

FUNCTIONAL SPECIFICITY AND DESIGN OF TRANSCRIPTION FACTORS:
THE OCT-1 HOMEODOMAIN

by

Joel Leonard Pomerantz

B.A. Brandeis University, Waltham, MA

Submitted to the

Department of Biology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy


at the

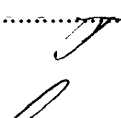
Massachusetts Institute of Technology

August, 1995

© Massachusetts Institute of Technology 1995

Λ Λ Λ

Signature of Author.....
 Department of Biology, August 11, 1995

Certified by.....
 Professor Phillip A. Sharp
Thesis Supervisor

Accepted by.....
 Professor Frank Solomon, Chairman
Departmental Committee on Graduate Studies

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

AUG 10 1995

LIBRARIES

1

Science

ABSTRACT

FUNCTIONAL SPECIFICITY AND DESIGN OF TRANSCRIPTION FACTORS: THE OCT-1 HOMEODOMAIN

by
Joel Leonard Pomerantz

Submitted to the Department of Biology on August 11, 1995
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

The development and differentiation of eukaryotic organisms depends upon transcriptional regulation exerted by homeodomain proteins; however, the mechanisms that determine the functional specificity of these factors remain largely undefined. The human Oct-1 homeodomain protein regulates Herpes Simplex Virus-1 (HSV) gene expression by participating in the formation of a multiprotein complex (C1 complex) that activates the HSV α (immediate early) genes. The C1 complex, composed of the viral α TIF protein, Oct-1, and the cellular C1 factor, forms on the α /IE element (5'-ATGCTAATGATATTCTTTGG-3'). A panel of Oct-1 homeodomain variants was generated and assayed for the ability to interact with α TIF and participate in C1 complex formation. The results indicated that in complex formation the homeodomain is recognized on the surface of helices 1 and 2 by the viral α TIF protein. The differing abilities of Oct-1 and the related Oct-2 protein to participate in complex formation correlated with a single amino acid difference in helix 1 of their homeodomains. This scenario models how protein-protein interactions at the surface of a DNA-bound homeodomain can determine functional specificity.

The Oct-1 homeodomain contacts DNA in the major groove through interactions which are common to many homeodomain factors. Asparagine 51 contacts an adenine and valine 47 contacts a thymine at conserved positions in a core homeodomain binding site (5'-TAATNN-3'). An Oct-1 POU domain bacterial expression library was generated in which residues 47 and 51 in the homeodomain were randomized. The library was screened with probes containing the wild-type octamer sequence (5'-ATGCAAAT-3'), and with probes containing substitutions at the recognized positions. Of many possibilities examined, the wild-type combination of residues and base-pairs provided interactions with the highest affinity and specificity. This suggests that the fold and the DNA-docking of the homeodomain constrain what residues can be used as determinants of DNA recognition.

Structure-based design was used to retarget the Oct-1 homeodomain to a novel sequence. Computer modeling predicted how the homeodomain might be fused to zinc fingers 1 and 2 from Zif268 to create a chimeric protein, ZFHD1. ZFHD1 displayed novel DNA-binding specificity in vitro, and, when fused to an activation domain, activated transcription in a sequence-specific manner in vivo. On the appropriate DNA element, ZFHD1 efficiently nucleated formation of a C1 complex in vitro and in vivo. This indicated that when targeted to DNA with high affinity and specificity, the homeodomain of Oct-1 mediates all the protein-protein interactions that are necessary to efficiently recruit α TIF and C1 factor into a functional enhancer complex.

Thesis Supervisor: Dr. Phillip A. Sharp, Salvador E. Luria Professor and Head,
Department of Biology

TABLE OF CONTENTS

Abstract	2
Dedication	4
Acknowledgments	5
Chapter I	
Introduction and Overview	6
Herpes Simplex Virus Gene Expression	8
α TIF	9
Oct-1	12
C1 factor	13
The C1 complex	14
The role of the Oct-1 homeodomain in C1 complex formation.....	15
Homeodomain Biology	17
Homeodomains and the body plan	18
Pit-1	22
$\alpha 1$ and $\alpha 2$	24
Transcription Factor Design	25
TBPm3	26
Selection of zinc finger variants by phage display	28
References	32
Chapter II	
Recognition of the Surface of a Homeodomain Protein	42
Chapter III	
Homeodomain Determinants of Major Groove Recognition	78
Chapter IV	
Structure-Based Design of Transcription Factors	107
Chapter V	
Analysis of Homeodomain Function by Structure-Based Design of a Transcription Factor	127
Epilogue	
Future Prospects for Designed Transcription Factors	151
Biographical Note	157

To my family

-

ACKNOWLEDGMENTS

To Phil Sharp, for providing me with the independence to learn what I could only have learned by myself, and the education in science that I could only have learned from Phil.

To Margarita Siafaca, for advice, for making everything work, for P-Day, for making the lab a home.

To David Baltimore, Carl Pabo, Bob Sauer, and Ranjan Sen, for serving on my thesis committee.

To Carl Pabo, for the education I didn't expect but for which I am grateful.

To Tom Kristie, for doing the work that made this thesis possible and for the tips on car painting and blood sacrifice.

To Marc Rothenberg, for turning me on to M.D. Ph.D. training, and getting me with the program.

To Ranjan Sen, for getting me started in molecular biology.

To Andrew "Sparkly" MacMillan, for the coffees, the tunes, the walks, the jokers and the perspective.

To Dan Chasman, for the biking tips, the life-outside-the-lab tips, the computer tips, and the Thanksgiving Keanu Reeves story.

To the people who welcomed me into the Sharp lab, Tom Kristie, Anna Gil, Rachel Meyers, Ben Shykind, Jorgen Kjems, Marc Timmers, Ken LeClair, and Jeff Parvin.

To Jae Kim, John Crispino, Lee Lim, Patrick McCaw, and Ben Shykind, the saner graduate students who populated Phil's lab with attitude, diversions, and man-eaters.

To the postdocs who helped transform the lab into a kinder and gentler one, Charles Query, Ben Blencowe, Karyn Cepek, Qiang Zhou, and Sean Harper.

To the Sharp lab technicians and their individual expertise, Bob Mulrenin, Sharon Jamison, Lana Parent, Courtney Wrenn, Robbyn Issner, and Yubin Qiu.

To the Housman, Jacks, Lees, Cochran, and Ruley labs, for flooding the Fifth Floor with personality.

To Biogen, Stockholm, Marathon, the Unibomber, Au Bon Pain, BPEC, P-Day, the darkroom construction, the 6th floor construction, 7.60, 7.02, the Whitehead cafeteria, the New Building pile drivers, Eric Nadler, Rebecca Scarr, James Brown, Tom Waits, the sequencing plate wrist, the fractured radial head, Ward, Amerigo, floor meeting, the Cheri Theatre, LaVerde's, retreats, the Charles River, Franz Kafka, and Albert Camus.

CHAPTER I

INTRODUCTION AND OVERVIEW

A fundamental characteristic of transcription factors is their modularity, the fact that they are composed of domains which are functionally and structurally independent. This was first discovered as the ability to cut and paste transcriptional activation domains onto heterologous DNA-binding domains (Brent and Ptashne, 1985), but domains have also been described which mediate other functions, including repression (e.g. Han and Manley, 1993), hetero- or homodimerization (e.g. Landschulz et al., 1988), and ligand binding (Evans, 1988). The function of a transcription factor is determined by the integration in one molecule of the separate functions of these modules. At present, there is no evidence that sequence-specific transcription factors catalyze a chemical reaction. Rather they function via intermolecular binding. The specificity of a transcription factor thus results from the molecular interactions specified at the surfaces of their domains.

The DNA-binding domain, or module, determines function in the simplest scenario by presenting a surface which docks against DNA and recognizes a particular sequence of base pairs. This interaction specifies which cis-elements in the genome the transcription factor can recognize and therefore which genes it will regulate. However, in many cases, the DNA-binding domain does more. This thesis describes the characterization of protein-DNA and protein-protein interactions through which the Oct-1 homeodomain achieves functional specificity. Four main themes emerge from the experiments presented in detail in the four succeeding chapters:

- 1) A DNA-binding domain can determine the functional specificity of a transcription factor through protein-protein interactions, a mechanism distinct from its DNA-binding specificity.

- 2) The fold of a DNA-binding domain and the manner in which it docks to DNA can constrain what amino acids are used as determinants of DNA recognition.

3) The modularity of transcription factors is a design tool. A structural understanding of domain-DNA interactions allows straightforward design of chimeric domains with novel specificities.

4) A designed transcription factor can be a research tool. The structure-based design of chimeric proteins allows a DNA-binding domain to be isolated for study from the rest of the protein from which it is derived, and the ability to change its DNA-binding specificity without introducing mutations at the protein-DNA interface.

The intention of this chapter is to place the experiments presented in the body of the thesis into a contextual framework in three relevant fields: 1) Herpes Simplex Virus gene expression; 2) the biology of the homeodomain family of regulators; 3) transcription factor design.

HERPES SIMPLEX VIRUS GENE EXPRESSION

The Herpes Simplex Virus (HSV)-1 is a 150kb double stranded DNA virus that infects a wide variety of cells in culture, replicating over the course of an 18 hour life cycle leading to the rapid destruction of the cells. In humans, after initiating a primary infection, the virus migrates through nerve axons to sensory ganglia, where it establishes latency. Reactivation of the virus occurs periodically over a lifetime in response to a variety of environmental factors (Roizman and Sears, 1991).

The HSV life cycle is controlled by a cascade of gene expression that begins upon infection with the transcription of the α or immediate early genes, the induction of which requires no new protein synthesis. The α genes encode regulatory proteins that transactivate the expression of β (early) genes, which encode proteins that are involved in

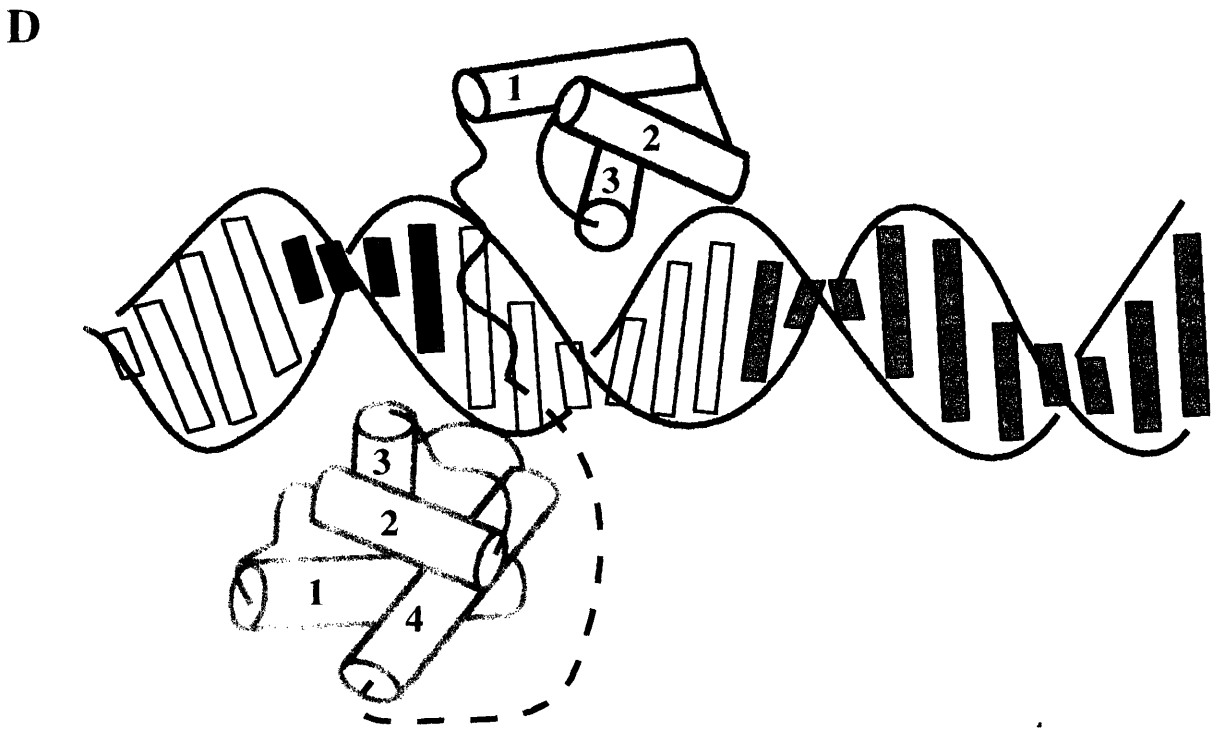
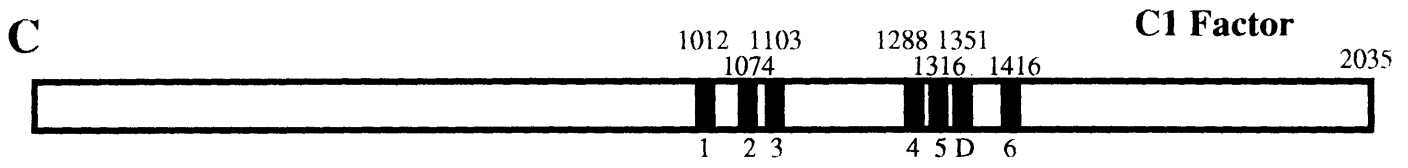
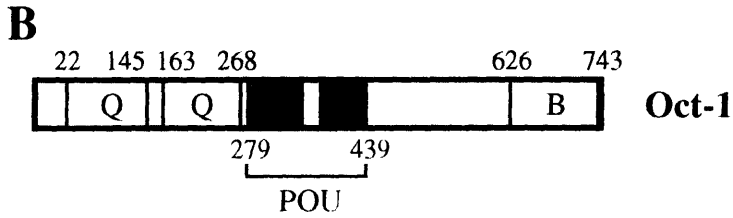
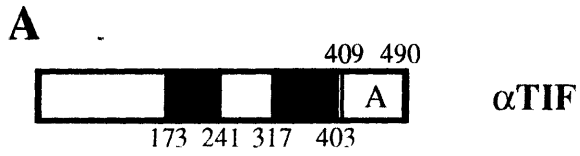
nucleic acid metabolism and are required for viral DNA synthesis. Following the onset of viral DNA replication the γ (late) genes, which encode the structural proteins of the virus, are transcribed.

Early studies demonstrated that the induction of α gene expression was dependent on the viral α TIF (α -trans-induction factor, also known as VP16, Vmw65, ICP25), a structural protein located in the tegument portion of the virus (between the capsid and the envelope) (Post et al., 1981; Campbell et al., 1984; Pellet et al 1985). After infection, α TIF enters the nucleus of the host cell and activates transcription through what was originally called the α TIC (α -trans-induction-cis site), an element found in one to three copies in the regulatory regions of the five α genes (Mackem and Roizman, 1982 a-c; Kristie and Roizman, 1984). This element is now more commonly referred to as the α /IE element or the TAATGARAT motif because of the central portion of its sequence consensus (5'-GYATGNTAATGARATTCYTTGNGGG-3'). Analysis of α TIF failed to demonstrate specific high affinity binding of the protein to the α /IE element (Marsden et al., 1987), and it was later found to associate only in the presence of cellular factors (Kristie and Roizman, 1987; Preston et al., 1988; O'Hare and Goding, 1988; Gerster and Roeder, 1988). One of these was identified as the previously described Oct-1 factor, and the other was purified and cloned on the basis of its role in α /IE association and is called the C1 factor. Oct-1, α TIF, and the C1 factor assemble into a complex on the α /IE element (C1 complex)(Figure 1). These factors and their role in C1 complex formation are briefly described below.

α TIF

α TIF is a 490 amino acid protein, best known for the extremely potent acidic transcriptional activation domain contained in its C-terminal 80 amino acids (Triezenberg et al., 1988; Sadowski et al., 1988). The activation domain is required for α /IE transactivation, but is dispensable for C1 complex formation. Two regions of the protein

Figure 1. Components and topology of the C1 complex. (A) A schematic representation of the viral α TIF protein. The regions shown in blue, when deleted, interfere with C1 complex formation. The C-terminal acidic activation domain is indicated (A). (B) A schematic representation of the Oct-1 protein. Glutamine rich regions of an activation domain in the N-terminal portion are indicated (Q). The activation domain that appears to be specialized for function at snRNA promoters is located at the C-terminus (B). The bipartite POU domain is composed of a POU-specific domain (residues 279-354, green), a homeodomain (residues 479-439, red), and a 24 residue linker in between. (C) A schematic representation of the C1 factor precursor protein showing the locations of the 20 amino acid repeats that are targets for proteolytic cleavage. The products of cleavage appear to remain associated (Wilson et al., 1993). The repeat core consensus is VCSNPPCETHETGTTN/HTATT. N-terminal sequencing of C1 factor subunits suggests that cleavage occurs between the underlined residues in repeats 2, 3, 5, and 6 (Kristie et al., 1993). The repeat indicated by a D contains a core sequence which diverges from the consensus. The numbers refer to the location of the first residue in the 20 amino acid core in each repeat. (D) A schematic representation of the C1 complex assembled on the α /IE element. The depicted interaction of the POU domain with the 5' portion of the element is based on the coordinates of Klemm et al. (1994). The POU-specific domain (green) and the homeodomain (red) are connected by a 24 residue linker that is not visible in the crystal structure and is therefore indicated by a dashed line. DNA subsites for the POU subdomains are indicated in the sequence of the α /IE element below in matching colors. The α TIF and C1 factors associate with the 3' portion of the element (shaded base-pairs) but their structures are unknown. α TIF is thought to contact the DNA in this region but it is not known whether the C1 factor makes any contact with the DNA.



5'-ATGCTAATGATATTCTTTGG-3'

are important for complex formation, one between residues 173 and 241, the other between 317 and 403 (Ace et al., 1988; Werstuck and Capone, 1989; Greaves and O'Hare, 1989, 1990). A short peptide comprising residues 360-367 has been shown to inhibit complex formation (Hayes and O'Hare, 1993), and Stern and Herr (1991) have demonstrated an interaction between the DNA-bound POU domain and a peptide containing residues 360-391. Independent DNA binding of α TIF to the TAATGARAT core of the α /IE element has been observed, but only at high protein concentrations (roughly 50 fold higher than is necessary to form the C1 complex) (Kristie and Sharp, 1990).

Oct-1

The Oct-1 transcription factor is a founding member of the POU domain family of factors (Herr et al., 1988), a subfamily of the larger homeodomain class of regulatory proteins. The protein is 743 amino acids long, composed of N- and C-terminal activation domains, and a central DNA-binding domain, the POU domain (Sturm et al., 1988). Oct-1 is ubiquitously expressed and has been implicated in the regulation of many cellular genes. An octamer element is critical for the ubiquitous expression of the small nuclear (sn)RNA genes (Ares et al., 1987; Bark et al., 1987; Carbon et al., 1987; Murphy et al., 1987), the cell-cycle-specific expression of the histone H2B gene (Sive et al., 1986; Fletcher et al., 1987; LaBella et al., 1988), and the tissue-specific expression of the interleukin-2 (Ullman et al., 1991) and immunoglobulin genes (Mizushima-Sugano and Roeder 1986; Staudt et al., 1986; Scheidereit et al., 1987; Wirth et al., 1987; LeBowitz et al., 1988; Muller et al., 1988; Gerster et al., 1987). Viral genomes also exploit the octamer element, which is found in the α /IE element of HSV, the SV40 enhancer (Davidson et al., 1986) and in the adenovirus origin of DNA replication (Pruijn et al., 1986, 1987; O'Neill et al., 1988).

Despite the implication of Oct-1 involvement in this wide spectrum of disparate regulatory scenarios, the activation domains of the protein appear to be specialized for function in the context of an snRNA promoter, and work only weakly in the context of an mRNA promoter (Tanaka et al., 1992). This suggests the possibility that other regions of the protein, namely the POU domain, are the functionally relevant portions in non-snRNA regulatory contexts. Indeed, the POU domain has been proven to be necessary and sufficient for activity at the origin of adenovirus DNA replication (Verrijzer et al. 1990a), and, as will be discussed, at the α /IE element. The POU domain of Oct-1 also has been shown to physically interact with several factors, including the glucocorticoid receptor (Kutroh et al., 1992), Pit-1 (Voss et al., 1991), TATA-Binding Protein (Zwilling et al., 1994), PTF (a regulator of snRNA genes; Murphy et al., 1992), and High Mobility Group protein-2 (Zwilling et al., 1995).

C1 factor

The C1 factor was identified biochemically as the cellular activity required for α TIF association with an Oct-1: α /IE element complex (Kristie et al., 1989; Xiao and Capone, 1990; Katan et al., 1990). The activity is present in extracts of insect cells, and every mammalian cell so far examined, suggesting a strong evolutionary conservation of function. The C1 factor was purified from HeLa nuclear extracts using a C1 complex formation assay and was found to consist of a group of polypeptides ranging in molecular weight from 68 to 155 kilodaltons (Kristie and Sharp, 1993; Wilson et al., 1993). Peptide sequencing of tryptic digests led to the isolation of a 8.2 kb cDNA which appears to encode a large precursor protein that is proteolytically processed (Wilson et al., 1993; Kristie et al., 1995). The precursor protein is 2035 amino acids in length and has no homology to any known sequence. The most remarkable feature of the protein is the presence of 7 repeats of a 20 amino acid motif in its central portion. These repeats appear to be targets for site-specific proteolytic cleavage that produces a heterogeneous set of N-

and C-terminal fragments which remain associated. The cellular function of the C1 factor is completely unknown.

The C1 complex

The 5' portion of the α /IE element is recognized by the POU domain of Oct-1, a bipartite DNA binding domain composed of a POU-specific domain (which binds the ATGC subsite) and a homeodomain (which binds the TAATGA subsite) (Kristie and Sharp, 1990; Verrijzer et al., 1990b). Oct-1 can associate with the element with high affinity in a stable protein-DNA complex, but neither α TIF nor the C1 factor display independent high affinity DNA binding. Evidence of the interaction of α TIF and C1 factors with the 3' portion of the element (ATGCTAATGATATTCITTGG) consists of the 3' extension of the Oct-1 footprint when α TIF and C1 factor are added to an Oct-1:DNA complex, and the effect of mutations in this 3' portion which abrogate complex formation without affecting Oct-1 affinity. The C1 factor associates with α TIF in solution, but neither factor appears to interact with Oct-1 in the absence of DNA (Kristie and Sharp, 1990).

The function of the C1 complex appears to be to tether the extremely potent activation domain of α TIF to the α /IE element only in the presence of the Oct-1 and C1 factors. Why does HSV set up such a scenario? It is clear that the ensemble of protein-protein and protein-DNA interactions in the C1 complex make α TIF tethering highly sequence specific for the α /IE element. The requirement for both the 5' portion of the element, an octamer site homolog, and the 3' portion, which is recognized by α TIF and possibly the C1 factor, may ensure that α TIF-mediated activation will only occur on viral α promoters and not at cellular promoters which may individually be recognized by octamer binding proteins or other cellular proteins that might act in concert with the C1 factor. Interestingly, Oct-1 DNA-binding activity has been shown to be regulated by phosphorylation in a cell-cycle-dependent manner. Phosphorylation occurs at the onset

of M phase at a residue (serine 7) in the N-terminal arm of the homeodomain, inactivating DNA-binding by the POU domain (Segil et al., 1991; Roberts et al., 1991). Perhaps the dependence of C1 complex formation on Oct-1 DNA-binding couples the cascade of viral gene expression to the cell cycle progression of the infected cell.

Most intriguing is to consider the dependence of C1 complex formation on the cellular C1 factor. In this light, it is striking to compare α TIF to the E1A gene product of adenovirus, another protein which transactivates viral gene expression but must rely upon protein-protein interactions with cellular factors to associate with a promoter. The study of E1A-associated proteins has yielded the characterization of a number of proteins which are fundamentally important for cell-cycle regulation and proliferation (Dyson and Harlow, 1992). Among those that directly bind to E1A are the retinoblastoma tumor suppressor pRb, and two related proteins p107 and p130. It is thought that binding of E1A to these cellular proteins inactivates their function, allowing progression of the cell into S phase. E1A association is independent of DNA-binding, as is the association of α TIF with the C1 factor. The conservation of C1 factor activity in insect and mammalian cells suggests a fundamental role for this protein in cellular function. The most exciting possibility is that the C1 factor represents a viral target as interesting and important as those targeted by adenovirus through E1A.

The role of the Oct-1 homeodomain in C1 complex formation

The availability of cDNA clones of Oct-1 allowed the dissection of which regions of the protein are necessary for C1 complex formation. Initial studies demonstrated that the POU domain alone possessed equivalent DNA-binding affinity and efficiency at complex formation as full length Oct-1 (Kristie et al., 1989; Stern et al., 1989). The homeodomain portion of the POU domain alone could nucleate C1 complex formation on the α /IE element, but only at high protein concentrations (100 fold higher than was necessary for the intact POU domain) (Kristie and Sharp, 1990). A third important

observation was that Oct-2, a related POU domain protein, formed the C1 complex with a 100-fold lower efficiency, despite a high homology to Oct-1 in the POU domain and an indistinguishable DNA binding specificity (Kristie et al., 1989). Finally, it was not known which component of the complex was responsible for the specific interaction with Oct-1.

The experiments described in Chapter 2 established that the surface of the DNA-bound Oct-1 homeodomain is recognized in C1 complex formation by the viral α TIF factor. The determinants for the interaction lie on homeodomain helices 1 and 2, and a single amino acid difference between Oct-1 and Oct-2 in helix 1 of the homeodomain accounts for their differing abilities to cooperatively interact with α TIF and form the C1 complex. This amino acid does not affect DNA-binding affinity or specificity. The experiments presented in Chapter 5 established that the 60 amino acid Oct-1 homeodomain contains all of the determinants responsible for the Oct-1-mediated protein-protein interactions that are required for complex formation. A designed transcription factor that can target the homeodomain to DNA with high affinity and specificity in the absence of the rest of the Oct-1 protein can nucleate formation of the C1 complex with an efficiency equivalent to the intact POU domain.

The differing abilities of Oct-1 and Oct-2 to participate in C1 complex formation provides a paradigm that explains how two transcription factors with extremely similar DNA-binding specificities could mediate distinct functional effects, and how a single cis-element could be important for the disparate regulation of many genes. It is possible that cellular analogs or homologues of α TIF associate with Oct-1 or Oct-2 to confer temporally or spatially restricted activity on the octamer element in the context of a particular promoter. There is evidence that this is the case for some promoters. The unique activity of the immunoglobulin octamer element in B cells has recently been attributed to the existence of the B-cell specific protein OCA-B (OBF-1, Bob-1) (Luo et

al., 1992; Gstaiger et al., 1995; Strubin et al., 1995). OCA-B associates with the POU domain of either Oct-1 or Oct-2 and contains an activation domain. OCA-B does not appear to have independent DNA-binding activity, and association with OCA-B does not affect the off-rate of Oct-1. The transcriptional induction of the interleukin-2 gene in response to antigenic stimulation of T cells is mediated by an antigen-receptor response element (ARRE-1) in the IL-2 enhancer. Oct-1 and an activity that is induced by activation, OAP⁴⁰ (octamer-associated protein 40), associate with the element (Ullman et al., 1991). OAP⁴⁰ possesses sequence specificity for the 5' portion of the ARRE-1 and decreases the off-rate of Oct-1. The factor has been identified as a mixture of Jun D and c-Jun factors (Ullman et al., 1993). It has yet to be determined which residues of Oct-1 mediate the interaction with OCA-B or OAP⁴⁰, but it would not be surprising if the homeodomain were involved.

In its role in HSV α /IE gene expression, the Oct-1 homeodomain can be thought of as a module with two surfaces, one that docks against DNA, the other which presents determinants for protein-protein interactions. This may be a general picture of how homeodomains function in other systems. As described in the following section, some examples of homeodomain function have discrete biochemical parallels with the HSV scenario, while others appear as though they will.

HOMEODOMAIN BIOLOGY

Homeodomain proteins comprise a large family of "master regulatory" proteins that control cell-type specification and organization of the body plan in eukaryotic organisms as evolutionarily distant as yeast and humans (Gehring, 1987; Scott et al., 1989; Affolter et al., 1990; Kenyon, 1994; Lawrence and Morata, 1994; Krumlauf, 1994). They are called "master regulatory" because in general, they are thought to act as transcription factors epistatic to cascades of gene expression which ultimately produce

the structural gene products that differentiate one cell type from another, or one region of the organism from adjacent tissues. Presence of the 60 amino acid homeodomain DNA-binding module defines the family, which includes hundreds of proteins that can be grouped into subfamilies by homology within the homeodomain.

The structure and DNA-binding properties of homeodomains have been vigorously studied. NMR and crystal structures of homeodomain proteins, most of them in the presence of DNA, have demonstrated that even minimally related homeodomains adopt a conserved structure and mode of docking to DNA (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991; Billeter et al., 1993; Klemm et al., 1994) . The domain is composed of an N-terminal arm which contacts DNA in the minor groove, and three α helices, the third of which makes base-specific contacts in the major groove. Residues throughout the domain contact the sugar-phosphate backbone of DNA. Many homeodomains display similar DNA sequence specificities, and bind to a sequence of the form 5'-TAATNN-3'. The first two base pairs are recognized in the minor groove (TAATNN), and the last four in the major groove (TAATNN) (Laughon, 1991). There are several classes of homeodomain factors which contain another DNA binding domain in addition to the homeodomain in the same protein. A given protein may contain a POU, paired, cut, Lim, or C2H2 zinc finger domain, or even other homeodomains (Laughon, 1991).

Homeodomains and the body plan

Homeodomain proteins were first discovered and perhaps, are best known, as the molecular basis for homeotic mutations, i.e. mutations which cause the transformation of one body part into another. The genes encoding these proteins are found in clusters in the genomes of nematodes, arthropods, and vertebrates (Kenyon, 1994). In *Drosophila*, eight genes comprise two complexes, the *Antennapedia* complex (ANT-C) and bithorax complex (BX-C). *C. elegans* contains a single cluster with four genes, while mice and

humans each possess four clusters containing a total of 38 genes. Based on sequence homology, many individual genes in each cluster can be grouped across species into 13 paralogous groups (Krumlauf, 1994).

In flies, the genes in the ANT-C and BX-C complexes are referred to as homeotic selector genes. These genes are responsible for the development of parasegments 1 through 14 in the embryo, which ultimately give rise to the posterior head, thorax and abdomen of the fly (Lawrence and Morata, 1994). In vertebrates, the genes are referred to in a *Hox* nomenclature, and mutations in *Hox* genes affect many tissues, including the axial skeleton, neural crest, central nervous system, and limbs (Krumlauf, 1994). The functional relationship between vertebrate and invertebrate homeodomain proteins has been highlighted by the observation that some mammalian *Hox* genes and their *Drosophila* homologues induce similar phenotypes when expressed ectopically in *Drosophila* embryos (Malicki et al., 1990; McGinnis et al., 1990).

The physical order of the genes in the cluster, 5' to 3', correlates with the expression pattern of the genes along the anterior-posterior axis, and with the temporal order of expression, from early to late in embryonic development. The genes have overlapping regions of expression, and each gene results in a particular phenotype when mutated or expressed ectopically. A key feature of the homeotic selector genes is that the effect of mutation or ectopic expression of a particular gene is dependent on what other homeotic genes are expressed in the particular region of the embryo affected. In both flies and mice, there appears to be a functional dominance of posterior genes over anterior genes expressed in the same domain, a phenomenon referred to as phenotypic suppression or posterior prevalence (Morata, 1993).

There are few precise molecular explanations for how these homeodomain proteins exert their specific actions (Hayashi and Scott, 1990; Kornberg, 1993). Some have begun to address this issue by first defining the domains of each protein which are important in an ectopic (ubiquitous) expression assay. This has been done by taking two

homeodomain proteins that demonstrate different phenotypes in the assay and observing the effects of swapping domains and amino acid differences. In several studies, the homeodomain has proven to be critical for functional specificity, as well as other regions C-terminal or N-terminal to the homeodomain (Kuziora and McGinnis, 1989; Gibson et al., 1990; Malicki et al., 1990; Mann and Hogness, 1990; McGinnis et al., 1990; Furukubo-Tokunaga et al., 1992).

Since the homeodomain serves as the DNA-binding domain for these proteins it is not surprising that this region would be critical for the specificity of function. The DNA-binding specificity of the domain might simply determine which downstream genes the protein will regulate, ultimately leading to a specific phenotype. However, many homeodomain proteins have extremely similar DNA-binding specificities (Laughon, 1991). All of the homeodomains used in the specificity studies are highly homologous in the third helix of the homeodomain which recognizes base-pairs in the major groove. Residues in this helix that correspond to those shown in the homeodomain:DNA crystal structures to be contacting DNA are identical in these proteins. Thus, it is likely that a mechanism other than intrinsic DNA-binding specificity is responsible for the different phenotypic effects of these proteins.

Four studies have performed swaps within the homeodomain to delineate the amino acids that are important for functional specificity. In the study that examined differences between Deformed (Dfd) and Ultrabithorax (Ubx) homeodomains (Lin and McGinnis, 1992), homeodomain residues -1, 1, 2, 4, 6, and 7 in the N-terminal arm were shown to be critical for determining phenotype. Swapping these residues appeared to have no effects on in vitro DNA-binding activity that correlated with in vivo activity. The four differences between Antennapedia (Antp) and Sex combs reduced (Scr) homeodomains, at residues 1, 4, 6, and 7, appeared to account for their differing functional specificities in two independent studies (Zeng et al., 1993; Furukubo-Tokunaga et al., 1993). Finally, in a study that examined the differences between Ubx

and Antp, residues -1, 2, 22, 24, 56, and the region C-terminal to the homeodomain were shown to be important (Chan and Mann, 1993).

Since the N-terminal arm makes contacts to DNA in the minor groove, the recurrence of N-terminal arm residues in these studies suggests the possibility that the DNA-binding properties of the homeodomain may contribute to functional specificity. Some studies have detected slight differences in DNA-binding affinity or specificity between homeotic selector gene products (Ekker et al., 1992, 1994). However, most of the critical residues in these biological specificity studies map in the three-dimensional structure to surfaces which would be accessible for the interaction with other proteins when the homeodomain binds DNA. Protein-protein interactions involving these residues could affect target specificity or effector function. As has been discussed, protein-protein interactions determine the differing abilities of Oct-1 and Oct-2 to participate in the C1 complex, and furthermore, even a single amino acid difference between the two proteins is sufficient to determine functional specificity for this regulatory event.

The elucidation of whether protein-protein interactions are as important for *Drosophila* homeodomains must wait in most cases until more of the target genes have been identified. However, the characterization of the functional differences between Ubx and Antp has resulted in the identification of extradenticle (exd) as a protein that differentially interacts with these homeodomain proteins to affect their target specificity. The product of the *extradenticle* gene had been thought to act as a "cofactor" for homeodomain function because mutations in this gene produced homeotic transformations in thoracic and abdominal segments without altering the expression patterns of homeotic selector genes (Peifer and Wieschaus, 1990). Two studies (Chan et al., 1994; van Dijk and Murre, 1994) demonstrated that exd selectively interacts with Ubx, as compared to Antp, in cooperative binding to the enhancer of the *decapentaplegic* gene. The interaction between exd and Ubx involved determinants of Ubx that were

shown to be critical in the ectopic expression assay. One of the critical determinants in Ubx was at homeodomain position 22 (numbering scheme of Qian et al., 1989). The critical difference between Oct-1 and Oct-2 in C1 complex formation lies at position 22.

