a1 homeodomain induces formation of this fourth helix.

Recently, Spolar and Record (1994) have pointed out that there is often ordering of protein residues as a result of the specific binding of the protein to its DNA target. Our observations suggest that protein–protein interactions may also serve to initiate this ordering process. Thus heterodimer formation may "prepay" some of the entropy cost associated with the ordering of the proteins in their DNA complex.

ACKNOWLEDGMENT

We thank Susan Baxter, David M. Gontrum, Andrew K. Vershon, and Arkady Mak for communication of unpublished results and constructive discussions. We also thank Dennis Hare and Hare Research for supplying the FELIX software.

REFERENCES

Ammerer, G. (1990) *Genes Dev. 4*, 299–312.
Dolan, J. W., & Fields, S. (1991) *Biochim. Biophys. Acta 1088* (2), 155–169.
Dranginis, A. M. (1990) *Nature 347*, 682–685.
Forman-Kay, J. D., Gronenborn, A. M., Kay, L. E., Wingfield, P. T., & Clore, M. G. (1990) *Biochemistry 29*, 1566–1572.
Fraenkel, Y., Navon, G., Aronheim, A., & Gershoni, J. M. (1990) *Biochemistry 29*, 2617–2622.
Gegner, J. A., & Dahlquist, F. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 750–754.
Goutte, C., & Johnson, A. D. (1988) *Cell 52*, 875–882.
Goutte, C., & Johnson, A. D. (1993) *J. Mol. Biol. 233*, 359–371.
Hall, M. N., & Johnson, A. D. (1987) *Science 237*, 1007–1012.
Harashima, S., Miller, A. M., Tanaka, K., Kusumoto, K., Tanaka, K., Mukai, Y., Nasmyth, K., & Oshima, Y. (1989) *Mol. Cell. Biol. 9*, 4523–4530.
Herskowitz, I. (1989) *Nature 342*, 749–757.
Johnson, A. D. (1992) *Transcription Regulation* (McKnight, S. L., & Yamamoto, K. R., Eds.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.
Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 5488–5490.
Kay, L. E., Marion, D., & Bax, A. (1989b) *J. Magn. Reson. 84*, 72–84.
Kelcher, C. A., Goutte, C., & Johnson, A. D. (1988) *Cell 53*, 927–936.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.
Levy, G. C., & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance in Spectroscopy*, J. Wiley & Sons, New York.
Mak, A., & Johnson, A. D. (1993) *Genes Dev. 7*, 1862–1870.
McIntosh, L. P., Wand, A. J., Lowry, D. A., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341–6362.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44–73.
O'Neil, K. T., Shuman, J. D., Ampe, C., & Delgado, W. F. (1991) *Biochemistry 30*, 9030–9034.
Pardi, A., Wagner, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 445–454.
Passmore, S., Elble, R., & Tye, B. K. (1989) *Genes Dev. 3*, 921–935.
Phillips, C. L., Vershon, A. K., Johnson, A. D., & Dahlquist, F. W. (1991) *Genes Dev. 5*, 764–772.
Porter, S. D., & Smith, M. (1986) *Nature 320*, 766–768.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell 59*, 573–580.
Ransone, L. J., & Verma, I. M. (1990) *Annu. Rev. Cell Biol. 6*, 539–557.
Sandstrom, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, New York.
Sauer, R. T., Smith, D. L., & Johnson, A. D (1988) *Genes Dev. 2*, 807–816.
Scott, M. P., Tamkun, J. W., & Hartzell, G. W. (1989) *Biochim. Biophys. Acta 89*, 25–48.
Spolar, R. S., & Record, T. M., Jr. (1994) *Science 263*, 777–784.
Sprague, G. F., Jr. (1990) *Adv. Genet. 27*, 33–62.
Strathern, J., Shafer, B., Hicks, J., & McGill, C. (1988) *Genetics 120*, 75–81.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Valensin, G., Kushnir, T., & Navon, G. (1982) *J. Magn. Reson. 46*, 23–29.
Vershon, A. K., & Johnson, A. D. (1992) *Cell 72*, 105–112.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517–528.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Zuiderweg, E. R. P. (1990) *J. Magn. Reson. 86*, 346–357.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387–2391.

APPENDIX B

Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA

Thomas Li, Martha R. Stark, Alexander D. Johnson, and Cynthia Wolberger

(published in Science *270*: 262-269, 1995)

Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA

Thomas Li, Martha R. Stark, Alexander D. Johnson, Cynthia Wolberger*

The *Saccharomyces cerevisiae* MAT**a**1 and MATα2 homeodomain proteins, which play a role in determining yeast cell type, form a heterodimer that binds DNA and represses transcription in a cell type–specific manner. Whereas the α2 and **a**1 proteins on their own have only modest affinity for DNA, the **a**1/α2 heterodimer binds DNA with high specificity and affinity. The three-dimensional crystal structure of the **a**1/α2 homeodomain heterodimer bound to DNA was determined at a resolution of 2.5 Å. The **a**1 and α2 homeodomains bind in a head-to-tail orientation, with heterodimer contacts mediated by a 16-residue tail located carboxyl-terminal to the α2 homeodomain. This tail becomes ordered in the presence of **a**1, part of it forming a short amphipathic helix that packs against the **a**1 homeodomain between helices 1 and 2. A pronounced 60° bend is induced in the DNA, which makes possible protein-protein and protein-DNA contacts that could not take place in a straight DNA fragment. Complex formation mediated by flexible protein-recognition peptides attached to stably folded DNA binding domains may prove to be a general feature of the architecture of other classes of eukaryotic transcriptional regulators.

Homeodomain proteins constitute a superfamily of DNA binding proteins that play critical roles in gene regulation and development in many eukaryotic species. These proteins have in common a conserved 60–amino acid DNA binding domain that has been well characterized structurally, biochemically, and genetically (*1*, *2*), whereas other portions of the proteins are quite divergent. The fold adopted by the homeodomain, as uncovered in structural studies of *Drosophila* (*3*–*5*), yeast (*6*, *7*), and human (*8*) homeodomains, consists of three α helices and an NH_2-terminal arm. Homeodomains bind DNA by inserting the third of these three α helices into the major groove of the DNA, while the NH_2-terminal arm contacts bases in the adjacent minor groove. Biochemical studies have shown that isolated homeodomains, which typically bind DNA as monomers, often exhibit only a relatively modest degree of DNA sequence selectivity (*9*–*11*). In the case of the yeast α2 homeodomain protein, we know that the specificity and affinity with which α2 binds DNA is augmented by its association with either of two partner proteins: the product of the MATa locus, a1, or the non–cell type–specific protein MCM1. The a1/α2 and α2/MCM1 complexes each bind to a distinct set of DNA sites in the yeast genome, causing repression of the adjacent genes. Combinatorial control by a1 and α2 provides a means to achieve cell type–specific regulation of a large set of genes since a1 and α2 are expressed together in only the a/α diploid yeast cell type (*12*).

One of the partner proteins of α2, a1, is also a member of the homeodomain superfamily (*13*). In the diploid a/α cell type, these two homeodomain proteins form a heterodimer that binds to sites upstream of haploid-specific genes (hsg) (*14*–*16*). In the absence of α2, the a1 protein exhibits no detectable specific binding to DNA (*11*). However, the presence of a1 in solution dramatically raises the affinity of α2 for hsg operators. The cooperative binding of α2 with a1 depends on the 21-residue COOH-terminal tail of α2, which is located immediately COOH-terminal to its homeodomain (*17*, *18*). Deletion of the COOH-terminal tail renders α2 incapable of cooperative binding with a1 in vitro and of repressing the haploid-specific genes in vivo (*18*). This tail is required for interaction with the a1 partner protein only, as its absence does not affect the cooperative interaction of α2 with MCM1 (*18*). The interaction of the tail of α2 with the a1 protein is quite specific; for example, the tail fails to interact with α2 itself, a homeodomain closely related to a1 (*17*). Splicing this peptide onto the *Drosophila* engrailed homeodomain renders engrailed capable of cooperative interaction with a1 (*17*). Additional interactions between a1 and α2 are mediated by the NH_2-terminal domains of the respective proteins, which have been proposed to contact one another by way of a coiled-coil interaction (*19*). Deletion analysis of a1 and α2 has shown that their cooperative binding to the hsg operator requires only the homeodomain of a1 and the homeodomain plus the tail of α2 (*20*). These fragments, which were used for

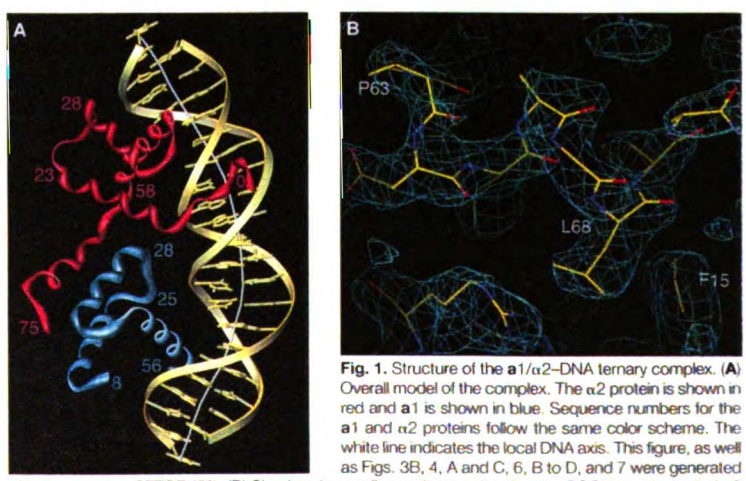


Fig. 1. Structure of the **a**1/α2–DNA ternary complex. (**A**) Overall model of the complex. The α2 protein is shown in red and **a**1 is shown in blue. Sequence numbers for the **a**1 and α2 proteins follow the same color scheme. The white line indicates the local DNA axis. This figure, as well as Figs. 3B, 4, A and C, 6, B to D, and 7 were generated with the program SETOR (*59*). (**B**) Simulated annealing omit map showing the COOH-terminal tail of α2 bound to **a**1. Residues 59 to 74 of α2 were deleted and the model was subjected to limited simulated annealing refinement in order to remove phase bias. The electron density map shown was calculated with $2F_o - F_c$ coefficients and phases from the resulting model and is contoured at 1.0 σ.

T. Li is in the Department of Biophysics and Biophysical Chemistry and C. Wolberger is in the Department of Biophysics and Biophysical Chemistry and the Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205–2185, USA. M. R. Stark and A. D. Johnson are in the Department of Microbiology and Immunology, University of California School of Medicine, San Francisco, CA 94143–0502, USA.

*To whom correspondence should be addressed

the structural studies described here, bind DNA with 10-fold reduced affinity as compared with the full-length proteins (*11*, *20*). We refer to the complex formed by these fragments as the a1/α2 heterodimer. The equilibrium constant for dissociation of this a1/α2 heterodimer into monomers is 2×10^{-8} M (*20*).

We previously described the x-ray crystal structure of the α2 homeodomain bound to DNA (*6*). In the absence of its partner proteins, the α2 homeodomain binds as a monomer to nearly straight B-form DNA. Although the fragment of α2 crystallized contained the COOH-terminal tail that mediates heterodimerization with a1, the 21 residues of the tail were disordered in the structure of the α2-DNA complex. Solution nuclear magnetic resonance (NMR) studies show that the COOH-terminal tail is unstructured in solution as well, becoming ordered only upon heterodimerization with a1, when part of it forms a short α helix (*7*, *20*).

In order to uncover how a flexible peptide tail mediates specific heterodimer formation between two homeodomains, we determined the three-dimensional structure of a ternary complex containing an a1/α2 heterodimer bound to DNA. This crystal structure, determined at a resolution of 2.5 Å, shows that the a1 and α2 homeodomains bind in tandem to a 21–base pair (bp) DNA fragment. In the ternary complex, 16 residues of the COOH-terminal tail of α2 undergo a conformational change, becoming ordered and contacting the a1 homeodomain at a surface that does not participate in DNA binding. The heterodimer induces a pronounced DNA bend that is required for contacts between the two proteins. The manner in which a1 and α2 heterodimerize represents a possible mechanism by which other transcriptional regulators can associate with one another on the DNA.