Interestingly, *exd* is a divergent homeodomain protein, related in sequence to the $\alpha 1$ and $\alpha 2$ homeodomain proteins of yeast. In addition to Ubx, *exd* also cooperatively interacts with abdominal-A and engrailed homeodomain proteins, through regions N-terminal to the homeodomain (for Ubx and *abd-A*), and C-terminal to the homeodomain (*engrailed*) (Van Dijk and Murre, 1994). The existence of *exd* homologs in humans suggests that the cooperative interaction between homeodomains may be a broadly conserved mechanism for determining the specificity of homeodomain action.

Although homeodomain proteins may be best known as critical players in the morphogenesis of animals, many homeodomain proteins control cell-type-specific functions. Two examples, one from human disease, the other from yeast, are illustrative. The first example, Pit-1, is pertinent to the discussion of the role of the Oct-1 in HSV gene regulation because it is a well-characterized POU domain family member and because like Oct-1, mutations have been found in the DNA-binding domain that alter effector function without compromising DNA binding. The second example, that of the $\alpha 2$ and $\alpha 1$ homeodomain proteins, is relevant because the system is one of the best examples of how protein-protein interactions with homeodomain proteins can enhance DNA target specificity, and because the homeodomain surface is important for one of these interactions.

Pit-1

The Pit-1 transcription factor is critical for anterior pituitary development (Voss and Rosenfeld, 1992). As is the case for many homeodomain proteins, natural mutations have been observed, leading to the description of a specific overt phenotype. In humans,

patients with combined pituitary hormone deficiency (CPHD) suffer from a spectrum of symptoms that often include irreversible mental and growth retardation. Some patients have been found to have alleles of Pit-1 with mutations in the POU domain. One of these is a nonsense mutation in the middle of the POU domain (Tatsumi et al., 1992). Another causes a substitution of tryptophan for arginine at position 58 in the homeodomain, which does not interfere with DNA binding but does abrogate the ability of the protein to transactivate in a tissue culture assay (Radovick et al., 1992). These patients have a complete absence of three cell types in the anterior pituitary: somatotrophs, which produce growth hormone, lactotrophs, which produce prolactin, and thyrotrophs, which produce thyroid-stimulating hormone. The same hormone deficiencies are observed in the Snell mouse dwarf mutation *dw*, which is caused by a mutation at homeodomain position 48 from tryptophan to cysteine, which destroys the ability of the protein to bind DNA (Li et al., 1990). Binding sites for Pit-1 are found in the promoter regions of the prolactin and growth hormone genes, and the expression of the growth hormone releasing factor receptor, which controls somatotroph proliferation, is reduced in Pit-1 deficient animals (Lin et al., 1992).

A distinct Pit-1 mutation has been discovered in two Dutch families. Affected (homozygous) individuals exhibit hypopituitarism without hypoplasia and possess an allele of Pit-1 which contains a proline for alanine substitute in helix 2 of the POU-specific domain (Pfaffle et al., 1992). This substitution does not affect DNA binding but does reduce the transactivation function. Thus, different mutations in the POU domain can lead to different phenotypes or clinical states. The Pit-1 mutations which do not affect DNA-binding are located on solvent-exposed surfaces of either the POU-specific domain or the homeodomain. There must be factors which recognize these residues and are required for transactivation in certain, perhaps as yet undefined, promoter contexts. The recognition of the surface of the Oct-1 homeodomain in C1 complex formation may provide a relevant model for how different mutations in Pit-1 might differentially affect

the expression of downstream genes that are critical for the development and function of the anterior pituitary.

$\alpha 2$ and $a 1$

Cell type-determination in *S. cerevisiae* is dependent on two homeodomain proteins, $\alpha 2$ and $a 1$ (Johnson, 1992). $\alpha 2$ is expressed in α cells, $a 1$ in a cells, and both are expressed in the diploid a/α cell. $\alpha 2$ binds cooperatively with the MCM1 protein at operators for a -specific genes (*asg*), leading to their repression in α and a/α cells. In a/α cells, $\alpha 2$ and $a 1$ bind cooperatively to the operators of haploid-specific genes (*hsg*), causing their repression.

In the recognition of the *asg* operator, MCM1 binds as a dimer in between two binding subsites for $\alpha 2$. In the absence of MCM1, $\alpha 2$ can bind as a dimer to artificial operators containing subsites in several different spacings and orientations (Smith and Johnson, 1992). It appears that cooperative interactions between $\alpha 2$ and MCM1 can occur only with the specific spacing and orientation of the subsites in the *asg* operator. The cooperative interaction with MCM1, thus enhances the target specificity of $\alpha 2$. The interaction involves a "hinge" region of $\alpha 2$ which is N-terminal to the homeodomain (Vershon and Johnson, 1993).

A similar cooperative interaction between $\alpha 2$ and $a 1$ determines target specificity of the two proteins for *hsg* operators (Goutte and Johnson, 1988; Mak and Johnson, 1993). Independent binding of $\alpha 2$ to the *hsg* operator is very weak, independent $a 1$ binding is undetectable. When the two proteins heterodimerize on DNA, the region of $\alpha 2$ C-terminal to the homeodomain (" $\alpha 2$ tail") undergoes a transition from disorder into a helical conformation (Phillips et al., 1994). This helix interacts with the solvent exposed surface of helices 1 and 2 of the $a 1$ homeodomain (T. Li and C. Wolberger, personal communication).

Thus, the efficient targeting of the $\alpha 2$ homeodomain protein to a-specific genes and haploid-specific genes depends upon cooperative protein-protein interactions with MCM1 and the $\alpha 1$ homeodomain. The interaction with $\alpha 1$ is similar to the α TIF-Oct-1 interaction because in both cases the surfaces of helices 1 and 2 are contacted. It would be interesting to compare the surfaces of $\alpha 1$ and Oct-1 homeodomains that are recognized by the $\alpha 2$ tail, and α TIF, respectively. Perhaps a structural transition in α TIF, concomitant with C1 complex formation, parallels the transition of the $\alpha 2$ tail as it interacts with $\alpha 1$. If this transition is triggered by binding of α TIF to the C1 factor, it would provide a satisfying explanation for the physical requirement for the C1 factor in complex assembly.

TRANSCRIPTION FACTOR DESIGN

The recognition of the Oct-1 homeodomain in C1 complex formation suggested the possibility that the surface of the Oct-1 homeodomain was recognized as a part of the regulation of cellular genes which rely on the octamer element. One way to address this experimentally would be to set up a cell-culture cotransfection system with an array of reporter vectors containing the appropriate enhancer or promoter regions of octamer-regulated genes. If one could establish an assay in which stimulation of reporter gene expression was dependent on the cotransfection of an Oct-1-expressing plasmid, one could assay the effects of mutations on the surface of the homeodomain in a straightforward manner. Such a system would allow the survey of many cellular promoters each of which might require a particular cell type for expression.

A barrier to this approach is the ubiquitous expression of Oct-1 in most, if not all human cell lines. Any transfected Oct-1 variant would have to compete for binding to the reporter with the wild-type, endogenous protein, which would already be present when the reporter vector entered the cell. This notion led to the attempts described in Chapters

3 and 4 to develop a variant of Oct-1 which could target the homeodomain to a DNA sequence that the wild-type Oct-1 could not recognize. This novel sequence would be inserted in the reporter vectors in place of the octamer element, and together with the altered-specificity Oct-1 variant, the system would be independent of endogenous octamer-binding activity. Two approaches were followed to obtain the altered-specificity variant. The first relied upon the assumption that specificity could be changed by the mutation of residues that mediate base-specific contacts. The second approach tested a structure-based strategy to design a chimeric domain composed of the Oct-1 homeodomain fused to two zinc fingers. It is useful to consider these efforts in the light of related experiments conducted to change DNA-binding specificity in other systems.

The attempt to change the DNA-binding specificity of a transcription factor may stem from several motives. These may include the desire to identify the residues in the protein which contact DNA, the evaluation of the flexibility of a domain for recognizing different sequences, and the development of proteins which will recognize a particular desired sequence for practical application in biological research or gene therapy. Both genetic and biochemical approaches have been fruitful (e.g. Youderian et al., 1983; Wharton and Ptashne, 1985; Ebright et al., 1987; Tzamarias et al., 1992; Huang et al., 1994). This section will discuss two examples. The first, the TATA-binding protein (TBP) m3 mutant, is one of the best examples of the use of an altered-specificity variant as a research tool. The second, the selection of zinc finger variants by phage display, may be the most powerful method for the development of proteins with desired sequence specificities.

TBPm3

A transcription factor variant with a novel specificity is perhaps most useful in an *in vivo* context where the wild-type variant is ubiquitous and necessary for the viability of the cell. The TATA-binding protein may represent the best example of an

indispensable transcription factor that is present in any in vivo eukaryotic context. The protein is necessary for transcription by polymerases I, II and III, and residues on the surface of the protein mediate interactions with other general transcription factors, TATA binding protein associated factors (TAFs), and sequence-specific transcription factors (Hernandez, 1993). Strubin and Struhl (1992) used a genetic selection to isolate a variant of TBP that could be manipulated in yeast without concern about cell viability. They isolated a mutant, TBPm3, that would bind to the sequence TGTA AAA, which is not bound efficiently by the wild-type TBP. The activity of TBPm3 could be monitored on an artificial promoter containing a TGTA AAA TATA box, while the wild-type endogenous TBP would remain in the background to support cell-viability but not interact with the artificial reporter. The altered-specificity mutations in this yeast protein also worked in the context of TBPs from other organisms, affording the opportunity to assay the functional differences between TBPs from different species. Although human TBP could not replace yeast TBP for cell viability, an altered-specificity human TBP was able to support Pol II transcription in yeast and respond to the GCN4 activator. This observation led to experiments that showed that human TBP is incompatible for transcription in yeast at Pol I, Pol II, and TATA-less Pol II promoters (Cormack et al., 1994). These studies allowed critical residues that differ between yeast and human TBPs to be mapped. Altered-specificity TBPs have also been used in plant (Heard et al., 1993) and human cells (Tansey et al., 1994). In the latter study, mutagenesis of TBP demonstrated that many regions on the surface of the protein participate in the response to activators of Pol II transcription. Finally, the altered-specificity yeast variant has been useful for experiments suggesting that the association of TBP with the promoter can be a rate-limiting step in the transcription initiation pathway in vivo, and that a sequence-specific transcription factor can stimulate transcription by enhancing TBP association with the TATA box (Klein and Struhl, 1994; Chatterjee and Struhl, 1995; Klages and Strubin, 1995).

TBPm3 was isolated without knowledge of the structure of TBP and without a firm idea of which residues in the protein mediated DNA-recognition. If such information is available for a particular domain, an approach becomes feasible in which the residues that contact DNA are randomized and a selection or screen is performed. One of the systems that applies this strategy and promises to deliver proteins with many applications involves the selection of zinc finger variants by phage display.

Selection of zinc finger variants by phage display

The phage display technology has offered the ability to select among up to a billion variants of a DNA-binding domain for one that has a desired sequence specificity. The essential feature of the system is that it allows a protein variant to be assayed *in vitro* for binding activity while it is physically associated with the DNA that encodes it (Smith and Scott, 1993). The domain being studied is fused to one of the proteins that is incorporated into the coat of a filamentous phage. When the phage is assembled, the DNA packaged into the phage will contain the gene for this fusion, either in the phage genome or in a phagemid. Phage are isolated for binding activity on solid supports containing the immobilized target DNA sequence, and the retained phage are eluted and amplified by infection of *E. coli*.

Several laboratories have applied this system to C2H2-type zinc fingers (Rebar and Pabo, 1994; Jamieson et al., 1994; Choo and Klug, 1994ab). Crystallographic analyses have defined the structure of zinc finger domains, their arrangement when bound to DNA, and the residues in the zinc finger which are responsible for base-specific interactions in the major groove (Pavletich and Pabo, 1991, 1993, Fairall et al., 1993). Using phage display, several studies have been able to select variants of the three finger domain of Zif268 (Rebar and Pabo, 1994; Jamieson et al., 1994; Choo and Klug, 1994ab). Each of the three fingers in this domain recognizes a triplet of base pairs in the 9 base-pair binding site. In these experiments, one of the three fingers has been

randomized in the context of the other wild-type fingers, and variants have been obtained which bind to a wide variety of triplets of base-pairs.

The modularity of the zinc finger domain has offered the opportunity to covalently link a set of three selected zinc fingers into a new composite domain. In this manner, Choo et al. (1994) have developed a protein that binds in a sequence-specific manner to a model target for therapeutic intervention, a BCR-ABL cDNA. The BCR-ABL gene product encoded in the cDNA corresponds to that which is produced in acute lymphoblastic leukemia as a result of a chromosomal translocation. In a model system, the stable transfection of this BCR-ABL cDNA confers IL-3 independence on a murine cell line in tissue culture. When this cell line was transiently transfected with a plasmid encoding the designed protein, IL-3 dependence was restored. The designed factor presumably inhibits transcription of the BCR-ABL cDNA by blocking elongation of RNA polymerase, although this has not been proven. Although it relies on a model system, this experiment highlights the potential therapeutic applications of proteins designed to bind to specific sequences.

Among the different classes of DNA-binding domains, zinc fingers may possess the widest repertoire for sequence recognition. Naturally occurring zinc fingers recognize a wide variety of sequences, and crystallographic analyses demonstrate that zinc fingers may dock to DNA in different ways (Pavletich and Pabo, 1993; Fairall et al., 1993). Other DNA-binding domains may not be as adaptable to different binding sites. Several families of transcription factors appear to have evolved to recognize a particular sequence motif, such as the basic-helix-loop helix (CANNTG) (Murre and Baltimore, 1992) and nuclear hormone receptor (AGNNCA, half site) (Evans, 1988) families.

The first attempt to develop an altered-specificity Oct-1 variant became an investigation of the plasticity of the Oct-1 homeodomain for recognizing different base-pairs in the major groove. The experiment, described in detail in Chapter 3, focused on residues 47 and 51, and involved an approach of randomization followed by biochemical

screening for binding activity. Valine 47 and asparagine 51 in the Oct-1 homeodomain mediate interactions that are highly conserved in the homeodomain family of factors. Valine 47 contacts the thymine that corresponds to the final residue in the TAAT core of many homeodomain binding sites (TAATTNN) (Klemm et al., 1994). Asparagine 51 is one of the most highly conserved homeodomain residues and contacts the second adenine (TAATNN)(Klemm et al., 1994).

The experiments demonstrated that among the many possible side-chain base interactions between positions 47 and 51 and the third and fourth base-pairs in the TAAT core, the wild-type combination of amino acids and base-pairs provided the interactions with the highest affinity and specificity. A variant was obtained that could recognize a sequence that the wild-type protein could not, but it could do so only with modest affinity and specificity. These results suggested that the folding of the homeodomain and the manner in which it docks to DNA constrain what residues can be used at positions 47 and 51 to interact specifically with DNA.

Several studies have shown that different residues can be introduced in the homeodomain at position 50 to change specificity for residues 3' to the TAAT core (TAATNNN) (Hanes and Brent, 1989, 1991; Treisman et al., 1989; Percival-Smith et al., 1990; Ades and Sauer, 1994). Within a particular domain, there may be two sets of residues which contact DNA, those that are inalterable, and those that can be changed to modulate specificity. Which set a particular residue may belong to will probably be determined by the orientation of that side chain as it is presented to the DNA. This orientation will determine what trajectories are possible for the presentation of functional groups in the amino acid side chain to their targets on the base-pairs or sugar-phosphate backbone. The position of the α -carbon backbone of the domain from which the side chain emanates as well as the adjacent space occupied by other residues in the protein are key constraints. Furthermore, the residues that mediate sequence recognition may be interdependent. A set of residues at the protein-DNA interface may directly interact to

position each other for the presentation of a surface complimentary to the DNA. A change in specificity may not be attainable through the manipulation of a residue that may only function as part of a network of recognition (Pabo and Sauer, 1992).

These realizations led to the pursuit of methods for changing the DNA-binding specificity of the Oct-1 homeodomain without introducing mutations at the protein-DNA interface. The high DNA-binding affinity and specificity of the Oct-1 POU domain derives from the cooperative recognition of the octamer site (5'-ATGCAAAT-3') by two domains which appear to be structurally independent yet connected by a 24 residue linker: the POU-specific domain (which recognizes ATGC) and the homeodomain (which recognizes AAAT) (Klemm et al., 1994). When examined individually, each subdomain demonstrates only modest affinity for its subsite. The configuration of the POU domain suggested that the homeodomain might be targeted to a novel DNA sequence if it were fused to a heterologous DNA-binding domain instead of the POU-specific domain. As detailed in Chapter 4, this prediction was tested by the structure-based design of ZFHD1, a chimeric protein in which the Oct-1 homeodomain was fused to zinc fingers 1 and 2 of Zif268. The designed protein was found to have sequence specificity distinct from that of either the parental Oct-1 or Zif268 proteins. When fused to an activation domain, ZFHD1 could activate reporter gene expression *in vivo* in a sequence-specific manner. Thus, the affinity and specificity of the designed factor allowed it to recognize a sequence *in vivo* which could not be recognized efficiently by Oct-1. ZFHD1 could thus provide an altered specificity Oct-1 variant without requiring any mutations at the homeodomain-DNA interface. The first test of the utility of this factor is described in Chapter 5. In these experiments, ZFHD1 was used to form the C1 complex on a novel sequence which was not efficiently recognized by Oct-1. This allowed the demonstration *in vitro* and *in vivo* that the Oct-1 homeodomain could provide all of the Oct-1 mediated protein-protein interactions necessary for recruitment of α TIF and C1 factors into a functional C1 complex.

REFERENCES

- Ace, C.I., M.A. Dalrymple, F.H. Ramsay, V.G. Preston, and C.M. Preston. 1988. Mutational analysis of the herpes simplex virus type 1 trans-inducing factor Vmw65. J. Gen. Virol. **69**: 2595-2605.
- Ades, S.E., and R.T. Sauer. 1994. Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50. Biochemistry **33**: 9187-9194.
- Affolter, M., A. Schier, and W.J. Gehring. 1990. Homeodomain proteins and the regulation of gene expression. Current Opinion in Cell Biology **2**: 485-495.
- Ares, M. Jr., J. Chung, L. Giglio, and A.M. Weiner. 1987. Distinct factors with SP1 and NFA specificities bind to adjacent functional elements of the human U2 snRNA gene enhancer. Genes & Dev. **1**: 808-817.
- Bark, C., P. Weller, J. Zabeilski, L. Janson, and U. Petterson. 1987. A distant enhancer element is required for transcription of a U6 RNA gene. Nature **328**: 356-362.
- Billeter, M., Y.Q. Qian, G. Otting, M. Muller, W.Gehring, and K. Wuthrich. 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. J. Mol. Biol. **234**: 1084-1097. .
- Brent, R. and M. Ptashne. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43**: 729-736.
- Campbell, M.E.M., J.W. Palfreyman, and C.M. Preston. 1984. Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. J. Mol. Biol. **180**: 1-19.
- Carbon, P., S. Murgo, J-P. Ebel, A. Krol, G. Tebb, and I.W. Mattaj. 1987. A common octamer motif binding protein is involved in the transcription of the U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. Cell **51**: 71-79.
- Chan, S.-K., L. Jaffe, M. Capovilla, J. Botas, and R.S. Mann. 1994. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. Cell **78**: 603-615.
- Chan, S.-K. and R.S. Mann. 1993. The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. Genes & Dev. **7**: 796-811.
- Chatterjee, S. and K. Struhl. 1995. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. Nature **374**: 820-822.
- Choo, Y. and A. Klug. 1994. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc. Natl. Acad. Sci. USA **91**: 11168-11172.
- Choo, Y. and A. Klug. 1994. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. USA **91**: 11163-11167.

Choo, Y., I. Sanchez-Garcia, and A. Klug. 1994. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature **372**: 642-645.

Cormack, B.P., M. Strubin, L.A. Stargell, and K. Struhl. 1994. Conserved and nonconserved functions of the yeast and human TATA-binding proteins. Genes & Dev. **8**: 1335-1343.

Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon. 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature **323**: 544-548.

Dyson, N. and E. Harlow. 1992. Adenovirus E1A targets key regulators of cell proliferation. In Cancer Surveys, Vol. 12: Tumor Suppressor Genes, the Cell Cycle and Cancer, (ed. A.J. Levine), pp 161-195, Cold Spring Harbor Laboratory Press, N.Y.

Ebright, R.H., A. Kolb, H. Buc, T.A. Kunkel, J.S. Krakow, and J. Beckwith. 1987. Role of glutamic acid-181 in DNA-sequence recognition by the catabolite activator protein (CAP) of Escherichia coli: altered DNA-sequence-recognition properties of [val181]CAP and [leu181]CAP. Proc. Natl. Acad. Sci. USA **84**: 6083-6087.

Ekker, S.C., D.G. Jackson, D.P. von Kessler, B.I. Sun, K.E. Young, and P.A. Beachy. 1994. The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J. **13**: 3551-3560.

Ekker, S.C., D.P. von Kessler, and P.A. Beachy. 1992. Differential sequence recognition is a determinant of specificity in homeotic gene action. EMBO J. **11**: 4059-4072.

Evans, R. 1988. The steroid and thyroid hormone receptor superfamily. Science **240**: 889-895.

Fairall, L., J.W.R. Schwabe, L. Chapman, J.T. Finch, and D. Rhodes. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature **366**: 483-487.

Fletcher, C., N. Heintz, and R.G. Roeder. 1987. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell **51**: 773-781.

Furukubo-Tokunaga, S. Flister and W.J. Gehring. 1993. Functional specificity of the Antennapedia homeodomain. Proc. Natl. Acad. Sci. USA **90**: 6360-6364.

Furukubo-Tokunaga, K., M. Muller, M. Affolter, L. Pick, U. Kloter, and W.J. Gehring. 1992. In vivo analysis of the helix-turn-helix motif of the fushi tarazu homeo domain of Drosophila melanogaster. Genes & Dev. **6**: 1082-1096.

Gehring, W.J. 1987. Homeo boxes in the study of development. Science **236**: 1245-1252.

Gerster, T. and R.G. Roeder. 1988. A herpesvirus trans-acting protein interacts with transcription factor OTF-1 and other cellular proteins. Proc. Natl. Acad. Sci. USA **85**: 6347-6351.

Gerster, T., P. Matthias, M. Thali, J. Jiricny, and W. Schaffner. 1987. Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer. EMBO J. **6**: 1323-1330.

Gibson, G., A. Schier, P. LeMotte, and W.J. Gehring. 1990. The specificities of sex combs reduced and antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell **62**: 1087-1103.

Goutte, C. and A.D. Johnson. 1988. $\alpha 1$ protein alters the DNA binding specificity of $\alpha 2$ repressor. Cell **52**: 875-882.

Greaves, R. and P.O'Hare. 1989. Separation of requirements for protein-DNA complex assembly from those for functional activity in the herpes simplex virus regulatory protein Vmw65. J. Virol. **63**:1641-1650.

Greaves, R. and P.O'Hare. 1990. Structural requirements in the herpes simplex type 1 transactivator Vmw65 for interaction with the cellular octamer-binding protein and target TAATGARAT sequences. J. Virol. **64**: 2716-2724.

Gstaiger, M., L. Knoepfel, O. Georgiev, W. Schaffner, and C.M. Hovens. 1995. A B-cell coactivator of octamer-binding transcription factors. Nature **373**: 360-362.

Han, K. and J.L. Manley. 1993. Functional domains of the Drosophila engrailed protein. EMBO J. **12**:2723-2733.

Hanes, S.D., and R. Brent. 1989. DNA specificity of the Bicoid activator protein is determined by the homeodomain recognition helix residue 9. Cell **57**: 1275-1283.

Hanes, S.D., and R. Brent. 1991. A genetic model for interaction of the homeodomain recognition helix with DNA. Science **251**: 426-430.

Hayashi, S. and M.P. Scott. 1990. What determines the specificity of action of Drosophila homeodomain proteins? Cell **63**: 883-894.

Hayes, S., and P. O'Hare. 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**: 852-862.

Heard, D.J., T. Kiss, and W. Filipowicz. 1993. Both Arabidopsis TATA binding protein (TBP) isoforms are functionally identical in RNA polymerase II and III transcription in plant cells: evidence for gene-specific changes in DNA binding specificity of TBP. EMBO J. **12**: 3519-3528.

Hernandez, N. 1993. TBP, a universal eukaryotic transcription factor? Genes & Dev. **7**: 1021-1032.

Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkin, and H.R. Horvitz. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes & Dev. **2**: 1513-1516.

Huang, L., T. Sera, and P.G. Schultz. 1994. A permutational approach toward protein-DNA recognition. Proc. Natl. Acad. Sci. USA **91**: 3969-3973.

Jamieson, A. C., S.-H. Kim, and J.A. Wells. 1994. In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry **33**:5689-5695.

- Johnson, A. 1992. A combinatorial regulatory circuit in budding yeast. In Transcriptional Regulation (eds. S.L. McKnight and K.R. Yamamoto), pp. 975-1006. Cold Spring Harbor Laboratory Press, N.Y.
- Katan, M. A. Haigh, C.P. Verrijzer, P.C. van der Vliet, and P. O'Hare. 1990. Characterization of a cellular factor which interacts functionally with Oct-1 in the assembly of a multicomponent transcription complex. Nucleic Acids. Res. **18**: 6871-6880.
- Kenyon, C. 1994. If birds can fly, why can't we? Homeotic genes and evolution. Cell **78**: 175-180.
- Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell **63**: 579-590.
- Klages, N. and M. Strubin. 1995, Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo. Nature **374**: 822-823.
- Klein, C. and K. Struhl. 1994. Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. Science. **266**: 280-282.
- Klemm, J.D., M.A. Rould, R. Aurora, W. Herr, and C. O. Pabo. 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.
- Kornberg, T.B. 1993. Understanding the homeodomain. J. Biol. Chem. **268**: 26813-26816.
- Kristie, T. M. and B. Roizman. 1984. Separation of sequences defining basal expression from those conferring alpha gene recognition within the regulatory domains of herpes simplex virus 1 alpha genes. Proc. Natl. Acad. Sci. USA **81**: 4065-4069.
- Kristie, T.M. and B. Roizman. 1987. Host cell proteins bind to the cis-acting site required for virion-mediated induction of herpes simplex virus 1 α genes. Proc. Natl. Acad. Sci. **84**: 71-75.
- Kristie, T. M., J.H. LeBowitz, and P.A. Sharp. 1989. The octamer-binding proteins form multi-protein-DNA complexes with the HSV alpha TIF regulatory protein. EMBO J. **8**: 4229-4238.
- Kristie, T.M., J.L. Pomerantz, T.C. Twomey, S. A. Parent, and P.A. Sharp. 1995. The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. J. Biol. Chem. **270**: 4387-4394.
- Kristie, T.M., and P.A. Sharp. 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV α -trans-activator protein. Genes & Dev. **4**: 2383-2396.
- Kristie, T.M., and P.A. Sharp. 1993. Purification of the cellular C1 factor required for the stable recognition of the Oct-1 homeodomain by the herpes simplex virus α trans-induction factor (VP16). J. Biol. Chem. **268**: 6525-6534.
- Krumlauf, R. 1994, Hox genes in vertebrate development. Cell **78**: 191-201.

- Kutoh, E., P.E. Stromstedt, and L. Poellinger. 1992. Functional interference between the ubiquitous and constitutive octamer transcription factor 1 (OTF-1) and the glucocorticoid receptor by direct protein-protein interaction involving the homeo subdomain of OTF-1. Mol. Cell. Biol. **12**: 4960-4969.
- Kuziora, M.A., and W. McGinnis. 1989. A homeodomain substitution changes the regulatory specificity of the deformed protein in *Drosophila* embryos. Cell **59**: 563-571.
- La Bella, F., H.L. Sive, R.G. Roeder, and N. Heintz. 1988. Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. Genes & Dev. **2**: 32-39.
- Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. Science **240**: 1759-1764.
- Laughon, A. (1991) DNA binding specificity of homeodomains. Biochemistry **30**: 11357-11367.
- Lawrence, P.A., and G. Morata. 1994. Homeobox genes: their function in *Drosophila* segmentation and pattern formation. Cell **78**: 181-189.
- LeBowitz, J. H., T. Kobayashi, L. Staudt, D. Baltimore, and P.A. Sharp. 1988. Octamer-binding proteins from B or HeLa cells stimulate transcription of the immunoglobulin heavy-chain promoter in vitro. Genes & Dev. **2**: 1227-1237.
- Li, S., E.B. Crenshaw, E.J. Rawson, D.M. Simmons, L.W. Swanson, and M.G. Rosenfeld. 1990. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU domain gene Pit-1. Nature **347**: 528-533.
- Lin, C., S.C. Lin, C.P. Chang, and M.G. Rosenfeld. 1992. Pit-1-dependent expression of the receptor for growth hormone releasing factor mediates pituitary cell growth. Nature **360**: 765-768.
- Lin, L. and W. McGinnis. 1992. Mapping functional specificity in the Dfd and Ubx homeodomains. Genes & Dev. **6**: 1071-1081.
- Luo, Y., H. Fujii, T. Gerster, and R.G. Roeder. 1992. A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. Cell **71**:231-242.
- Mackem, S. and B. Roizman. 1982. Regulation of alpha genes of herpes simplex virus: the alpha 27 promoter-thymidine kinase chimeric gene is positively regulated in converted L cells. J. Virol. **43**: 1015-1023.
- Mackem, S. and B. Roizman. 1982. Structural features of the alpha gene 4, 0, and 27 promoter-regulatory sequences which confer alpha regulation on chimeric thymidine kinase genes. J. Virol. **44**: 939-949.
- Mackem, S. and B. Roizman. 1982. Differentiation between alpha promoter and regulator regions of herpes simplex virus I: The functional domains and sequence of a movable alpha regulator. Proc. Natl. Acad. Sci. USA **79**: 4917-4921.

- Mak, A. and A.D. Johnson. 1993. The carboxy-terminal tail of the homeo domain protein $\alpha 2$ is required for function with a second homeo domain protein. Genes & Dev. **7**: 1862-1870.
- Malicki, J., K. Schughart, and W. McGinnis. 1990. Mouse Hox-2.2 specifies thoracic segmental identity in *Drosophila* embryos and larvae. Cell **63**: 961-967.
- Mann, R.S., and D.S. Hogness. 1990. Functional dissection of ultrabithorax proteins in *D. melanogaster*. Cell **60**: 597-610.
- Marsden, H.S., M.E.M. Campbell, L. Haarr, M.C. Frame, D.S. Parris, M. Murphy, R.G. Hope, M.T. Muller, and C.M. Preston. 1987. The 65,000-Mr DNA-binding and virion trans-inducing proteins of herpes simplex virus type 1. J. Virol. **61**: 2428-2437.
- McGinnis, N., M.A. Kuziora, and W. McGinnis. 1990. Human Hox-4.2 and *Drosophila* deformed encode similar regulatory specificities in *Drosophila* embryos and larvae. Cell **63**: 969-976.
- Mizushima-Sugano, J. and R.G. Roeder. 1986. Cell-type-specific transcription of an immunoglobulin k light chain in vitro. Proc. Natl. Acad. Sci. USA **83**: 8511-8515.
- Morata, G. 1993. Homeotic genes of *Drosophila*. Curr. Opin. Gen. Dev. **3**: 606-614.
- Muller, M. M., S. Ruppert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature **336**: 544-551.
- Murphy, S., C. Di Leigro, and M. Melli. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell **51**: 81-87.
- Murphy, S., J.-B. Yoon, T. Gerster, and R.G. Roeder. 1992. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. Mol. Cell. Biol. **12**: 3247-3261.
- Murre, C. and D. Baltimore. 1992. The helix-loop-helix motif: structure and function. In Transcriptional Regulation (eds. S.L. McKnight and K.R. Yamamoto), pp. 861-879.. Cold Spring Harbor Laboratory Press, N.Y.
- O'Hare, P. and C.R. Goding. 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virus transactivation. Cell **52**: 435-445.
- O'Neill, E.A., C. Fletcher, C.R. Burrow, N. Heintz, R.G. Roeder, and T.J. Kelly. 1988. Transcription factor OTF-1 is functionally identical to the DNA replication factor NFIII. Science **241**: 1210-1213.
- Otting, G., Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring, and K. Wuthrich. 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. **9**: 3085-3092.
- Pabo, C.O., and R.T. Sauer. 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61**:1053-1095.

- Pavletich, N.P. and C.O. Pabo. 1991. Crystal structure of a five finger GLI-DNA complex: new perspectives on zinc fingers. Science **261**: 1701-1707.
- Pavletich, N.P. and C.O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science **252**: 809-817.
- Peifer, M. And E. Wieschaus. 1994. Mutations in the Drosophila gene extradenticle affect the way specific homeodomain proteins regulate segmental identity. Genes. Dev. **4**: 1209-1223.
- Pellett, P. E., J.L.C. McKnight, F.J. Jenkins, and B. Roizman. 1985. Nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of trans-inducing alpha genes. Proc. Natl. Acad. Sci. USA **82**: 5870-5874.
- Percival-Smith, A., M. Mueller, M. Affolter, and W.J. Gehring. 1990. The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. EMBO J. **9**: 3867-3974.
- Pfaffle, R.W., G.E. DiMattia, J.S. Parks, M.R. Brown, J.M. Wit, M. Jansen, H. van der Nat, J.L. van den Brande, M.G. Rosenfeld, and H.A. Ingraham. 1992. Mutation in the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. Science **257**: 1118-1121.
- Phillips, C.L., M.R. Stark, A.D. Johnson, and F.W. Dahlquist. 1994. Heterodimerization of the yeast homeodomain transcriptional regulators $\alpha 2$ and $\alpha 1$ induces an interfacial helix in $\alpha 2$. Biochemistry **33**: 9292-9302.
- Post, L.E., S. Mackem, and B. Roizman. 1981. The regulation of α genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with α gene promoters. Cell **24**: 555-565.
- Preston, C. M., M.C. Frame, M.E.M. Campbell. 1988. A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. Cell **52**: 425-434.
- Pruijn, G.J.M., W. van Driel, T. van Miltenburg, and P.C. van der Vliet. 1986. Nuclear factor III, a novel sequence-specific DNA binding protein from HeLa cells stimulates adenovirus DNA replication. Nature **322**: 656-659.
- Pruijn, G.J.M., W. van Driel, and P.C. van der Vliet. 1987. Promoter and enhancer elements containing a conserved sequence motif are recognized by nuclear factor III, a protein stimulating adenovirus DNA replication. EMBO J. **6**: 3771-3778.
- Qian, Y.Q., M. Billeter, G. Otting, M. Mueller, W.J. Gehring, and K. Wuthrich. 1989. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell **59**: 573-580.
- Radovick, S., M. Nations, Y. Du, L.A. Berg, B.D. Weintraub, and F.E. Wondisford. 1992. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. Science **257**: 1115-1118.
- Rebar, E. and C.O. Pabo. 1994. Zinc finger phage: Affinity selection of fingers with new DNA-binding specificities. Science **263**: 671-673.

Roberts, S. B., N. Segil, and N. Heintz. 1991. Differential phosphorylation of the transcription factor Oct-1 during the cell cycle. Science **253**: 1022-1026.

Roizman, B. and A.E. Sears. 1991. Herpes Simplex Viruses and their replication. In Fundamental Virology, (eds. B. N. Fields and D.M. Knipe) pp.849-896, Raven Press, N.Y.