Overview of the ternary complex. The a1/α2–DNA complex (Fig. 1A) contains the a1 and α2 proteins bound to a 21-bp fragment of bent duplex DNA. The structure of the complex was determined as described (Table 1); a representative view showing the fit of the model to the electron density map is shown in Fig. 1B. The sequence of the DNA site was derived from the sequences of 14 in vivo a1/α2 binding sites (Fig. 2C) and differs in the four central base pairs from the consensus sequence. Of a number of different DNA sequences and oligonucleotide lengths, this DNA site yielded the best crystals. The sequences of the proteins and the DNA, and the numbering scheme used to describe them, are shown in Fig. 2. The a1 and α2 homeodomains bind in a tandem orientation to one face of the DNA, burying 2300 $Å^2$ of protein and DNA surface area. The tandem binding of the two homeodomains, which had been predicted on the basis of chemical protection experiments (*16*), is in agreement with the observed pattern of DNA protection from hydroxyl radical attack and methylation in the presence of bound a1/α2 (Fig. 2D).

The α2 protein contacts the a1 homeodomain with a peptide tail located COOH-terminal to the α2 homeodomain. As compared with the structure of α2 alone bound to DNA, an additional 16 residues are ordered in the ternary complex. This COOH-terminal tail, consisting of residues 59 to 74, extends from the end of helix 3 of the α2 homeodomain (Fig. 1A). Residues 59 to 62 contain a short stretch of extended chain, followed by two turns of an amphipathic helix (residues 63 to 69), which contacts the a1 homeodomain on the face opposite to that which binds DNA. The helix in the tail of α2 packs between helices 1 and 2 of a1, with the axes of all three helices roughly parallel. The α2 tail helix is somewhat distorted, with deviations from ideal hydrogen bonding geometry at residues 64 and 65. As predicted in biochemical and NMR studies (*20*), all contacts with the a1 homeodomain are mediated by the COOH-terminal extension of α2; the homeodomain of α2 (residues 0 to 59) forms no direct contacts with the a1 protein.

The model of the ternary complex contains residues 0 to 74 of α2, including the

Table 1. Crystallographic analysis. The purification and crystallization of the a1/α2–DNA complex and the method of flash-freezing the crystals have been described (*51*). The complex crystallizes in space group P6, with unit cell dimensions $a = b = 132.8$ Å and $c = 45.71$ Å. The crystals form with one complex per asymmetric unit and contain 68 percent solvent. Diffraction from these crystals extends to a maximum resolution of 2.4 Å, with diffraction along the *c* axis falling off in intensity beyond 2.7 Å. Heavy atom derivatives were prepared by substituting 5-iodouracil for thymine at base 1, 11, or 22 of the DNA (IdU^1, IdU^{11}, and IdU^{22}). X-ray diffraction data collection, processing, and multiple isomorphorous replacement (MIR) phase calculation and refinement were as described (*52*). The MIR phases were used to calculate an electron density map at 2.8 Å resolution, which was improved by one round of solvent flattening (*53*). A nearly complete model of the complex was fit to this map by means of the O graphics program (*54*). A statistically random selection of 10 percent of the total reflection data was excluded from the refinement and used to calculate the free *R* factor (R_{free}) as a monitor of model bias (*55*). The model was subjected to several rounds of positional and simulated-annealing refinement with X-PLOR (*56*). Several missing side chains and residues were added following inspection of $2F_o - F_c$ and $F_o - F_c$ maps and a phase-combined map calculated with SIGMAA (*57*). The analysis was continued with 120 cycles of positional and constrained temperature factor refinement, yielding an *R* factor at 2.8 Å resolution of 22.4 percent and an R_{free} of 29.2 percent. The model was then further refined at 2.5 Å resolution against the derivative data set, IdU^{11}. In the resolution range from 2.8 to 2.5 Å, the IdU^{11} data set contained 94 percent of the expected data; of those reflections recorded, 81 percent of the intensities were greater than 2σ. After rigid body, positional, and simulated annealing refinement, an additional five residues in the COOH-terminal tail of α2 were fit to $2F_o - F_c$ and $F_o - F_c$ maps. Simulated annealing omit maps (*58*) were calculated at many stages of the refinement to verify the placement of residues in the electron density map. Water molecules were included at the final stage of refinement, based on the presence of peaks in difference electron density maps of at least 3σ in significance. All water molecules have *B* factors of less than 50 $Å^2$ and participate in at least one hydrogen-bonding interaction. The model of the complex presented here contains 1851 atoms and 58 water molecules. The average atomic *B* factor for the proteins, excluding water molecules and the COOH-terminal tail of α2, is 39.1 $Å^2$; the average *B* factor of all atoms in the COOH-terminal tail of α2 is 60.4 $Å^2$.

	Native	IdU^1	IdU^{22}	IdU^{11}
Resolution (Å)	2.74	2.40	2.65	2.40
Measured reflections	74,037	83,570	35,860	76,841
Unique reflections	12,014	16,357	10,715	18,106
Completeness (%)	92.8	90.4	80.0	94.0
Overall $I/\sigma(I)$	9.01	11.7	10.5	14.7
R_{merge} (%)*	13.1	7.4	6.8	5.5
R_{iso} (%)†		11.6	8.2	8.3
R_{cullis}‡		0.57	0.65	0.60
Phasing power§		1.86	0.92	1.82
Mean overall figure of merit (20–2.8Å)‖		0.54		
	Refinement statistics			
Resolution range	6–2.8 Å			6–2.5 Å
R factor (%)¶	22.4			22.5
R_{free} factor (%)#	29.2			29.8
Refined geometry	*Overall*	*Protein*	*DNA*	
rmsd bond length (Å)	0.018	0.016	0.019	
rmsd bond angle (°)	1.93	1.83	2.02	

*$R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma I$, *I* observed intensity, $\langle I \rangle$ average intensity of multiple observations of symmetry-related reflections. †$R_{iso} = \Sigma||F_{PH} - F_P||/\Sigma F_P$, F_{PH} and F_P are the observed derivative and native structure factor amplitude. ‡$R_{cullis} = \Sigma||F_{PH} \pm F_P| - F_{H(calc)}||/\Sigma|F_{PH} - F_P|$. §Phasing power = $\Sigma|F_{H(calc)}|^2/\Sigma||F_{H(calc)}| - |F_{PH(obs)}||^2$. ‖$|F_{H(calc)}| - |F_{H(obs)}|$ is the lack-of-closure error. ‖Mean figure of merit = $\langle|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|\rangle$, α: phase, $P(\alpha)$: phase probability distribution. ¶$R = \Sigma||F_o| - |F_{H(calc)}||/\Sigma F_o$. #*R* factor for a subset of 10 percent of the reflection data that were not included in the crystallographic refinement.

NH_2-terminal arm (residues 0 to 7), helices 1 (residues 10 to 23), 2 (residues 28 to 37), and 3 (residues 42 to 58), and the COOH-terminal tail (59 to 74). An additional five COOH-terminal residues, which are disordered in the structure, are not required for **a**1/α2 repression (*18*). Of the **a**1 protein, residues 8 to 56 are included in the model. The first 12 amino acids in the protein, including the NH_2-terminal arm, are disordered, as is the COOH-terminal residue of helix 3, Lys^{57}. The **a**1 and α2 homeodomains, which are 21 percent identical in sequence, are similar in structure and superpose with a 1.0 Å root-mean-square difference in Cα positions (residues 8 to 57).

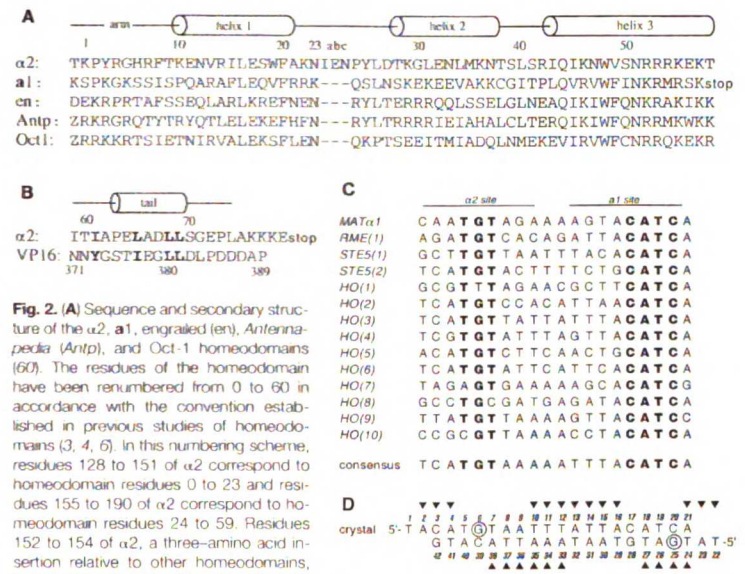


Fig. 2. (**A**) Sequence and secondary structure of the α2, **a**1, engrailed (en), *Antennapedia* (*Antp*), and Oct-1 homeodomains (*60*). The residues of the homeodomain have been renumbered from 0 to 60 in accordance with the convention established in previous studies of homeodomains (*3*, *4*, *6*). In this numbering scheme, residues 128 to 151 of α2 correspond to homeodomain residues 0 to 23 and residues 155 to 190 of α2 correspond to homeodomain residues 24 to 59. Residues 152 to 154 of α2, a three–amino acid insertion relative to other homeodomains, are labeled a, b, and c. The **a**1 residues 69 to 126 have been renumbered 0 to 57. (**B**) Alignment of the COOH-terminal tail of α2 with a segment of the herpes virus activator protein, VP16, which interacts with the Oct-1 homeodomain. The COOH-terminal tail of α2, residues 191 to 210, have been renumbered 60 to 79. Absolute residue numbers for VP16 are shown below the amino acid sequence. (**C**) Naturally occurring and synthetic DNA binding sites of **a**1/α2. DNA sequences are written 5′ to 3′; invariant and highly conserved bases are highlighted in bold. Sequences are shown of 14 in vivo binding sites for **a**1/α2 located upstream of haploid-specific genes (*16*); the consensus sequence is derived from the list shown. (**D**) Sequence of the double-stranded oligonucleotide used in this crystallographic study and results of in vitro chemical protection experiments (*16*). Circled guanine bases are protected from methylation by binding to DNA of the **a**1/α2 heterodimer. Triangles indicate bases that are protected from hydroxyl radical attack by **a**1/α2.

DNA structure and crystal packing. The DNA in the **a**1/α2–DNA complex contains a marked overall bend of 60° (Figs. 1A and 3). In contrast is the nearly straight DNA in the complex of α2 alone bound to DNA as discussed below, which contains a bend of 8.8° within a single homeodomain binding site and an overall bend of 7°. Our observations are in agreement with solution studies that have shown that the **a**1/α2 heterodimer introduces a bend in the DNA estimated at 100°, while α2 alone does not (*21*). The bend in the **a**1/α2 binding site occurs without dramatic local distortion or kinking of the B-DNA helix. Rather, the DNA helix is smoothly bent, most noticeably at the center of the DNA fragment and in the **a**1 half of the binding site. The bend is largely the result of a variation in base roll, which adopts negative values near the center of the DNA site and positive values in flanking base pairs (Fig. 3A). A consequence of the bend is a narrowing of the minor groove at the center of the DNA fragment, between the **a**1 and α2 proteins (Fig. 3A). The continued bending of the DNA in the **a**1 half of the site leads to a widening of the minor groove and a narrowing of the major groove. The DNA in the ternary complex bends toward the **a**1/α2 heterodimer, facilitating interactions between the two proteins. The COOH-terminal tail of α2 spans the gap between the **a**1 and α2 homeodomains at the point where the minor groove of the DNA helix is at its narrowest. Without the observed bend in the DNA, the tail of α2 could not reach its binding site on the back of the **a**1 homeodomain. The protein-protein interactions at the heterodimer interface, as well as the