Sadowski, I., J. Ma, S. Triezenberg, and M. Ptashne. 1988. GAL4-VP16 is an unusually potent transcriptional activator. Nature **335**: 563-564.

Scheidereit, C., A. Heguy, and R.G. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. Cell **51**: 783-793.

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

Segil, N., S. B. Roberts, and N. Heintz. 1991. Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct-1 binding activity. Science **254**: 1814-1816.

Sive, H. L., N. Heintz, R.G. Roeder. 1986. Multiple sequence elements are required for maximal in vitro transcription of a human histon H2B gene. Mol. Cell Biol. **6**: 3329-3340.

Smith, D.L. and A.D. Johnson. 1992. A molecular mechanism for combinatorial control in yeast; MCM1 protein sets the spacing and orientation of the homeodomains of an $\alpha 2$ dimer. Cell **68**: 133-142.

Smith, G.P. and J.K. Scott. 1993. Libraries of peptides and proteins displayed on filamentous phage. Methods in Enzymology **217**: 228-257.

Staudt, L. M., H. Singh, R. Sen, T. Wirth, P.A. Sharp, and D. Baltimore. 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature **323**: 640-643.

Stern, S. and W. Herr. 1991. The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain. Genes & Dev. **5**: 2555-2566.

Stern, S., M. Tanaka, and W. Herr. 1989. The Oct-1 homeo domain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature **341**: 624-630.

Strubin, M., J.W. Newell, and P. Matthias. 1995. OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. Cell **80**: 497-506.

Strubin, M., and K. Struhl. 1992. Yeast and human TFIID with altered DNA-binding specificity for TATA elements. Cell **68**: 721-730.

Sturm, R. A., G. Das, and W. Herr. 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes & Dev. **2**: 1582-1599.

Tanaka, M., J.-S. Lai, and W. Herr. 1992. Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter. Cell **68**: 755-767.

Tansey, W.P., S. Ruppert, R. Tjian, and W. Herr. 1994. Multiple regions of TBP participate in the response to transcriptional activators in vivo. Genes & Dev. **8**: 2756-2769.

Tatsumi, K., K. Miyai, T. Notomi, K. Kaibe, N. Amino, Y. Mizuno, and H. Kohno. 1992. Cretinism with combined hormone deficiency caused by a mutation in the PIT1 gene. Nature Genetics **1**: 56-58.

Treisman, J., P. Gonczy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell **59**: 553-562.

Triezenberg, S.J., R.C. Kingsbury, and S.L. McKnight. 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. Genes & Dev. **2**: 1764-1778.

Tzamarias, D., W.T. Pu, and K. Struhl. 1992. Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. Proc. Natl. Acad. Sci. USA **89**: 2007-2011.

Ullman, K.S., W.M. Flanagan, C.A. Edwards, and G.R. Crabtree. 1991. Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, OAP⁴⁰. Science **254**: 558-562.

Ullman, K.S., J.P. Northrop, A. Admon, and G.R. Crabtree. 1993. Jun family members are controlled by a calcium-regulated, cyclosporin A-sensitive signaling pathway in activated T lymphocytes. Genes & Dev. **7**: 188-196.

Van Dijk, M.A. and C. Murre. 1994. extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell **78**: 617-624.

Verrijzer, C.P., A.J. Kal, and P.C. van der Vliet. 1990a. The DNA binding domain (POU domain) of transcription factor Oct-1 suffices for stimulation of DNA replication. EMBO J. **9**: 1883-1888.

Verrijzer, C.P., A.J. Kal, and P.C. van der Vliet. 1990b. The Oct-1 homeo domain contacts only part of the octamer sequence and full Oct-1 DNA binding activity requires the POU-specific domain. Genes & Dev. **4**: 1964-1974.

Vershon, A.K., and A.D. Johnson. 1993. A short, disordered protein region mediates interactions between the homeodomain of the yeast $\alpha 2$ protein and the MCM1 protein. Cell **72**: 105-112.

Voss, J.W., and M.G. Rosenfeld. 1992. Anterior pituitary development: short tales from dwarf mice. Cell **70**: 527-530.

Voss, J.W., L. Wilson, and M.G. Rosenfeld. 1991. POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. Genes & Dev. **5**: 1309-1320.

Werstuck, G. and J.P. Capone. 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**: 5509-5513.

Wharton, R.P. and M. Ptashne. 1985. Changing the binding specificity of a repressor by redesigning an α -helix. Nature **316**: 601-605.

Wilson, A.C., K. LaMarco, M.G. Peterson, and W. Herr. 1993. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. Cell **74**: 115-125.

Wirth, T., L. Staudt, and D. Baltimore. 1987. An octamer oligonucleotide upstream of a TATA element is sufficient for lymphoid-specific promoter activity. Nature **329**: 174-177.

Wolberger, C., A.K. Vershon, B. Liu, A.D. Johnson, and C.O. Pabo. 1991. Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67**: 517-528.

Xiao, P. and J.P. Capone. 1990. A cellular factor binds to the herpes simplex virus type 1 transactivator Vmw65 and is required for Vmw65-dependent protein-DNA complex assembly with Oct-1. Mol. Cell. Biol. **10**: 4974-4977.

Youderian, P., A. Vershon, S. Bouvier, R.T. Sauer, and M.M. Susskind. 1983. Changing the DNA-binding specificity of a repressor. Cell **35**: 777-783.

Zeng, W., D.J. Andrew, L.D. Mathies, M.A. Horner, and M.P. Scott. 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**: 339-352.

Zwilling, S., A. Annweiler, and T. Wirth. 1994. The POU domains of Oct-1 and Oct-2 transcription factors mediate specific interaction with TBP. Nucleic Acids Res. **22**: 1655-1662.

Zwilling, S., H. Konig, and T. Wirth. 1995. High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. EMBO J. **14**: 1198-1208.

CHAPTER II

RECOGNITION OF THE SURFACE OF A HOMEODOMAIN PROTEIN

This chapter originally appeared in *Genes & Development* **6**: 2047-2057 (1992).

ABSTRACT

Homeodomain proteins exhibit distinct biological functions with specificities that cannot be predicted by their sequence specificities for binding DNA. Recognition of the surface of the Oct-1 POU-homeodomain provides a general model for the contribution of selective protein-protein interactions to the functional specificity of the homeodomain family of factors. The assembly of Oct-1 into a multiprotein complex on the Herpes Simplex Virus α /IE enhancer is specified by the interactions of its homeodomain with ancillary factors. This complex (C1 complex) is composed of the viral α TIF protein (VP16), Oct-1, and one additional cellular component, the C1 factor. Variants of the Oct-1 POU-homeodomain were generated by site-directed mutagenesis which altered the residues predicted to form the exposed surface of the domain-DNA complex. Proteins with single amino acid substitutions on the surface of either helices 1 or 2 of the Oct-1 POU-homeodomain had decreased abilities to form the C1 complex. The behavior of these mutants in a cooperative DNA-binding assay with α TIF suggested that the Oct-1 POU-homeodomain is principally recognized by α TIF in the C1 complex. The preferential recognition of Oct-1 over the closely related Oct-2 protein is critically influenced by a single residue on the surface of helix 1 since the introduction of this residue into the Oct-2 POU-homeodomain significantly enhanced its ability to form a C1 complex.

INTRODUCTION

Homeodomain proteins comprise a broad family of regulatory factors which are important determinants for morphogenesis, cell-type determination, and cell-type-specific functions in a wide evolutionary range of organisms (for reviews see Affolter et al., 1990; Gehring, 1987; Scott et al., 1989). The homeodomain motif consists of a DNA-binding domain that contains three α -helices and the determination of the structures of three homeodomain-DNA complexes has revealed a general conservation of domain folding and DNA recognition among proteins with minimal amino acid sequence similarity (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991). Most characterized homeodomains specifically recognize a DNA sequence containing a 5'-TAAT-3' core (Laughon, 1991), but despite having similar DNA-binding specificities, individual homeodomain proteins confer extremely specific regulatory actions (for review see Hayashi and Scott, 1990). Furthermore, several studies suggest that the functional specificity of homeodomain proteins is largely determined by a minimal region containing the homeodomain (Kuziora and McGinnis, 1989; Gibson et al., 1990; Malicki et al., 1990; Mann and Hogness, 1990; McGinnis et al., 1990; Lin and McGinnis, 1992; Furukubo-Tokunaga et al., 1992). For example, the replacement of the homeobox sequences of the Deformed gene of *Drosophila melanogaster* with the homeobox sequences in the Ultrabithorax gene targets regulation by the chimeric gene in vivo to those sites which are normally controlled by Ultrabithorax. This change in specificity occurs even though the DNA-recognition helices of these two homeodomains are essentially identical (Kuziora and McGinnis, 1989; Lin and McGinnis, 1992). A critical question, therefore, concerns how homeodomain regulatory specificity is determined.

One possible mechanism for the determination of the biological specificity of homeodomains invokes interactions with other regulatory proteins. While a single homeodomain may have the potential to bind to many target promoter sites, it may be

specifically recruited into a functional complex at only a subset of those sites by selective protein-protein interactions.

An excellent system for studying interactions between homeodomains and other proteins involves the differing abilities of two highly homologous proteins, Oct-1 and Oct-2, to assemble into a multiprotein complex on the Herpes Simplex Virus α /IE enhancer element. This element (5'ATGCTAATGATATTCTTTGG3') is required for the transcriptional regulation of the five viral α genes (Mackem and Roizman 1982a-c; Kristie and Roizman, 1984; Gaffney et al., 1985; Bzik and Preston, 1986; for review, see McKnight et al., 1986) and is recognized by a multiprotein complex (C1 complex) that contains Oct-1, the viral factor α TIF (VP16, Vmw65, ICP25) (McKnight et al, 1987; Preston et al, 1988; O'Hare and Goding, 1988; Gerster and Roeder, 1988; Stern et al., 1989; Kristie et al., 1989), and at least one additional cellular factor, the C1 factor (Kristie et al., 1989; Xiao and Capone, 1990). The 5' portion of the enhancer element consists of a homolog of the consensus octamer site (5'-ATGCAAAT-3') and is specifically recognized by the Oct-1 POU domain (Kristie and Sharp, 1990), which contains POU-specific and POU-homeo subdomains characteristic of the POU domain subclass of homeodomain factors (Herr et al., 1988). The remainder of the element is recognized by α TIF, and possibly by components of the cellular C1 factor (Kristie and Sharp, 1990).

Initial analyses of the C1 complex suggested that most, if not all, of the Oct-1 determinants that mediate protein-protein interactions in the formation of the C1 complex are contained in the POU-homeodomain of the protein (Kristie and Sharp, 1990). However, despite a high degree of homology with Oct-1 in the region of the POU-homeodomain, and an apparently equivalent ability to bind to the DNA element, Oct-2 had a 100 fold lower potential to form a C1 complex (Kristie et al., 1989). Although interactions between Oct-1 and α TIF have been demonstrated in the absence of the C1 factor (Kristie and Sharp, 1990; Stern and Herr, 1991), it remains unclear which

component of the C1 complex is responsible for the selective recognition of the Oct-1 POU-homeodomain.

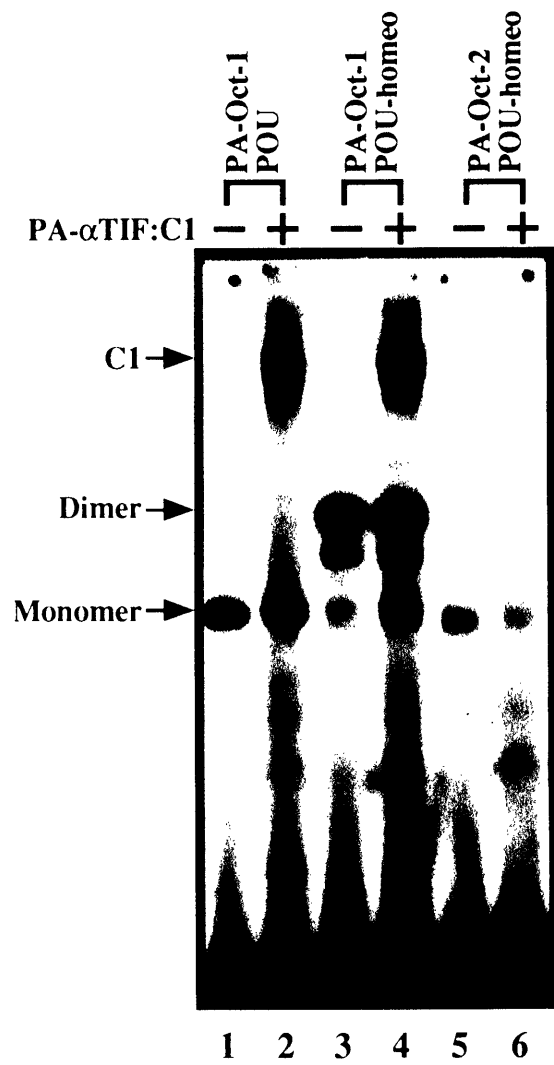
Site-directed mutagenesis has been used to determine which amino acids are important for the recognition of the Oct-1 POU-homeodomain in the formation of a C1 complex. The results indicate that residues on the surface of both putative helices 1 and 2 are important in the selective protein-protein interactions and that a single amino acid substitution in helix 1 of the Oct-2 POU-homeodomain promotes significant interaction of this protein with components of the C1 complex. Additionally, evidence is presented that α TIF is the component of the C1 complex which is principally responsible for the recognition of the Oct-1 POU-homeodomain. This study demonstrates that individual amino acid differences between highly related homeodomains can dictate functional specificity through specific protein-protein interactions with regulatory factors.

RESULTS

Differing abilities of the Oct-1 and Oct-2 POU-homeodomains to participate in C1 complex formation

The Oct-1 and Oct-2 POU-homeodomains were produced as Staphylococcus protein A (PA) fusion proteins and compared in an electrophoretic mobility shift assay for their potential to form the C1 complex (Figure 1). As previously observed (Kristie and Sharp, 1990), the Oct-1 POU-homeodomain binds to the HSV α 0 probe, which contains the α /IE element, cooperatively as a homodimer (lane 3), and forms the multiprotein C1 complex upon addition of α TIF (PA- α TIF) and a chromatographic fraction of HeLa cell nuclear extract containing the C1 factor (lane 4). The Oct-2 POU-homeodomain, however, generates a complex on the α /IE element with the expected mobility of a monomeric protein-DNA complex (lane 5), and does not form a C1 complex in the presence of α TIF and the C1 factor (lane 6). Thus, the Oct-1 and Oct-2 POU-homeodomains differ in both the potential to dimerize on the α /IE element and in the ability to interact with the α TIF and

Figure 1. Comparison of Oct-1 and Oct-2 POU-homeodomains in the assembly of a C1 complex. DNA-protein binding reactions were done as described in Materials and Methods. The probe DNA, HSV α 0, contains the α /IE element from -168 to -142 of the α 0 promoter. Protein A fusion proteins containing the Oct-1 POU domain (100 pg), the Oct-1 POU-homeodomain (100 ng), or the Oct-2 POU-homeodomain (100 ng) were incubated in the absence (-) or presence (+) of 15 ng PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor as indicated. The positions of the multiprotein C1 complex as well as the monomeric and homodimeric Oct-1- or Oct-2-DNA complexes are indicated with arrows.



C1 factors. As will be shown later, the ability to bind as a dimer is not related to the potential to form a C1 complex. Furthermore, both proteins bind as monomers to a probe containing a consensus octamer site with comparable affinity (data not shown). Clearly, the POU-homeodomains of Oct-1 and Oct-2 contain determinants which mediate their discrimination by α TIF, the C1 factor, or both.

Residues in the Oct-1 POU homeodomain involved in C1 complex formation

Amino acid residues in the Oct-1 POU-homeodomain which are most likely to mediate specific protein-protein interactions would be expected to be located on the physically available surface of the DNA-bound domain and to be divergent among members of the POU domain family of factors. A model of the Oct-1 POU-homeodomain (Figure 5) was generated by aligning its sequence with that of the engrailed homeodomain and positioning its amino acids based upon the crystal structure of the engrailed homeodomain-DNA complex (Kissinger et al., 1990). A comparison of the POU-homeodomain amino acid sequences among representatives of the different classes (He et al., 1989) of POU domain proteins is depicted in Figure 2A. The regions that contain the most highly conserved residues within this family are boxed. The amino acid sequence of the engrailed homeodomain is listed for comparison, and the positions of residues which have been noted in the structural studies to be involved in packing of the hydrophobic core or in homeodomain-DNA interactions are indicated. Most of these positions lie within regions that are highly conserved among POU-domain factors, as would be expected if homeodomain structure and DNA recognition is generally conserved. The unboxed or divergent amino acids could be responsible for factor-specific protein-protein interactions that contribute to the functional specificities of these proteins. Helical wheel representations (Figure 2B) of helices 1 and 2 of the Oct-1 POU-homeodomain predict that the divergent residues, including those that differ between Oct-1 and Oct-2, comprise the surface expected to remain physically accessible upon DNA binding.

Figure 2. (A) Comparison of POU-homeodomain amino acid sequences among several POU-domain factors. Amino acids which are highly conserved in the POU-homeodomain of proteins representing the different subclasses (He et al., 1989) of POU-domain factors are boxed. The seven amino acids which differ between Oct-1 and Oct-2 are indicated (-). The positions occupied by residues which have been noted in structural studies to contact DNA (dark rectangles) or to participate in interactions within the core of the domain (dark circles), as well as those which are highly conserved among all homeodomains (open circles) are indicated below the sequence of the engrailed homeodomain as adapted from Laughton, 1991. For ease of comparison with other homeodomain proteins, the amino acids are numbered according to the scheme of Qian et al., 1989. The POU-homeodomain amino acid sequences have been previously compiled by Rosenfeld, 1991. (B) Helical wheel representations of helices 1 and 2 of the Oct-1 POU-homeodomain. The divergent amino acids are indicated in bold type while the conserved amino acids are in outline type. The amino acids which differ between Oct-1 and Oct-2 are underlined. The dashed line divides the helices into two halves, the lower of which is predicted to pack against helix 3 and face the DNA in a protein-DNA complex.

A
OCT 3/4
UNC 86
BRN 3
TST 1
BRN 1
PIT 1
OCT 2
OCT 1
ENGR

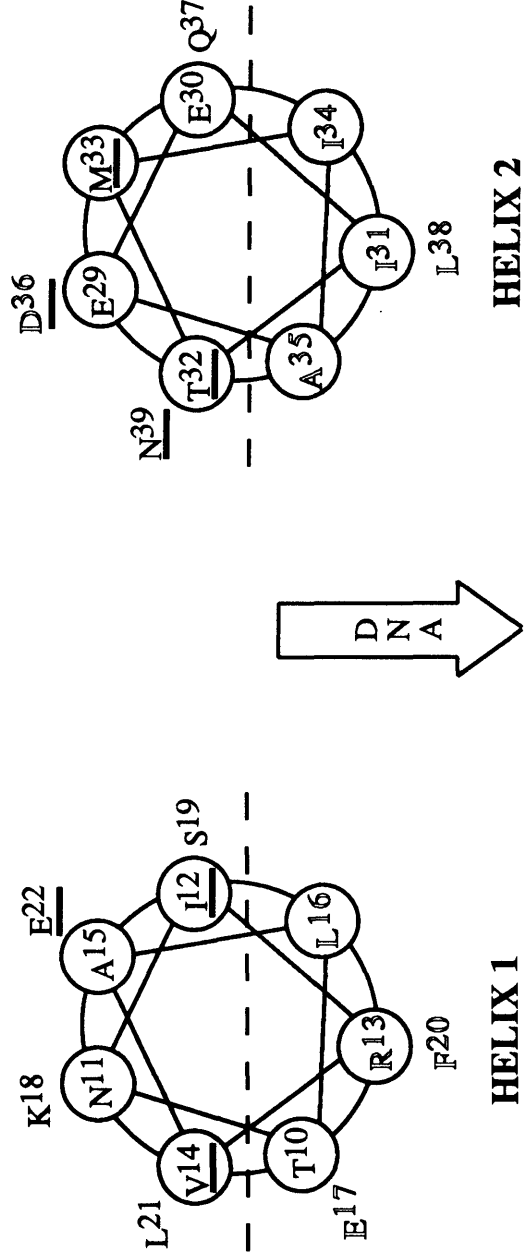
1	RRKKRTSI	ENRV	R	WS	LE	TM	F	LKCP	KPS	LQQ	I	TH	IAN	Q	L	G	LEKDVVRVWF	CNRRQKGKR
	KRRKRTSI	AAPE	K	RE	LE	QF	F	KQOP	RPS	GER	I	AS	IAD	R	L	D	LKKNVVRVWF	CNQRQKQKR
	KRRKRTSI	AAPE	K	RS	LE	AY	F	AVQP	RPS	SEK	I	AA	IAE	L	D	D	LKKNVVRVWF	CNQRQKQKR
	KRRKRTSI	EVG	K	GA	LE	SH	F	LKCP	KPS	AHE	I	TG	LAD	L	Q	Q	LEKEVVRVWF	CNRRQKEKR
	KRRKRTSI	EVS	K	GA	LE	SH	F	LKCP	KPS	SQE	I	TN	LAD	L	Q	Q	LEKEVVRVWF	CNRRQKEKR
	KRRRTTI	SIAA	K	DA	LE	RH	F	GEHS	KPS	SQE	I	MR	MAE	L	N	N	LEKEVVRVWF	CNRRQREKR
	RRKKRTSI	ETNV	R	FA	LE	KS	F	LANQ	KPT	SEE	I	LL	IAE	L	H	H	MEKEVIRVWF	CNRRQKEKR
	RRKKRTSI	ETNI	R	VA	LE	KS	F	LENQ	KPT	SEE	I	TM	IAD	L	N	N	MEKEVIRVWF	CNRRQKEKR
	EKRPRTAF	SSEQ	L	AR	LK	RE	F	NENR	YLT	ERR	R	QQ	LSS	L	G	G	LNEAQIKIWF	ONKRAKIKK

HELIIX 3

HELIIX 2

HELIIX 1

B



HELIIX 1

HELIIX 2

Thus, both the modeled structure of the Oct-1 POU-homeodomain and the sequence comparison of several POU domain proteins define a common set of amino acids which are likely to be responsible for protein-protein interactions, and which were therefore targeted for site-directed mutagenesis. The substitution of these amino acids would not be expected to disrupt either the domain's structure or its ability to bind DNA. Since the amino acids that are different between Oct-1 and Oct-2 in the region of the POU-homeodomain must contribute to the differential ability of these proteins to interact with components of the C1 complex, the appropriate Oct-1 residues were substituted with the corresponding Oct-2 residues. Other amino acids in the selected set were mutated so as to remove potentially important functional groups or to alter the character of a particular residue (i.e from positive to negative, hydrophobic to polar, and vice versa). All of the constructed variants of the Oct-1 POU-homeodomain were expressed in *E. coli* as PA fusions and purified by chromatography on IgG sepharose. In order to test their structural integrity and their capacity to bind DNA, all mutants were titrated into reactions containing an octamer consensus site probe. All mutants bound with affinities that were comparable to that of the wild type protein (data not shown).

Each mutant polypeptide was tested for the ability to bind to the α /IE element and to interact with the α TIF and C1 factors in a C1 complex assembly assay. The proteins were titrated into reactions containing the α /IE element and were found to cooperatively form homodimers in a manner similar or identical to that of the wild type protein (data not shown). Several amino acid substitutions significantly reduced the ability of the Oct-1 POU-homeodomain to form a C1 complex. As illustrated in Figure 3, mutant proteins were incubated with α TIF and a chromatographic fraction of HeLa cell nuclear extract containing the C1 factor and assayed in an electrophoretic mobility shift assay. For quantitative analysis, each protein was assayed for its ability to form a C1 complex at identical DNA binding activity (Table 1). Substitution of lysine 18 with glutamic acid (K18E) (Figure 3, lane 8), serine 19 with cysteine (S19C) (lane 10), or glutamic acid 22

Figure 3. Effects of amino acid substitutions in helices 1 and 2 of the Oct-1 POU-homeodomain on the formation of a C1 complex. Wild type (WT) (50 ng) and mutant PA-Oct-1-POU-homeodomain fusion proteins (30-80 ng) were incubated in DNA-protein binding reactions in the absence (-) or presence (+) of 15 ng PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor as indicated. The positions of the C1 and homodimeric POU-homeodomain-DNA complexes are indicated with horizontal arrows. Vertical arrows indicate the Oct-1 POU-homeodomain variants in which the Oct-1 residues which differ from those of Oct-2 in the POU-homeodomain have been substituted with the corresponding Oct-2 amino acids.

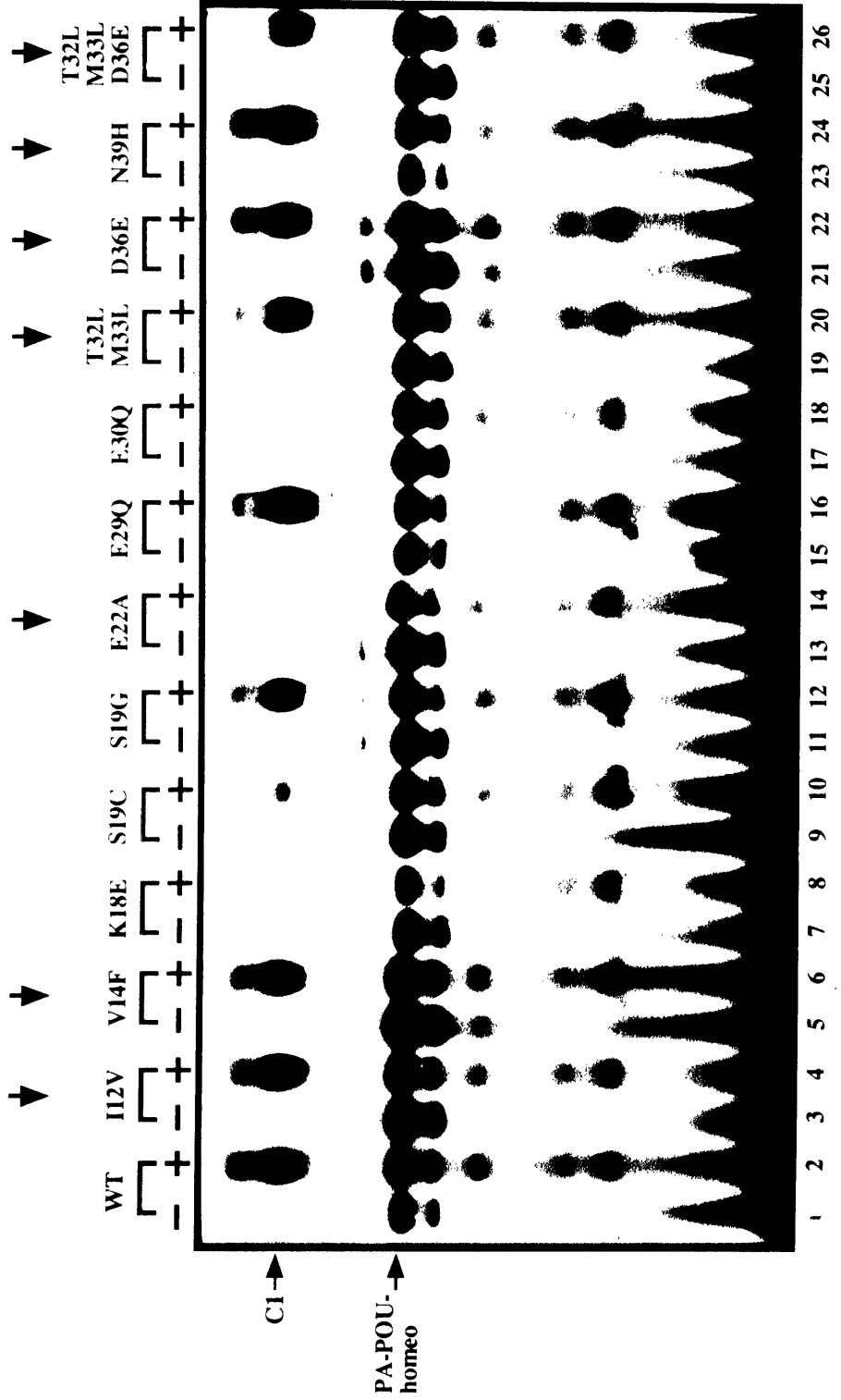


TABLE 1. Phenotypes associated with amino acid substitutions in the Oct-1 POU-homeodomain.

	SUBSTITUTION	ABILITY TO FORM C1 (%WT) ^a	ABILITY TO INTERACT WITH α TIF ^b
Helix 1	N11 to G11	123	+
	N11 to A11	108	+
	I12 to V12	109	+
	I12 to G12	92	+
	V14 to G14	114	+
	V14 to F14	89	+
	A15 to G15	75	+
	K18 to E18	13	-
	S19 to G19	81	+
	S19 to C19	13	-
	L21 to G21	99	+
	E22 to A22	16	-
	Helix 2	E29 to Q29	101
E30 to Q30		5	-
T32M33 to L32L33		77	+/-
T32M33 to V32V33		34	-
T32M33 to L32V33		39	-
D36 to E36		95	+
Q37 to E37		108	+
Q37 to A37		105	+
N39 to H39		98	+
N39 to D39		145	+
N39 to Y39	113	+	
T32M33D36 to L32L33E36	68	+/-	
S28T32M33 to L28L32L33	73	+/-	

^aPA-Oct-1 POU-homeodomain fusion proteins were incubated in DNA-protein binding reactions with subsaturating amounts of PA- α TIF and C1 factor as described in Materials and Methods and in the legend to Figure 3. The amount of C1 complex formed with each mutant protein was compared to that formed with the wild type at equivalent DNA binding activity in reactions in which 0.1-1.0% of the probe was bound by POU-homeodomain homodimer. In all cases, the concentration of mutant protein that was required to achieve equivalent DNA binding to that of the wild-type protein did not vary by more than twofold of the wild-type protein concentration. The numbers correspond to the average of at least three experiments for each mutant. The substitutions are grouped according to location in helices 1 and 2. Substitutions that were made in the intervening loop reduced the apparent DNA binding affinity of the domain and were therefore not pursued.

^bPA-Oct-1 POU-homeodomain fusion proteins were incubated in DNA-protein binding and UV-crosslinking reactions as described in Materials and Methods and in the legend to Figure 4. + indicates that the Oct-1 POU-homeodomain variant stimulated the efficiency of UV crosslinking of PA- α TIF to DNA by 4-6 fold or to a level equivalent to that observed with the wild type protein. +/- indicates 2-3 fold stimulation while - indicates <1.5 fold stimulation.

with alanine (E22A) (lane 14) in helix 1, or glutamic acid 30 with glutamine (E30Q) (lane 18) in helix 2 reduced the capacity of the Oct-1 POU-homeodomain to form a C1 complex to that of 13%, 13%, 16% and 5% of wild type, respectively. In contrast to the substitution of serine 19 with cysteine, substitution of serine 19 with glycine (S19G) resulted in a protein with only a mildly reduced ability (81% wild type) to form a C1 complex (lane 12). Oct-1 and Oct-2 differ in the POU-homeodomain at positions 12, 14, 22, 32, 33, 36, and 39. Of these, only substitution of the Oct-1 residue at position 22 with the corresponding Oct-2 residue dramatically affected the protein's ability to form a C1 complex (Figure 3, compare lanes 4, 6, 14, 20, 22, and 24, see Table 1). Stern et al. (1989) previously demonstrated that the simultaneous substitution of residues threonine 32, methionine 33, and aspartic acid 36 in the Oct-1 helix 2 with the corresponding residues from the Oct-2 POU-homeodomain reduced the efficiency of C1 complex formation. In this analysis the same combination of substitutions (T32L, M32L, D36E) had a relatively mild effect (68% wild type) (Figure 3, lane 26) compared to that which resulted from the aforementioned individual amino acid substitutions.

It is clear from these data that residues in both helices 1 and 2 of the Oct-1 POU-homeodomain mediate protein-protein interactions which are required for the efficient assembly of the C1 complex. Among those amino acids which, when substituted, result in a polypeptide which is significantly disabled with respect to the formation of a C1 complex are those which are likely to be specifically and directly involved in the recognition of the Oct-1 POU-homeodomain by the α TIF and C1 factors.

Interaction of Oct-1 POU-homeodomain variants with α TIF

The reduced ability of a mutant Oct-1 POU-homeodomain protein to participate in the formation of a C1 complex could be due to its diminished capacity to interact with α TIF, the C1 factor, or both. Although an Oct-1: α TIF:DNA intermediate complex is not readily detected in a native gel electrophoretic mobility shift assay, the cooperative DNA

binding interaction between the Oct-1 POU-homeodomain and α TIF in the absence of the C1 factor can be detected in a UV-induced crosslinking assay (Kristie and Sharp, 1990). Therefore, this assay was used to address whether the panel of mutant proteins could interact directly with α TIF. As illustrated in Figure 4, the wild type POU-homeodomain stimulated the efficiency of UV-crosslinking of α TIF to the α /IE element by 4-5 fold as compared to reactions containing only the α TIF protein (compare lanes 2 and 3). Interestingly, each of the mutant proteins which contained amino acid substitutions that significantly reduced the ability of the protein to form a C1 complex also failed to stimulate the crosslinking of α TIF to DNA (lanes 6, 7, 9, and 11). The mutant proteins which exhibited relatively milder phenotypes with respect to their assembly into C1 complexes also exhibited diminished capacities to stimulate the cooperative DNA-binding of α TIF (Table 1). Conversely, those Oct-1 POU-homeodomain variants which exhibited nearly wild type phenotypes in a C1 complex formation assay also retained the ability to stimulate the crosslinking of α TIF (lanes 4, 5, 8, 10, 12, and 13, for example, Table 1). Thus, the phenotype exhibited by all mutants which were deficient in the assembly of a C1 complex can be accounted for by their decreased abilities to interact specifically with the viral α TIF factor.

A single amino acid substitution in the Oct-2 POU-homeodomain confers the ability to efficiently participate in the formation of a C1 complex and to stimulate the cooperative DNA-binding of α TIF

Figure 5 depicts the surface of the Oct-1 POU-homeodomain which should be accessible for protein-protein interactions. The arrows denote residues 18, 19, 22, and 30, which, when substituted individually, resulted in mutant proteins with significantly reduced potential to form the C1 complex and to cooperatively interact with α TIF. These residues are thus expected to be involved in the critical interactions of the Oct-1 POU-homeodomain

Figure 4. Cooperative interactions of α TIF with variants of the Oct-1 POU-homeodomain. DNA-protein binding and UV-crosslinking reactions were done as described in Materials and Methods. The reactions contained 1100 ng SDS-PAGE purified PA- α TIF and 500 ng wild type (WT) or mutant PA-Oct-1 POU-homeodomain fusion proteins as indicated. The positions of the PA- α TIF (74 kD) and PA-Oct-1 POU-homeodomain (38 kD) proteins are indicated with arrows. The 27 kD species is a degradation product of the PA-Oct-1 POU homeodomain proteins. The bands at 17 kD and 36 kD are present in control reactions in the absence of any PA fusion protein and are thus likely to be incompletely digested DNA. The migrations of 14 C protein molecular weight markers are indicated at the right.

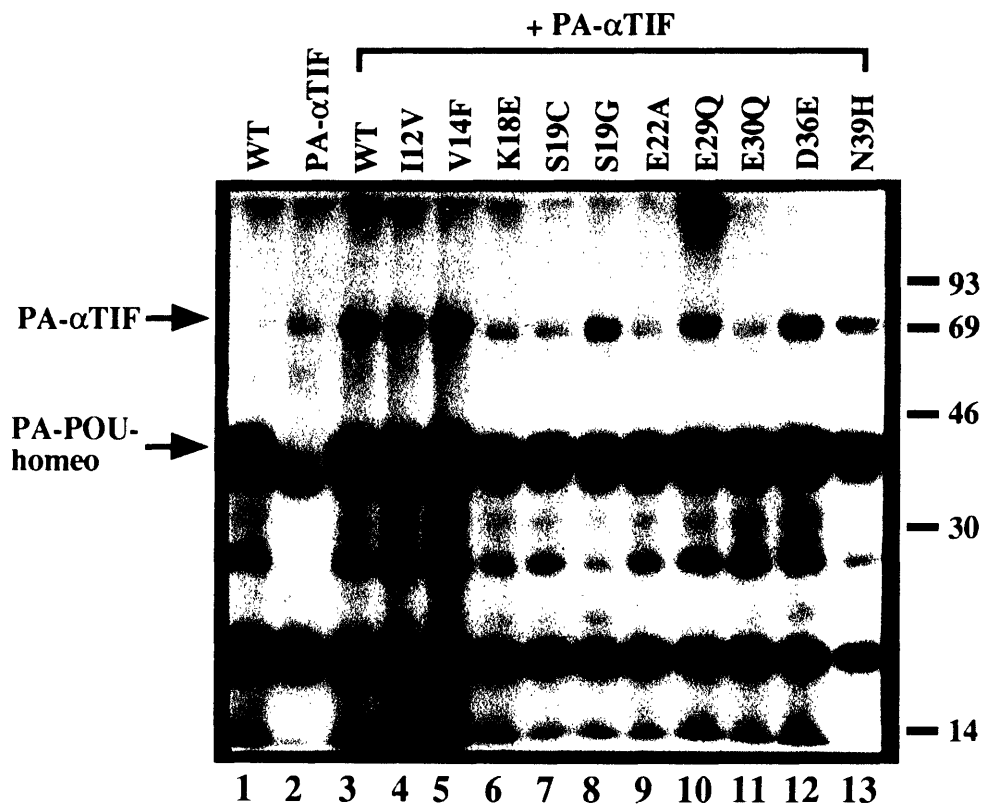
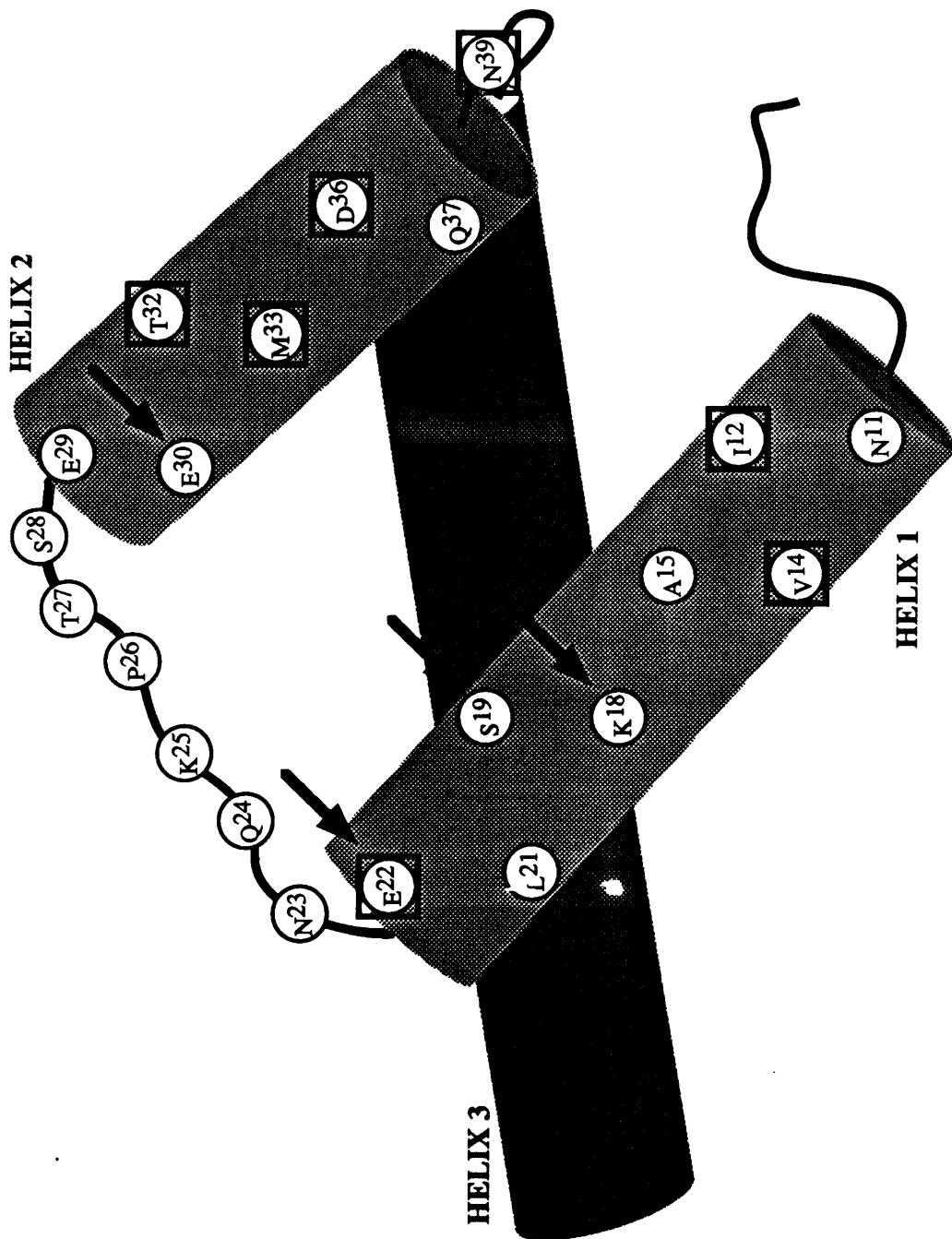


Figure 5. The surface of the Oct-1 POU-homeodomain which is specifically recognized in the assembly of the C1 complex. The Oct-1 POU-homeodomain was modeled according to the crystal structure of the engrailed homeodomain-DNA complex (Kissinger et al., 1990). The displayed amino acids are expected to occupy positions in helices 1, 2, and the intervening loop which are on the surface of the domain that is opposite to the DNA binding surface. The residues which, when substituted, resulted in the most pronounced reductions in the ability to form the C1 complex and to stimulate the cooperative DNA binding of α TIF are indicated with arrows. The amino acids which differ between Oct-1 and Oct-2 in the POU-homeodomain are boxed.



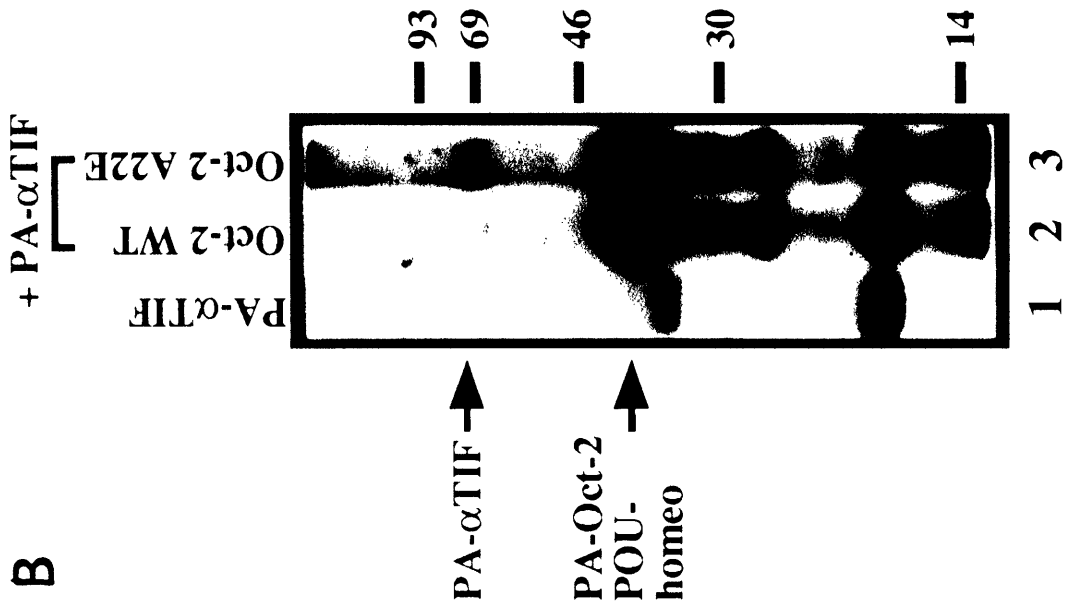
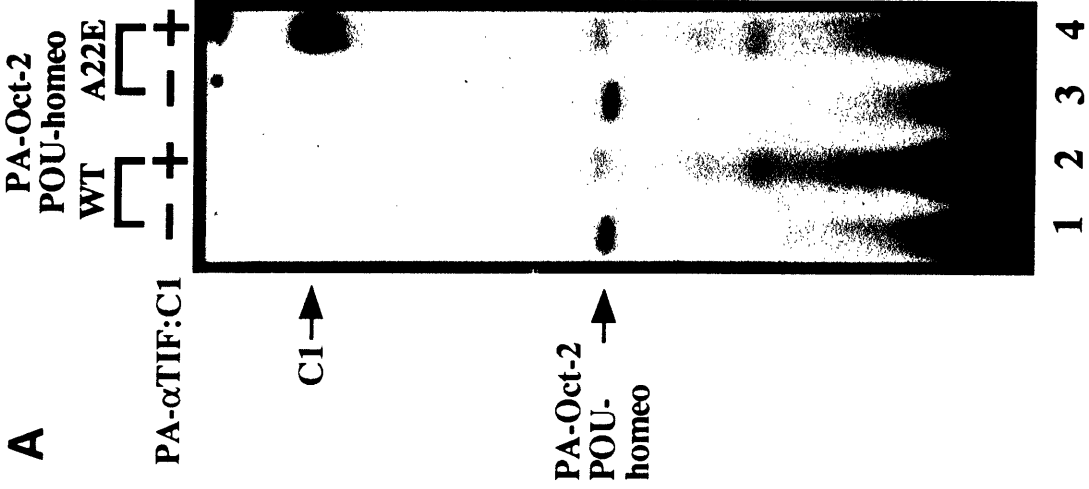
with α TIF in the assembly of a C1 complex. The amino acids which differ between the Oct-1 and Oct-2 POU-homeodomains are indicated by boxes.

Of the seven amino acid differences between Oct-1 and Oct-2 in the POU-homeodomain, substitution of the glutamic acid at position 22 in Oct-1 for the corresponding alanine in Oct-2 produced a protein which displayed the most significant phenotype. This suggested that the amino acid at this position would not only be a key determinant for the recognition of Oct-1 in the formation of a C1 complex but also one that might account for the significantly lower affinity of the Oct-2 POU-homeodomain for components of the C1 complex. Therefore, a variant of the Oct-2 POU-homeodomain was produced which contained a substitution of the Oct-2 alanine at position 22 with the Oct-1 glutamic acid. As shown in Figure 6A, this single substitution now confers upon the Oct-2 POU-homeodomain the ability to efficiently interact with components of the C1 complex (compare lanes 2 and 4). This Oct-2 variant bound to the α /IE element as a monomeric protein with an affinity that was comparable to that of the wild type Oct-2 POU-homeodomain (lanes 1 and 3, data not shown). Based on the results obtained with the Oct-1 variants, it was expected that the enhanced capacity of this protein to form a C1 complex would be concomitant with an increased ability to directly interact with α TIF. As shown in Figure 6B, the Oct-2 POU-homeodomain variant does, in fact, have an increased ability to stimulate the binding of α TIF to DNA in the UV-induced crosslinking assay as compared to the wild type Oct-2 POU-homeodomain protein (compare lanes 2 and 3).

DISCUSSION

Although it is clear that homeodomain proteins are critical for many highly regulated processes, the mechanisms by which these proteins act with exquisite functional specificity have been largely undefined. The functions of homeodomain proteins have been most extensively investigated in *Drosophila* morphogenesis, in which the regulatory specificity

Figure 6. The substitution of alanine 22 with glutamic acid in the Oct-2 POU-homeodomain enhances the ability of the protein to form a C1 complex and interact with α TIF. (A) The wild type (WT) PA-Oct-2 POU-homeodomain (50 ng) or a mutant containing the substitution of alanine 22 with glutamic acid (A22E) (50 ng) were incubated in DNA-protein binding reactions in the absence (-) or presence(+) of 15 ng PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor as indicated. The positions of the C1 and monomeric Oct-2 POU-homeodomain-DNA complexes are indicated with arrows. (B) DNA-protein binding and UV-crosslinking reactions were done using 930 ng SDS PAGE-purified PA- α TIF, alone or in the presence of 500 ng of either the wild type (WT) PA-Oct-2 POU homeodomain or the A22E mutant. The positions of the PA- α TIF (74 kD) and PA-Oct-2 POU-homeodomain (38 kD) are indicated with arrows. The migrations of 14 C protein molecular weight markers are indicated at the right.



of several proteins has been mapped to a minimal region containing the homeodomain (Kuziora and McGinnis, 1989; Gibson et al., 1990; Malicki et al., 1990; Mann and Hogness, 1990; McGinnis et al., 1990; Lin and McGinnis, 1992; Furukubo-Tokunaga et al., 1992). This specificity cannot solely be explained by the DNA-binding properties of these homeodomains since proteins with distinct biological actions exhibit extremely similar DNA sequence specificities. We propose that the biological specificity contained in the homeodomain can be primarily determined by the recognition of the surface of the DNA-bound domain by other proteins. This is clearly the mechanism by which Oct-1 is selectively assembled into the regulatory C1 complex on the α /immediate early enhancer of the Herpes Simplex Virus. The recognition of Oct-1 is mediated principally by specific interactions between the viral α TIF factor and amino acid residues on the surface of helices 1 and 2 of the Oct-1 POU-homeodomain.

Amino acid residues in helix 1 and helix 2 of the Oct-1 POU-homeodomain are critical for the assembly of the C1 complex

A panel of variants of the Oct-1 POU-homeodomain was generated by site-directed mutagenesis. In order to preserve the Oct-1 POU-homeodomain's structure and its ability to bind DNA, the mutagenesis was limited to those amino acids which are divergent and which are predicted to occupy positions on the surface of the domain. In most cases, the character of an individual side chain residue was radically changed by the substitution, and therefore a mutation which did not result in a dramatic phenotype indicates that the wild type residue at that position is unlikely to mediate a critical interaction. Conversely, a mutation which did result in a pronounced phenotype indicates that the wild type residue is either directly interacting with components of the C1 complex or is sterically constrained in the assembled complex. Specifically, the individual substitution of lysine 18, serine 19, glutamic acid 22, and glutamic acid 30 resulted in the most dramatic reductions in the ability

of the Oct-1 POU-homeodomain to participate in the formation of a C1 complex. Although lysine 18 may be specifically recognized in the C1 complex, the phenotype may also result from a charge or steric incompatibility of the glutamic acid introduced at that position. Serine 19 is unlikely to mediate a direct, specific interaction since only mild effects were observed upon substitution of this residue with glycine. However, a more dramatic phenotype resulted from the substitution of this residue with cysteine, which represents the replacement of a hydroxyl group with a sulfhydryl group. This suggests that the side chain of this residue is sterically constrained in the assembled complex. The result of the replacement of glutamic acid 30 with glutamine indicates that this residue is in intimate proximity to and may, in fact, mediate a direct contact with components of the C1 complex. Finally, it is likely that glutamic acid 22 is specifically recognized in the C1 complex via a direct interaction since the removal of much of its side chain (substitution with alanine) resulted in a significant phenotype. More significantly, the replacement of alanine with glutamic acid at position 22 in the Oct-2 POU-homeodomain resulted in a variant with a dramatically enhanced ability to form a C1 complex.

Recognition of the Oct-1 POU-homeodomain by α TIF

Each amino acid substitution which compromised the formation of the C1 complex also reduced the ability of the Oct-1 POU-homeodomain to cooperatively stimulate the binding of α TIF to DNA, indicating that the surface of the Oct-1 POU-homeodomain is principally recognized by the viral α TIF protein. To date, the specific role of the C1 factor in the stabilization of the C1 complex is unclear. This multicomponent factor binds to α TIF in the absence of DNA or Oct-1 (Kristie and Sharp, 1990). A related activity is also present in insect cells (Kristie et al., 1989), suggesting that the C1 factor is evolutionarily conserved.

The importance of glutamic acid 22 in the Oct-1 POU-homeodomain

Stern et al. (1989) described a qualitative loss of the ability of the full Oct-1 protein to form a C1 complex when residues at positions 32, 33, and 36 in helix 2 were simultaneously substituted with the corresponding Oct-2 residues. The interpretation of this data was that helix 2 of the Oct-1 POU-homeodomain contained the critical determinants for the assembly of Oct-1 into a C1 complex. In contrast, the equivalent construct in this analysis (T32L, M33L, D36E) exhibited only a mildly reduced ability (68 % wild type) to form the C1 complex. Other mutants which contained simultaneous substitutions of threonine 32 and methionine 33 had intermediate phenotypes in the C1 complex formation assay (34-77% wild type) and exhibited diminished abilities to stimulate the crosslinking of α TIF to DNA. However, the data presented here suggest that the most significant individual determinant for the discrimination between Oct-1 and Oct-2 lies not in helix 2 but in helix 1 at position 22. Substitution of the glutamic acid in Oct-1 for the alanine in Oct-2 produced an Oct-1 variant with a dramatically reduced ability to form a C1 complex and to interact with α TIF. The reciprocal exchange (Oct-2 A22E) resulted in an Oct-2 POU-homeodomain protein with a significantly enhanced ability to form a C1 complex and to directly interact with α TIF. This strongly implicates the glutamic acid at position 22 as a critical determinant for the selective recognition of Oct-1 in the formation of the C1 complex.

The cooperative binding of the Oct-1 POU-homeodomain

In contrast to the intact Oct-1 POU domain, the Oct-1 POU-homeodomain can bind the α /IE element cooperatively as a homodimer (discussed in Kristie and Sharp, 1990). The difference in the abilities of the isolated Oct-1 and Oct-2 POU-homeodomains to cooperatively form a homodimer on the α /IE element was surprising but it is unlikely to be

related to the proteins' respective capacities to interact with other components in the formation of a C1 complex. The amino acid substitutions in Oct-1 which diminished its ability to form a C1 complex did not affect its potential to form a homodimer on the α /IE element. Conversely, the mutation in Oct-2 (A22E) which enhanced its capacity to form a C1 complex did not confer the ability to form a homodimer. The differences which account for the different potentials of the Oct-1 and Oct-2 POU-homeodomains to cooperatively homodimerize are presently under investigation. The homodimerization of the Oct-1 protein may reflect unique capabilities of this protein for homeodomain-homeodomain interactions with other proteins. Consistent with this notion, Voss et al. (1991) have demonstrated that Oct-1 and the pituitary-specific POU domain factor Pit-1 bind as a heterodimer to elements in the rat prolactin promoter and that the two proteins also associate in solution via an interaction mediated, in part, by the POU-homeodomain of Pit-1. Of note, Treacy et al. (1992) have described the interesting regulatory consequences of the contrasting abilities of the *Drosophila* I-POU and Twin of I-POU proteins to interact with the Cfl-a protein via POU-homeodomain-POU-homeodomain interactions.

The interaction of Oct-1 and α TIF provides a model for the determination of homeodomain functional specificity by protein-protein interactions

It is possible that the POU-homeodomains of Oct-1 and Oct-2 are recognized by cellular factors which are analogous to α TIF and which serve to modulate their functional specificities. The existence of such factors would explain how the octamer element, which is recognized by both of these proteins, is important for the regulation of a wide variety of disparately controlled genes including the constitutively expressed snRNA genes (Ares et al., 1987; Bark et al., 1987; Carbon et al., 1987; Murphy et al., 1987), the cell-cycle-specifically expressed histone H2B gene (Sive et al., 1986; Fletcher et al., 1987; LaBella et al., 1988), and the tissue-specifically expressed interleukin-2 (Ullman et al., 1991) and

immunoglobulin genes (Mizushima-Sugano and Roeder 1986; Staudt et al., 1986; Scheidereit et al., 1987; Wirth et al., 1987; LeBowitz et al., 1988; Muller et al., 1988; Gerster et al., 1987).

It is likely that the recognition of the homeodomain surface by other proteins is a common mechanism by which the regulatory specificities of this family of factors are determined. In *Drosophila*, where the developmental regulatory potential of a particular factor can be readily assayed, minimal amino acid differences between two homeodomains can determine the distinct biological actions of the proteins (Lin and McGinnis, 1992). The observation that this biological specificity may not be related to DNA-binding specificity is readily explained by a model based upon the recognition of the surface of the homeodomain such as that proposed here for the Oct-1 POU-homeodomain. In this system, the four critical residues for C1 complex formation are clustered in the C-terminal half of helix 1 and the N-terminal third of helix 2. Therefore, this surface of the Oct-1 POU-homeodomain must be exposed to components of the C1 complex when the Oct-1 protein binds DNA *in vivo*. This spatial architecture is probably a general feature of homeodomain proteins: the surface of the domain is exposed in the protein-DNA complex, providing a target for regulatory proteins which contribute to homeodomain functional specificity.

MATERIALS AND METHODS

Mutagenesis and production of PA fusion proteins. Constructs encoding the *S. aureus* protein A fusion proteins which contained the Oct-1 POU domain [amino acids 270-441, (Sturm et al., 1988)] and α TIF [amino acids 1-412, (Pellett et al., 1985)] have been described (Kristie and Sharp, 1990). Plasmid pO1HSS was constructed by cloning the DNA fragment encoding the Oct-1 POU-homeodomain (amino acids 368-441) into the vector pBS(+) (Stratagene) using the *E. coli* TG1 strain. The resultant transformant was grown with R408 Helper Phage (Stratagene) at 37°C for 18 hr. Single stranded DNA was

isolated from the culture supernatant following the addition of 0.25 volume 20% PEG (8000)/3.5M NH₄Ac, incubation at 4°C for 30 minutes, and centrifugation at 10000 g for 20 minutes. The single stranded DNA pellet was extracted five times with phenol/chloroform [1 vol:1vol], precipitated with ethanol and resuspended for use as a substrate for mutagenesis. Mutagenesis was performed using the Oligonucleotide-directed in-vitro mutagenesis system version 2 (Amersham) according to the manufacturers instructions. Oligonucleotides used for the mutagenesis were from 20 to 30 nucleotides long and contained approximately 10 nucleotides on each side of the mutation-specific base(s). The products of the mutagenesis reactions were screened by dideoxynucleotide sequencing. The fragments which contained the desired mutations were isolated and cloned into pRIT2T(Pharmacia) so as to generate in frame fusions with the PA gene.

Plasmid pO2HWT was constructed by cloning the DNA fragment encoding the Oct-2 POU-homeodomain [amino acids 286-356, (Clerc et al., 1988)] into pRIT2T so as to generate an in frame fusion with the PA gene. pO2HA317E, a construct encoding the Oct-2 POU-homeodomain mutant (alanine 317 to glutamic acid) was generated by making a base substitution via Recombinant PCR (Higuchi, 1990). The sequences of both Oct-2 constructs were verified by dideoxynucleotide sequencing.

Protein A fusion proteins were expressed in E. coli N4830 strain and were purified by affinity chromatography on IgG-Sepharose (Pharmacia) as described (Kristie and Sharp, 1990). The concentration and purity of each PA fusion protein were determined by densitometric analysis of Coomassie-staining SDS-PAGE-resolved proteins. The fusion proteins were judged to be 30-90% pure.

Electrophoretic Mobility Shift Assays. The HSV α 0 probe (α /IE element: 5'-GTGCATGCTAATGATATTCTTTGGGG-3') used in the C1 complex formation assays has been described (Kristie and Sharp, 1990). DNA-protein binding reactions contained 0.4-0.8 ng DNA probe, 300 ng poly[d(I-C)]/poly[d(I-C)], 10mM HEPES pH 7.9, 0.5 mM

EDTA, 30-90 mM KCl, 0.75 mM DTT, 4% Ficoll 400, 300 µg/ml bovine serum albumin, and the appropriate purified proteins or chromatographic fraction in a total volume of 10 µl. Reactions were incubated at 30°C for 30 minutes and were resolved in 4% nondenaturing polyacrylamide gels using 0.5x Tris-glycine electrophoresis buffer as described (Fried and Crothers, 1981; Garner and Revzin, 1981). Chromatographic fractions which contained the C1 factor were prepared by fractionation of a nuclear extract of HeLa cells (Dignam et al., 1983). The extract was applied to a Mono S FPLC column in buffer A +100 mM KCl [40 mM HEPES pH 7.9/0.5 mM EDTA/0.5 mM DTT/20% (vol/vol) glycerol]. The column was washed with 10 column volumes of buffer A + 100 mM KCl, and the adsorbed proteins were eluted in buffer A with a linear gradient of 100-700 mM KCl. Fractions which contained the C1 factor activity were applied to a Mono Q FPLC column in buffer A + 50 mM KCl. The column was washed with 10 column volumes of buffer A + 50 mM KCl, and the adsorbed proteins were eluted in buffer A with a linear gradient of 50-700 mM KCl. The fractions which contained the C1 factor activity were combined and aliquots were incubated with potato acid phosphatase (SIGMA) for 15 minutes at 25°C prior to addition to protein-DNA binding reactions. In order to compare DNA binding affinities, wild type and mutant PA-POU-homeodomain proteins were titrated into DNA-protein binding reactions which were performed under the conditions described above using a probe containing a consensus octamer site (5'ATGCAAAT3') (Kristie and Sharp, 1990), 30 ng poly[d(I-C)]/poly[d(I-C)], and between 5 and 500 ng of octamer binding protein. The protein-DNA complexes and the free DNA were quantitated after electrophoresis using a phosphorimager (Molecular Dynamics) with ImageQuant 3.0 and 3.15 software.

UV-induced crosslinking reactions. Body labeled DNA probes were prepared as previously described (Kristie and Sharp, 1990). The DNA-protein binding reactions were performed under the conditions described above and contained 40-150 ng poly[d(I-C)]/poly[d(I-C)], 200-500 ng PA-POU-homeodomain protein, and 200-1100 ng PA-

α TIF. Reactions were incubated at 30°C for 30 minutes and were irradiated with a Fotodyne UV lamp (254nm) at 3000 μ W/cm² in a 15°C water bath. The reactions were brought to 7.5 mM CaCl₂ and digested with 10 μ g DNase 1 and 5 units micrococcal nuclease for 30 minutes at 37°C. The digested products were resolved in an 11% SDS-denaturing gel and transferred to nitrocellulose in the presence of 0.1% SDS. The PA- α TIF used in these reactions was purified by preparative SDS-PAGE as described (Kristie and Sharp, 1990). The amount of crosslinking of α TIF to DNA was quantitated using a phosphorimager (Molecular Dynamics) with ImageQuant 3.15 software.

ACKNOWLEDGEMENTS

We thank A. Gil, D. Chasman, P. McCaw, L. Johnson, R. Meyers, and C. Pabo for provocative discussions and critical review of this manuscript; M. Siafaca for her ever-present assistance; L. Parent and C. Wrenn for indispensable technical assistance; and members of the Sharp laboratory for their continual support. J.L.P. is supported by the Harvard Medical School M.D.-Ph.D. program by Public Health Service National Research Service Award 2 T32 GM07753-13 and by the M.I.T. Department of Biology. T.M.K. is a Leukemia Society of America Special Fellow. This work was supported by U.S. Public Health Service grant PO1-CA42063 from the N.I.H., by cooperative agreement CDR-8803014 from the National Science Foundation to P.A.S. and partially by N.C.I. Cancer Center Support (core) grant P30-CA14051.

REFERENCES

- Affolter, M., A. Schier, and W.J. Gehring. 1990. Homeodomain proteins and the regulation of gene expression. Current Opinion in Cell Biology 2: 485-495.
- Ares, M. Jr., J. Chung, L. Giglio, and A.M. Weiner. 1987. Distinct factors with SP1 and NFA specificities bind to adjacent functional elements of the human U2 snRNA gene enhancer. Genes & Dev. 1: 808-817.
- Bark, C., P. Weller, J. Zabeilski, L. Janson, and U. Petterson. 1987. A distant enhancer element is required for transcription of a U6 RNA gene. Nature 328: 356-362.
- Bzik, D. J. and C.M. Preston. 1986. Analysis of DNA sequences which regulate the transcription of herpes simplex virus immediate early gene 3: DNA sequences required for enhancer-like activity and response to trans-activation by a virion polypeptide. Nucleic Acid Res. 14: 929-943.
- Carbon, P., S. Murgo, J-P. Ebel, A. Krol, G. Tebb, and I.W. Mattaj. 1987. A common octamer motif binding protein is involved in the transcription of the U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. Cell 51: 71-79.
- Clerc, R. G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore, and P.A. Sharp. 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. Genes & Dev. 2: 1570-1581.
- Dignam, J.D., R.M. Lebovitz, and R. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11: 1475-1489.
- Fletcher, C., N. Heintz, and R.G. Roeder. 1987. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell 51: 773-781.
- Fried, M. and D. Crothers. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9: 6505-6525.
- Furukubo-Tokunaga, K., M. Muller, M. Affolter, L. Pick, U. Kloter, and W.J. Gehring. 1992. In vivo analysis of the helix-turn-helix motif of the fushi tarazu homeo domain of *Drosophila melanogaster*. Genes & Dev. 6: 1082-1096.
- Gaffney, D. F., J. McLauchlan, J.L. Whitton, and J.B. Clements. 1985. A modular system for the assay of transcription regulatory signals: the sequence TAATGARAT is required for herpes simplex virus immediate early gene activation. Nucleic Acids Res. 13: 7874-7862.
- Garner, M.N. and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions. Nucleic Acids Res. 9: 3047-3060.
- Gehring, W.J. 1987. Homeo boxes in the study of development. Science 236: 1245-1252.
- Gerster, T. and R.G. Roeder. 1988. A herpesvirus trans-acting protein interacts with transcription factor OTF-1 and other cellular proteins. Proc. Natl. Acad. Sci. USA 85: 6347-6351.

Gerster, T., P. Matthias, M. Thali, J. Jiricny, and W. Schaffner. 1987. Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer. EMBO J. **6**: 1323-1330.

Gibson, G., A. Schier, P. LeMotte, and W.J. Gehring. 1990. The specificities of sex combs reduced and antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell **62**: 1087-1103.

Hayashi, S. and M.P. Scott. 1990. What determines the specificity of action of Drosophila homeodomain proteins? Cell **63**: 883-894.

He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson, and M.G. Rosenfeld. 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature **340**: 35-42.

Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkin, and H.R. Horvitz. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes & Dev. **2**: 1513-1516.

Higuchi, R. 1990. Recombinant PCR. In PCR protocols: a guide to methods and applications (ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White), pp. 177-183. Academic Press, Inc., San Diego, California.

Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell **63**: 579-590.

Kristie, T. M. and B. Roizman. 1984. Separation of sequences defining basal expression from those conferring alpha gene recognition within the regulatory domains of herpes simplex virus 1 alpha genes. Proc. Natl. Acad. Sci. USA **81**: 4065-4069.

Kristie, T. M., J.H. LeBowitz, and P.A. Sharp. 1989. The octamer-binding proteins form multi-protein-DNA complexes with the HSV alpha TIF regulatory protein. EMBO J. **8**: 4229-4238.

Kristie, T.M., and P.A. Sharp. 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV α -trans-activator protein. Genes & Dev. **4**: 2383-2396.

Kuziora, M.A., and W. McGinnis. 1989. A homeodomain substitution changes the regulatory specificity of the deformed protein in Drosophila embryos. Cell **59**: 563-571.

La Bella, F., H.L. Sive, R.G. Roeder, and N. Heintz. 1988. Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. Genes & Dev. **2**: 32-39.

Laughon, A. (1991) DNA binding specificity of homeodomains. Biochemistry **30**: 11357-11367.

- LeBowitz, J. H., T. Kobayashi, L. Staudt, D. Baltimore, and P.A. Sharp. 1988. Octamer-binding proteins from B or HeLa cells stimulate transcription of the immunoglobulin heavy-chain promoter in vitro. Genes & Dev. **2**: 1227-1237.
- Lin, L. and W. McGinnis. 1992. Mapping functional specificity in the Dfd and Ubx homeodomains. Genes & Dev. **6**: 1071-1081.
- Mackem, S. and B. Roizman. 1982. Regulation of alpha genes of herpes simplex virus: the alpha 27 promoter-thymidine kinase chimeric gene is positively regulated in converted L cells. J. Virol. **43**: 1015-1023.
- Mackem, S. and B. Roizman. 1982. Structural features of the alpha gene 4, 0, and 27 promoter-regulatory sequences which confer alpha regulation on chimeric thymidine kinase genes. J. Virol. **44**: 939-949.
- Mackem, S. and B. Roizman. 1982. Differentiation between alpha promoter and regulator regions of herpes simplex virus I: The functional domains and sequence of a movable alpha regulator. Proc. Natl. Acad. Sci. USA **79**: 4917-4921.
- Malicki, J., K. Schughart, and W. McGinnis. 1990. Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. Cell **63**: 961-967.
- Mann, R.S., and D.S. Hogness. 1990. Functional dissection of ultrabithorax proteins in D. melanogaster. Cell **60**: 597-610.
- McGinnis, N., M.A. Kuziora, and W. McGinnis. 1990. Human Hox-4.2 and Drosophila deformed encode similar regulatory specificities in Drosophila embryos and larvae. Cell **63**: 969-976.
- McKnight, J. L. C., T.M. Kristie, and B. Roizman. 1987. The binding of the virion protein mediating alpha gene induction in herpes simplex virus 1 infected cells to its cis site requires cellular proteins. Proc. Natl. Acad. Sci. USA **84**: 7061-7065.
- McKnight, J. L. C., T.M. Kristie, S. Silver, P.E. Pellett, G. Mavromara-Nazos, G. Campadelli-Fiume, M. Arsenakis, and B. Roizman. 1986. Regulation of herpes simplex virus 1 gene expression: the effect of genomic environments and its implication for model systems, In Cancer Cell, vol IV, pp. 163-173, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mizushima-Sugano, J. and R.G. Roeder. 1986. Cell-type-specific transcription of an immunoglobulin k light chain in vitro. Proc. Natl. Acad. Sci. USA **83**: 8511-8515.
- Muller, M. M., S. Ruppert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature **336**: 544-551.
- Murphy, S., C. Di Leigro, and M. Melli. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell **51**: 81-87.
- O'Hare, P. and C.R. Goding. 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virus transactivation. Cell **52**: 435-445.

Otting, G., Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Gehring, and K. Wuthrich. 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. **9**: 3085-3092.

Pellett, P. E., J.L.C. McKnight, F.J. Jenkins, and B. Roizman. 1985. Nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of trans-inducing alpha genes. Proc. Natl. Acad. Sci. USA **82**: 5870-5874.

Preston, C. M., M.C. Frame, M.E.M. Campbell. 1988. A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. Cell **52**: 425-434.

Qian, Y.Q., M. Billeter, G. Otting, M. Mueller, W.J. Gehring, and K. Wuthrich. 1989. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell **59**: 573-580.