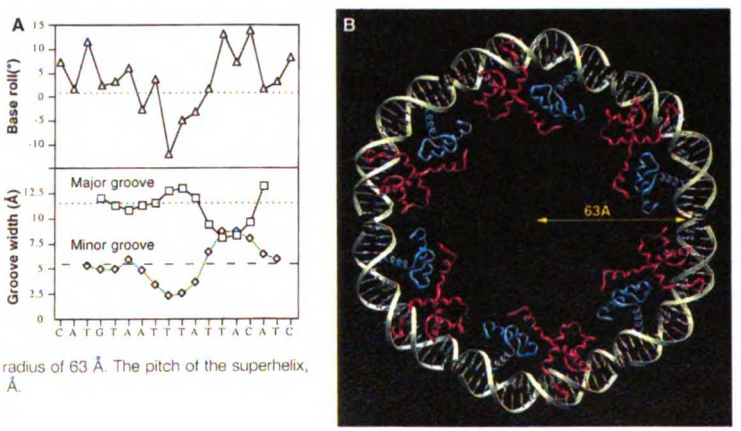


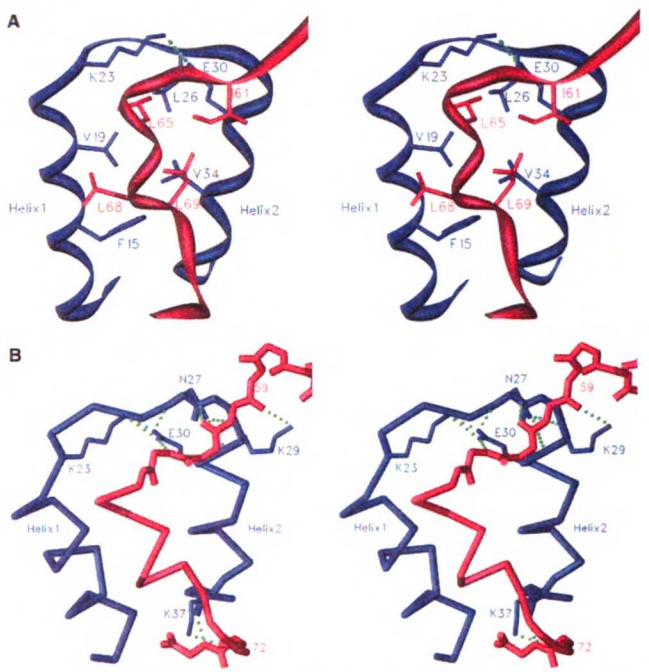
Fig. 3. (**A**) Plot of DNA parameters. A plot is shown of base roll, major groove width, and minor groove width as a function of DNA sequence; horizontal dotted or dashed lines indicate average values for B-DNA. The sequence of the top strand only is shown. Groove widths are measured as the minimum in phosphate-phosphate distance less 5.8 Å for the van der Waals radii of the phosphate groups. All parameters were calculated with the program, CURVES (*61*). (**B**) Packing of complexes in the **a**1/α2–DNA crystals; view looking down the crystallographic 6_1 axis. Adjacent complexes stack end-to-end in the crystal, with the dinucleotide 5′-ApT overhang in one complex forming base-pairing interactions with the complementary 5′-overhanging bases in the DNA of the adjacent complex. With each complex containing a 21-bp DNA fragment that has an overall bend of 60°, the stacked complexes form a superhelical spiral with a radius of 63 Å. The pitch of the superhelix, corresponding to the crystallographic c axis, is 45.71 Å.

contacts between a1 and α2 and their respective DNA subsites (both described below), probably play a role in stabilizing the bend. In addition, we observed a spine of hydration in the minor groove of the DNA, where it narrows and base roll angles are negative, and in the major groove of the DNA, where bending results in a minimum in major groove width and a maximum positive value of base pair roll angles (Fig. 3A). This hydration may contribute to the stability of the bent DNA conformation, in addition to participating in water-mediated hydrogen bonds between the proteins and the DNA bases.

The DNA bending gives rise to an unusual packing of complexes in the crystal. As has been observed in other crystals of protein-DNA complexes (*4*, *6*, *22*–*26*), the DNA stacks end-to-end in the crystal, forming a pseudocontinuous helix with Watson-Crick base pairing between the overhanging 5′-ApT at the end of one complex and the complementary unpaired bases at the end of the adjacent complex. In the a1/α2–DNA crystals, each successive complex bends in the same direction, resulting in a superhelical spiral of complexes that obeys the 6_1 screw symmetry of the space group (Fig. 3B). As measured from the projection shown in Fig. 3B, the radius of curvature of the DNA in the ternary complex is 63 Å. There are crystal contacts between the α2 homeodomain in one complex and the a1 protein in the neighboring complex that may limit the degree of bending that can occur at the ends of the DNA. The observed 60° bend may therefore represent an underestimate of the bend induced by a1/α2 in a single binding site embedded in a longer fragment of DNA.

Heterodimer interface. The COOH-terminal tail of α2 is unfolded in the monomer, even when bound to DNA. On interaction with the a1 protein, the tail of α2 folds to form a complementary surface to its binding site on the a1 homeodomain. The heterodimer is stabilized primarily by hydrophobic interactions, in addition to the presence of several hydrogen bonds (Fig. 4, A and B). A hydrophobic patch on the a1 homeodomain is formed by Val^{19} of helix 1, Val^{34} of helix 2, and Leu^{26} in the strand that connects helices 1 and 2. These three side chains are partly exposed and lie at the floor of a depression in the surface of the a1 homeodomain (Fig. 4, A and C). This depression is flanked at one end by a salt bridge between Lys^{23} and Glu^{30}, and at the other by Phe^{15}.

In the tail of α2, residues 63 to 69 adopt a helical conformation with three leucine side chains (Leu^{65}, Leu^{68}, and Leu^{69}) projecting from one face of the helix (Fig. 4A). In addition, Ile^{61} packs against Leu^{65} and Leu^{69} and helps to stabilize the short helix. The amphipathic helix binds to the a1 homeodomain, inserting the three leucine residues into the hydrophobic patch between helices 1 and 2 of a1 (Fig. 4, A and C). The a1/α2 heterodimer is further stabilized by hydrogen bonds with the a1 homeodomain that are mediated by main chain atoms in the COOH-terminal tail of α2 flanking the amphipathic helix (Fig. 4B). The two antiparallel helices of a1 and the helix contributed by the tail of α2 form a three-helix bundle, with the helix in α2 parallel to helix 2 of a1. The total buried surface area is 754 $Å^2$.



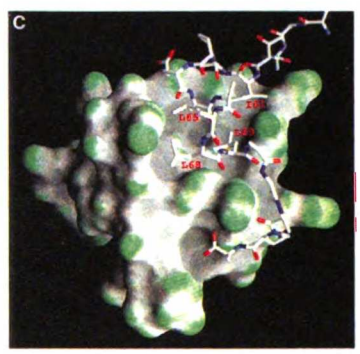


Fig. 4. (**A**) Stereo view of the **a**1/α2 heterodimer interface. The **a**1 protein is shown in blue and the α2 protein is shown in red; helix 3 of the **a**1 homeodomain has been deleted for clarity. In the **a**1 homeodomain, residues Phe^{15}, Val^{19}, Leu^{26}, and Val^{34} form a hydrophobic patch between helices 1 and 2. Residues 63 to 69 in the tail of α2 form a distorted amphipathic helix with residues Leu^{65}, Leu^{68}, and Leu^{69} exposed on one face. Ile^{61}, located NH_2-terminal to the amphipathic helix, packs against Leu^{65} and Leu^{69}. The hydrophobic face of the α2 tail forms favorable van der Waals contact with the hydrophobic patch between helices 1 and 2 of the **a**1 homeodomain, as well as with the salt bridge formed by Lys^{23} and Glu^{30}. (Additional contacts, not shown, are formed with the aliphatic side chain of Arg^{22}.) (**B**) Stereo view of the hydrogen-bonding interactions between **a**1 and α2. The main chain NH of Ala^{62} in α2 donates a hydrogen bond to the Oε of Glu^{30} in **a**1, and the second Oε accepts a hydrogen bond from the peptide NH of Asn^{27} of **a**1. The side chain of Asn^{27} in **a**1 forms a bridging contact, donating a hydrogen bond to the peptide N of Thr^{60} in α2 and capping helix 2 of **a**1 by accepting a hydrogen bond from the peptide N of Lys^{29}. The side chain of Lys^{29} donates a hydrogen bond to the carbonyl of Ile^{59}. COOH-terminal to the amphipathic helix in the α2 tail, the carbonyl of Glu^{72} hydrogen bonds with the Nζ of Lys^{37} in **a**1, while the amide N of Glu^{72} forms a hydrogen bond with the carbonyl O of Lys^{37}. (**C**) Depiction of the molecular surface of the **a**1 homeodomain, with a stick model of the COOH-terminal tail of α2. The **a**1 surface is color-coded such that the most convex part of the surface is green, the most concave part is gray, and planar surfaces are white [the figure was made with the program GRASP (*62*)].

The results of mutagenesis studies of the α2 protein verify the importance of the heterodimer contacts we observed. Point mutations that change a wild-type residue to alanine were introduced at various positions in the α2 tail and assayed for their effect on the ability of the intact α2 and a1 proteins to repress transcription in yeast (Fig. 5). The strongest effect of alanine substitution was observed at Ile^{61}, Leu^{65}, Leu^{68}, and Leu^{69}, the four hydrophobic residues that mediate key interactions in the a1/α2 complex. Substitution of side chains in the α2 tail that are not involved in heterodimer contacts have a negligible effect on repression (Fig. 5). In prior studies substitution of Leu^{65} with Ser was found to disrupt the ability of a1/α2 to repress haploid-specific genes in vivo (*27*) and the affinity of intact a1/α2 for DNA in vitro was 200 times lower (*28*).

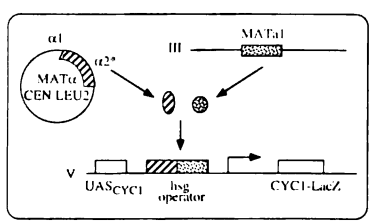
Protein-DNA interactions. The a1/α2 heterodimer forms an extensive set of contacts with a DNA binding site that spans 18 bp (Fig. 6A). Each homeodomain contacts both the bases and the sugar-phosphate backbone by means of a combination of direct side chain contacts and water-mediated hydrogen bonds. Common features of a1 and α2 binding include a conserved set of phosphate contacts and one base contact that serve to stabilize the homeodomain on its binding site. A key residue is Asn^{51}, invariant among all homeodomains (*2*), which forms a bidentate contact with an adenine base (Fig. 6, A, B, and D) that has been observed in the engrailed, Oct-1, and α2-DNA complexes (*4*, *6*, *8*). Flanking this base, a1 and α2 contact four phosphates in the same way (Fig. 6A). These contacts are mediated by four residues that are identical in both proteins—Leu^{26}, Gln^{44}, Trp^{48}, and Arg^{53}—and by the main chain NH of residue 8. DNA sequence recognition presumably arises from contacts with the DNA bases that are mediated by side chains that differ between the two homeodomains.

The α2 protein contacts a total of 7 bp in either the major or minor groove. Four base pairs in the major groove are contacted by three side chains in helix 3: Asn^{51}, Arg^{54}, and Ser^{50} (Fig. 6B). Of these, Ser^{50}, which contributes to DNA binding specificity differences among homeodomains (*29–31*), forms water-mediated hydrogen bonds with 2 bp, and Arg^{54} donates hydrogen bonds to a base (Gua^{6}) and to a side chain, Asn^{51}, which contacts Ade^{38}. The Arg^{54} contact accounts for the observation that Gua^{6} is the sole guanine in the α2 site that is protected from methylation by a1/α2 binding. In the minor groove, 5 bp are contacted by three residues in the NH_2-terminal arm of α2—Arg^{4}, Gly^{5}, and Arg^{7} (Fig. 6C). Two of the base pairs contacted in the minor groove are also contacted by side chains in the major groove, thereby accounting for the high DNA sequence conservation at these positions (Fig. 2C). In addition to contacting 2 bp, Arg^{7} hydrogen bonds to a water molecule that is part of the spine of hydration observed where the minor groove narrows. Further stabilization of the α2 homeodomain on its binding site is provided by a set of hydrogen bonds, salt bridges, and van der Waals interactions with the sugar-phosphate backbone of the DNA that are diagrammed in Fig. 6A. Most of these arise from side chains in helix 3 that form contacts with the DNA backbone flanking helix 3.

The a1 homeodomain is positioned on the DNA in a manner similar to that of the α2 protein, contacting bases in the major groove with five residues in helix 3—Val^{47}, Ile^{50}, Asn^{51}, Met^{54}, and Arg^{55} (Fig. 6D). With the exception of the invariant Asn^{51}, the identities of DNA-contacting residues 47, 50, 54, and 55 differ from those in α2 and hence mediate different base contacts. The Ile^{50} and Met^{54} residues are in van der Waals contact with, respectively, Thy^{15} and Cyt^{17}. Also, Arg^{55} forms two hydrogen bonds with Gua^{16}, thereby accounting for the protection of this base from methylation when a1/α2 binds to DNA; Arg^{55} donates an additional hydrogen bond to Thy^{19}, which is base-paired with the adenine (Ade^{26}) that is contacted by the invariant Asn^{51}. The Val^{47} residue forms additional van der Waals contacts with Ade^{26}. Furthermore, 10 residues participate in contacts with the DNA backbone (Fig. 6A). The observed major groove contacts are consistent with the sequence conservation in the a1 subsite (Fig. 2C): All four invariant DNA positions are contacted in the major groove and two of the base pairs (Thy^{19}·Ade^{26} and Cyt^{20}·Gua^{25}) are contacted by more than one side chain.

In addition to direct side chain–DNA contacts, we see a network of five water molecules immobilized in the major groove at the interface between helix 3 and the DNA (Fig. 6D). These water molecules participate in hydrogen bond interactions with a total of 4 bp, and are stabilized by a hydrogen bond formed by one of the water molecules with the guanidinium of Arg^{46}, which also forms a salt bridge with a phosphate. Formation of this network of water-mediated contacts requires the observed local DNA curvature. An extensively hydrated interface between helix 3 and the DNA has also been observed by solution NMR in the Antennapedia-DNA complex (*32*).

Unlike that of the α2 homeodomain, the NH_2-terminal arm of a1 is disordered in the crystal. This result was unanticipated, as our present structure of the α2 homeodomain and earlier structural studies of homeodomains (*4–6*, *8*) show that the NH_2-



	β-gal units	Pheromone Production a-factor	Pheromone Production α-factor	Mating Behavior as a	Mating Behavior as α
vector only	45.0	+++	--	+	--
MAT α2	5.4	+	--	+	--
T58A	6.5	+	--	+	--
I59A	7.0	+	--	+	--
T60A	6.3	+	--	+	--
I61A	22.7	+	+++	+	+
P63A	7.4	+	--	+	--
E64A	7.8	+	--	+	--
L65A	44.1	+	+++	+	+
D67A	6.1	+	--	+	--
L68A	19.5	+	+++	+	+
L69A	32.4	+	+++	+	+
E72A	5.9	+	--	+	--

Fig. 5. Effects of point mutations in the COOH-terminal tail of α2 on a1/α2-mediated repression in vivo. Alanine was substituted for the wild-type residue at various positions in the COOH-terminal tail of α2. Low copy (CEN) plasmids containing the MATα locus with specific site-directed mutations in the MATα2 gene (α2* in the figure) were transformed into a MAT**a** strain and repression by a1/α2 was monitored: (i) by expression of a test promoter whose expression is controlled by a single hsg operator, (ii) by the ability of transformants to produce mating pheromones, and (iii) by the ability of the transformants to mate. The mutations in the COOH-terminal tail of α2 are shown in the first column on the left. The second column shows the expression of a CYC-LacZ reporter construct that contains an hsg operator. In a cell carrying a wild-type MATα2 gene on a plasmid, the β-galactosidase reporter gene is expressed at approximately one-ninth the level of a cell containing the plasmid vector only [this level of repression is not complete, probably because of plasmid loss—see (*28*)]. Four of the α2 tail mutants fail to repress efficiently the test promoter, with Leu^{65}→Ala having the strongest effect. The third and fourth columns show the production of mating pheromones by the transformants. For all of the mutants, production of **a** factor is repressed as compared to the vector control, indicating that the α2 mutant proteins are synthesized in the cell and can still function with MCM1 to repress the **a**-specific genes (*18*). The fourth column indicates that the four mutants that show significant derepression of the test promoter fail to repress STE12, a haploid-specific gene required for α-factor production. The fifth and sixth columns show the mating behavior of the transformants. The phenotype of the four mutants that enable the yeast to mate as α cells are consistent with a defect in **a**1/α2–mediated repression.

terminal arm of the homeodomain binds in the minor groove of the DNA. Chemical modification experiments that probe the protection of the DNA backbone from hydroxyl radical attack in the presence of the **a**1/α2 heterodimer are consistent with the binding of the NH_2-terminal arm of **a**1 in the minor groove (*16*). It is possible that the NH_2-terminal arm of **a**1 has been displaced from the minor groove in the **a**1/α2–DNA crystal, where the NH_2-terminal arm of **a**1 would be expected to bind near the junction between two double-stranded synthetic oligonucleotides. The 5′ phosphate group that would be present at that junction in a continuous DNA strand could be of importance in stabilizing binding of the NH_2-terminal arm of **a**1 to DNA. If there are indeed further DNA contacts formed with bases in the minor groove, some may contribute further to the observed DNA sequence preference of the **a**1 protein.

Because of the extensive set of contacts formed by the **a**1 homeodomain with DNA, it is surprising that **a**1 does not bind DNA detectably on its own. There are several possible explanations for this phenomenon. One is that DNA bending is required for the protein-DNA contacts we observe. The energy to induce and stabilize this bend might be too great for **a**1 alone, but may be facilitated by the presence of α2 on the DNA and the interaction between the α2 tail and the **a**1 homeodomain. Another explanation may be that the binding of the tail of α2 to **a**1 induces a small conformational change in the **a**1 homeodomain that increases its overall DNA binding energy. Solution NMR studies of the **a**1 homeodomain have shown that, on heterodimerization with α2, the resonances in **a**1 that are most perturbed on complex formation cor-

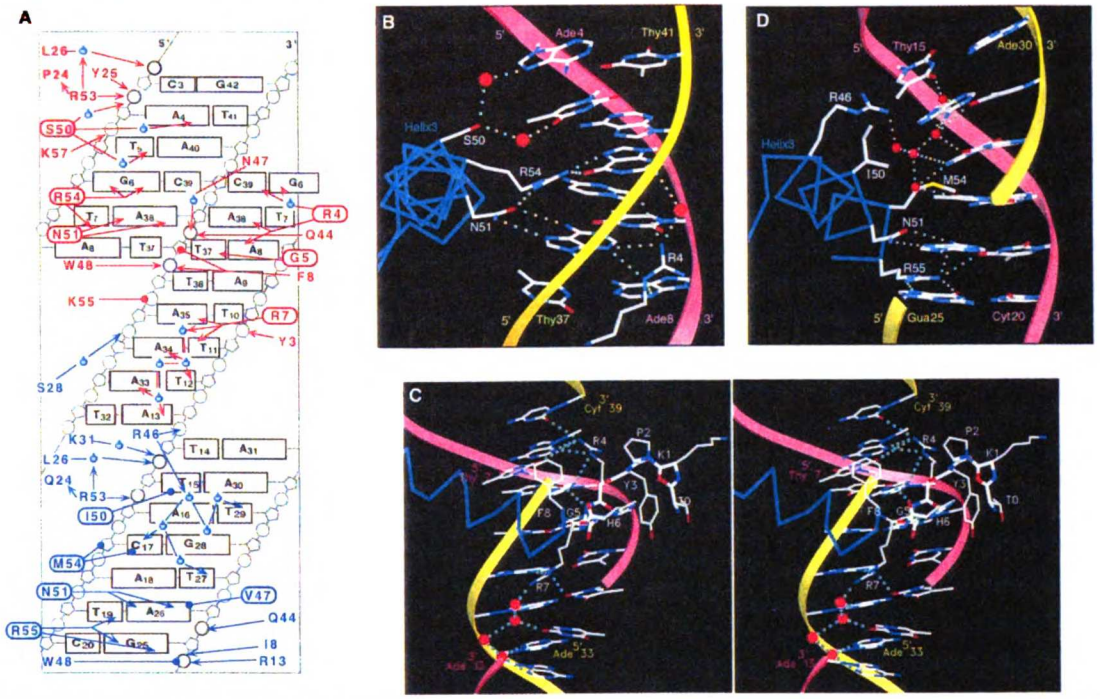


Fig. 6. Protein-DNA interactions. (**A**) Schematic diagram summarizing contacts formed by both **a**1 and α2 with the DNA. The α2-mediated contacts are shown in red; **a**1-mediated contacts are shown in blue. Residues encircled with ovals form base contacts. Hydrogen bonds and salt bridges are indicated by arrows; van der Waals interactions are indicated by lines that end in circles. Bond criteria are as follows: for hydrogen bonds, less than or equal to 3.3 Å separation of a donor-acceptor pair; van der Waals interactions, less than 4.0 Å distance between contacting groups. Residues involved in base contacts are encircled with ovals. The shaded phosphate groups and adenine bases are contacted in a conserved manner by both **a**1 and α2. The complex equilibrium constant for dissociation of the ternary complex into free **a**1, α2, and DNA is 10^{-15} M^2 for the **a**1 and α2 fragments used in this study and 10^{-16} M^2 for the intact proteins (*11, 20*). (**B**) Contacts formed between helix 3 of α2 and bases in the major groove. Asn^{51} contacts Ade^{-8}, with the Oδ of the side chain accepting a hydrogen bond from the N6 of the base and whose Nδ is in contact with the N7 (dist. = 3.4 Å). Arg^{54} hydrogen bonds to the N7 and O6 of Gua^{6} with its NH1, while its NH2 group forms a bridging hydrogen bond to the Oδ of Asn^{51}. Ser^{50} forms water-mediated hydrogen bonds with the N7 of Ade^{4} and the O4 of Thy^{5}. (**C**) Stereo view of contacts between the NH_2-terminal arm of α2 and bases in the minor groove of the DNA. A total of five bases are contacted by three side chains in the NH_2-terminal arm of α2. Arg^{4} contacts 2 bp directly and a third via water-mediated hydrogen bonds. The Nε of Arg^{4} is within hydrogen bonding distance of the N3 of Ade^{8} and the N3 of Ade^{-8}, whereas the NH_2 contacts the Gua^{6}-Cyt^{39} base pair via water-mediated hydrogen bonds. The peptide NH of Gly^{5} donates a hydrogen bond to the O2 of Thy^{17}, while the guanidinium group of Arg^{7} donates hydrogen bonds to the N3 of Ade^{-5} and the O2 of Thy^{11}. In addition, the NH_2 of Arg^{7} helps to stabilize the spine of hydration in the minor groove. (**D**) Contacts between the **a**1 homeodomain and bases in the major groove. Ile^{50} is in van der Waals contact with the methyl of Thy^{15}, while Asn^{51} forms two hydrogen bonds with Ade^{-6}, whose Oδ accepts a hydrogen bond from the N6 of the base and whose Nδ donates a hydrogen bond to the N7. The Cγ of Met^{54} is in van der Waals contact with the C6 of the Cyt^{17} pyrimidine, while the Cε is in van der Waals contact with the adjoining subdeoxyribose. The Arg^{55} side chain forms contacts with two base pairs, its NH1 donating a hydrogen bond to the N7 of Gua^{25} and its NH2 donating hydrogen bonds to the O6 of Gua^{25} and the O4 of Thy^{19}. In addition, there is a network of five bound water molecules in the major groove that hydrogen bond to the guanidinium of Arg^{46}, as well as 3 bp. Not shown in this depiction is Val^{47}, whose Cγ is in van der Waals contact with the C8 of Ade^{26}.

respond to residues that lie in portions of helices 1 and 2 and in the loop connecting them (*33*). A likely set of contacts that could be affected by a conformational change in the loop are those mediated by the invariant homeodomain residue Arg^{53}, which participates in an extensive set of contacts with the sugar-phosphate backbone, a bound water molecule, and the backbone of **a**1 in the loop between helices 1 and 2 (Fig. 6A) (*34*). Very small changes in the loop conformation could easily disrupt this network of contacts, thereby diminishing the affinity of **a**1 for DNA in the absence of α2. Since Arg^{53} is invariant among all homeodomains and mediates a conserved set of contacts in **a**1, α2, and other homeodomains, it is likely to be crucial for formation of favorable interactions between **a**1 and its DNA binding site.

In previous studies, the homeodomain was found to be positioned on the DNA in a highly conserved manner by DNA contacts that are mediated by conserved residues (*4*–*6*, *8*). A pertinent question is whether the orientation of the homeodomain on the DNA and the contacts it forms are perturbed as a result of heterodimerization. We find that, despite the additional protein-protein interactions and the concomitant distortion in the DNA, the α2 homeodomain in the heterodimer binds DNA in a manner essentially identical to that found in the structure of α2 alone bound to DNA (Fig. 7). This similarity extends to the side chain contacts formed by α2 with its binding site and the local structure of the DNA. The few differences between the two structures are most likely the result of differences in the resolution of the structure determinations (*35*). Because of the similarity in the way α2 is positioned on the DNA in the presence and absence of **a**1, it is very surprising that a mutant α2 protein with Ala substituted for Ser^{50}, Asn^{51}, and Arg^{53} exhibits no significant difference in the affinity of the **a**1/α2 heterodimer for DNA while greatly reducing the affinity of α2 alone for DNA (*28*). It is possible, however, that this triple mutation reduces the DNA sequence discrimination of the **a**1/α2 heterodimer without impairing its affinity for the DNA.