Rosenfeld, M.G. 1991. POU-domain transcription factors: pou-er-ful developmental regulators. Genes & Dev. **5**: 897-907.

Scheidereit, C., A. Heguy, and R.G. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. Cell **51**: 783-793.

Scott, M. P., J.W. Tam Kun, and G.W. Hartzell, III. 1989. The structure and function of the homeodomain. BBA Rev. Cancer **989**: 25-48.

Sive, H. L., N. Heintz, R.G. Roeder. 1986. Multiple sequence elements are required for maximal in vitro transcription of a human histon H2B gene. Mol. Cell Biol. **6**: 3329-3340.

Staudt, L. M., H. Singh, R. Sen, T. Wirth, P.A. Sharp, and D. Baltimore. 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature **323**: 640-643.

Stern, S. and W. Herr. 1991. The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain. Genes & Dev. **5**: 2555-2566.

Stern, S., M. Tanaka, and W. Herr. 1989. The Oct-1 homeo domain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature **341**: 624-630.

Sturm, R. A., G. Das, and W. Herr. 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes & Dev. **2**: 1582-1599.

Treacy, M.N., L.I. Neilson, E.E. Turner, X. He, and M.G. Rosenfeld. 1992. Twin of I-POU: a two amino acid difference in the I-POU homeodomain distinguishes an activator from an inhibitor of transcription. Cell **68**: 491-505.

Ullman, K.S., W.M. Flanagan, C.A. Edwards, and G.R. Crabtree. 1991. Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, OAP⁴⁰. Science **254**: 558-562.

Voss, J.W., L. Wilson, and M.G. Rosenfeld. 1991. POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. Genes & Dev. **5**: 1309-1320.

Wirth, T., L. Staudt, and D. Baltimore. 1987. An octamer oligonucleotide upstream of a TATA element is sufficient for lymphoid-specific promoter activity. Nature **329**: 174-177.

Wolberger, C., A.K. Vershon, B. Liu, A.D. Johnson, and C.O. Pabo. 1991. Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67**: 517-528.

Xiao, P. and J.P. Capone. 1990. A cellular factor binds to the herpes simplex virus type 1 transactivator Vmw65 and is required for Vmw65-dependent protein-DNA complex assembly with Oct-1. Mol. Cell. Biol. **10**: 4974-4977.

CHAPTER III

HOMEODOMAIN DETERMINANTS OF MAJOR GROOVE RECOGNITION

This chapter originally appeared in *Biochemistry* **33**: 10851-10858 (1994).

ABSTRACT

The homeodomain is a highly conserved structural module that binds DNA and participates in protein-protein interactions. Most homeodomains contain residues at positions 47 and 51 which mediate recognition of a TAAT core binding sequence in the major groove. The constraints imposed on the identity of these residues by homeodomain structure and DNA-docking have been examined in the context of the POU domain of the Oct-1 transcription factor. A bacterial library, in which POU-homeodomain residues 47 and 51 have been randomized, was probed on nitrocellulose filters for the binding of DNA fragments containing the consensus octamer sequence. The residues which provide for the highest affinity interaction with the octamer consensus sequence, and the greatest specificity, are the highly conserved wild-type residues valine 47 and asparagine 51. Interestingly, a class of variants containing arginine at position 51 was also detected in the screen and found to have moderate affinity for the consensus sequence but reduced specificity compared to the wild-type protein. A single variant containing arginine at both positions 47 and 51 was detected when the library was probed with fragments containing nucleotide substitutions at positions expected to be contacted by residues 47 and 51. This variant was used to alter the DNA-binding specificity of a transcriptional regulatory complex which depends upon Oct-1 for DNA recognition. These findings suggest that homeodomain structure and DNA-docking constrain the versatility of the domain in that only a limited set of amino acid determinants can endow the domain with specific, high affinity DNA binding.

INTRODUCTION

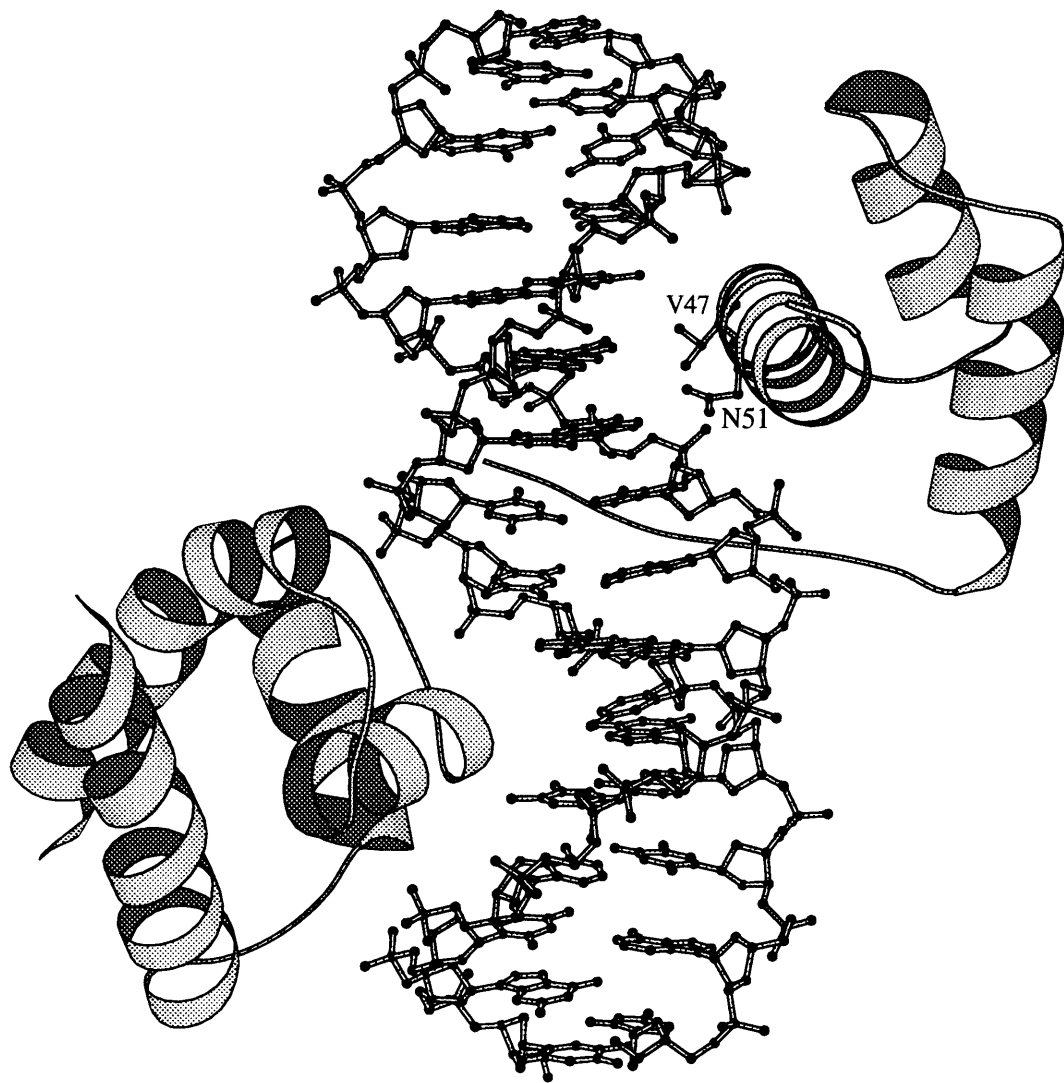
The homeodomain is a structural module that determines the specificity of action of a wide variety of transcription factors. This specificity is conferred by both its DNA-binding surface as well as by surfaces that are targets for protein-protein interactions with other transcriptional regulators. Structural analyses of homeodomain-DNA complexes have revealed a highly conserved structure and mode of docking DNA (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991; Klemm et al., 1994). These studies have complemented extensive biochemical and genetic experiments addressing the interaction of the homeodomain with its conserved 5'-TAATNN-3' binding sequence (Laughon, 1991). The homeodomain is composed of an N-terminal arm which makes contacts with bases (5'-TAATNN-3') in the minor groove of DNA, and three α helices, the third of which makes base contacts in the major groove (5'-TAATNN-3'). For several homeodomains, specificity for nucleotides 3' to the TAAT core is determined by residue 50 in helix 3 (Hanes and Brent, 1989, 1991; Treisman et al., 1989; Percival-Smith et al., 1990). Residues that contact the sugar-phosphate backbone are distributed throughout the domain.

The highly conserved nature of homeodomain-DNA interactions raises the issue of how different homeodomain proteins can exert specific regulatory effects, especially when functional specificity maps to the homeodomain itself (Hayashi & Scott, 1990). The observation that different homeodomains can determine dramatically different biological actions despite little or no difference in DNA-binding specificity has led to the description of mechanisms of specificity other than that provided by monomeric DNA-binding. These involve homo- and hetero-dimerization of homeodomains and the cooperative interaction of the homeodomain with other regulatory proteins (Smith & Johnson, 1992; Pomerantz et al., 1992; Lai et al., 1992; Vershon & Johnson, 1993; Wilson, D. et al., 1993).

The limited variation in the specificity of homeodomains for a TAAT core binding site emphasizes the question of how the highly conserved tertiary structure of the domain constrains its DNA-binding specificity for this core sequence. Most of the amino acid determinants in the homeodomain that form specific contacts with bases in the TAAT core are highly conserved among homeodomains (Laughon, 1991). This raises the possibility that only a limited set of amino acids can occupy these positions and make specific base contacts in the context of homeodomain structure and DNA docking .

The POU-homeodomain of the Oct-1 transcription factor participates in prototypical homeodomain-DNA interactions in that specific recognition in the major groove is mediated by residues asparagine (N) 51 and valine (V) 47 (Klemm et al., 1994). Oct-1 is a founding member of the POU domain subfamily of homeodomain transcription factors (Herr et al., 1988) which recognizes the 5' half of the consensus octamer site (5'-ATGCAAAT-3') with the POU-specific domain and the 3' half (5'-ATGCAAAT-3') with the POU-homeodomain (Figure 1). In the DNA-protein complex, the adenine at position 7 (5'-ATGCAAAT-3') is contacted by POU-homeodomain residue N51. This interaction corresponds to that observed in the engrailed:DNA (Kissinger et al., 1990) and MAT α 2:DNA (Wolberger et al., 1991) complexes, in which the asparagine accepts a hydrogen bond from the N6 of adenine and donates a hydrogen bond to the N7. Asparagine 51 is one of the most highly conserved residues in all homeodomains, and its interaction with the core adenine is believed to be a signature feature of homeodomain-DNA binding. The thymine at position 8 (5'-ATGCAAAT-3') interacts with V47 via a van der Waals contact involving the thymine methyl group (Klemm et al., 1994). Homeodomain residue 47 is most often either an isoleucine or a valine, and in the engrailed:DNA cocrystal structure, isoleucine 47 contacts the counterpart thymine (5'-TAAT-3') via an analogous interaction. Thus, Oct-1 POU-homeodomain residues V47 and N51 represent determinants which are used by most homeodomains for base-specific contacts to the TAAT core sequence in the major groove of DNA.

Figure 1. The Oct-1 POU domain bound to DNA. Residues 47 and 51 in helix 3 of the POU-homeodomain are shown in ball-and-stick representation interacting with adenine 7 (5'-ATGCAAAT-3') and thymine 8 (5'-ATGCAAT-3') of the octamer sequence, respectively. This figure was generated by MOLSCRIPT (Kraulis, 1991) using the coordinates of Klemm et al. (1994).



3'
T 8
A 7
A 6
A 5
C 4
G 3
T 2
A 1
5'

In order to examine how the structural context of homeodomain-DNA binding constrains the nature of amino acids involved in specific interactions with the TAAT core sequence, an Oct-1 POU domain bacterial expression library was generated in which residues 47 and 51 in the POU-homeodomain were randomized. This library was prepared so that the DNA-binding of a large number of POU domain variants could be screened by the direct binding of radioactive DNA probes on nitrocellulose filters. This library has been used to investigate what residues at position 47 and 51 are consistent with binding to the octamer consensus site. In addition, the library has been used to detect Oct-1 POU domain variants which have novel DNA-binding specificities, and which can alter the binding specificity of a transcriptional regulatory complex which depends upon Oct-1 for DNA recognition.

MATERIALS AND METHODS

Construction and screening of the bacterial PA-Oct-1 POU domain expression library. A Protein A-Oct-1 POU-domain expression library was constructed by a strategy we refer to as Deletion Insertion Randomization (DIR). DIR is useful when restriction sites flanking a region desired to be randomized are not available and cannot be engineered silently. Oligonucleotide-mediated mutagenesis (Sayers et al., 1992) is employed using a population of oligonucleotides which has been synthesized to randomize codons. In order to avoid selective hybridization of the oligonucleotides that contain bases which are complementary to the wild-type ssDNA substrate, a previous round of mutagenesis is performed that deletes the region of desired randomization. The construct containing the deletion is used to prepare a ssDNA substrate for the insertion randomization mutagenesis. A fragment encoding a portion of the Oct-1 POU domain (amino acids 345-441 (Sturm et al., 1988)) was cloned into the vector pBS+(Stratagene). ssDNA was prepared using VCSM13 helper phage (Stratagene) and XL1-Blue host strain according

to the manufacturer's protocol. This ssDNA was the substrate for mutagenesis performed using the oligonucleotide-directed mutagenesis system version 2 (Amersham) according to the manufacturer's instructions. Deletion mutagenesis, designed to delete the region of DNA encoding POU-homeodomain residues 45-53 (homeodomain numbering scheme of Qian et al., 1989) in helix 3, was accomplished using the oligonucleotide 5'-CAATATGGAAAAAGAGGTGCAGAAAGAAAAAAGAATCC-3'. From the resulting construct, Δ 45-53, ssDNA was generated and used as a template for insertional mutagenesis designed to replace the deleted region with the codons for homeodomain residues 47 and 51 randomized as NNG/C. The mutagenic oligonucleotide used for this insertion was 5'-ATATGGAAAAAGAGGTGATTCGT NNG/CTGGTTCTGTNNG/CCGCCGCCAGAAAGAAAAAAGAATCAACC-3'. The products of the insertional mutagenesis reaction were used for transformation of XL1-Blue strain. 9500 transformed colonies were scraped from LB-agarose ampicillin (100ug/ml) plates, pooled, and plasmids were prepared by standard protocols. The fragment encoding the POU domain was excised from the plasmid pool and ligated into vector pRIT2T (Pharmacia) so as to generate an in frame fusion of the full POU domain (residues 270-411 (Sturm et al., 1988)) with the Protein A gene product of *S. aureus*. The ligation products were used to transform *E. coli* strain N4830 which was plated on LB-agar ampicillin plates for library screening. A random distribution of nucleotides at homeodomain codons 47 and 51 was confirmed by dideoxysequencing of a number of individual expression plasmids and by sequencing a sample of the total pooled expression library.

After plating, colonies were allowed to grow for 24 hr at 30°C before transfer to nitrocellulose filters (132mm, BA85/23 Schleicher and Schuell). After marking the position of filters on plates, the filters were lifted and incubated colony side up on a new set of plates at 42°C for 2 hr to induce fusion protein expression. Colony lysis was achieved by exposure to chloroform vapor for 10 min and then immersion in lysis buffer

(100 mM Tris-HCl (pH 7.8), 150 mM NaCl, 5 mM MgCl₂, 1.5% bovine serum albumin Fraction V (heat shock), 400 µg/ml lysozyme) (25ml/filter) for 10 min at 25°C. After air drying, filters were then processed through a denaturation/renaturation cycle. Filters were gently shaken in buffer J (25 mM Hepes (pH 7.9), 25 mM NaCl, 5 mM MgCl₂, 0.5 mM DTT) plus 6M Guanidine HCl for 5 min at 4°C. The Guanidine HCl concentration was diluted two fold in each of 5 steps with the removal of half of the buffer and the addition of an equal volume of buffer J, after which filters were washed twice with buffer J. Each step was incubated for 5 min at 4°C. Filters were then incubated in blocking buffer (50 mM Tris-HCl (pH 7.9), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% BSA Fraction V) (50 ml/filter) with gentle shaking for 1 hr at 25°C, then rinsed twice, 5 min each, with binding buffer (20 mM Hepes (pH 7.9), 50 mM KCl, 1 mM EDTA, 0.7 mM DTT, 0.025% NP40). During the lysis, denaturation/renaturation, and blocking steps, each filter was placed in a separate petri dish for the incubation. Filters were then gently shaken in binding buffer plus 5 µg/ml denatured sonicated salmon sperm DNA and ³²P-labelled probe at a final 1-2 x 10⁶ cpm/ml (~10⁻¹⁰ M) for 1 hr at 25°C. Filters were then washed twice for a total of 15 min with binding buffer (500 ml for up to 4 filters), blotted dry, and then exposed to Kodak X-OMAT AR film at -70°C with an intensifying screen for 12-24 hr. Positive colonies were picked, patched onto new plates and rescreened. Plasmids were isolated from colonies which rescreened positive and sequenced by dideoxysequencing to determine the residues encoded at positions 47 and 51 of the homeodomain. Probes used for screening were derived from cloning the fragment

5'-GATCCTATGCAANNACC-3'
3'-GGATACGTTNNCTGGAGCT-5'

into the XhoI and BamHI sites of pBSKII+ (Stratagene). All 16 variants were obtained, verified by dideoxysequencing and excised for use as probes by digesting with XbaI and Asp718. Fragments were labelled by the large (Klenow) fragment of *E. coli* DNA Pol I

in the presence of dGTP, dCTP, dTTP, and α -³²P-dATP, and then gel-purified on nondenaturing polyacrylamide gels.

5000 colonies were screened with the probe containing the octamer consensus sequence. For screening with the pools of mutant probes, 4000 colonies were screened by each probe pool, with each of the four probes in the pool at an equal concentration of $1-2 \times 10^6$ cpm/ml ($\sim 10^{-10}$ M).

Expression of fusion proteins. Selected variants were expressed in *E. coli* N4830 strain and were purified by affinity chromatography on IgG-Sepharose as described previously (Kristie & Sharp, 1990). The concentration of each PA fusion protein was determined by densitometric analysis of Coomassie-stained SDS-PAGE-resolved proteins using bovine serum albumin (Boehringer Mannheim) as standard.

Electrophoretic Mobility Shift Assays. Relative affinities of variants were determined by electrophoretic mobility shift assays using the same probes as those used for colony screening. DNA-protein-binding reactions contained 3-30 pg of DNA probe, 10 ng of poly[d(I-C)]/poly[d(I-C)], 10 mM Hepes (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 0.75 mM DTT, 4% Ficoll-400, 300 μ g/ml of bovine serum albumin, and 1-8000 pg of PA-Oct-1 POU domain variant in a total volume of 10 μ l. The concentration of DNA probe was always at least one order of magnitude below the apparent dissociation constant. Reactions were incubated at 30°C for 30 minutes and resolved in 4% nondenaturing polyacrylamide gels (Pomerantz et al., 1992). The protein-DNA and free DNA complexes were quantitated using a Molecular Dynamics Phosphorimager with ImageQuant 3.15 and 3.22 software. Apparent dissociation constants were determined as the inverse of the slope of the line derived from plotting fraction of probe bound/fraction of probe unbound vs. total PA-POU domain protein concentration. In all cases the lines consisted of at least four points. C1 complex formation assays were performed using the HSV α 0 probe (HSV α /IE element: 5'-GTGCATGCTAATGATATTCTTTGGGG-3') (Kristie et al., 1989) or a mutated version (HSV α /IE "GG" element : 5'-

GTGCATGCTAGGGATATTCTTTGGGG-3') that was generated by the oligonucleotide directed mutagenesis system version 2 (Amersham) according to the manufacturer's instructions. DNA-protein-binding reactions were performed as described above using 0.4-0.8 ng of DNA probe, 300 ng of poly[d(I-C)]/poly[d(I-C)], and, where indicated, 15 ng of PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor (Pomerantz et al., 1992).

Random Binding Site Selection. The probe used for random binding site selection was generated by annealing the following two oligonucleotides and polymerizing with Klenow in the presence of dGTP, dCTP, dTTP, and α -³²P-dATP : Primer R: 5'-GGCTGAGTCTGAACGGATCCN₁₃CCTCGAGACTGAGCGTTCG-3'; Primer A: 5'-CGACGCTCAGTCTCGAGG-3'. For the first round of selection 50 pg of PA-POU domain variant was incubated with 5 ng of probe in DNA-protein-binding reactions under the conditions described above in the absence of any poly[d(I-C)]/poly[d(I-C)]. In each round, reactions were electrophoresed as described above, gels were dried, exposed to film, and the DNA protein-complexes were excised from the dried gel for elution and PCR amplification of bound fragments (Blackwell & Weintraub, 1990) using Primer A above and Primer B: 5'-GGCTGAGTCTGAACGGATCC-3'. A contamination control was processed in parallel starting from a gel slice containing no protein-DNA complex. In all cases this control did not produce any detectable PCR products. Approximately 1 ng of amplified product was used in the binding reaction of the next round of selection. For the second and third rounds, 50 pg of PA-POU domain variant was used; 10 pg was used in the fourth round. The amplified products of the fourth round of selection were digested with BamHI and XhoI and ligated into the vector pBSKII+. Plasmids were derived from transformants and sequenced by dideoxysequencing.

RESULTS

A bacterial expression library was generated in which POU-homeodomain residues 47 and 51 were simultaneously randomized. In this library, the Oct-1 POU domain was fused to protein A (PA) of *S. aureus* to facilitate the expression, purification and analysis of individual variants. The library was screened for the binding of radioactive DNA probes using conditions similar to those originally described for the screening of cDNA expression libraries by Singh et al. (1988) and Vinson et al. (1988) (see Materials and Methods). A similar procedure has been described by Lorimer et al. (1992).

The nature of residues at positions 47 and 51 which can participate in the specific recognition of an octamer site was determined by probing the library with a DNA fragment containing the octamer consensus sequence (5'-ATGCAAATGA-3'). As shown in Table 1, the combination of residues that was most efficiently detected in the screen was the combination found in the wild-type protein, valine at position 47 and asparagine at position 51. All of the variants detected can be segregated into classes according to the residue at position 51. Surprisingly, in addition to a class containing asparagine at position 51, a class containing arginine at this position was also detected, as well as a single variant containing glutamine at position 51 and valine at position 47.

The relative effects of amino acid substitutions on binding affinity were determined by comparison of a representative panel of variants to the wild-type protein (Table 1). Of those analyzed, the wild-type protein exhibited the highest affinity for the fragment. Variants containing only substitutions of V47 had minor reductions in affinity, the largest a 6.5-fold reduction for the G47 N51 variant. The greatest reduction in binding affinity was observed for the V47 Q51 variant, which had a 1100-fold reduction in affinity. Two representatives of the R51 class of variants, G47 R51 and R47 R51, had

intermediate affinities with reductions of 23- and 32-fold, respectively, compared to the wild-type protein.

TABLE 1 - RESULTS OF SCREENING WITH
5'-ATGCAAATGA-3'

<u>47</u>	<u>51</u>	<u>#isolates^a</u>	<u>#predicted^b</u>	<u>isol./pred.</u>	<u>affinity^c</u>
V ^d	N ^d	8	10	0.80	1
R	N	5	15	0.33	3.0
T	N	2	10	0.20	nd
N	N	1	5	0.20	2.8
I	N	1	5	0.20	nd
C	N	2	5	0.40	nd
G	N	1	10	0.10	6.5
H	N	1	5	0.20	4.7
R	R	13	44	0.30	32
G	R	7	30	0.23	23
S	R	2	44	0.05	nd
N	R	1	15	0.07	nd
A	R	2	30	0.07	nd
C	R	1	15	0.07	nd
H	R	1	15	0.07	nd
T	R	2	20	0.10	nd
V	Q	2	10	0.20	1100

^a~5000 colonies screened

^bthe expected frequency of that variant in 5000 colonies of the library. codons 47 and 51 were randomized as NNG/C; 32 possible codons at each position, total complexity of library is 1024

^crelative dissociation constant, normalized to that of the wild-type protein

^dresidue found in wild-type protein

The large variability of residues at position 47 detected in the screen suggested only a moderate contribution of V47 to the binding affinity and specificity of the POU-domain. The minimal effects on affinity observed for its substitution with chemically diverse side chains were consistent with this notion. On the other hand, the greater apparent selectivity at position 51 suggested a more critical role for N51, which was supported by the 1100-fold reduction in binding affinity upon substitution with the chemically similar glutamine. The intermediate affinity of variants containing arginine at 51 suggested that their specificity might be quite different from that of the wild-type protein. Therefore, the specificity of variants with substitutions at 51 for the adenine

residue (adenine 7, Fig. 1) contacted by N51 of the wild-type protein (5'-ATGCAAATGA-3') was examined.

The V47 Q51, R47 R51, and G47 R51 variants were compared to the wild-type protein (V47 N51) for binding to fragments containing nucleotide substitutions of adenine 7 (Table 2). All variants exhibited a preference for adenine at nucleotide 7; however, the wild-type protein had a much higher specificity for this residue. Specifically, the substitution of adenine with guanine, thymine, and cytosine resulted in reductions in affinity of 300-, 1100-, and 45000-fold, respectively, for the wild-type protein. In comparison, for the V47 Q51 variant, these substitutions reduced affinity by 49-, 75-, and 87-fold, respectively, while for the R47 R51 variant, reductions of 1.4-, 20-, and 19-fold, respectively, were observed. Thus, asparagine at position 51 is not only required for the highest affinity interaction with the octamer site, it also determines the greatest selectivity for adenine at nucleotide position 7.

TABLE 2 - SPECIFICITY OF VARIANTS FOR NUCLEOTIDE 7

	<u>V47N51</u> ^a	<u>V47Q51</u> ^a	<u>R47R51</u> ^a	<u>G47R51</u> ^a
5'-ATGCAAATGA-3'	1	1100	32	23
5'-ATGCAAGTGA-3'	300	54000	44	69
5'-ATGCAACTGA-3'	45000	96000	610	500
<u>5'-ATGCAATTGA-3'</u>	<u>1100</u>	<u>83000</u>	<u>650</u>	<u>700</u>

^arelative dissociation constant, normalized to that of the wild-type protein for the octamer consensus sequence

The randomized expression library was also used to explore what other combinations of homeodomain residues at 47 and 51 and nucleotides at positions 7 and 8 could provide high affinity homeodomain-DNA interactions. The library was probed with fragments of DNA containing nucleotide substitutions at these positions. Since much of the specificity of binding was determined by interactions mediated by nucleotide 7, pools of probes were grouped according to the identity of this residue. For example,

the "CN" pool consisted of probes 5'-ATGCAACAGA-3', 5'-ATGCAATGA-3', 5'-ATGCAACCGA-3', and 5'-ATGCAAGGA-3'. Interestingly, when the library was probed with either the "CN" pool or the "TN" pool, no positive colonies were observed. This suggested that neither cytosine nor thymine at position 7 could mediate an interaction that could contribute enough affinity for detection by this screen. When the library was probed with the "GN" pool, a single variant, R47 R51, was detectable at a low frequency (4 positives, 35 predicted (4000 screened)). This variant was characterized further.

It was possible that the arginine substitutions at positions 47 and 51 in the R47 R51 variant mediated novel interactions with nucleotides other than those at positions 7 and 8. This variant had much less selectivity for nucleotide 7 than the wild-type protein (Table 2) and yet was detected in both the "GN" pool screen and in the initial screen with the unsubstituted octamer consensus sequence. To directly compare its nucleotide preferences at all positions in the binding site with that of the wild-type protein, a random binding site selection assay was employed (Blackwell & Weintraub, 1990; Pollock & Triesman, 1990). The wild-type protein (V47 N51) and the R47 R51 variant were challenged in four rounds of binding site selection, along with the R47 N51 variant as a control for the effects of the substitution of V47 with arginine. At least twenty five sequences were determined for each protein from the pool of sites that were selected in the fourth round (Figure 2A, C, E). Consensograms (Wilson, D. et al., 1993) were derived from these sequences (Figure 2B, D, F). As shown previously (Verrijzer et al., 1992), the wild-type POU domain selected a consensus octamer site (5'-TATGCAAAT-3') with strict preferences for nucleotides at positions -1 through 8 (Fig. 2A, B). In contrast, the R47 R51 variant selected sites with reduced stringency, especially at positions 7 and 8 (Fig. 2C, D). Although this variant prefers adenine at position 7, this preference is not as strong as it is for the wild-type protein: guanosine was selected at that position 6/26 times for R47 R51 and 0/25 for V47 N51. Thus, substitution of N51 with

Figure 2. Binding site sequences selected by the wild-type Oct-1 POU domain (V47 N51) and the R47 R51 and R47 N51 variants. (A, C, E) Sequences of sites isolated after four rounds of selection. (B, D, F) Consensograms (Wilson, D. et al., 1993) derived from the selected sequences.

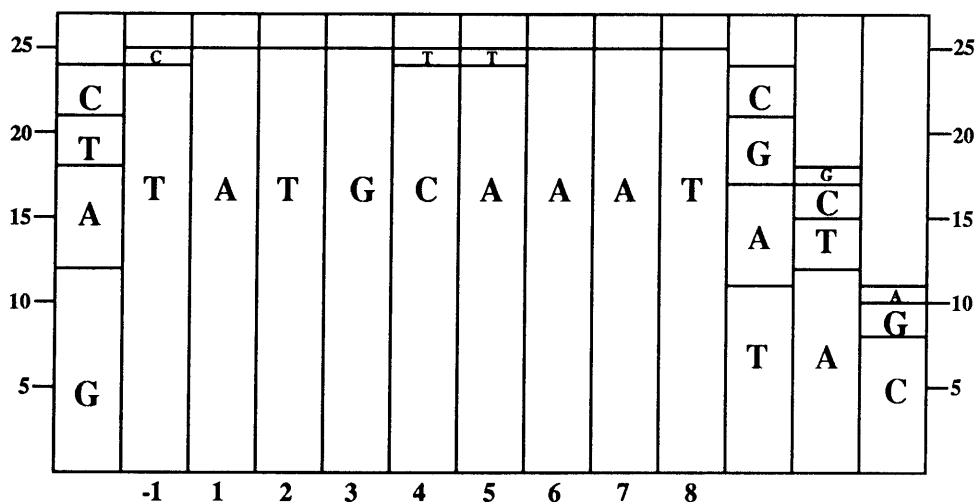
A

```

GG GGATATGCAATA GG
CC TCGTATGCAAAATG CC
GG GGGTATGTAAATT GG
GG CCTATGCAAAATCT GG
CC TTATGCAAAATAG CC
GG GTATGCAAAATTAA GG
GG GTCATATGCAAAAT GG
GG ATATGCAAAATCAC GG
CC AGCATGCAAAATGC CC
CC TTATGCAAAATTAC CC
GG GTATGCAAAATATC GG
GG GGTATGCAAAATTT GG
GG GCATATGCAAAATA GG
GG CGTATGCAAAATTA GG
GG GGTATGCAAAATTA GG
GG GTTATGCAAAATCA GG
CC CTATGTAAATATC CC
GG GGTATGCAAAATGA GG
GG GCTATGCAAAATTA GG
CC TATGCAAAATAGCC CC
CC ATATGCAAAATTAC CC
GG GTATGCAAAATGAG GG
GG GGATATGCAATA GG
GG GGTATGCTAATAC GG
CC GTATGCAAAATTC CC
  
```

B

V47 N51



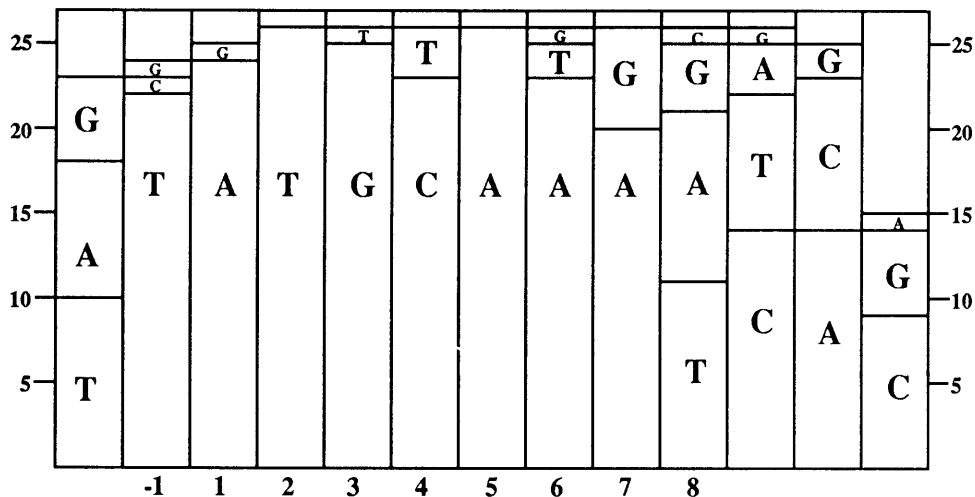
C

```

CC ATATGCAAAATCCC CC
CC ATATGCAAAATAC CC
CC CATATGCAAGTAC CC
CC ATATGCAAAATCCC CC
CC GGTATTCAAAACA CC
GG CTCATGCAAAACA GG
CC AGATGCAAAAGTAG CC
CC ATTATGCAAGACA CC
CC AATATGTAAAGCA CC
GG GTTATGTAAATCA GG
GG GGTATGCAAAACA GG
GG GTTATGTAAATTC GG
CC GTATGCAGATGCG CC
CC TATGCAAAATGAC CC
CC TTTATGCAAAATCG CC
GG GTTTATGCAAAAC GG
CC GTATGCAAGTAG CC
CC ATATGCAAAATCCC CC
CC TTATGCAAAATCCC CC
CC GTATGCAAAACAC CC
CC ATATGCAAGGTAC CC
CC TTATGCAAGACAC CC
CC TTATGCAAAACAC CC
CC TGCATATTAGGAC CC
CC ATGCAEACTGGGC CC
CC TTTGTGCAAGTAC CC
  
```

D

R47 R51



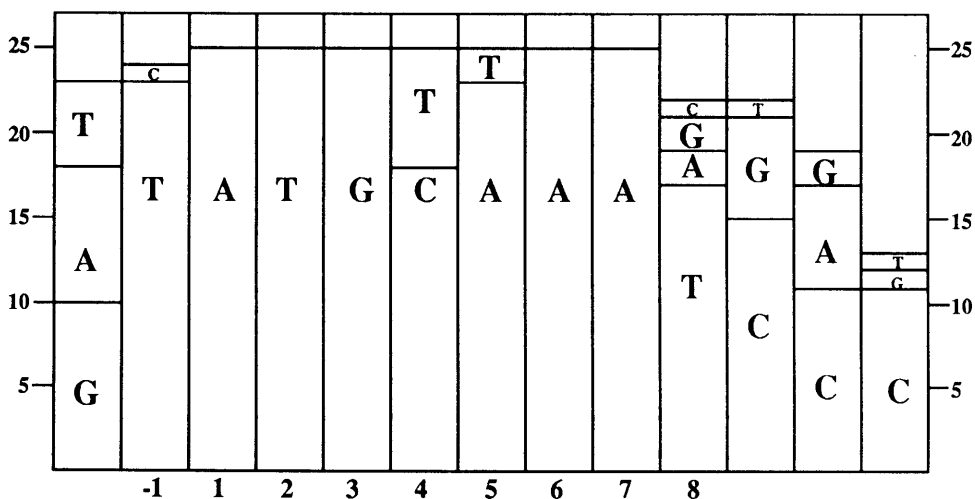
E

```

CC TATGCAAAATGGCC CC
CC TTATGCAAAATCCC CC
GG TGTATGCAAAATGA GG
GG GGATATGCAAAATC GG
GG GTCCGTATGCAAA GG
CC TTATGCAAAAGCGC CC
CC GATATGCAAAACC CC
CC AAATGCAAAATCC CC
CC TTATGTAAATCCC CC
CC AATGTAAATCCC CC
GG GCGGATATGCTAA GG
GG GCGCTATGCAAA GG
GG GGTATGCAAAATCC GG
GG GAATATGCAAAATC GG
CC ATATGTAAATGCC CC
GG GGTATGCTAATGA GG
CC TTATGTAAACGAC CC
CC ATGCAAAATCCGGGCC CC
CC AATGTAAATCCC CC
CC GTATGCAAAATTAC CC
CC GTATGCAAAATGAC CC
CC TGTATGTAAATCC CC
GG CCGCATGCAAAATC GG
CC GTATGTAAAGCAC CC
GG GTATGCAAAACCT GG
  
```

F

R47 N51



arginine results in a change in the "adenine requirement" at position 7 to a "purine requirement." The R47 R51 variant also exhibits much less stringency of selection at position 8 as compared to the wild-type. Adenine and thymine were selected equally by the variant with less preference for guanosine and cytosine while only thymine was selected by the wild-type protein. The effect at position 8 is partially attributable to the substitution of V47 with arginine (Fig. 2E, F). It appears from the sequences selected, that the R47 R51 variant did not select any nucleotide at any position with greater stringency than did the wild-type protein. This suggests that arginine, either at position 47 or 51, does not make unique nucleotide-specific contacts at any position in the selected site.