The structure of the **a**1/α2–DNA complex shows how the binding specificity of a homeodomain can be raised by complex formation with a second protein. The heterodimerization of the α2 and **a**1 homeodomains, each with only modest affinity and specificity for DNA, results in a heterodimer that binds DNA in a highly specific manner, preferring its own binding site over random DNA by a ratio of at least 10^5 (*11*). The specificity of this interaction derives from at least three sources: the specificity of the interactions between the COOH-terminal tail of α2 and the **a**1 homeodomain, summation of the DNA sequence preferences of the individual proteins, and the precise binding site spacing imposed by the nature of the heterodimer interface. It is clear from the **a**1/α2–DNA structure that insertion or deletion of base pairs between the **a**1 and α2 binding sites would disrupt heterodimer contacts, and this observation is supported by studies of **a**1/α2 binding to altered DNA sites (*16*, *35a*). A final contributing factor to ternary complex stability may be the relative deformability of a given sequence of DNA in that DNA bending is required for formation of the complex that we observe. Thus the sequence at even noncontacted bases may contribute to overall complex stability. The net result of all these considerations is a larger set of criteria that must be met for a stable **a**1/α2–DNA complex to form, and hence a much higher specificity of binding than is observed for the monomeric α2 or **a**1 proteins.

Multi-protein complex formation in transcription. The work presented above shows how the interaction between two homeodomain proteins can be mediated by a flexible tail that becomes ordered on complex formation. There are now numerous examples of cooperative interactions involving homeodomain proteins, some of which might be mediated by the same types of interactions observed in the **a**1/α2 heterodimer. The *Caenorhabditis elegans* homeodomain protein MEC-3 heterodimerizes with the UNC-86 POU domain protein, which also contains a homeodomain (*36*). This interaction is reminiscent of **a**1/α2 heterodimer formation in that it is dependent upon the 16 amino acids COOH-terminal to the MEC-3 homeodomain. Cooperative interactions have been observed between the *Drosophila* extradenticle (exd) homeodomain protein and the Ultrabithorax (Ubx), engrailed, and abdominal-A homeodomain proteins (*37*–*38*). These interactions require either NH_2- or COOH-terminal extensions to the homeodomain, one as short as 15 residues (*37*). Pbx1, a human homeodomain protein closely related to exd (*39*), similarly has been observed to bind DNA cooperatively with several Hox homeodomain proteins (*40*). The Pbx-Hox interactions are dependent on a short NH_2-terminal peptide in the Hox proteins and a COOH-terminal tail in Pbx. The homeodomain most closely related to both Pbx and exd is **a**1, with which they share 40 percent sequence identity (*41*); in contrast, **a**1 and α2 share only 21 percent sequence identity. The sequence similarity between **a**1, Pbx, and exd may imply the existence of a common mechanism used by these proteins to recognize their homeodomain partners.

The asymmetric nature of **a**1/α2 heterodimer formation raises the possibility that a similar type of interaction could occur between a homeodomain protein and a nonhomeodomain protein containing a peptide similar to the tail of α2. An example of this may be the mammalian Oct-1 POU homeodomain protein, which binds DNA cooperatively with the herpes virus VP16 transcriptional activator (*42*). Point mutations in Oct-1 that disrupt the Oct-1–VP16 complex are located in helices 1 and 2 of the Oct-1 homeodomain. As shown above and noted in the NMR study of **a**1/α2 (*33*), the location of the Oct-1 mutations is analogous to the region on the **a**1 homeodomain that is contacted by the tail of α2 (*43*, *44*). If Oct-1 is acting as the analog of **a**1, what part of VP16 may play a role similar to the COOH-terminal tail of α2? The results of deletion and mutagenesis studies have identified a 12-residue region of VP16 responsible for interaction with the Oct-1 homeodomain (*45*, *46*). This region (residues 376 to 387), which has been proposed to form an amphipathic helix, contains sequence similarity to the helix in the α2 tail that contacts **a**1 (Fig. 2B). Model-

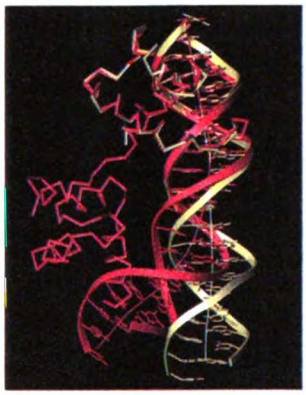


Fig. 7. Interaction of α2 with DNA in the **a**1/α2–DNA ternary complex as compared with the structure of α2 alone bound to DNA. The two structures were aligned by performing a least-squares superposition of the α2 homeodomain in the **a**1/α2–DNA complex with one of the α2 homeodomains in the structure of α2 alone bound to DNA (*6*). Since the latter structure contains two α2 homeodomain monomers bound to a 21-bp DNA fragment, the homeodomain chosen for the alignment is the one bound to the identical 9-bp sequence to which α2 is bound in the ternary complex. The second α2 homeodomain contained in the α2-DNA structure, which does not contact the other α2 monomer, has been omitted for clarity. The conformation of the α2 homeodomains, the local structure of the DNA, and the protein-DNA contacts are nearly identical in the two complexes. There are pronounced differences outside the α2 binding sites due to the overall 60° bend in the DNA induced by the **a**1/α2 heterodimer.

building studies show that this peptide from VP16, when folded in the conformation of the α2 tail, could pack between helices 1 and 2 of Oct-1 in a manner similar to that observed in the a1/α2 heterodimer.

It is likely that there are transcription factors from other structural classes that are also bound in protein-protein complexes by a flexible peptide that adopts a distinctive conformation only upon complex formation. There are strong parallels between the manner in which α2 interacts with a1 and the way in which it forms a complex with the MADS box protein (47), MCM1 (48, 49). A short region NH_2-terminal to the α2 homeodomain that is unstructured in the free protein is required for cooperative binding with MCM1 (50). This NH_2-terminal peptide specifies complex formation with MCM1, as splicing it to the engrailed homeodomain confers on engrailed the ability to bind DNA cooperatively with MCM1 (50). Thus α2 has evolved as a DNA binding protein capable of interaction with two structurally distinct partners by acquiring peptide extensions both NH_2- and COOH-terminal to its DNA binding domain that specify complex formation with different partners. The presence of flexible protein-recognition peptides that extend from stably folded DNA binding domains may prove to be a general feature of the architecture of other classes of eukaryotic transcriptional regulators.

REFERENCES AND NOTES

1. W. J. Gehring *et al.*, *Cell* **78**, 211 (1994).
2. A. Laughon, *Biochemistry* **30**, 11357 (1991).
3. Y. Q. Qian *et al.*, *Cell* **59**, 573 (1989).
4. C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Kornberg, C. O. Pabo, *ibid.* **63**, 579 (1990).
5. G. Otting *et al.*, *EMBO J.* **9**, 3085 (1990).
6. C. Wolberger, A. K. Vershon, B. Liu, A. D. Johnson, C. O. Pabo, *Cell* **67**, 517 (1991).
7. C. L. Phillips, A. K. Vershon, A. D. Johnson, F. W. Dahlquist, *Genes Dev.* **5**, 764 (1991).
8. J. D. Klemm, M. A. Rould, R. Aurora, W. Herr, C. O. Pabo, *Cell* **77**, 21 (1994).
9. M. Affolter, A. Percival-Smith, M. Muller, W. Leupin, W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4093 (1990).
10. T. Hoey and M. Levine, *Nature* **332**, 858 (1988).
11. C. Goutte and A. D. Johnson, *J. Mol. Biol.* **233**, 359 (1993).
12. A. D. Johnson, in *Transcriptional Regulation*, S. L. McKnight, K. R. Yamamoto, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1992), vol. 2, pp. 975.
13. J. C. W. Shepherd, W. McGinnis, A. E. Carrasco, E. M. De Robertis, W. J. Gehring, *Nature* **310**, 70 (1984).
14. C. Goutte and A. D. Johnson, *Cell* **52**, 875 (1988).
15. A. M. Drangova, *Nature* **347**, 682 (1990).
16. C. Goutte and A. D. Johnson, *EMBO J.* **13**, 1434 (1994).
17. M. R. Stark and A. D. Johnson, *Nature* **371**, 429 (1994).
18. A. Mak and A. D. Johnson, *Genes Dev.* **7**, 1862 (1993).
19. C. Y. Ho, J. G. Adamson, R. S. Hodges, M. Smith, *EMBO J.* **13**, 1403 (1994).
20. C. Phillips, M. R. Stark, A. D. Johnson, F. W. Dahlquist, *Biochemistry* **33**, 9294 (1994).
21. D. L. Smith, A. B. Desai, A. D. Johnson, *Nucleic Acids Res.* **23**, 1239 (1995).
22. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988).
23. J. A. Anderson, M. Ptashne, S. C. Harrison, *Nature* **326**, 846 (1987).
24. S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **253**, 1001 (1991).
25. C. Wolberger, Y. Dong, M. Ptashne, S. C. Harrison, *Nature* **335**, 789 (1988).
26. J.-A. Feng, R. C. Jonnson, R. E. Dickerson, *Science* **263**, 348 (1994).
27. J. Strathern, B. Shafer, J. Hicks, C. McGill, *Genetics* **120**, 75 (1988).
28. A. K. Vershon, Y. Jin, A. D. Johnson, *Genes Dev.* **9**, 182 (1995).
29. S. D. Hanes and R. Brent, *Science* **251**, 426 (1991).
30. A. Percival-Smith, M. Mueller, M. Affolter, W. J. Gehring, *EMBO J.* **9**, 3967 (1990).
31. J. Treisman, P. Gonczy, M. Vashishtha, E. Harris, C. Desplan, *Cell* **59**, 553 (1989).
32. M. Billeter *et al.*, *J. Mol. Biol.* **234**, 1084 (1993).
33. S. M. Baxter, D. M. Gontrum, C. L. Phillips, A. F. Roth, F. W. Dahlquist, *Biochemistry* **33**, 15309 (1994).
34. Arg^{54} donates two hydrogen bonds to phosphate P4 via its Nε and NH2 groups. The NH1 group donates one hydrogen bond to the main chain carbonyl of residue 24 and a second hydrogen bond to a bound water molecule. The water, which is found in an identical position in both the a1 and α2 subsites, in turn accepts a hydrogen bond from the peptide NH of Leu^{51} and donates a hydrogen bond to phosphate P3.
35. As compared with the structure of α2 alone bound to DNA, the α2 homeodomain in the heterodimer contains five additional ordered residues in the NH_2-terminal arm, some of which mediate minor groove contacts. In addition, no water molecules were placed in the α2-DNA structure, which was determined at a resolution of 2.7 Å.

35a. Y. Jin, J. Mead, T. Li, C. Wolberger, A. K. Vershon, *Science* **270**, 290 (1995).

36. D. Xue, Y. Tu, M. Chalfie, *ibid.* **261**, 1324 (1993).
37. F. B. Johnson, E. Parker, M. A. Krasnow, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 739 (1995).
38. S.-K. Chan, L. Jaffe, M. Capovilla, J. Botas, R. S. Mann, *Cell* **78**, 603 (1994); M. A. van Dijk and C. Murre, *ibid.*, p. 617.
39. J. Nourse *et al.*, *ibid.* **60**, 535 (1990); M. P. Kamps, C. Murre, X. Sun, D. Baltimore, *ibid.*, p. 547; C. Rauskolb, M. Peifer, E. Wieschaus, *ibid.* **74**, 1101 (1993).
40. C.-P. Chang *et al.*, *Genes Dev.* **9**, 663 (1995).
41. W. A. Flegel *et al.*, *Mechan. Dev.* **41**, 155 (1993).
42. S. Stern, M. Tanaka, W. Herr, *Nature* **341**, 624 (1989).
43. J. L. Pomerantz, T. M. Kristie, P. A. Sharp, *Genes Dev.* **6**, 2047 (1992).
44. J. S. Lai, M. A. Cleary, W. Herr, *ibid.*, p. 2058.
45. R. F. Greaves and P. O'Hare, *J. Virol.* **64**, 2716 (1990).
46. S. Stern and W. Herr, *Genes Dev.* **5**, 2555 (1991).
47. H. Ma, M. F. Yanofsky, E. M. Meyerowitz, *ibid.*, p. 484.
48. C. A. Keleher, C. Goutte, A. D. Johnson, *Cell* **53**, 927 (1988).
49. R. T. Sauer, D. L. Smith, A. D. Johnson, *Genes Dev.* **2**, 807 (1988).
50. A. K. Vershon and A. D. Johnson, *Cell* **72**, 105 (1993).
51. T. Li, M. Stark, A. D. Johnson, C. Wolberger, *Proteins* **21**, 161 (1995).
52. Crystals were flash-frozen and maintained at −179°C. x-ray diffraction data were collected with an *R*-AXIS II image plate detector mounted on a Rigaku RU-200 rotating anode x-ray generator, with CuKα radiation filtered by a graphite monochromator. The reflection data were processed, merged, and scaled with the *R*-AXIS software (Molecular Structure). Subsequent phase refinement and map calculations were performed with the PHASES suite of programs [W. Furey and S. Swaminathan, *American Crystallographic Association Meeting Abstracts* PA33 (1990)]. Iodine positions were located by Patterson and difference Fourier methods, and heavy atom parameters were calculated and refined to produce a set of MIR phases.
53. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
54. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta. Crystallogr.* **A47**, 110 (1991).
55. A. T. Brunger, *Nature* **355**, 472 (1992).
56. ———, *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).
57. R. J. Read, *Acta. Crystallogr.* **A48**, 900 (1990).
58. A. Hodel, S.-H. Kim, A. T. Brunger, *ibid.* **A48**, 851 (1992).
59. S. V. Evans, *J. Mol. Graph.* **11**, 134 (1993).
60. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
61. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 63 (1988).
62. A. Nicholls, K. Sharp, B. Honig, *Proteins*, **11**, 281 (1991).

63. We thank M. Glover, J. Kim, M. Rould, M. Bianchet, A. Gitts, J. Klemm, and W. Weis for advice and suggestions and M. Amzel, D. Leahy, and J. Berg for comments on the manuscript. Time on a Cray supercomputer was provided by the National Cancer Institute–Frederick Biomedical Supercomputing Center. Supported by NSF grant MCB-9304526 (C.W.), the David and Lucile Packard Foundation (C.W.), and the Lucille P. Markey Charitable Trust (C.W.) and by a National Institutes of Health grant GM-37049 (A.D.J.). The coordinates have been deposited in the Protein Data Bank in Brookhaven, NY, with accession number 1YRN.

26 July 1995; accepted 13 September 1995

APPENDIX C

Accessibility of α2-Repressed Promoters to the Activator Gal4

Michael J. Redd, Martha R. Stark, and Alexander D. Johnson

(published in Molecular and Cellular Biology *16*: 2865-2869, 1996)

Accessibility of α2-Repressed Promoters to the Activator Gal4

MICHAEL J. REDD,[1] MARTHA R. STARK,[1] AND ALEXANDER D. JOHNSON[1,2]*

Department of Biochemistry and Biophysics[1] and Department of Microbiology and Immunology,[2] University of California, San Francisco, California 94143-0414

Received 11 January 1996/Returned for modification 1 March 1996/Accepted 22 March 1996

It has been proposed that eukaryotic repressors of transcription can act by organizing chromatin, thereby preventing the accessibility of nearby DNA to activator proteins required for transcription initiation. In this study, we test this idea for the yeast α2 repressor using a simple, artificial promoter that contains a single binding site for the activator protein Gal4 and a single binding site for the repressor α2. When both the repressor and the activator are expressed in the same cell, the artificial promoter is efficiently repressed. In vivo footprinting experiments demonstrate that Gal4 can occupy its binding site even when the promoter is repressed. This result indicates that α2-directed repression must result from interference with some stage in transcription initiation other than activator binding to DNA.

Negative regulation of transcription in eukaryotes occurs by a variety of mechanisms. Some repressors act by preventing the DNA binding of activators, some bind DNA and interact with nearby activators, "quenching" their activation surface, and some communicate directly with the general transcription machinery, blocking its function or assembly (for reviews, see references 14, 16, 18, and 26). Still other repressors appear to organize repressive forms of chromatin that block the accessibility of proteins to DNA (for reviews, see references 31, 33, and 45). For some repressors, more than one of these mechanisms is thought to function simultaneously, resulting in a very low level of gene expression under repressing conditions.

One case in which two mechanisms of repression have been proposed is that of the yeast α2 protein. This protein is responsible for repressing the expression of two sets of cell-type-specific genes, **a**-specific genes and haploid-specific genes (for reviews, see references 7, 15, and 17). To repress **a**-specific genes, α2 binds cooperatively with the Mcm1 protein to a 34-bp DNA sequence called the **a**-specific gene operator. α2/Mcm1 binds a second protein complex composed of the Tup1 and Ssn6 proteins. Tup1 and Ssn6 are required for the repression of at least five sets of yeast genes and have been proposed to function as a general repression machine in *Saccharomyces cerevisiae*, recruited to DNA by a variety of sequence-specific DNA-binding proteins (21, 24, 41, 42).

The **a**-specific gene operator will bring about repression when placed in many positions upstream of a target gene, and models for repression by α2/Mcm1/Ssn6/Tup1 (referred to as the α2 repression complex) must account for this action at a distance (20, 32). One model proposes that the α2 repression complex interacts directly with the general transcription machinery at the promoter, blocking its assembly or maturation (13, 20). A second model proposes that the α2 repression complex positions nucleosomes over promoter elements, blocking the accessibility of nearby DNA to proteins (23, 34, 35, 37). In this work, we wished to determine whether an α2-repressed promoter is accessible to Gal4, a yeast activator protein that binds DNA.

* Corresponding author. Phone: (415) 476-8783. Fax: (415) 476-0939.

MATERIALS AND METHODS

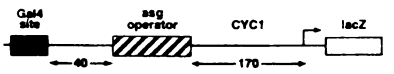
Plasmids. The a-specific gene operator used in this study is derived from *STE6* (20). The Gal4-binding site is the consensus site (CGGAGGACTGTCCTCCGT GCA) (44). The Gal4-binding site and the *STE6* operator were ligated into the *Pst*I site and the *Sal*I site, respectively, of the Bluescript polylinker and were subsequently subcloned into the blunted *Sal*I site of pΔSS (19) in either orientation to produce $pASG_{2\mu}$ and $pGAL_{2\mu}$. Promoter regions were then sequenced. Integrating plasmids were constructed by removing the 2μm sequences, resulting in $pASG_{int}$ and $pGAL_{int}$.

Yeast strains and β-galactosidase assays. All four yeast strains used in this study are derivatives of EG123 (*MATa trp1 leu2 ura3 his4*). *matΔ* is KT23αx8, created by deletion of *MATα* from 246-1-1 (*MATα trp1 leu2 ura3 his4*) (36, 39). Plasmid pSJ4*LEU* was used to make a deletion insertion of *LEU2* at the *GAL4* gene (10). Plasmids $pASG_{int}$ and $pGAL_{int}$ were integrated into the *ura3-52* allele. Integrations were confirmed by Southern analysis (38). β-Galactosidase assays were performed as described by Goutte and Johnson (12). Cells were grown initially on synthetic medium minus uracil plus 2% glucose and then transferred to synthetic medium minus uracil plus 2% galactose, 2% ethanol, and 3% glycerol for several cell doublings.

Competitive PCR for quantitation of mRNA. The levels of repression of an a-specific gene, STE2, were compared at the RNA level between *MATα* and *matΔ* cells. Quantitative PCR (9) was used to detect the very low levels of **a**-specific gene mRNA present in α cells. Briefly, RNA was isolated from cells, reverse transcribed (Superscript II; BRL) by using a *STE2*-specific primer, and added to PCR mixtures containing known amounts of a competitor DNA that was amplified with the same *STE2* primers as the cDNA but that resulted in a smaller PCR product due to an internal deletion in the *STE2* gene. The relative amounts of target cDNA versus competitor can be measured by direct scanning of ethidium-stained gels (1-D Multi-Lane Scan, IS-1000 Digital Imaging System), and these amounts can be compared between *MATα* and *matΔ* cells to determine the level of repression of an **a**-specific gene.

Genomic footprinting. In vivo footprinting was performed as previously described, with modifications (1). Yeast strains were grown in 100 ml of synthetic medium minus uracil plus 2% galactose, 3% glycerol, and 2% ethanol to a density of 10^7 cells per ml. The cells were pelleted and resuspended in ice-cold medium to a final volume of 1 ml. A 5-μl volume of dimethyl sulfate was added with vigorous mixing. The cells were incubated at 20°C for 5 min, after which the reaction was quenched with 50 ml of ice-cold 10 mM Tris (pH 7.5)–1 mM EDTA. The cells were pelleted and resuspended in 900 μl of lysis buffer (50 mM morpholinepropanesulfonic acid [pH 7.0], 200 mM NaCl, 5 mM EDTA, 0.5% Triton X-100). The cells were lysed with glass beads (0.5-mm diameter) for 45 s in a bead beater (Biospec Products). The lysate was removed from the glass beads and diluted in 3.5 ml of additional lysis buffer. The lysate was treated with RNase A (250 μg/ml) and proteinase K (100 μg/ml) for 1 h at 37°C. The cellular debris was pelleted (12,000 × *g* for 20 min), and genomic DNA was prepared by loading the supernatant onto a Qiagen column (Qiagen Inc., Studio City, Calif.). DNA was then digested with *Hae*III, phenol chloroform extracted, ethanol precipitated, and resuspended in 100 μl of Tris-EDTA. Finally, the DNA was dialyzed against water (12,000- to 14,000-Da exclusion limit) for 2 h.

Methylated bases were detected by multiple rounds of primer extension with *Taq* polymerase. A 0.5-μg amount of DNA from cells with 2μm plasmids or 10 μg from cells with single-copy reporters, 1 pmol of end-labeled primer, 1 U of *Taq* polymerase, 200 μM each deoxynucleoside triphosphate, and 1× *Taq* buffer (40 mM NaCl, 10 mM Tris [pH 8.9], 5 mM $MgCl_2$, 0.01% gelatin [30]) were combined in a total volume of 50 μl. Mineral oil was layered over the samples,



	$pASG_{2\mu}$	$pASG_{int}$
MATα	3	0.15
matΔ	2500	125
MATα gal4::LEU2	2	0.17
matΔ gal4::LEU2	220	8

	$pGAL_{2\mu}$	$pGAL_{int}$
MATα	88	11
matΔ	3900	217
MATα gal4::LEU2	2	0.14
matΔ gal4::LEU2	30	2

FIG. 1. α2 represses test constructs activated by Gal4. The test constructs are diagrammed at the top of the figure. Each construct consists of a single Gal4-binding site and a single **a**-specific gene (asg) operator upstream of a *CYC1 lacZ* promoter fusion. The distances in base pairs between the promoter elements are indicated. At the bottom of the figure are the results of β-galactosidase activity assays performed with four different strains. Values are the averages of assays performed in duplicate on three independent transformants.

which were then subjected to 10 to 20 rounds of thermal cycles (1 min at 94°C, 2 min at 55 to 63°C, and 1 min at 72°C). The mineral oil was extracted with chloroform, and the samples were ethanol precipitated. The pellets were washed with 70% ethanol, dried briefly, and resuspended in 4 μl of formamide loading buffer. The primer extension products were then electrophoresed through a 6% polyacrylamide sequencing gel. The gels were dried and exposed to Kodak XAR-5 film for 12 to 24 h. Note that many methylated guanines appear as doublets by *Taq* polymerase primer extension because of the variable addition of an extra nucleotide. This does not affect the interpretation of these results.

Plasmid DNA was methylated in vitro as described by Maxam and Gilbert (28), and 10 ng was used for primer extension as described above. Neither the methylated plasmid DNA nor the genomic DNA was treated with piperidine, since this step is unnecessary (4).

The primers used in this study were as follows. For plasmid $pASG_{2\mu}$, the bottom-strand primer (5'-ATCCACGCTATATACACGCCTGGC-3') anneals to top-strand sequences in the *CYC1* promoter from positions −236 to −212 with respect to the first codon. The $pGAL_{2\mu}$ primer (5'-CTAAAGTTGCCTGGCCA TCCACGC-3') anneals to the top strand of the *CYC1* promoter from positions −220 to −196 with respect to the first codon. The primers used for the coding and noncoding strands of plasmid $pGAL_{2\mu}$ were 5'-AACTGTATTATAAGTAA ATGCATG-3' and 5'-TGCCATATGATCATGTGTCGTCGC-3', respectively. For the integrating constructs, primers were designed that hybridized to sequences in both the *CYC1* promoter ($pASG_{int}$) and the *URA3* gene ($pGAL_{int}$), as well as in the *STE6* operator, in order to avoid background from the native yeast genes. For $pASG_{int}$, the primer used was 5'-CGGATCTGCTCGACGA GCGTGTAA-3'. The primer used for $pASG_{2\mu}$ yielded the same results. For $pGAL_{int}$, the primer used was 5'-TCAGTTATTACCCTCGACCTCGTCG-3'.