The relaxed specificity of the R47 R51 variant suggested its potential to redirect the formation of Oct-1-dependent transcriptional regulatory complexes to novel sequences that are not efficiently recognized by the wild-type protein. We sought to test this possibility using as a model system the formation of the multiprotein C1 complex on the Herpes Simplex Virus (HSV) α or immediate-early (α /IE) enhancer element. Formation of the C1 complex is dependent on the binding of the element by the Oct-1 POU domain and the viral α TIF protein (VP16, Vmw65, ICP25)(McKnight et al., 1987; Gerster & Roeder, 1988; O'Hare & Goding, 1988; Preston et al., 1988; Kristie et al., 1989; Stern et al., 1989), on the specific recognition of the Oct-1 POU-homeodomain surface by α TIF (Pomerantz et al., 1992; Lai et al., 1992), and on the presence of the cellular C1 factor (HCF) (Kristie & Sharp, 1993; Wilson, A.C. et al., 1993). Some nucleotide substitutions in the octamer-related sequence in this element should impair binding of the wild-type Oct-1 protein while having only a minimal effect on the binding of the R47 R51 variant. Formation of the regulatory complex on this novel element would then be dependent on the R47 R51 variant, and not possible with the wild-type protein. The combination of nucleotide substitutions at positions 7 and 8 which would provide the greatest discrimination in binding between the wild-type protein and the R47

R51 variant was determined. The R47 R51 variant was directly compared to the wild-type protein for binding to 16 probes containing all combinations of nucleotides at positions 7 and 8. As shown in Table 3, the sequence with the greatest preference (260 fold) for the variant contained guanine at both positions 7 and 8 (5'-ATGCAAGGGA-3').

The double guanine substitution was incorporated into the HSV α /IE element (5' -ATGCTAGGGGATATTCTTTGG-3' (HSV α /IE "GG")) and tested for the formation of the C1 complex in the presence of α TIF, the C1 factor, and either the wild-type or variant Oct-1 POU domain. The R47 R51 POU domain variant was first compared to the wild-type protein for formation of the C1 complex on the unsubstituted HSV α /IE element (Figure 3). The R47 R51 variant was clearly capable of forming the complex (cf. lanes 2 and 4), although 14 fold more protein was required to attain similar levels of DNA binding. When normalized on the basis of DNA binding activity, the variant was found to have only a 2.4-fold lower ability to form the complex than the wild-type protein

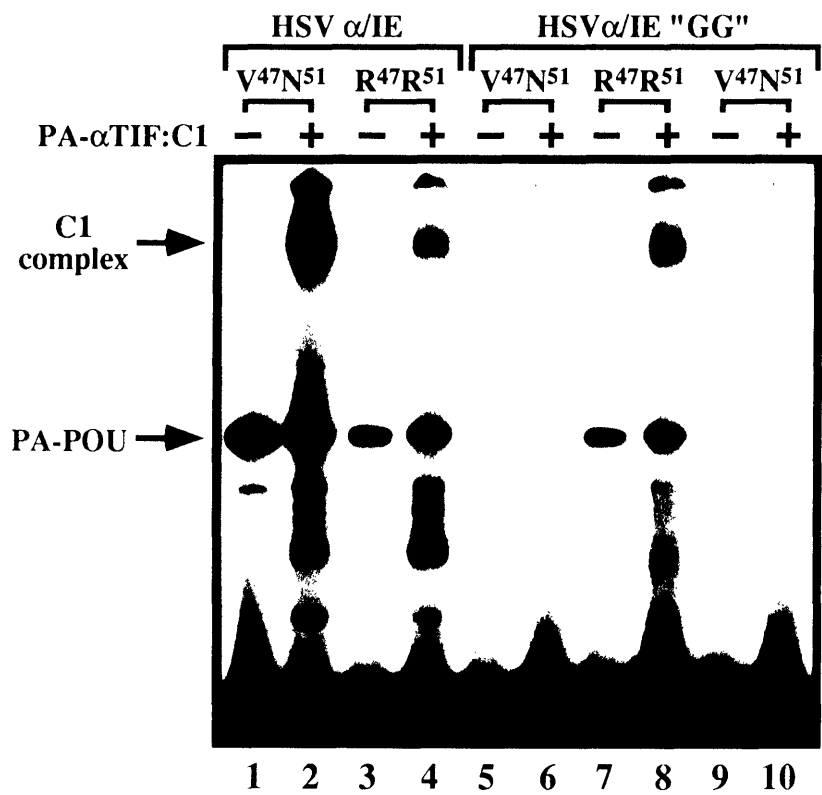
TABLE 3 - RELATIVE SPECIFICITY OF V47 N51 AND R47 R51

	<u>V47N51</u> ^a	<u>R47R51</u> ^a	<u>VN/RR</u> ^b
5'-ATGCAAATGA-3'	1	32	0.031
GT	300	44	6.8
CT	45000	610	74
TT	1100	650	1.7
AC	5.3	120	0.04
AA	7.1	79	0.09
AG	16	130	0.12
GC	37000	1000	37
GA	3900	100	39
GG	39000	150	260
CC	29000	130	220
CA	39000	810	48
CG	75000	710	110
TC	1700	1100	1.5
TA	5300	980	5.4
TG	4200	1200	3.5

^arelative dissociation constant, normalized to that of the wild-type protein binding to the octamer consensus

^bthe dissociation constant for the V47 N51 protein divided by that for the R47 R51 variant

Figure 3. The R47 R51 variant changes the DNA-binding specificity of an Oct-1-dependent regulatory complex. Wild-type (V47 N51, 25pg in lanes 1, 2, 5 and 6, 350 pg in lanes 9 and 10) and variant (R47 R51, 350 pg) PA-Oct-1 POU domain fusion proteins were incubated in DNA-protein-binding reactions in the absence (-) or presence (+) of 15 ng of PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor as indicated. Reactions included either the unsubstituted HSV α /IE element (5'-ATGCTAATGATATTCTTTGG-3') (lanes 1-4) or the double guanosine substituted HSV α /IE "GG" element (5'-ATGCTAGGGATATTCTTTGG-3') (lanes 5-10) as probe. The positions of the multiprotein C1 and POU domain:DNA complexes are indicated with arrows.



on the unsubstituted element. When the HSV α /IE "GG" element was used, no binding of the wild-type POU domain was detectable even at concentrations 14 fold higher than that required to bind the unsubstituted element (lanes 5 and 9), and upon addition of α TIF and the C1 factor, no complex formation was evident (lanes 6 and 10). In contrast, the R47 R51 variant bound the HSV α /IE "GG" element to an extent similar to that for the unsubstituted element (lane 7) and was capable of efficiently forming the C1 complex with α TIF and the C1 factor (lane 8). Therefore, the R47 R51 variant can be used to alter the DNA-binding specificity of a transcriptional regulatory complex which depends upon Oct-1 for DNA recognition.

DISCUSSION

Transcription factors utilize structural modules to recognize specific DNA sequences (Pabo & Sauer, 1992). Some modules, such as the zinc finger originally discovered in TFIID (Miller et al., 1985), accommodate different sets of amino acid determinants so that DNA binding specificity can vary within the framework of a conserved domain structure (Pavletich & Pabo, 1991, 1993; Fairall et al., 1993). Other modules appear to be specialized for the recognition of particular sequences, such as those of the basic-helix-loop-helix (CANNTG) (Murre & Baltimore, 1992) and nuclear hormone receptor (AGNNCA (half site)) (Evans, 1988) families. The homeodomain has apparently evolved with a constrained specificity for a binding site with a TAAT core. The Oct-1 POU-homeodomain has been used to test the stringency of determinants that mediate recognition of the TAAT core in the major groove.

Among all possible interactions determined by Oct-1 POU-homeodomain residues 47 and 51 and the nucleotides at position 7 and 8 in the octamer binding sequence (5'-ATGCAAAT-3'), the N51-adenine 7 and V47-thymine 8 interactions provide the highest affinity and greatest degree of specificity. The structure of the homeodomain and its mode of docking against DNA probably impose greater constraints

on the amino acid 51-nucleotide 7 interaction than on the amino acid 47-nucleotide 8 interaction. Fewer residues were detected in the screens at position 51 than at 47, and substitution of either N51 or adenine 7 had a larger effect on affinity and specificity than did substitution of V47 or thymine 8. The fact that all variants analyzed had higher affinities for sites containing adenine at position 7 than those containing other bases at that position suggests that adenine at the third position of the homeodomain subsite (AAAT) is most compatible with the spatial architecture of the homeodomain even when residue 51 is not the highly conserved asparagine. In accord with these results, Botfield et al. (1994) have found that in the homologous Oct-2 POU-homeodomain, N51 provides the highest affinity interaction with the wild-type octamer sequence when compared to 19 substituted variants.

The ability to achieve reasonable affinity ($\sim 10^{-9}$ M) for DNA with arginine at position 51 was surprising, given its lack of chemical similarity to asparagine. In addition, arginine is not found in this position in any natural homeodomain. Arginine at 51 is probably not making a nucleotide-specific contact and it is possible that this residue contributes to binding affinity by ionic interaction with the sugar-phosphate backbone. Alternatively, the purine requirement at position 7 that is observed for the R47 R51 variant may reflect an interaction of arginine with the N7 of the base. A similar specificity, determined by an arginine positioned in the major groove by an α helix, has been invoked based upon the Hin recombinase:DNA recombination half site crystal structure (Feng et al., 1994). The relaxed sequence specificity of the R47 R51 variant allows it to nucleate the formation of a transcriptional regulatory complex on a DNA sequence element which is not efficiently bound by wild-type Oct-1. Since formation of the C1 complex is critically dependent on the presentation of residues on the surface of helices 1 and 2 (Pomerantz et al., 1992; Lai et al., 1992) we conclude that substitution of residues 47 and 51 with arginine does not drastically perturb homeodomain structure.

Since the R47 R51 variant binds with reasonable affinity and can participate in the formation of regulatory complexes, it is reasonable to question the absence of this variant set of amino acids in the known sequences of homeodomains. Both the reduced affinity and specificity of the variant, relative to the wild-type protein, may preclude its utility *in vivo*. The R47 R51 variant binds the octamer sequence with thirty-fold lower affinity than the wild-type protein. In addition, its reduced specificity allows it to recognize many more sequence variants at a given protein concentration than the greater specificity of the wild-type protein would allow. The partitioning of the variant between its "specific" and "nonspecific" sites may make the occupancy of an individual target sequence without inappropriate action at other sites impossible at physiological levels of expression of the protein.

Other studies have altered the binding specificity of homeodomains by substitution of residue 50 since this residue is an important determinant of specificity for nucleotides 3' to the TAAT core (Hanes & Brent, 1989, 1991; Treisman et al., 1989; Percival-Smith et al., 1990). However, for POU domain proteins the residue at this position (cysteine) does not confer sequence selectivity and can be substituted without effect (Ingraham et al., 1990; Verrijzer et al., 1992). Consistent with this is that the POU domain does not select nucleotides 3' to the octamer site with a high degree of preference. The 3' boundary of sequence recognition by POU-homeodomains appears to be much more critically dependent on residues 47 and 51. It is presently unclear why cysteine 50 is absolutely conserved among POU domain proteins.

The interactions mediated by residues 47 and 51 do not appear to be independent. For example, the effect of the substitution of N51 with arginine is dependent on what residue is at position 47. This substitution reduces the affinity for the octamer consensus sequence by 11-fold when residue 47 is arginine but only by 3.5 fold when residue 47 is glycine (Table 1). In addition, the residue at position 51 also influences which nucleotide is selected at position 8 when residue 47 is arginine. The R47 N51 variant has less

selectivity for thymine at position 8 than does the wild-type protein, and this selectivity is further reduced for the R47 R51 variant (Figure 2). We also note the lack of detection of a variant containing valine at position 47 and arginine at position 51. The frequency of this variant in the library is predicted (29/5000) to be larger than most of the variants that were detected in the screen performed with the unsubstituted octamer sequence as probe. This implies that this variant should have been detected if its affinity (off-rate) for the octamer sequence fell within the large range of affinities exhibited by those variants that were detected. The combination of valine at position 47 and arginine at position 51 may somehow be incompatible with homeodomain structure or DNA-binding.

The interaction between residues 47 and 51 is probably related to the steric constraints that are imposed upon them as they pack into the major groove. The advantage of having wild-type residues V47 and N51 is not only due to their individual potentials for base-specific contacts but also to their ability to be sterically accommodated in the protein-DNA complex. It is striking that the V47 Q51 variant, which has the identical functional groups for nucleotide recognition and which differs from the wild-type only by a methylene group, has a 1100-fold reduced affinity for the octamer consensus. The substitution of either residue 47 or 51 with the much larger arginine may introduce steric interactions that reduce the density of packing of the side chains of helix 3 into the major groove, and therefore alter the complementarity of the protein surface to the DNA. This is evident in the consensograms of Figure 2. The wild-type consensogram presents a profile of selected nucleotides with nearly absolute preferences and clear boundaries on either side of the binding site. In contrast, the consensograms of the R47 N51 and R47 R51 variants exhibit a breakdown of binding-site stringency. The arginine substitutions affect the selection of nucleotides at positions other than 7 and 8, which may reflect the inability to pack the arginine side chain in a way which does not perturb other interactions. For example, even the nucleotide selected at position 4, which is within the POU-specific domain subsite, is influenced by the substitution of 47 and 51.

This implies that the packing of residues in helix 3 also impinges upon the binding specificity of the POU-specific domain, and reveals an interdependence of the POU-specific and POU-homeo subdomains that must arise from the spacing of subsites that is strictly preferred by the POU domain. Although there are no protein-protein interactions observed between the subdomains in the Oct-1 POU domain:DNA complex, and the linker between them is disordered, changing the spacing of subsites reduces the affinity of the POU domain 10-100 fold (Klemm et al., 1994).

The surprising detection of R51 variants attests to the utility of the randomization/screening approach. This technique relies upon the detection of direct binding to radiolabelled DNA probes and allows for the screening of a large number of proteins without competition between variants for binding to the probe. Such an approach should complement others (Youderian et al., 1983; Rebar & Pabo, 1994) for the isolation of transcription factors with new DNA-binding specificities that will provide useful tools and advance the understanding of protein-DNA interactions.

We conclude that it is probably not possible to alter homeodomain DNA-binding specificity by the substitution of residues 47 and 51 without sacrificing affinity and the ability to discriminate between sites. The homeodomain emerges as a module which is limited in its ability to recognize different DNA sequences and which has evolved the capacity to participate in other mechanisms of regulatory specificity such as protein-protein interactions. The constraints on homeodomain residues 47 and 51 emphasize that in protein-DNA recognition, the potential interactions that a particular amino acid residue may specify are critically dependent on the structural context determined by the folding and DNA-docking of the module in which it is found. Therefore, the spatial architecture of a DNA-binding domain may greatly constrain which amino acid residues can serve as its determinants of DNA recognition.

ACKNOWLEDGEMENTS

We thank Juli Klemm and Carl Pabo for the coordinates of the Oct1:DNA complex and for provocative discussions; Carl Pabo, Andrew MacMillan, Lee Lim, Juli Klemm, John Crispino, and Dan Chasman for critical reading of the manuscript ;Tom Kristie and members of the Sharp lab for their continual support; M. Siafaca for her ever present assistance; and R. Issner and Y. Qiu for indispensable technical assistance.

REFERENCES

- Blackwell, T.K., & Weintraub, H. (1990) *Science* 250, 1104-1110.
- Botfield, M.C., Jancso, A., & Weiss, M.A. (1994) *Biochemistry* 33, 6177-6185.
- Evans, R.M. (1988) *Science* 240, 889-895.
- Fairall, L., Schwabe, J.W. R., Chapman, L., Finch, J.T., & Rhodes, D. (1993) *Nature* 366, 483-487.
- Feng, J.-A., Johnson, R.C., & Dickerson, R.E. (1994) *Science* 263, 348-355.
- Gerster, T. , & Roeder, R. G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6347-6351.
- Hanes, S.D., & Brent, R. (1989) *Cell* 57, 1275-1283.
- Hanes, S.D., & Brent, R. (1991) *Science* 251, 426-430.
- Hayashi, S., & Scott, M.P. (1990) *Cell* 63, 883-894.
- Herr, W., Sturm, R.A., Clerc,R.G., Corcoran,L.M., Baltimore,D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M. , Ruvkin,G., & Horvitz,H.R.. (1988) *Genes & Dev.* 2, 1513-1516.
- Ingraham, H.A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M.S., Wilson, L., & Rosenfeld, M.G. (1990) *Cell* 61, 1021-1033.
- Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., & Pabo,C.O. (1990) *Cell* 63, 579-590.
- Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., & Pabo, C.O. (1994) *Cell* 77, 21-32.

- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950.
- Kristie, T. M., LeBowitz, J.H., & Sharp, P.A. (1989) *EMBO J.* 8, 4229-4238.
- Kristie, T.M, & Sharp, P.A. (1990) *Genes & Dev.* 4, 2383-2396.
- Kristie, T.M, & Sharp, P.A. (1993) *J. Biol. Chem.* 268, 6525-6534.
- Lai, J.-S., Cleary, M.A., & Herr, W. (1992) *Genes & Dev.* 6, 2058-2065.
- Laughon, A. (1991) *Biochemistry* 30, 11357-11367.
- Lorimer, I. A. J., Ho, C.-Y., & Smith, M. (1992) *Biotechniques* 12, 536-543.
- McKnight, J. L. C., Kristie, T.M. , & Roizman, B. (1987) *Proc. Natl. Acad. Sci. USA* 84, 7061-7065.
- Miller, J., McLachlan, A.D., & Klug, A. (1985) *EMBO J.* 4, 1609-1614.
- Murre, C., & Baltimore, D. (1992) in *Transcriptional Regulation* (McKnight, S.L., & Yamamoto, K., R., Eds.) pp. 861-879, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- O'Hare, P. , & Goding, C.R. (1988) *Cell* 52, 435-445.
- Otting, G., Qian, Y.Q. , Billeter, M. , Muller, M. , Affolter, M. , Gehring, W.J. , & Wuthrich, K. (1990) *EMBO J.* 9, 3085-3092.
- Pabo, C.O., & Sauer, R.T. (1992) *Annu. Rev. Biochem.* 61, 1053-1095.
- Pavletich, N.P., & Pabo, C.O. (1991) *Science* 252, 809-817.
- Pavletich, N.P., & Pabo, C.O. (1993) *Science* 261, 1701-1707.
- Percival-Smith, A., Muller, M., Affolter, M., & Gehring, W. (1990) *EMBO J.* 9, 3967-3974.
- Pollock, R., & Treisman, R. (1990) *Nucleic Acids Res.* 18, 6197-6204.
- Pomerantz, J.L., Kristie, T.M., & Sharp, P.A. (1992) *Genes & Dev.* 6, 2047-2057.
- Preston, C. M., Frame, M.C. , Campbell, M.E.M. (1988) *Cell* 52, 425-434.
- Qian, Y.Q., Billeter, M. , Otting, G., Mueller, M., Gehring, W.J. , & Wuthrich, K. (1989) *Cell* 59, 573-580.
- Rebar, E.J. & Pabo, C.O. (1994) *Science* 263, 671-673.

Sayers, J.R., Krekel, C. & Eckstein, F. (1992) *Biotechniques* 13, 592-596.

Singh, H., LeBowitz, J.H., Baldwin, A.S., & Sharp, P.A. (1988) *Cell* 52, 415-423.

Smith, D. L., & Johnson , A.D. (1992) *Cell* 68, 133-142.

Stern, S., Tanaka, M. , & Herr, W. (1989) *Nature* 341, 624-630.

Sturm, R. A., Das, G. , & Herr,W. (1988) *Genes & Dev.* 2, 1582-1599.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., & Desplan, C. (1989) *Cell* 59, 553-562.

Verrijzer, C. P., Alkema, M.J., van Weperen, W.W., Van Leeuwen, H.C., Strating, M. J. J., & van der Vliet, P.C. (1992) *EMBO J.* 11, 4993-5003.

Vershon, A. K., & Johnson , A.D. (1993) *Cell* 72, 1-20.

Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H., & McKnight, S.L. (1988) *Genes & Dev.* 2, 801-806.

Wilson, A.C., LaMarco, K., Peterson, M.G., & Herr, W. (1993) *Cell* 74, 115-125.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., & Desplan, C. (1993) *Genes & Dev.* 7, 2120-2134.

Wolberger, C., Vershon, A.K. , Liu, B. , Johnson, A.D. , & Pabo, C.O. (1991) *Cell* 67, 517-528.

Youderian, P., Vershon, A., Bouvier, S., Sauer, R.T., & Susskind, M.M. (1983) *Cell* 35, 777-783.

CHAPTER IV

STRUCTURE-BASED DESIGN OF TRANSCRIPTION FACTORS

This chapter originally appeared in *Science* **267**: 93-96 (1995).

ABSTRACT

Computer modeling suggested that transcription factors with novel sequence specificities could be designed by combining known DNA-binding domains. This structure-based strategy was tested by constructing a fusion protein, ZFHD1, that contained zinc fingers 1 and 2 from Zif268, a short polypeptide linker, and the homeodomain from Oct-1. The fusion protein bound optimally to a sequence containing adjacent homeodomain (TAATTA) and zinc-finger (NGGGNG) subsites. When fused to an activation domain, ZFHD1 regulated promoter activity in vivo in a sequence-specific manner. Analysis of known protein-DNA complexes suggests that many other DNA-binding proteins could be designed in a similar fashion.

Transcription factors are critical regulators of gene expression. The rational design of transcription factors with novel DNA-binding specificities and regulatory activities will provide powerful reagents for biological research and gene therapy. The recent determination of a series of structures of protein-DNA complexes has facilitated a design strategy that uses computer modeling to predict how DNA-binding domains could be combined to generate novel specificities. We have explored this strategy by designing and testing a zinc finger-homeodomain fusion protein.

Computer modeling studies were used to visualize how zinc fingers might be fused to the Oct-1 homeodomain. The known crystal structures of the Zif268-DNA (1) and Oct-1-DNA (2) complexes were aligned by superimposing the double helices in several different registers. Two arrangements were particularly interesting. In one alignment the COOH-terminal end of zinc finger 2 was 8.8 Å away from the NH₂-terminal arm of the homeodomain (Fig. 1), suggesting that a short polypeptide linker could connect these domains. In this model the fusion protein would bind a hybrid DNA site with the sequence 5'-AAATNNTGGGCG-3'. The Oct-1 homeodomain would recognize the AAAT subsite, zinc finger 2 would recognize the TGG subsite, and zinc finger 1 would recognize the GCG subsite. There was no possibility for steric interference between the zinc fingers and the homeodomain in this arrangement. Superimposing the DNA duplexes in other registers generated a second plausible arrangement for a hybrid protein (3); however, this model was not as favorable since there was a risk of steric interference between the zinc fingers and the homeodomain.

The design strategy was tested by construction of a fusion protein, ZFHD1, that contained fingers 1 and 2 of Zif268, a glycine-glycine-arginine-arginine linker, and the Oct-1 homeodomain (Fig. 2A). A glutathione S-transferase domain was added to facilitate expression and purification, and the DNA-binding activity of this fusion protein was determined by selecting binding sites from a random pool of oligonucleotides. After four rounds of selection, 16 sites were cloned and sequenced (Fig. 2B). Comparing these

Figure 1. Model of a zinc finger - homeodomain hybrid. Finger 1 of Zif268 is depicted in purple, finger 2 in yellow, and the Oct-1 homeodomain in red. The DNA is blue with the base pairs in the AAAT and TGGGCG subsites highlighted in cyan; the hybrid protein recognizes a sequence of the form 5'-AAATNNTGGGCG-3'. The C_α of Gly 59 (COOH-terminus of finger 2) is 8.8Å from the C_α of Arg 2 (the first homeodomain residue visible in the crystal structure) (13). This figure was generated with Insight II (Biosym Technologies, San Diego).

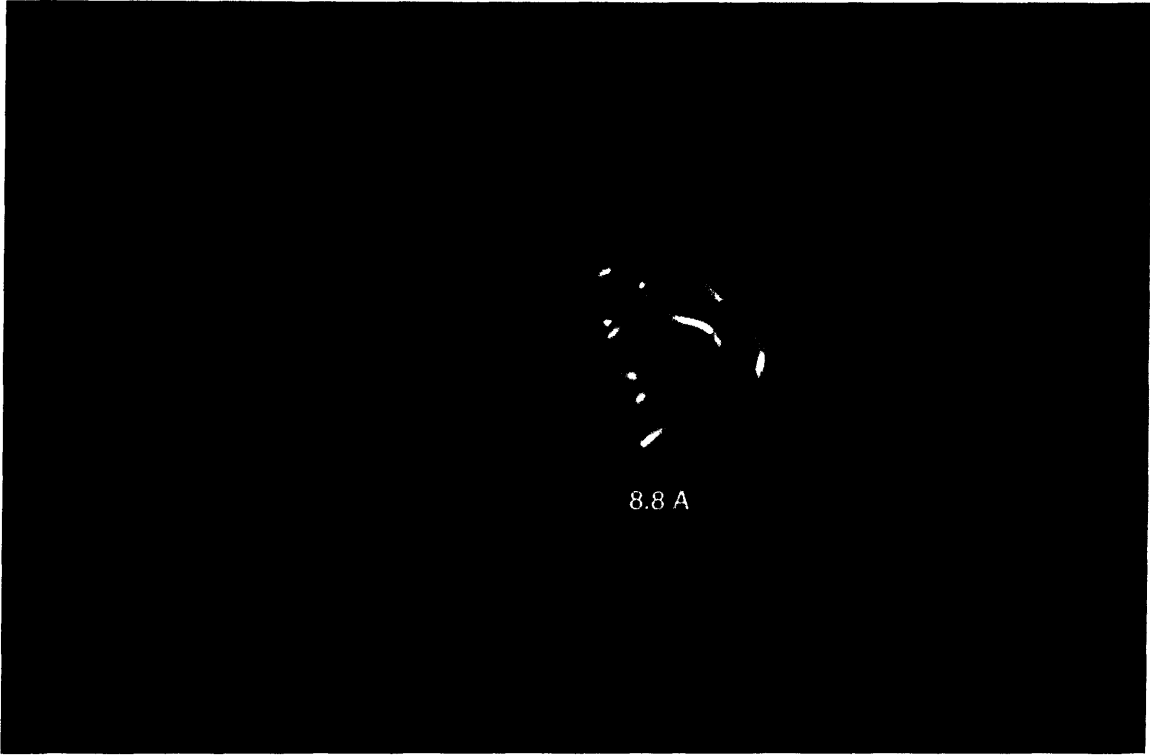
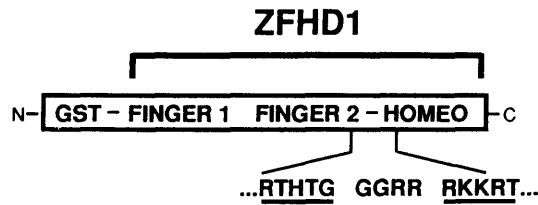


Figure 2. Selection by ZFHD1 of a hybrid binding site from a pool of random oligonucleotides. **(A)** Structure of the fusion protein used to select binding sites (14). The underlined residues are from the Zif268-DNA (1) and Oct-1-DNA (2) crystal structures and correspond to the termini used in the computer modeling studies. The linker contains two glycines that were included for flexibility and the two arginines that are present at positions -1 and 1 of the Oct-1 homeodomain. **(B)** Sequences of 16 sites isolated after four rounds of binding site selection(15). **(C)** Consensogram derived from the sequences in (B) which indicates the percent occurrence of each nucleotide at each position. **(D)** Schematic diagram illustrating the two possible orientations of the homeodomain subsite relative to the zinc finger subsite suggested by the consensus. Mode 1 corresponds to the configuration depicted in Figure 1.

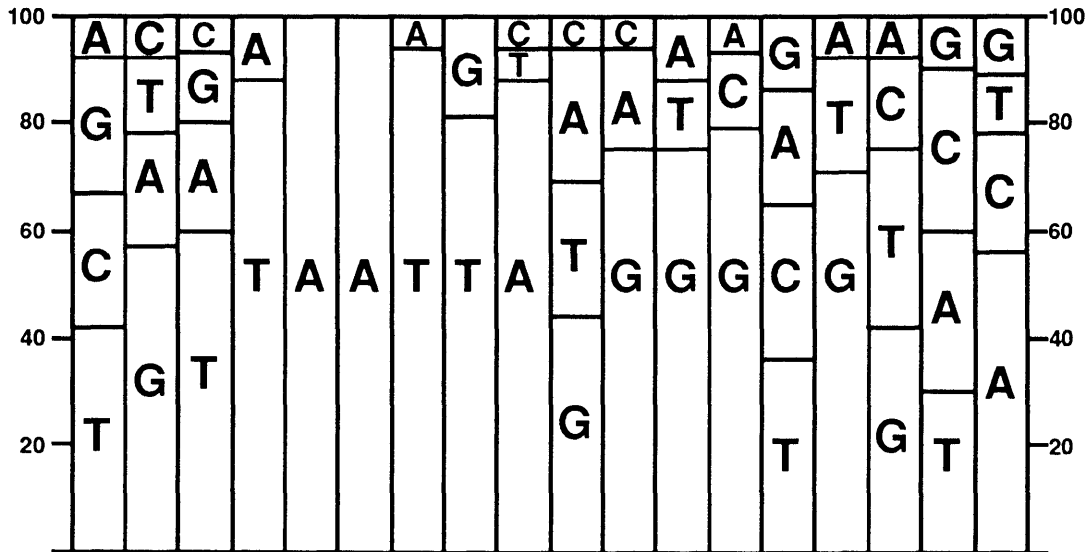
A



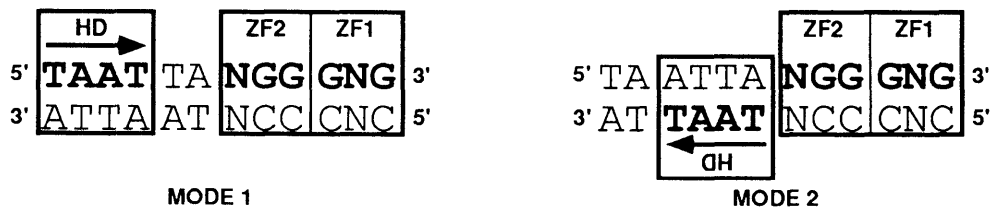
B

GTTTGGCACCTGACTAATTTAAGGAG
 GCGTTAATTAAGGGAGGTAAGGCC
 CTCGGCCGTTAATGAGGGGTGTTG
 TAATTATGGGCGGGATCGAATAGCC
 GGCAATAATCAATCCTTTAATTATGG
 GGCCGTACCTCATGAAATTAGGGCG
 GTTAATTATGGGGTAATAATGGTGC
 GTCGGGCTCTGTAAATTATGGGTGG
 GGATAATTACGGGTGGCATTAGGC
 GATAAATAGGGCGTCCCATCCCGT
 TAAATTAGGGCTTTAATTACGGTC
 TCATTAGAGTGTTAATGAGATCCGC
 TAGTTGCTAATTTGTATTAATTAAAG
 AGTTATTAATTAAGAATGTTAATTA
 GTGTGATAATGAGCTGGTCCGTCCC
 ATATTAAGGCGTAATTCGGACAAGA

C



D

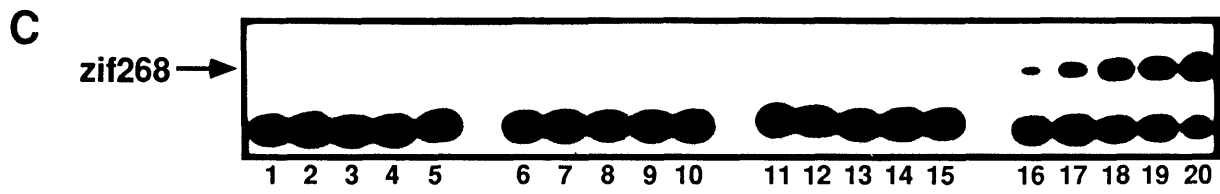
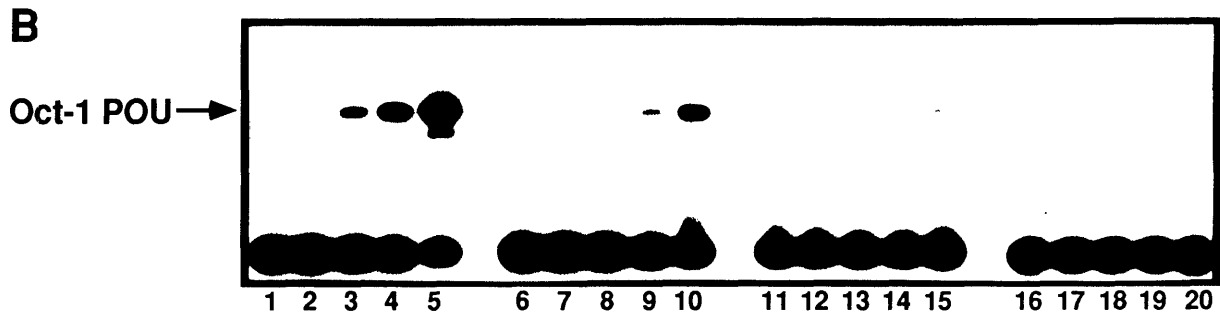
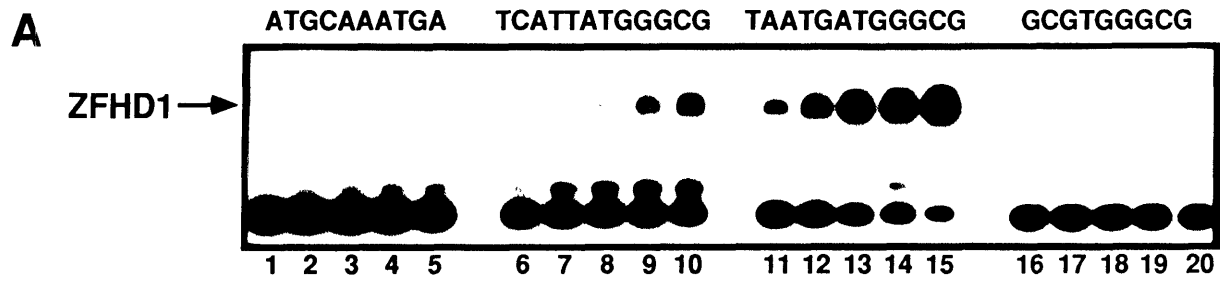


sequences revealed the consensus binding site 5'-TAATTANGGGNG-3' (Fig. 2C). The 5' half of this consensus, TAATTA, resembled a canonical homeodomain binding site TAATNN (4) and matched the site (TAATNA) that is preferred by the Oct-1 homeodomain in the absence of the POU-specific domain (5). The 3' half of the consensus, NGGGNG, resembled adjacent binding sites for fingers 2 (TGG) and 1 (GCG) of Zif268. The guanines were more tightly conserved than the other positions in these zinc finger subsites, and the crystal structure shows that these are the positions of the critical side chain-base interactions (1).