Isolation and analysis of chromatin. Chromatin was isolated from four strains (*MATα*, *matΔ*, *MATα gal4::LEU2*, and *matΔ gal4::LEU2*) containing promoter constructs $pGAL_{2\mu}$, $pGAL_{int}$, or $pASG_{2\mu}$, according to the Nonidet P-40-permeabilized spheroplast method (22). Briefly, the cells were grown in the medium used for the β-galactosidase assays to an optical density (A_{660}) of 0.8, washed with 1 M sorbitol, and digested with 0.5 mg of Zymolyase T100 (ICN) per ml. Nuclei were washed and resuspended in buffer containing 1 M sorbitol, 50 mM NaCl, 10 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM β-mercaptoethanol, and 0.075% Nonidet P-40. The nuclei were digested for 5 min at 37°C with micrococcal nuclease (MNase) (Worthington Biochemical Corp.) concentrations ranging from 0 to 25 U/ml. DNA was purified by phenol extraction after digestion with proteinase K and RNase A. Naked DNA was prepared in this manner before MNase digestion with 7.5, 15, or 30 U/ml for 1 min at 37°C. Indirect end label analysis was used to determine the positions of nuclease-sensitive regions according to the method described by Thoma et al. (40). Chromatin and naked DNA were cut with a variety of restriction enzymes that cut either in the *lacZ* gene or in the *URA3* gene. The enzymes used that cut in *lacZ* (with the distance from the start of the **a**-specific gene operator in pGAL constructs or from the start of the Gal4-binding site in pASG constructs indicated in parentheses) were *Hpa*I (853 bp), *Dde*I (534 bp), and *Fsp*I (451 bp). The enzymes used that cut in *URA3* were *Stu*I (441 bp) and *Dde*I (160 bp). Probes were generated by PCR and varied in length from 50 to 238 bp.

RESULTS

α2 represses Gal4-activated promoters. The chromatin reorganization model for repression predicts that DNA near the operator should be less accessible to proteins than is naked DNA. To determine whether an α2-repressed promoter is accessible to Gal4, hybrid promoters containing a single Gal4-binding site and a single **a**-specific gene operator upstream of a *CYC1* β-galactosidase promoter fusion were constructed (Fig. 1). The Gal4-binding site was placed either upstream (pASG) or downstream (pGAL) of the **a**-specific gene operator with respect to the *CYC1* promoter. The plasmid names reflect the DNA element, either the Gal4-binding site or the **a**-specific gene operator, that lies adjacent to the *CYC1* promoter. Promoter constructs either were placed on multicopy 2μm yeast plasmids ($pASG_{2\mu}$ and $pGAL_{2\mu}$) or were integrated into the chromosome at the *URA3* locus ($pASG_{int}$ and $pGAL_{int}$). To assess whether these test promoters were activated by Gal4 and whether activated transcription could be repressed by α2, the constructs were transformed into the following four different cell types: cells containing both α2 and Gal4 (*MATα GAL4*), cells containing only Gal4 (*matΔ GAL4*) or only α2 (*MATα gal4::LEU2*), and cells lacking both proteins (*matΔ gal4::LEU2*). In the presence of galactose, the promoters are activated 10- to 130-fold by Gal4 (Fig. 1; compare values from *matΔ GAL4* cells with those from *matΔ gal4::LEU2* cells). Furthermore, α2 represses transcription approximately 800-fold relative to the activated level when the operator is positioned between the Gal4 site and the *CYC1* promoter ($pASG_{2\mu}$ and $pASG_{int}$) and about 30-fold when the operator is positioned upstream of the Gal4-binding site ($pGAL_{2\mu}$ and $pGAL_{int}$; compare expression from *MATα GAL4* cells with that from *matΔ GAL4* cells). These results indicate that α2 is capable of efficiently repressing activated transcription from these constructs. The fact that the repression is greater when the operator is between the Gal4-binding site and the promoter than when the operator is upstream of the Gal4-binding site is consistent with the behavior of the operator in other test constructs (19). The expression of the constructs in *matΔ gal4::LEU2* strains is presumably due to activation by the MCM1 protein bound to the **a**-specific gene operator (2, 20).

The level of repression of an a-specific gene correlates with the repression of the hybrid reporters. We wished to know whether the strong repression (20- to 800-fold) of the test promoters is comparable to that of a bona fide **a**-specific gene. To determine the magnitude of α2 repression of the **a**-specific gene *STE2*, we employed quantitative RNA PCR analysis (9).

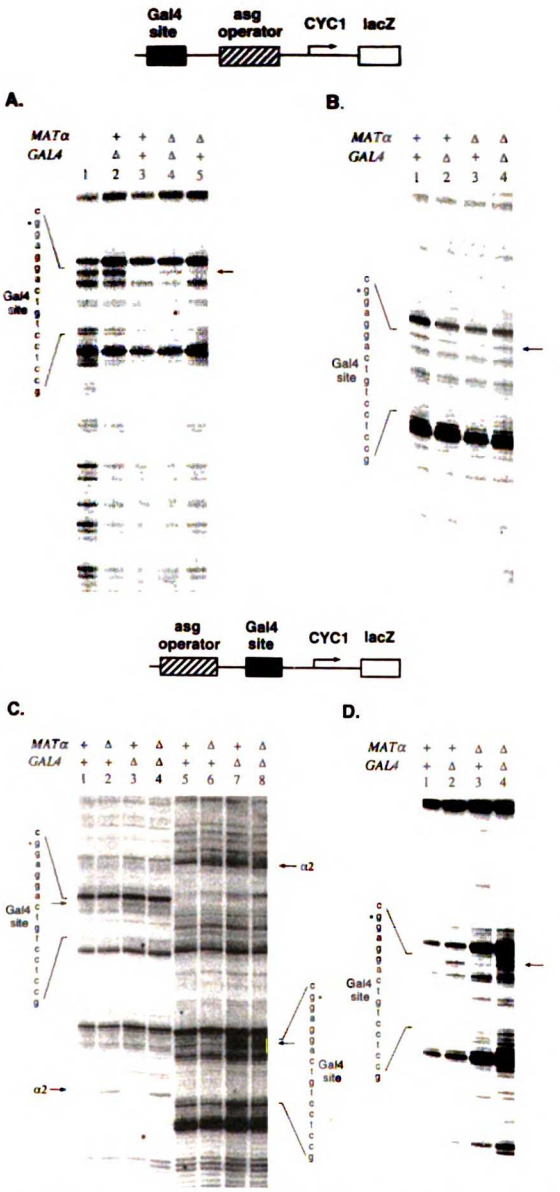


FIG. 2. Gal4 can occupy its site when the test constructs are repressed. Each panel shows the primer extensions from in vivo methylated DNA of the indicated test promoters in four different cell types: *MATα*, *matΔ*, *MATα gal4::LEU2*, and *matΔ gal4::LEU2*. The Gal4-binding site is indicated. The strong bands that bracket the Gal4-binding site are sequence-specific stops for *Taq* polymerase. The Gal4 footprint is clearly detected in *GAL4*$^+$ strains and is indicated by an arrow. The constructs are diagrammed over the appropriate panels. (A) Primer extension of the noncoding strand of promoter construct $pASG_{2\mu}$. Lane 1, extension products from in vitro-methylated plasmid DNA. Coding strand primer extension yields similar results (not shown). (B) Primer extension of the coding strand of promoter construct pASG integrated at *URA3*. (C) Primer extension of the coding strand (lanes 1 to 4) and the noncoding strand (lanes 5 to 8) of promoter construct $pGAL_{2\mu}$; (the α2 footprint is indicated by an arrow). (D) Primer extension of the noncoding strand of promoter construct pGAL integrated at *URA3*. asg, **a**-specific gene.

The results indicate that *STE2* transcription is repressed 200-fold in α cells relative to **a** cells (which lack α2), a result that is comparable to that observed in the test promoters, in which the α2 operator is located between the Gal4-binding site and the promoter (data not shown). This result indicates that the test promoters used in this study provide a legitimate model system in which to analyze α2 repression.

Gal4 can occupy its site when the test constructs are repressed. In principle, α2 repression of the test promoters could result either from interference with Gal4 DNA binding or from interference with a subsequent step in transcription initiation. In order to determine whether α2 interferes with Gal4 DNA binding in vivo, we performed dimethyl sulfate footprinting experiments on growing yeast cells. When bound to DNA, Gal4 protects a single guanine on each strand of its binding site from methylation by dimethyl sulfate (11). This protection can be seen in Fig. 2A by comparing the results from DNA isolated from strains that contain Gal4 (lanes 3 and 5) with those that lack it (lanes 2 and 4). In the case of constructs $pASG_{2\mu}$ and $pGAL_{int}$, a Gal4 footprint can be detected both in the activated state (*matΔ GAL4* cells) and in the repressed state (*MATα GAL4* cells) (compare lanes 3 and 5 in Fig. 2A and lanes 1 and 3 in Fig. 2D). For construct $pGAL_{2\mu}$, a clear Gal4 footprint is visible when the construct is active, and a weaker footprint is visible under repressed conditions (Fig. 2C; compare lanes 5 and 6). In the case of construct $pASG_{int}$, a Gal4 footprint is seen in *matΔ* cells but cannot be detected in α cells (Fig. 2B; compare lanes 1 and 3). In three of four of the test promoters (including the most strongly repressed), Gal4 occupies its binding site under conditions in which transcription is tightly repressed (*MATα GAL4* cells). These results indicate that α2 must repress transcription by some means other than preventing the DNA binding of activator proteins. We do not know the reason why Gal4 fails to occupy one of the repressed templates; however, the results obtained with the other three templates prove that repression can occur even though Gal4 is bound. We also note that the α2 footprint can be seen in these experiments (Fig. 2C, lanes 1, 3, 5, and 7, as indicated).

Nucleosomes are not positioned over test promoters. It has been observed that α2 bound to DNA positions nucleosomes adjacent to it, and it has been proposed that this positioning can contribute to transcriptional repression. In contrast to the behavior of α2, DNA-bound Gal4 is able to disrupt binding of the core histone particle both in vitro and in vivo (29, 46). To assess the role of nucleosome positioning in transcriptional repression of the test constructs used in this study, we mapped the distribution of nucleosomes over these constructs in both active and repressed states. Chromatin was isolated and digested with MNase, and the relevant regions of the DNA were displayed by indirect end labeling (40). Digestion patterns across the test promoter $pGAL_{2\mu}$ resembled those of the naked DNA controls (Fig. 3), indicating a lack of positioned nucleosomes even when Gal4 is absent (*MATα gal4::LEU2*). Moreover, the digestion patterns across test construct $pGAL_{2\mu}$ were not observably different in the presence or absence of α2, even though α2 had a dramatic effect on the expression of this construct. In the same chromatin preparations, positioned nucleosomes were seen across the *URA3* gene (in accordance with reference 3), which is located immediately upstream of the test promoter (Fig. 3; note the patterns of enhanced and protected bands in the chromatin preparations which are indicative of positioned nucleosomes [lanes 1 to 4] compared with naked DNA [lane 5]). This last observation indicates that the experiments shown in Fig. 3 are of sufficient resolution to detect positioned nucleosomes. Moreover, we detected positioned nucleosomes across the promoter of the **a**-specific gene

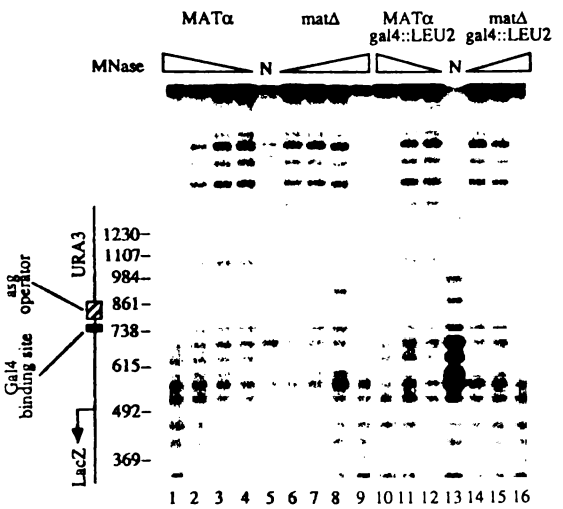


FIG. 3. MNase mapping of the $pGAL_{2\mu}$ promoter region. The indirect-end-labeling method was used to display the results of MNase digestion of chromatin isolated from four strains. Chromatin and naked DNA were cut with *Hpa*I after digestion with MNase. *Hpa*I cuts in the *lacZ* gene, 853 bp downstream of the beginning of the a-specific gene (asg) operator. The labeled primer used for indirect end labeling is 238 bp long, extending from the *Hpa*I site in *lacZ* toward the a-specific gene operator. Lanes 1 to 4, chromatin isolated from the *MATα* strain and digested with decreasing amounts of MNase (6, 3, 1.5, and 0.75 U/ml); lanes 6 to 9, chromatin isolated from the *matΔ* strain and digested with the same but increasing amounts of MNase; lanes 10 to 12, chromatin isolated from the *MATα gal4::LEU2* strain and digested with decreasing amounts of MNase (6, 1.5, and 0.75 U/ml); lanes 14 to 16, chromatin isolated from the *matΔ gal4::LEU2* strain and digested with increasing amounts of MNase (0.75, 1.5, and 3 U/ml). N, naked DNA digested with 15 U of MNase (lane 5) or 30 U of MNase (lane 13) per ml. Size markers in base pairs are indicated on the left, along with a diagram indicating the positions of the a-specific gene operator and the Gal4-binding site, as well as the *lacZ* and *URA3* genes.

STE2 (in accordance with the results described by Ganter et al. [8]), again suggesting that the failure to observe positioned nucleosomes across the artificial promoters is not due to a problem in detecting nucleosomes (data not shown). We repeated nucleosome mapping with the additional promoters ($pGAL_{int}$ and $pASG_{2\mu}$) and, in agreement with the results of Fig. 3, observed no evidence of positioned nucleosomes over any of the hybrid promoters in any of the four strains used in this work (data not shown).

DISCUSSION

This study demonstrates that α2 can efficiently repress transcription of a simple, artificial test promoter while still allowing access of the activator protein GAL4 to its binding site on the DNA. Thus, the α2 repressor must block transcription at a step subsequent to activator binding. On the surface, the presence of GAL4 on the DNA of the repressed promoters seems at odds with the proposal that α2 represses transcription by positioning nucleosomes around its binding site. On the basis of experiments performed in vivo and in vitro (29, 46), DNA-bound GAL4 appears to disrupt nucleosomes. One might have predicted that GAL4 would prevent the nucleosome positioning on the constructs described in this article. This idea was tested experimentally, and the results indicate a lack of specifically positioned nucleosomes regardless of whether GAL4 is present on the DNA.

The failure to detect positioned nucleosomes in the absence of Gal4 was initially surprising in light of the strong nucleosome positioning produced by α2 on native a-specific genes. However, the test promoter differs from those of a-specific genes in several ways. The TATA boxes and the transcription start site of the hybrid promoters are derived from the *CYC1* promoter. One feature of the *CYC1* promoter that might explain the absence of positioned nucleosomes is the constitutive binding of TBP to the TATA box of this promoter as proposed by Chen et al. (6). These investigators found that a derivative of the *CYC1 lacZ* promoter lacking upstream repressor or activator sites was free of positioned nucleosomes. Furthermore, in vivo footprinting indicated that TBP was bound to the TATA elements of this silent *CYC1 lacZ* promoter (also see reference 5). Our results could be explained by the model that TBP is bound to the TATA elements and prevents the *CYC1* promoter from being packaged in nucleosomes. With respect to TBP binding, the *CYC1* promoter may differ from other yeast promoters, including those of some a-specific genes. Despite this fact, the *CYC1* promoters used in this study were very strongly activated by Gal4 and were strongly repressed by α2, suggesting that the differences in initial TBP binding among promoters is relatively unimportant for regulation by these proteins. Finally, if TBP bound to the *CYC1* TATA elements does prevent nucleosomes from forming over this promoter, one might have predicted that a repressor that acts solely by nucleosome positioning would be unable to repress the *CYC1* promoter. As shown here and elsewhere (19, 21), α2 can tightly repress this promoter and the level of repression can be even higher than that of a bona fide a-specific gene.

If α2 does not repress transcription by controlling access of activator proteins to DNA, how does it work? Since α2 can repress basal transcription in vitro (13), it has been proposed that the α2 repression complex may act directly on the basal transcription machinery, interfering with a step in transcription initiation. In further support of this model is the discovery that components of the RNA polymerase II holoenzyme are required for efficient α2 repression (25, 27, 43). Direct interference with the basal transcription machinery seems an apt mechanism for a repressor such as α2 that must efficiently repress a large number of genes that utilize a variety of activator proteins.

ACKNOWLEDGMENTS

We thank Edward Giniger for his invaluable help both in formulating and motivating these experiments. We thank Kelly Komachi, Erin O'Shea, David Morgan, and especially Ramon Tabtiang for insightful comments on the manuscript. We thank Burkhard Braun and Rebecca Smith for advice and the rest of the Johnson laboratory for their support and useful discussions. We also thank Sharon Roth for advice and helpful comments.

This work was supported by a grant from the National Institutes of Health to A.D.J.

REFERENCES

1. **Axelrod, J. D., and J. Majors.** 1989. An improved method for photofootprinting yeast genes in vivo using Taq polymerase. Nucleic Acids Res. **17:** 171–183.
2. **Bender, A., and G. F. Sprague.** 1987. MATα1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. Cell **50:**681–691.
3. **Bernardi, F., M. Zatchej, and F. Thoma.** 1992. Species specific protein—DNA interactions may determine the chromatin units of genes in S. cerevisiae and in S. pombe. EMBO J. **11:**1177–1185.
4. **Brewer, A. C., P. J. Marsh, and R. K. Patient.** 1990. A simplified method for in vivo footprinting using DMS. Nucleic Acids Res. **18:**5574.
5. **Chatterjee, S., and K. Struhl.** 1995. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. Nature (London) **374:**820–822.

6. **Chen, J., M. Ding, and D. S. Pederson.** 1994. Binding of TFIID to the CYC1 TATA boxes in yeast occurs independently of upstream activating sequences. Proc. Natl. Acad. Sci. USA **91**:11909–11913.
7. **Dolan, J. W., and S. Fields.** 1991. Cell-type-specific transcription in yeast. Biochim. Biophys. Acta **1088**:155–169.
8. **Ganter, B., S. Tan, and T. J. Richmond.** 1993. Genomic footprinting of the promoter regions of STE2 and STE3 genes in the yeast Saccharomyces cerevisiae. J. Mol. Biol. **234**:975–987.
9. **Gilliland, G., S. Perrin, and H. F. Bunn.** 1990. Competitive PCR for quantitation of mRNA, p. 60–69. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, Calif.
10. **Giniger, E., and M. Ptashne.** 1988. Cooperative DNA binding of the yeast transcriptional activator GAL4. Proc. Natl. Acad. Sci. USA **85**:382–386.
11. **Giniger, E., S. M. Varnum, and M. Ptashne.** 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell **40**:767–774.
12. **Goutte, C., and A. D. Johnson.** 1988. a1 protein alters the DNA binding specificity of α2 repressor. Cell **52**:875–882.
13. **Herschbach, B. M., M. B. Arnaud, and A. D. Johnson.** 1994. Transcriptional repression directed by the yeast α2 protein in vitro. Nature (London) **370**:309–311.
14. **Herschbach, B. M., and A. D. Johnson.** 1993. Transcriptional repression in eukaryotes. Annu. Rev. Cell Biol. **9**:479–509.
15. **Herskowitz, I., J. Rine, and J. Strathern.** 1992. Mating-type determination and mating-type interconversion in Saccharomyces cerevisiae, p. 583–656. *In* E. W. Jones, J. R. Pringle, and J. R. Broach (ed.), The molecular and cellular biology of the yeast Saccharomyces. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
16. **Jackson, M. E.** 1991. Negative regulation of eukaryotic transcription. J. Cell Sci. **100**:1–7.
17. **Johnson, A. D.** 1992. A combinatorial regulatory circuit in budding yeast, p. 975–1006. *In* S. L. McKnight and K. R. Yamamoto (ed.), Transcriptional regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
18. **Johnson, A. D.** 1995. The price of repression. Cell **81**:655–658.
19. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MAT α2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell **42**:237–247.
20. **Keleher, C. A., C. Goutte, and A. D. Johnson.** 1988. The yeast cell-type-specific repressor α2 acts cooperatively with a non-cell-type-specific protein. Cell **53**:927–936.
21. **Keleher, C. A., M. J. Redd, J. Schultz, M. Carlson, and A. D. Johnson.** 1992. Ssn6-Tup1 is a general repressor of transcription in yeast. Cell **68**:709–719.
22. **Kent, N. A., L. E. Bird, and J. Mellor.** 1993. Chromatin analysis in yeast using NP-40 permeabilised sphaeroplasts. Nucleic Acids Res. **21**:4653–4654.
23. **Kladde, M. P., and R. T. Simpson.** 1994. Positioned nucleosomes inhibit Dam methylation in vivo. Proc. Natl. Acad. Sci. USA **91**:1361–1365.
24. **Komachi, K., M. J. Redd, and A. D. Johnson.** 1994. The WD repeats of Tup1 interact with the homeo domain protein α2. Genes Dev. **8**:2857–2867.
25. **Kuchin, S., P. Yeghiayan, and M. Carlson.** 1995. Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. Proc. Natl. Acad. Sci. USA **92**:4006–4010.
26. **Levine, M., and J. L. Manley.** 1989. Transcriptional repression of eukaryotic promoters. Cell **59**:405–408.
27. **Liao, S. M., J. Zhang, D. A. Jeffery, A. J. Koleske, C. M. Thompson, D. M. Chao, M. Viljoen, H. J. van Vuuren, and R. A. Young.** 1995. A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature (London) **374**:193–196.
28. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
29. **Morse, R. H.** 1993. Nucleosome disruption by transcription factor binding in yeast. Science **262**:1563–1566.
30. **Mueller, P. R., and B. Wold.** 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science **246**:780–786. (Erratum, **248**: 802, 1990.)
31. **Paranjape, S. M., R. T. Kamakaka, and J. T. Kadonaga.** 1994. Role of chromatin structure in the regulation of transcription by RNA polymerase II. Annu. Rev. Biochem. **63**:265–297.
32. **Patterton, H. G., and R. T. Simpson.** 1994. Nucleosomal location of the STE6 TATA box and Matα2p-mediated repression. Mol. Cell. Biol. **14**:4002–4010.
33. **Roth, S. Y.** 1995. Chromatin-mediated transcriptional repression in yeast. Curr. Opin. Genet. Dev. **5**:168–173.
34. **Roth, S. Y., A. Dean, and R. T. Simpson.** 1990. Yeast α2 repressor positions nucleosomes in TRP1/ARS1 chromatin. Mol. Cell. Biol. **10**:2247–2260.
35. **Shimizu, M., S. Y. Roth, C. Szent-Gyorgyi, and R. T. Simpson.** 1991. Nucleosomes are positioned with base pair precision adjacent to the α2 operator in Saccharomyces cerevisiae. EMBO J. **10**:3033–3041.
36. **Siliciano, P. G., and K. Tatchell.** 1984. Transcription and regulatory signals at the mating type locus in yeast. Cell **37**:969–978.
37. **Singh, J., and A. J. Klar.** 1992. Active genes in budding yeast display enhanced in vivo accessibility to foreign DNA methylases: a novel in vivo probe for chromatin structure of yeast. Genes Dev. **6**:186–196.
38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
39. **Tatchell, K., K. A. Nasmyth, B. D. Hall, C. Astell, and M. Smith.** 1981. In vitro mutation analysis of the mating-type locus in yeast. Cell **27**:25–35.
40. **Thoma, F., L. W. Bergman, and R. T. Simpson.** 1984. Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions. J. Mol. Biol. **177**:715–733.
41. **Treitel, M. A., and M. Carlson.** 1995. Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. Proc. Natl. Acad. Sci. USA **92**:3132–3136.
42. **Tzamarias, D., and K. Struhl.** 1995. Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 corepressor complex to differentially regulated promoters. Genes Dev. **9**:821–831.
43. **Wahi, M., and A. D. Johnson.** 1995. Identification of genes required for α2 repression in Saccharomyces cerevisiae. Genetics **140**:79–90.
44. **Webster, N., J. R. Jin, S. Green, M. Hollis, and P. Chambon.** 1988. The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of GAL4 trans-activator. Cell **52**:169–178.
45. **Wolffe, A. P.** 1994. Transcription: in tune with the histones. Cell **77**:13–16.
46. **Workman, J. L., and R. E. Kingston.** 1992. Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science **258**: 1780–1784.