The ZFHD1 consensus sequence (5'-TAATTANGGGNG-3') matched the model that appeared most structurally feasible (6), but because of the internal symmetry of the TAATTA subsite this sequence was also consistent with the homeodomain binding in another orientation (Fig. 2D, compare mode 1 and mode 2). This alternative arrangement, in which the critical TAAT is on the other strand and directly juxtaposed with the zinc finger (TGGGCG) subsites, was considered unlikely since modeling had suggested that this arrangement required a linker to span >20 Å between the COOH-terminus of finger 2 and the NH₂-terminus of the homeodomain. To determine how the homeodomain bound to the TAATTA sequence in the 5' half of the consensus, ZFHD1 was tested for binding to probes (5'-TAATGATGGGCG-3' and 5'-TCATTATGGGCG-3') designed to distinguish between these orientations. ZFHD1 bound to the 5'-TAATGATGGGCG-3' probe with a dissociation constant of 8.4×10^{-10} M, and preferred this probe to the 5'-TCATTATGGGCG-3' probe by a factor of 33 (Fig. 3A compare lanes 6-10 and 11-15). This suggested that the first four bases of the consensus sequence form the critical TAAT subsite that is recognized by the homeodomain (mode 1) and that ZFHD1 binds as predicted in the model shown in Figure 1.

We compared ZFHD1, Oct-1 and Zif268 for their abilities to distinguish among the Oct-1 site 5'-ATGCAAATGA-3', the Zif268 site 5'-GCGTGGGCG-3', and the hybrid binding site 5'-TAATGATGGGCG-3'. The fusion protein ZFHD1 preferred the optimal

Figure 3. Comparison of the DNA binding specificity of ZFHD1, the Oct-1 POU domain (which contains a homeodomain and a POU-specific domain), and the three zinc fingers from Zif268 (16). (A) The GST-ZFHD1 protein was titrated into DNA-binding reactions containing the probe listed at the top of each set of lanes. Lanes 1, 6, 11, and 16 contain the protein at 9.8×10^{-11} M; Protein concentration was increased in 3 fold increments in subsequent lanes of each set. The position of the protein-DNA complex is indicated by the arrow. (B) The PA-Oct-1 POU fusion protein (17) was titrated into parallel DNA-binding reactions as in (A), but lanes 1, 6, 11, 16 contain the protein at 2.1×10^{-12} M. The position of the protein-DNA complex is indicated by the arrow. (C) A peptide containing Zif268 fingers 1, 2 and 3 (1) was titrated into parallel DNA-binding reactions with lanes 1, 6, 11, 16 containing the peptide at 3.3×10^{-11} M. The position of the protein-DNA complex is indicated by the arrow.



hybrid site to the octamer site by a factor of 240 (Figure 3A compare lanes 1-5 and 11-15), and did not bind to the Zif site (lanes 16-20). The POU domain of Oct-1 (Fig. 3B) bound to the octamer site with a dissociation constant of 1.8×10^{-10} M (lanes 1-5), preferring this site to the hybrid sequences by factors of 10 (lanes 6-10) and 30 (lanes 11-15), and did not bind to the Zif site (lanes 16-20) (7). The three fingers of Zif268 (Fig. 3C) bound to the Zif site with a dissociation constant of 3.3×10^{-10} M (lanes 16-20), and did not bind to the other three sites (lanes 1-15). These experiments proved that ZFHD1 bound tightly and specifically to the hybrid site and displayed DNA-binding specificity that was clearly distinct from that of either of the original proteins.

We fused ZFHD1 to a transcriptional activation domain and used transfection experiments to determine whether the novel DNA-binding protein could function *in vivo*. An expression plasmid encoding ZFHD1 fused to the COOH-terminal 81 amino acids of the Herpes Simplex Virus VP16 protein (ZFHD1-VP16) was co-transfected into 293 cells with reporter constructs containing the SV40 promoter and the firefly luciferase gene. To determine whether the fusion protein could specifically regulate gene expression, we tested reporter constructs containing two tandem copies of either the ZFHD1 site 5'-TAATGATGGGCG-3', the octamer site 5'-ATGCAAATGA-3', or the Zif site 5'-GCGTGGGCG-3' inserted upstream of the SV40 promoter. When the reporter contained two copies of the ZFHD1 site, the ZFHD1-VP16 protein stimulated the activity of the promoter in a dose-dependent manner (8). Furthermore, the stimulatory activity was specific for the promoter containing the ZFHD1 binding sites (Fig. 4). At levels of protein which stimulated this promoter by 44 fold, no stimulation above background was observed for promoters containing either the octamer or Zif sites. Thus, ZFHD1 efficiently and specifically recognized its target site *in vivo*.

This structure-based strategy of fusing known DNA-binding modules may provide a general method for designing transcription factors with novel DNA-binding specificities. Computer modeling suggests a number of other plausible arrangements for

Figure 4. Regulation of promoter activity in vivo by ZFHD1. The 293 cells were co-transfected with 5ug of reporter vector, 10ug of expression vector, and 5 ug of pCMV-hGH (gift of J. Parvin) used as an internal control (18). The reporter vectors contained two tandem copies of either the ZFHD1 site (TAATGATGGGCG), the Oct-1 site (ATGCAAATGA), the Zif site (GCGTGGGCG), or no insert. The expression vector encoded the ZFHD1 protein fused to the COOH-terminal 81 amino acids of VP16 (+ bars), and the empty expression vector Rc/CMV was used as control (-bars). The level of luciferase activity obtained, normalized to hGH production, was set to 1.0 for the co-transfection of Rc/CMV with the no-insert reporter pGL2-Promoter. Bar graphs represent the average of three independent experiments. Actual values and standard deviation reading from left to right are: $1.00 \pm .05$, $3.30 \pm .63$; $0.96 \pm .08$, 42.2 ± 5.1 ; $0.76 \pm .07$, $2.36 \pm .34$; $1.22 \pm .10$, 4.22 ± 1.41 . Fold induction refers to the level of normalized activity obtained with the ZFHD1-VP16 expression construct divided by that obtained with Rc/CMV.

TARGET: no insert ZFHD1 site Oct-1 site zif site

FOLD INDUCTION: 3.3 44.0 3.1 3.5



ZFHD1-VP16: - + - + - + - +

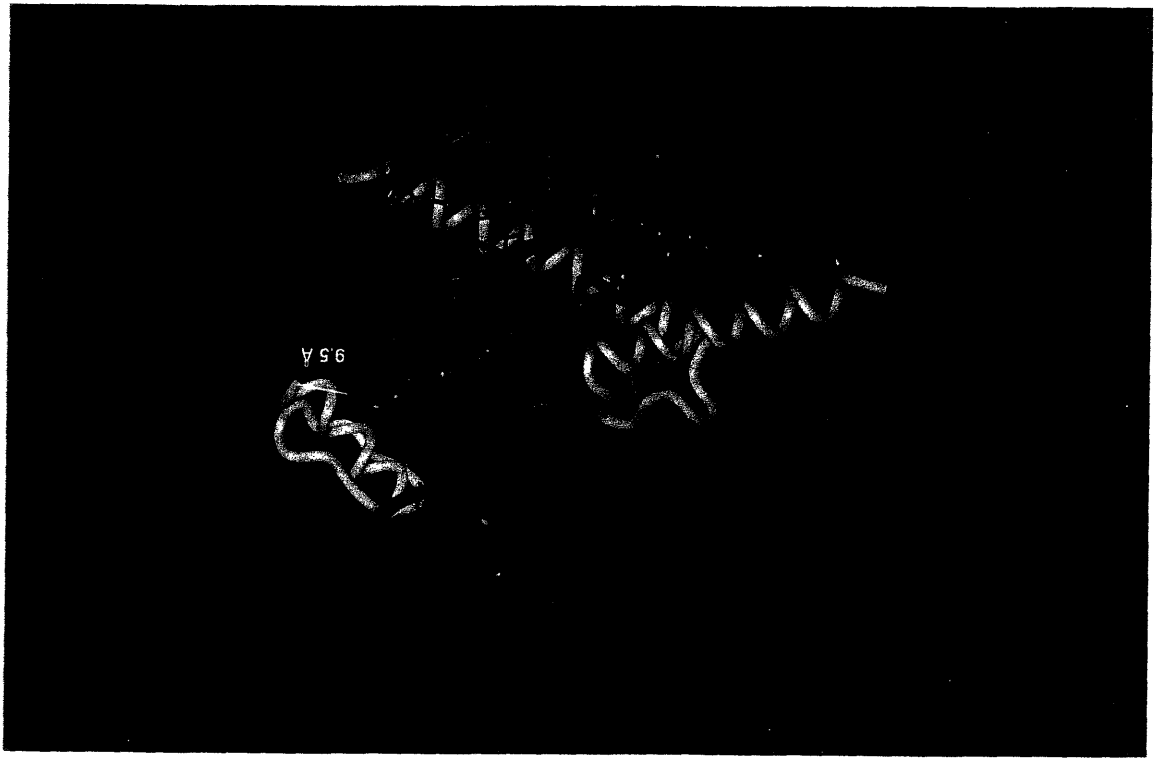
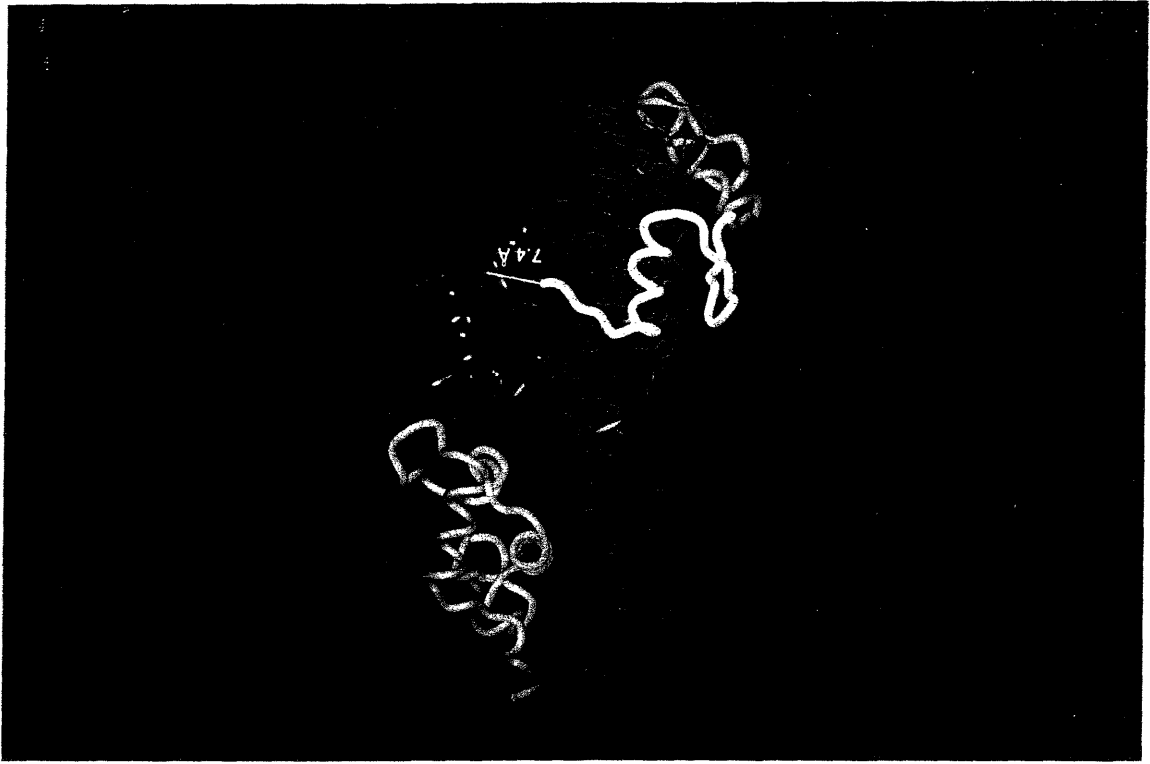
hybrid proteins. Figure 5 illustrates models of a zinc finger-basic-helix-loop-helix fusion protein (Fig. 5A) and a zinc finger-steroid receptor fusion protein (Fig. 5B) that should recognize hybrid binding sites. In each case, the modules can be fused by a short polypeptide linker without steric interference between the domains. This strategy could also be extended by varying the length and sequence of the polypeptide linkers and then using selection methods to optimize the binding affinity and specificity of the hybrid protein.

The strategy of fusing modules can also be combined with those designed for changing the sequence specificity of individual modules. Several DNA-binding domains are amenable to mutational strategies for changing sequence specificity (9,10), and zinc fingers may offer the most versatility (11). Combining structure-based design with mutational changes in specificity would greatly expand the range of sequences that could be targeted by hybrid domains.

The high affinity of ZFHD1 for its optimal site and the fact that ZFHD1, Oct-1, and Zif268 all clearly preferred different sites illustrate the success of the combinatorial approach. The specificity of the hybrid transcription factor depends upon the relatively moderate affinity, but high sequence specificity, for the binding of a single module and the chelate effect (12) provided by the covalent linkage of modules. The design criteria that allowed the construction of ZFHD1 included the short length of polypeptide linker that was required to fuse the DNA-binding domains and the absence of steric interference between these domains.

Designed transcription factors will be useful for the targeted regulation of specific cellular genes. Using particular DNA-binding domains in a hybrid (or adding other domains) may allow a protein to interact with other cellular factors or to be modulated by a particular regulatory pathway. The structure-based design of hybrid transcription factors should facilitate the development of efficient and specific reagents for biological research and gene therapy.

Figure 5. Models of fusions of other DNA-binding modules. **(A)** A putative zinc finger-basic-helix-loop-helix fusion. Finger 1 of Zif268 is depicted in purple, finger 2 in yellow, and the MyoD bHLH region (19) in red and gray. The distance in angstroms between the COOH-terminus of finger 2 and the NH₂-terminus of the basic region of the bHLH domain is indicated. **(B)** A putative zinc finger-steroid receptor fusion. Finger 1 of Zif268 is depicted in purple, finger 2 in yellow, and the glucocorticoid receptor (20) in red and gray. The distance in angstroms between the NH₂-terminus of finger 2 and the COOH-terminus of the glucocorticoid receptor is indicated. These figures were generated with Insight II (Biosym Technologies, San Diego).



REFERENCES AND NOTES

1. N. P. Pavletich and C. O. Pabo, Science 252, 809 (1991).
2. J. D. Klemm, M.A. Rould, R. Aurora, W. Herr, C. O. Pabo, Cell 77, 21 (1994).
3. This alternative arrangement would also have a short (<10 Å) linker connecting zinc finger 2 to the homeodomain, but the subsites are arranged so that the predicted binding sequence is 5'-CGCCCANNAAAT-3'.
4. A. Laughon, Biochemistry 30, 11357 (1991).
5. C. P. Verrijzer et al., EMBO J. 11, 4993 (1992).
6. No selected sites matched the binding sequence predicted for the alternative arrangement (3).
7. The relatively high affinity of the POU domain for the hybrid sites may reflect nonspecific contacts made by the POU-specific domain.
8. J.L.Pomerantz, unpublished results.
9. P. Youderian, A. Vershon, S. Bouvier, R. T. Sauer, M. M. Susskind, Cell 35, 777 (1983); R. P. Wharton and M. Ptashne, Nature 316, 601 (1985); M. Suckow, A. Madan, B. Kisters-Woike, B. von Wilcken-Bergmann, B. Muller-Hill, Nucleic Acids Research 22, 2198 (1994).
10. J. L. Pomerantz and P. A. Sharp, Biochemistry 33, 10851 (1994).
11. E. J. Rebar and C. O. Pabo, Science 263, 671 (1994); A. C. Jamieson, S.-H. Kim, J. A. Wells, Biochemistry 33, 5689 (1994); R. Desjarlais and J. M. Berg, Proc. Natl. Acad. Sci. USA 90, 2256 (1993).
12. W.P. Jencks, Proc. Natl. Acad. Sci. USA 78, 4046 (1981).
13. Each model of a hybrid protein (Figures 1 and 5) was constructed by juxtaposing portions of two different crystallographically determined protein-DNA complexes. Models were initially prepared by superimposing the double helices in various registers and were analyzed to see how the polypeptide chains might be connected. Superimposing sets of phosphates typically gave root mean squared distances of 0.5-1.5

Å between corresponding atoms. These distances give some perspective on the error limits involved in modeling and were one of the reasons we chose a flexible linker containing several glycines in these initial studies.

14. A fragment encoding Zif268 residues 333-390 (21), two glycines, and Oct-1 residues 378-439 (22) was generated by PCR (polymerase chain reaction), confirmed by dideoxysequencing, and cloned into the BamHI site of pGEX2T (Pharmacia) to generate an in frame fusion to glutathione S-transferase (GST). The GST-ZFHD1 protein was expressed by standard methods (23), purified on Glutathione Sepharose 4B (Pharmacia) according to the manufacturer's protocol, and stored at -80°C in 50 mM Tris pH 8.0, 100 mM KCl, and 10% glycerol. Protein concentration was determined by densitometric scanning of coomassie-stained SDS PAGE-resolved proteins using bovine serum albumin (Boehringer Mannheim) as standard.

15. The probe used for random binding site selection contained the sequence 5'-GGC TGAGTCTGAACGGATCCN₂₅CCTCGAGACTGAGCGTCG-3'. Four rounds of selection were performed as described (10), except that 100 ng poly[d(I-C)]/poly[d(I-C)] and 0.025% Nonidet P-40 were included in the binding reactions. Selections used 5 ng randomized DNA in the first round and approximately 1 ng in subsequent rounds. Binding reactions contained 6.4 ng of GST-ZFHD1 in round 1, 1.6 ng in round 2, 0.4 ng in round 3, and 0.1 ng in round 4.

16. DNA-binding reactions contained 10mM Hepes (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 0.75 mM DTT, 4% Ficoll-400, 300 µg/ml of bovine serum albumin, with the appropriate protein and binding site in a total volume of 10 µl. The concentration of binding site was always lower than the apparent dissociation constant by at least a factor of 10. The PA-Oct-1 POU fusion has been described (17). The purified three finger Zif268 peptide (1) was a kind gift from Monica Elrod-Erickson. Reactions were incubated at 30°C for 30 minutes and resolved in 4% nondenaturing polyacrylamide gels (17). Apparent dissociation constants were determined as described (10). Probes were

derived by cloning the following fragments into the Kpn I and Xho I sites of pBSKII+ (Stratagene) and excising the fragment with Asp718 and Hind III: 5'-CCTCGAGGTCATTATGGGCGCTAGGTACC-3', 5'-CCTCGAGGCGCCCATCATTA CTAGGTACC-3', 5'-CCTCGAGGCGCCCACGCCTAGGTACC-3', 5'-CCTCGAGGT CATTTCATA CTAGGTA CC-3'.

17. J. L. Pomerantz, T. M. Kristie, P.A. Sharp, Genes & Development 6, 2047 (1992).

18. The ZFHD1-VP16 expression vector was constructed by cloning a fragment encoding the epitope MYPYDVPDYA, ZFHD1, and VP16 residues 399-479 (24) into the Not I and Apa I sites of Rc/CMV (Invitrogen). Reporter vectors were constructed by cloning into the Xho I and Kpn I sites of pGL2-Promoter (Promega) the following fragments: 5'-GGTACCAGTATGCAAATGACTGCAGTATGCAAATGACCTCGAG-3', 5'-GGTAC CAGGCGTGGGCGCTGCAGGCGTGGGCGCCTCGAG-3', 5'-GGTACCAGTAATGA TGGGCGCTGCAGTAATGATGGGCGCCTCGAG-3'. The 293 cells were transfected using calcium phosphate precipitation with a glycerol shock as described (23). Quantitation of hGH production was performed using the Tandem-R HGH Immunoradiometric Assay (Hybritech Inc., San Diego, CA) according to the manufacturer's instructions. Cell extracts were made 48 hours after transfection (23) and luciferase activity was determined using 10 µl of 100 µl total extract/10 cm plate and 100 µl of Luciferase Assay Reagent (Promega) in a ML2250 Luminometer (Dynatech Laboratories, Chantilly, VA) using the enhanced flash program and integrating for 20 seconds with no delay.

19. P.C.M. Ma, M.A. Rould, H. Weintraub, C. Pabo, Cell 77, 451 (1994).

20. B. F. Luisi et al., Nature 352, 497 (1991).

21. B. A. Christy, L. F. Lau, D. Nathans, Proc. Natl. Acad. Sci. USA 85, 7857 (1988).

22. R. A. Sturm, G. Das, W. Herr, Genes and Development 2, 1582 (1988).

23. F. M. Ausubel et al., Eds., Current Protocols in Molecular Biology (John Wiley & Sons, New York, 1994).

24. P.E. Pellett, J.L.C. McKnight, F. J. Jenkins, B. Roizman, Proc. Natl. Acad. Sci. USA 82 5870 (1985).

25. We thank J. Crispino, A. MacMillan, and J. Kim for critical reading of the manuscript; D. Shaywitz and J. Klemm for helpful discussions; members of the Sharp lab for their continual support; M. Siafaca for her ever present assistance; L. Nekludova for help with modeling and programming; W. Xu and R. Sauer for help using Insight; and Y. Qiu and R. Issner for indispensable technical assistance. This work was supported by U.S. Public Health Service Grant PO1-CA42063 from the National Institutes of Health, by cooperative agreement CDR-8803014 from the National Science Foundation to P.A.S., and partially by the National Cancer Institute Cancer Center Support (core) Grant P30-CA14051. Graphics were prepared using equipment purchased with support from the PEW Charitable Trusts and the Howard Hughes Medical Institute. C.O.P. is in the Howard Hughes Medical Institute. J.L.P. is partially supported by the Harvard Medical School M.D.-Ph.D. program.

CHAPTER V

ANALYSIS OF HOMEODOMAIN FUNCTION BY STRUCTURE-BASED DESIGN OF A TRANSCRIPTION FACTOR

ABSTRACT

The homeodomain is a 60 amino acid module which mediates critical protein-DNA and protein-protein interactions for a large family of regulatory proteins. We have used structure-based design to analyze the ability of the Oct-1 homeodomain to nucleate an enhancer complex. The Oct-1 protein regulates Herpes Simplex Virus (HSV) gene expression by participating in the formation of a multiprotein complex (C1 complex) which regulates α (immediate early) genes. We recently described the design of ZFHD1, a chimeric transcription factor containing zinc fingers 1 and 2 of Zif268, a four residue linker, and the Oct-1 homeodomain. In the presence of α TIF and C1 factors, ZFHD1 efficiently nucleates formation of the C1 complex in vitro and specifically activates gene expression in vivo. The novel sequence specificity of ZFHD1 recruits C1 complex formation to a novel enhancer element. ZFHD1 function depends on the recognition of the Oct-1 homeodomain surface. The results indicate that the Oct-1 homeodomain mediates all the protein-protein interactions that are required to efficiently recruit α TIF and C1 factors into a C1 complex. The structure-based design of transcription factors should provide valuable tools for dissecting the interactions of DNA-bound domains in other regulatory circuits.

INTRODUCTION

Homeodomain proteins play central roles in the development and differentiation of eukaryotic organisms, but a mechanistic understanding of their biological specificity has only been achieved in a few cases (1). Like other transcription factors, homeodomain proteins are modular, containing functionally and structurally independent domains which determine their sequence specificity and regulatory action (activation or repression). The DNA-binding of these proteins relies upon the 60 amino acid homeodomain which has been extensively characterized biochemically and at the structural level (2-6). In several studies, functional differences between two regulatory proteins have mapped to their homeodomains, even though their DNA-binding properties may be indistinguishable (7-12). These observations have suggested that the homeodomain serves not only as a DNA-binding module, but also as a target for protein-protein interactions with other factors that enhance its target specificity and effector function (13-16).

The role of the human Oct-1 protein in the regulation of Herpes Simplex Virus gene expression exemplifies how protein-protein interactions with the homeodomain can determine functional specificity. Viral α or immediate early (α /IE) gene expression is controlled by the assembly of a multiprotein complex (C1 complex) composed of Oct-1, the viral α TIF protein (VP16, Vmw65, ICP25), and the cellular C1 factor (HCF), on the α /IE element (5'-ATGCTAATGATATTCTTTGG-3') (17-25). The 5' portion of the element is recognized by the Oct-1 POU domain, which is a bipartite DNA-binding domain. The POU-specific domain binds the ATGC subsite and the associated homeodomain binds the TAATGA subsite (26, 27). The 3' portion of the element is recognized by α TIF and possibly by the C1 factor (26). Oct-2 has identical DNA-binding specificity to Oct-1 but is 100-fold less efficient at complex formation (21, 22). This difference between Oct-1 and Oct-2 in the regulation of HSV correlates with a single amino acid difference on the surface of their homeodomains which allows α TIF to

distinguish between them in complex assembly (13, 14). The selective recognition of the Oct-1 homeodomain surface by α TIF exemplifies how protein-protein interactions can confer upon two homeodomain proteins with identical DNA-binding specificity dramatically different abilities to regulate a particular gene.

Previous studies established that the POU domain, composed of the POU-specific domain, a 24 residue linker, and the homeodomain, was as efficient as the full length Oct-1 protein at binding the HSV α /IE element and forming the C1 complex (21). The isolated Oct-1 homeodomain was capable of nucleating complex formation, but only functioned at high concentrations (13, 26). The low affinity and modest specificity of the homeodomain for DNA precluded an *in vivo* assay of complex formation in the absence of the POU-specific domain.

We have used structure-based design to develop a chimeric protein, ZFHD1 (28), that has allowed us to test whether the homeodomain of Oct-1 can efficiently mediate all of the protein-protein interactions necessary to recruit α TIF and the C1 factor into a functional enhancer complex. ZFHD1 contains the Oct-1 homeodomain fused to two zinc finger domains. The chimeric protein possesses high affinity and specificity for a novel DNA site, permitting the study of homeodomain function in physical isolation from the POU-specific domain and free of competition *in vivo* with the endogenous Oct-1 protein.

MATERIALS AND METHODS

Gel electrophoretic mobility shift assays. DNA binding reactions contained 0.4 - 0.8 ng DNA probe, 75 ng sonicated salmon sperm DNA, 10 mM Hepes (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 0.75 mM DTT, 4% Ficoll-400, 300 μ g/ml of bovine serum albumin, and the appropriate purified proteins or chromatographic fraction in a total volume of 10 μ l. Reactions were incubated at 30°C for 30 min and resolved in 4% nondenaturing

polyacrylamide gels (13). Protein-DNA complexes and the free DNA were quantitated using a Molecular Dynamics PhosphorImager with ImageQuant 3.22 software.

Production of recombinant proteins. The design and characterization of the GST-ZFHD1 fusion has been described (28). GST-ZFHD1 variants containing point mutations on the surface of the homeodomain were generated by PCR amplification of appropriate fragments using the expression vectors described below as substrates, followed by cloning the amplified fragments into the BamHI site of pGEX2T (Pharmacia) to generate in-frame fusions to GST. GST-ZFHD1 variants were expressed and purified as described (28).

Transient Transfection Assays. Reporter vectors were constructed by cloning the following fragments into the Xho I and Kpn I sites of pGL2-Promoter (Promega): α /IE: 5'-GGTACCATGCTAATGATATTCTTTGGCTGCAGATGCTAATGATATTCTTTGGCTCGAG-3'; α /IE-ZF: 5'-GGTACCGCCCTAGTAATGATATTCTTTGGCTGCAGCCCTAGTAATGATATTCTTTGGCTCGAG-3'; α /IE-ZF-3'MT: 5'-GGTACCGCCCTAGTAATGCTGTTCTTTGGCTCGAG-3'. These reporters were subsequently digested with Bgl II and Dsa I, end-filled with the Klenow fragment of E. Coli DNA Pol I, and religated to remove the promoter region upstream of the TATA box that contains the 21 base pair repeat elements of the SV40 early promoter (29). The ZFHD1 expression vector was constructed by cloning a fragment encoding the hemagglutinin epitope MYPYDVPDYA (30) and ZFHD1 (28) into the Not I and Apa I sites of Rc/CMV(Invitrogen). Vectors expressing variants of ZFHD1 that contained mutations on the homeodomain surface were constructed as follows: The ZFHD1-encoding fragment was cloned into the Not I and Apa I sites of pBSKII+(Stratagene) and ssDNA was produced to be used as the substrate for oligonucleotide-directed mutagenesis (31) using VCSM13 helper phage (Stratagene) according to the manufacturer's protocol. Mutagenesis was performed as described (31) except that T7 DNA Polymerase was used in the initial polymerization step (32).

Oligonucleotides used for mutagenesis contained ~10 nucleotides on each side of the mutation-specific bases. Products of the mutagenesis reactions were screened by dideoxysequencing, and the desired fragments were isolated and cloned into the Not I and Apa I sites of Rc/CMV. The α TIF expression vector, pCMV1TIF1, containing the α TIF gene under control of the CMV promoter was a gift of J.L.C. McKnight. The 293 cells were transfected and the results quantitated as described (28).

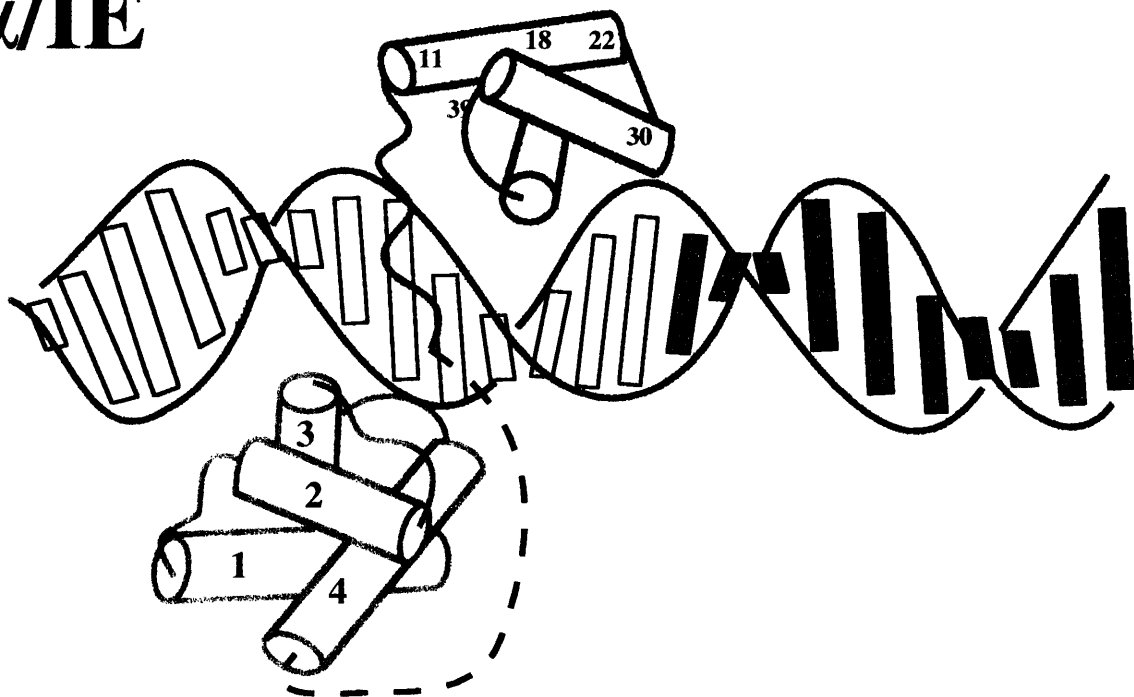
RESULTS

Recently we described the design and characterization of ZFHD1, a transcription factor composed of zinc fingers 1 and 2 from Zif268 fused to the Oct-1 homeodomain with a four residue linker (28). ZFHD1 displayed DNA-binding specificity in vitro that was distinct from that of either parental protein and, when fused to an acidic activation domain, activated transcription in a sequence specific manner in vivo. Although the optimal ZFHD1 binding site (28) has a different arrangement, computer modeling suggested that the linker between finger 2 and the homeodomain might also permit the ZFHD1 protein to bind DNA tightly in a configuration that would allow the homeodomain to be accessible for nucleation of a C1 complex. In this arrangement, finger 1 would bind the CNC triplet (CNCCCNNNTAATNN), finger 2 the CCN triplet (CNCCCNNNTAATNN), and the homeodomain would recognize the TAATNN sequence (CNCCCNNNTAATNN). This putative arrangement was used to design the novel α /IE-ZF element (Figure 1), which represents a fusion of this putative ZFHD1 binding site with the 3' portion of the α /IE element.

The Oct-1 POU domain and ZFHD1 were tested for binding to the α /IE and α /IE-ZF elements, and for the ability to nucleate C1 complex formation (Figure 2). As expected, the POU domain bound the α /IE element efficiently and readily formed the C1 complex upon addition of the α TIF and C1 factors (lanes 1 and 2). In contrast, the POU domain had a significantly lower affinity for the α /IE-ZF element, and did not efficiently nucleate the C1 complex on this site (lanes 3 and 4). At high protein concentrations, the extent of POU domain binding to the the α /IE-ZF element was comparable to that observed on the α /IE element; however, C1 complex formation was significantly less efficient on the α /IE-ZF element (lanes 9-12). This probably reflects binding of the POU-specific domain to sequences in the 3' portion of the α /IE-ZF element, analogous to that which has been observed for a related α /IE element which also lacks an ATGC binding site (34). Binding of the POU-specific domain to 3' sequences would be expected to

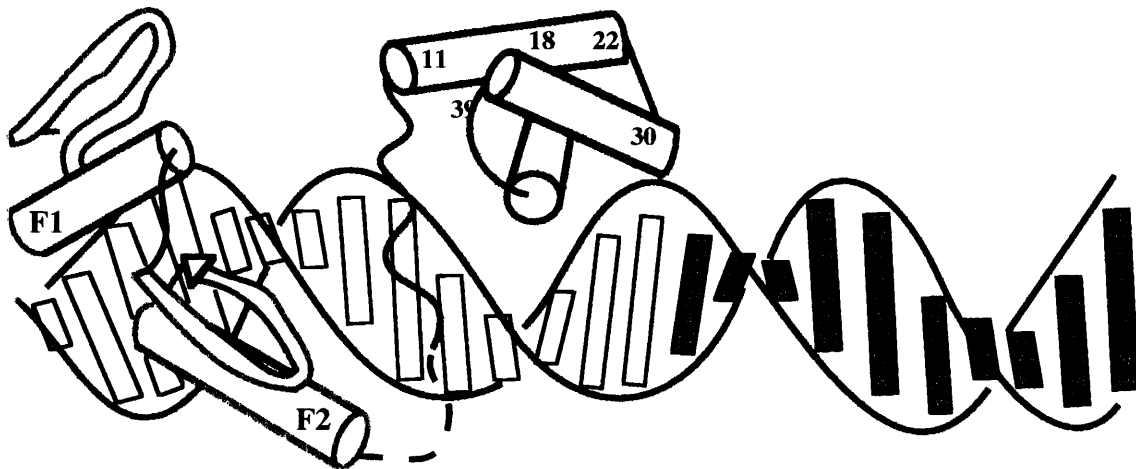
Figure 1. Schematic representation of the C1 complex. In the natural α /IE element (top), the Oct-1 POU-specific domain (green) recognizes the ATGC subsite, the homeodomain (red) binds the TAATGA subsite, and the remainder of the element (shaded base-pairs) is recognized by the α TIF and C1 factors. Residues on the surface of the homeodomain which were mutated in this study are shown on helix 1 (N11, K18, and E22), helix 2 (E30), and in the loop between helices 2 and 3 (N39). The α /IE element was converted to the α /IE-ZF element (bottom) by replacing the subsite for the POU-specific domain with binding sites for zinc finger 1 (purple) and zinc finger 2 (green) of ZFHD1 which recognize the CGC and CCT triplets, respectively. The depicted topology of protein-DNA complexes is based on the coordinates of Klemm et al. (1994) (6), Pavletich and Pabo (1991) (33), and computer modeling (28).

α/IE



5'-ATGCTAATGATATTCTTTGG-3'

$\alpha/IE-ZF$



5'-CGCCCTAGTAATGATATTCTTTGG-3'

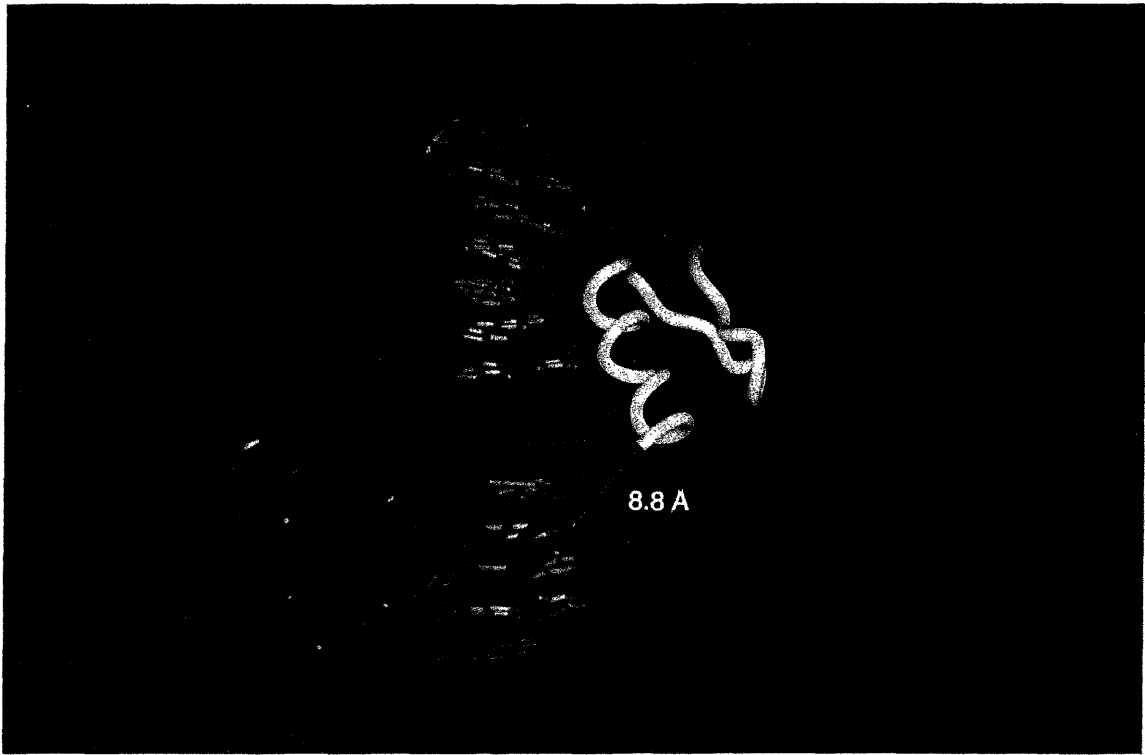
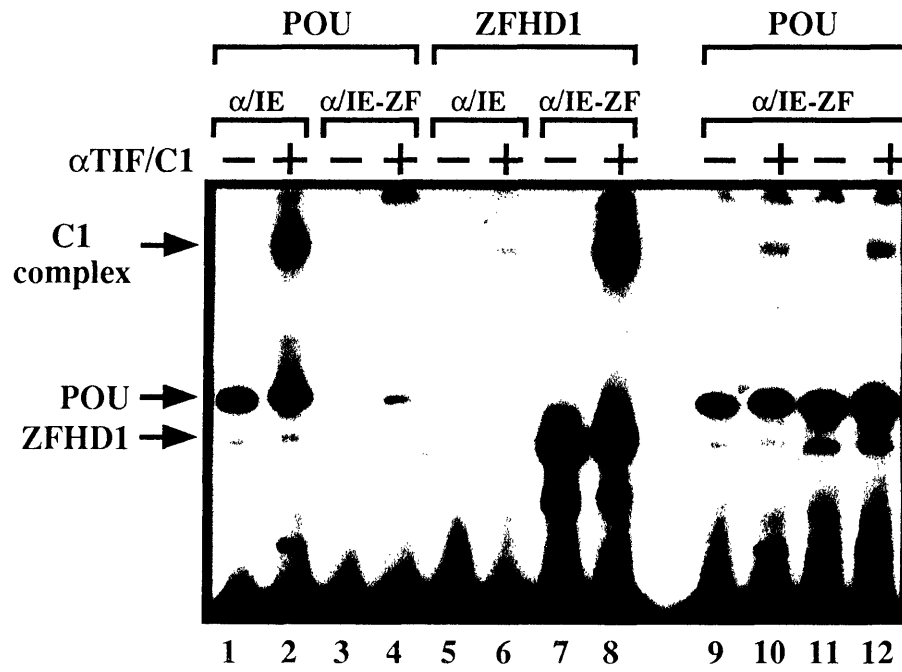


Figure 2. Comparison of the POU domain and ZFHD1 for DNA-binding specificity and ability to participate in C1 complex formation. DNA-protein binding reactions were done as described in Materials and Methods with a probe containing either the α /IE (5'-GTGCATGCTAATGATATTCTTTGG-3') (HSV α 0 probe (21)) or the α /IE-ZF (5'-CGC CCTAGTAATGATATTCTTTGG-3') element as indicated. The Protein A-Oct-1 POU fusion protein (200 pg, lanes 1-4; 500 pg lanes 9 and 10; 1000 pg lanes 11 and 12) or the GST-ZFHD1 fusion protein(200 pg) were incubated in the absence (-) or presence (+) of 15 ng PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor (13) as indicated. The reactions contained subsaturating concentrations of α TIF and C1 factor. The positions of the multiprotein C1 complex as well as the POU-DNA and ZFHD1-DNA complexes are indicated with arrows.



sterically interfere with α TIF and C1 factor association.

ZFHD1 displayed a clear preference for the α /IE-ZF element. The designed factor did not bind the natural α /IE element or nucleate complex formation at this site (lanes 5 and 6), but ZFHD1 efficiently bound the α /IE-ZF element and, upon addition of α TIF and C1, efficiently nucleated complex formation (lanes 7 and 8). Most importantly, comparable concentrations of the POU domain and ZFHD1 generated comparable DNA-binding activity and efficiency at forming the C1 complex at their preferred sites. This argues strongly that the 60 amino acids in the homeodomain provide all the Oct-1-mediated protein-protein interactions necessary for efficient C1 complex formation. Furthermore, the results clearly indicate that the designed chimeric protein could be used to recruit the enhancer complex to a novel site which is not efficiently recognized by the wild-type Oct-1 protein.

To determine whether ZFHD1 could target enhancer complex formation *in vivo* to the α /IE-ZF element, transient transfection experiments were performed (Figure 3). 293 cells were cotransfected with a vector expressing α TIF and a reporter construct containing two tandem copies of either the α /IE or α /IE-ZF element upstream of a minimal promoter and the luciferase gene (see Materials and Methods). Cotransfection of the α TIF expression vector and the α /IE reporter resulted in a 31-fold activation, reflecting assembly of the C1 complex with the endogenous Oct-1 and C1 factors. Both of these cellular factors are quite abundant and are probably not limiting for activity. In contrast, cotransfection of α TIF and the α /IE-ZF reporter resulted in only a 4-fold activation, presumably reflecting the reduced ability of the endogenous Oct-1 to bind to the α /IE-ZF element and form a C1 complex. Cotransfection of vectors expressing ZFHD1 and α TIF with the α /IE-ZF reporter resulted in a 34-fold activation. Further controls confirmed that activation was a consequence of α TIF association and C1 complex assembly. A third reporter contained a mutant α /IE-ZF element, designed to support ZFHD1 binding but containing nucleotide substitutions (21) expected to

abrogate α TIF binding and complex assembly (α /IE-ZF-3'MT:CGCCCTAGTAATGCTGTTCTTT). Cotransfection of α TIF and ZFHD1 with this reporter resulted in only a 4-fold activation. These results demonstrate the affinity and specificity of ZFHD1 for the α /IE-ZF element and the ability of ZFHD1 to recruit α TIF and C1 factors in vivo through interactions with the homeodomain.

The ability of ZFHD1 to nucleate complex formation and activate transcription at the α /IE-ZF element should be dependent on the recognition of the Oct-1 homeodomain surface that is solvent exposed when the domain binds DNA (13, 14). Several mutations on the surfaces of helices 1 and 2 of the homeodomain have been previously characterized in vitro for their effect on the cooperative interaction with α TIF and for their effect on C1 complex assembly (13). Five of these mutations were individually introduced into ZFHD1 and then assayed, in vivo and in vitro, for complex formation on the α /IE-ZF element. Single amino acid substitutions at positions 18, 22, and 30 on the homeodomain surface have been shown to severely reduce the ability of Oct-1 to participate in complex assembly, while substitutions at positions 11 and 39 have milder effects. As expected, these mutations had similar effects in the context of the chimeric ZFHD1 protein without affecting the DNA-binding characteristics of the designed protein (Figure 4A and 4B), confirming that the arrangement of the homeodomain in the ZFHD1/DNA complex mimics its arrangement in the POU domain/DNA complex. As compared to the 34-fold activation observed with the wild-type ZFHD1, the activation observed with the variants was 18-fold (N11A), 4-fold (K18E), 4-fold (E22A), 4-fold (E30Q), and 30-fold (N39H) (Figure 4A). The extent of in vitro complex formation for the variants directly paralleled their activity in vivo (Figure 4B). As compared to the level seen with the wild-type ZFHD1, the extent of C1 complex formed was 65% (N11A), 4% (K18E), 4% (E22A), undetectable (E30Q), and 102% (N39H) (Figure 4B). The results confirm that recognition of the surface of the Oct-1 homeodomain is critical for the assembly of the C1 complex in vivo.

Figure 3. Transcriptional activity of α TIF and ZFHD1 in vivo. The 293 cells were cotransfected with 5 μ g of reporter vector, 10.05 μ g total of expression vector, and 5 μ g of pCMV-hGH used as an internal control. Where indicated, 10 μ g of a vector expressing ZFHD1 (+) or the equivalent amount of the empty Rc/CMV vector (-) were cotransfected with (+) or without (-) 50 ng of pCMV1TIF1 (α TIF) . Reporter vectors contained two tandem copies of the α /IE element (5'-ATGCTAATGATATTCTTTGG-3'), the α /IE-ZF element (5'-CGCCCTAGTAATGATATTCTTTGG-3'), or the α /IE-ZF-3'MT element (5'-CGCCCTAGTAATGCTGTTCTTTGG-3'). The amount of luciferase activity obtained, normalized to hGH production, was set to 1.0 for the cotransfection of Rc/CMV with the α /IE reporter. Each bar represents the average of three independent experiments. Actual values and standard deviation, reading from left to right, are: 1.00 ± 0.26 ; 31.00 ± 4.17 ; 0.97 ± 0.26 ; 3.97 ± 0.46 ; 0.83 ± 0.03 ; 28.17 ± 4.53 ; 0.57 ± 0.07 ; and 2.37 ± 0.26 . Fold induction refers to the ratio of normalized activity obtained in the presence and absence of α TIF expression.

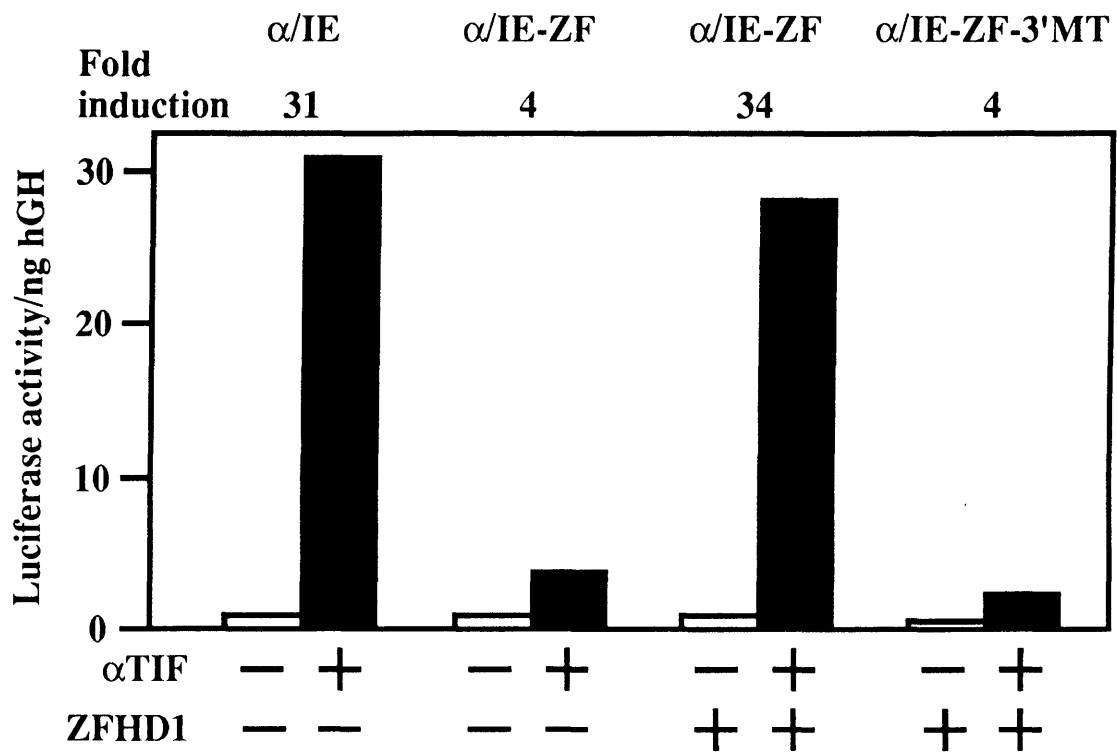
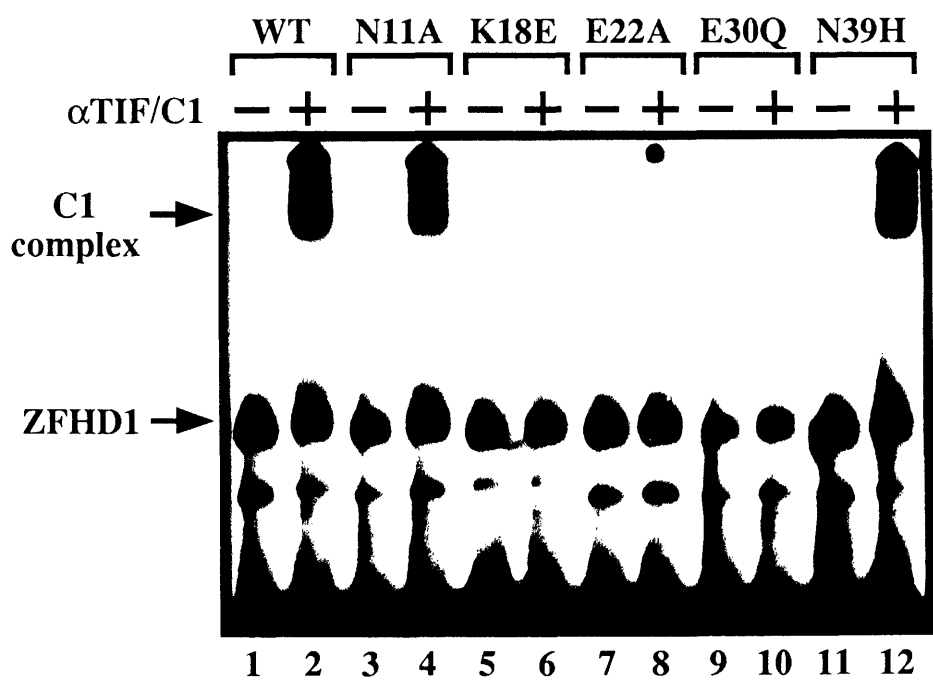
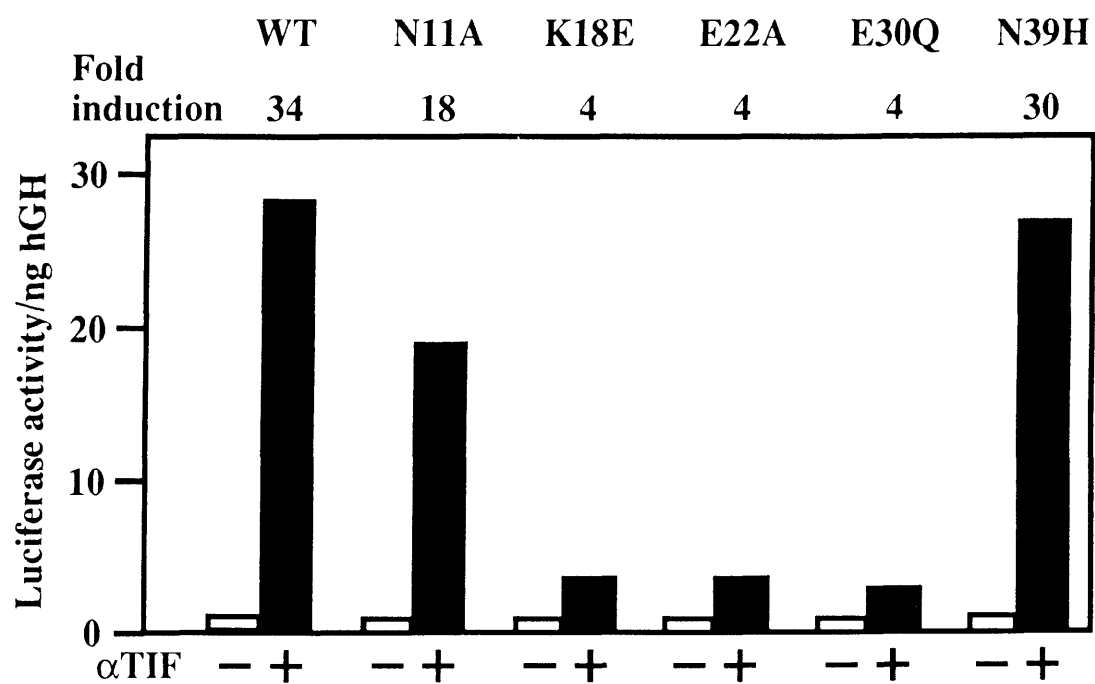


Figure 4. ZFHD1-mediated C1 complex formation depends upon recognition of the homeodomain surface. (A) The 293 cells were cotransfected with 5 μ g of reporter vector, 10.05 μ g total of expression vector, and 5 μ g of pCMV-hGH used as an internal control. Ten μ g of ZFHD1 expression vector encoding wild-type ZFHD1 (WT) or the indicated variants were cotransfected with the reporter containing the α /IE-ZF elements with (+) or without (-) 50 ng of pCMV1TIF1 (α TIF). Luciferase activity was normalized as in Figure 2. Each bar represents the average of three independent experiments. Actual values and standard deviation, reading from left to right, are: 0.83 ± 0.03 ; 28.17 ± 4.53 ; 1.03 ± 0.30 ; 18.73 ± 2.63 ; 0.77 ± 0.33 ; 3.30 ± 0.93 ; 0.80 ± 0.23 ; 3.23 ± 0.57 ; 0.73 ± 0.26 ; 2.63 ± 0.28 ; 0.90 ± 0.15 ; and 26.83 ± 3.10 . Fold induction refers to the ratio of normalized activity obtained in the presence and absence of α TIF expression. Comparable levels of expression of all ZFHD1 variants were confirmed by the presence of a gel shift activity, present only in extracts of transfected cells, that was reactive to the α HA monoclonal antibody 12CA5 (gift of K. Moberg). (B) The GST-ZFHD1 fusion protein (200pg) or variants containing the indicated substitutions were incubated in DNA-protein binding reactions in the absence (-) or presence (+) of 15 ng PA- α TIF and 1 μ l of a chromatographic fraction containing the HeLa cell C1 factor. The probe contained the α /IE-ZF (5'-CGCCCTAGTAATGATATTCTTTGG-3') element, and the reactions contained subsaturating concentrations of α TIF and C1 factor. The positions of the multiprotein C1 complex and of the ZFHD1-DNA complexes are indicated with arrows.



DISCUSSION

Structure-based design can provide valuable tools for studying the biological functions of transcription factors. We have used this strategy to dissect the Oct-1 POU domain and examine the role of the homeodomain in HSV α /IE gene regulation. The ability of the chimeric ZFHD1 protein to nucleate the C1 complex on the α /IE-ZF element *in vitro* with efficiency comparable to that of the intact POU domain is the best evidence to date that the Oct-1 homeodomain alone provides all of the protein-protein interactions that Oct-1 contributes to C1 complex formation. The primary role of the POU-specific domain in the formation of the wild-type C1 complex is to enhance the affinity and specificity of homeodomain-DNA association. The POU-specific domain can be replaced by the two zinc fingers in ZFHD1 without compromising the ability to nucleate C1 complex formation.

Both Oct-1 and C1 factor are abundant nuclear proteins that are expressed in most, if not all, cell types. The use of the α /IE-ZF element and the designed ZFHD1 factor permitted the assay of C1 complex formation *in vivo*, independent of endogenous Oct-1 activity. ZFHD1 stimulated gene expression *in vivo* through the α /IE-ZF element to a level (34-fold) comparable to that observed at the α /IE element with the endogenous Oct-1 protein (31-fold). This activity was dependent on 1) cotransfection of ZFHD1 with α TIF, 2) sequences 3' to the ZFHD1 binding site in the α /IE-ZF element which are recognized by α TIF and possibly the C1 factor, and 3) on the appropriate amino acids on the surface of the homeodomain which are recognized by α TIF (13). This specificity indicates that of the 743 amino acids in the Oct-1 protein, the 60 amino acid homeodomain is sufficient to nucleate C1 complex formation *in vivo* when efficiently targeted to the appropriate DNA sequence.

The structure-based design of ZFHD1 has permitted this analysis because the unique DNA binding specificity of the designed factor targets the homeodomain to a specific DNA site without disrupting the homeodomain-DNA interaction. Related

structure-based strategies should provide useful new approaches to the study of gene regulation in many other systems. The use of ZFHD1 for the analysis of homeodomain function provides a first example of how these design strategies can be used to characterize the biological activity of DNA-bound domains.

The chimeric ZFHD1 protein may also be useful for dissecting the role of the Oct-1 homeodomain in other regulatory contexts. Oct-1 is important for the regulated expression of the snRNA genes which are ubiquitously expressed, the histone H2B gene which is expressed in a cell-cycle-specific fashion, and the interleukin-2 and immunoglobulin genes which are expressed only in lymphoid tissues (35). It is possible that regulatory specificity in some or all of these contexts will be determined by recognition of the homeodomain surface by cellular homologues of α TIF. Indeed, the unique activity of the Oct-1 binding site in B cells has recently been attributed to a B-cell specific factor, OCA-B (Bob1, OBF-1), which associates with the Oct-1 POU domain and contains an activation domain (36-38).

The recognition of the homeodomain surface mediates functional specificity in a number of other systems. In *Drosophila*, for example, the differing abilities of the Ultrabithorax (Ubx) and Antennapedia homeodomain proteins to regulate the decapentaplegic gene is determined by their differing potentials for cooperative enhancer binding with the extradenticle (exd) protein (15, 16). The interaction of Ubx with exd is dependent on residues 22, 24, and 56 on the Ubx homeodomain surface (numbering scheme used in the homeodomain structural studies), and on the region C-terminal to the Ubx homeodomain (15). Cooperative interaction with exd also appears to modulate the target specificity of the abdominal-A and engrailed homeodomains (16). In human cells, the surfaces of helices one and two of the Phox1 homeodomain are important for the ability of the protein to recruit SRF and signal-responsive accessory factors to the c-fos serum response element (39). These studies underscore the importance of being able to dissect the specific protein-DNA and protein-protein interactions that occur at the

homeodomain surfaces. Because the structure and DNA-docking of homeodomains are so highly conserved (4), it may be possible to design ZFHD1-analogs with other homeodomains.

The study of transcription factor function has benefited greatly from the construction of chimeric proteins. For example, investigations of activation domains were greatly facilitated by fusing them to heterologous DNA-binding domains which would target the effector domain to a "neutral" element that would not support the binding of competing activities (40, 41). The success of these studies depended on the ability of activation domains to function when presented in a variety of spatial contexts. For the targeting of a DNA-binding domain to a novel site, one must insure that the chimeric protein permits the specific orientation that is required for DNA binding. Structure-based design provides a powerful technique for retargeting individual DNA binding domains in a stereochemically precise way.

An attractive feature of the structure-based design strategy is that it offers the opportunity to change binding specificity without introducing any mutations at the protein-DNA interface. For some domains, such mutations may be difficult to obtain (42), or may inadvertently influence the interaction with other factors (43-45). Computer modeling studies (28) suggest that the rational design of chimeric DNA-binding domains should be possible for many different DNA-binding modules, facilitating characterization of the protein-protein interactions that may define their biological activity.

ACKNOWLEDGEMENTS

We thank K. Cepek, J. Crispino, J. Kim, and A. MacMillan for critical reading of the manuscript; J. Klemm for help with graphics; M. Siafaca for her ever-present assistance; and Y. Qiu and R. Issner for indispensable technical assistance. This work was supported by U.S. Public Health Service Grant PO1-CA42063 from the National

Institutes of Health, by cooperative agreement CDR-8803014 from the National Science Foundation to P.A.S., and partially by the National Cancer Institute Cancer Center Support (core) grant P30-CA14051. J.L.P. is partially supported by the Harvard Medical School M.D.-Ph.D. program.

REFERENCES

1. Kornberg, T.B. (1993) *J. Biol. Chem.* **268**, 26813-26816.
2. Laughon, A. (1991) *Biochemistry* **30**, 11357-11367.
3. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., & Pabo, C.O. (1990) *Cell* **63**, 579-590.
4. Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., & Pabo, C.O. (1991) *Cell* **67**, 517-528.
5. Billeter, M., Qian, Y.Q., Otting, G., Muller, M., Gehring, W., & Wuthrich, K. (1993) *J. Mol. Biol.* **234**, 1084-1097.
6. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., & Pabo, C.O. (1994) *Cell* **77**, 21-32.
7. Kuziora, M.A., & McGinnis, W. (1989) *Cell* **59**, 563-571.
8. Gibson, G., Schier, A., LeMotte, P., & Gehring, W.J. (1990) *Cell* **62**, 1087-1103.
9. Mann, R.S., & Hogness, D.S. (1990) *Cell* **60**, 597-610.
10. Furukubo-Tokunaga, K., Muller, M., Affolter, M., Pick, L., Kloter, U., & Gehring, W.J. (1992) *Genes & Dev.* **6**, 1082-1096.
11. Lin, L. & McGinnis, W. (1992) *Genes & Dev.* **6**, 1071-1081.
12. Chan, S.K., & Mann, R.S. (1993) *Genes & Dev.* **7**, 796-811.
13. Pomerantz, J.L., Kristie, T.M., & Sharp, P.A. (1992) *Genes & Dev.* **6**, 2047-2057.
14. Lai, J.-S., Cleary, M.A. & Herr, W. (1992) *Genes & Dev.* **6**, 2058-2065.
15. Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., & Mann, R.S. (1994) *Cell* **78**, 603-615.
16. van Dijk, M.A. & Murre, C. (1994) *Cell* **78**, 617-624.
17. McKnight, J. L. C., Kristie, T.M., & Roizman, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7061-7065.
18. Gerster, T. & Roeder, R.G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6347-6351.
19. O'Hare, P. & Goding, C. R. (1988) *Cell* **52**, 435-445.
20. Preston, C. M., Frame, M.C., & Campbell, M.E.M. (1988) *Cell* **52**, 425-434.

21. Kristie, T. M., LeBowitz, J.H. & Sharp, P.A. (1989) *EMBO J.* **8**, 4229-4238.
22. Stern, S., Tanaka, M., & Herr, W. (1989) *Nature* **341**, 624-630.
23. Kristie, T.M. & Sharp, P.A. (1993) *J. Biol. Chem.* **268**, 6525-6534.
24. Wilson, A.C., LaMarco, K., Peterson, M.G., & Herr, W. (1993) *Cell* **74**, 115-125.
25. Kristie, T.M., Pomerantz, J.L., Twomey, T.C., Parent, S.A., & Sharp, P.A. (1995) *J. Biol. Chem.* **270**, 4387-4394.
26. Kristie, T.M. & Sharp, P.A. (1990) *Genes & Dev.* **4**, 2383-2396.
27. Verrijzer, C.P., Kal, A.J., & van der Vliet, P.C. (1990) *Genes & Dev.* **4**, 1964-1974.
28. Pomerantz, J.L., Sharp, P.A., & Pabo, C.O. (1995) *Science* **267**, 93-96.
29. Fromental, C., Kanno, M., Nomiyama, H., & Chambon, P. (1988) *Cell* **54**, 943-953.
30. Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A., & Wigler, M. (1988) *Mol. Cell. Biol.* **8**, 2159-2165.
31. Sayers, J.R., & Eckstein, F. (1991) in *Directed Mutagenesis: A Practical Approach*, ed. McPherson, M. (IRL Press, Oxford), pp. 49-69.
32. Sayers, J.R., Krekel, C., & Eckstein, F. (1992) *Biotechniques* **13**, 592-596.
33. Pavletich, N.P., & Pabo, C.O. (1991) *Science* **252**, 809-817.
34. Cleary, M.A. & Herr, W. (1995) *Mol. Cell. Biol.* **15**, 2090-2100
35. Herr, W. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K.R. (Cold Spring Harbor Laboratory Press, New York), pp. 1103-1135.
36. Luo, Y., Fujii, H., Gerster, T., & Roeder, R. G. (1992) *Cell* **71**, 231-241.
37. Strubin, M., Newell, J.W. & Matthias, P. (1995) *Cell* **80**, 497-506.
38. Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W., & Hovens, C.M., (1995) *Nature* **373**, 360-362.
39. Grueneberg, D.A., Simon, K.J., Brennan, K, & Gilman, M. (1995) *Mol Cell Biol.* **15**, 3318-3326.
40. Brent, R. & Ptashne, M. (1985) *Cell* **43**, 729-736.
41. Tanaka, M, Lai, J.-S., & Herr, W. (1992) *Cell* **68**, 755-767.

42. Pomerantz, J.L., & Sharp, P.A. (1994) *Biochemistry* **33**, 10851-10858.
43. Davis, R.L., & Weintraub, H. (1992) *Science* **256**, 1027-1030.
44. Ma, P.C.M., Rould, M.A., Weintraub, H., & Pabo, C.O. (1994) *Cell* **77**, 451-459.
45. Miner, J.N., & Yamamoto, K.R. (1992) *Genes & Dev.* **6**, 2491-2501.

EPILOGUE

FUTURE PROSPECTS FOR DESIGNED TRANSCRIPTION FACTORS

The most powerful potential application for designed DNA-binding proteins is the targeted regulation of specific genes. The use of such factors will only be made practical by the development of generalizable techniques that address several issues, including specificity, efficiency, regulatability, and delivery (i. e. for gene therapy). Because of the modularity of transcription factors, domains that mediate effector function (activation or repression) or regulatability (e.g. ligand-responsive nuclear localization) can be borrowed from natural factors. Therefore, designing a specific transcription factor may, to a first approximation, amount to attaching such domains to a DNA binding domain with a desired sequence specificity.

Can a domain be designed or selected to recognize any sequence? At present, the answer is an optimistic maybe. The first wave of zinc finger phage display experiments demonstrated that new fingers with new specificities could be selected (Rebar and Pabo, 1994; Jamieson et al., 1994; Choo and Klug, 1994ab). But they also showed that a finger could not be selected for the recognition of some base-pair triplets. These results may reflect an inherent limitation in the sequence-recognition capability of C2H2 zinc fingers, but it is too early to judge whether other factors (such as contextual effects) were at play. Similar experiments with zinc fingers from different proteins should illuminate how wide a range of sequences can be used to select fingers that bind with high affinity and specificity. It is also possible that selection experiments conducted with other domains might expand the diversity of sequences that can be recognized.

The structure-based design of transcription factors provides an alternative approach. Because the technique does not disrupt the wild-type protein-DNA interface it can take advantage of the natural DNA-binding specificity that a given domain has acquired during evolution under selection for function. A particular target sequence could be inspected for the presence of a proper arrangement of binding sites for structurally-characterized domains. These domains could then be fused and the affinity

and specificity of the chimeric domain tested. This strategy should complement the selection schemes if the technique can be extended.

The use of other DNA-binding domains to design chimeric proteins should be attempted. It would be valuable to test whether homologous domains behave in similar ways in chimeric domains, and whether there is generality in the structural prediction. Is ZFHD1 a model for the fusion of any two C2H2 zinc fingers to any homeodomain? It is unlikely that this is universally true, but there may be some possibility for establishing rules for fusing domains. For example, all homeodomains may not behave similarly in a chimeric protein, but perhaps all POU-homeodomains will because they are much more homologous to each other than to other homeodomains. One would hope that the structural and biochemical characterization of a few members of a transcription factor family would allow the application of general principles to many members of the family.

The computer modeling presented in Chapter 4 suggested that the design of chimeric proteins may be possible for other domains. Many DNA-binding domains fold into compact structures that have termini accessible for fusion when bound to DNA. As structural information accumulates for many DNA-binding proteins, it should be straightforward to test which combinations of domains are compatible.

The affinity and specificity of a designed chimeric domain might be modulated by residues in the linker that connects two domains. Since the linker might be juxtaposed to the DNA, residues may be able to contact the sugar-phosphate backbone, or make base-specific contacts, in either the major or minor grooves. The linker may also determine specificity by constraining the relative orientations of the fused domains such that the chimeric protein would be specific for a particular arrangement of subsites. The linker may accomplish this by imposing a length restriction between domains, and longer linkers may position the domains through direct protein-protein interactions. One way to explore this possibility would be to randomize the linker in ZFHD1 between finger 2 and the homeodomain, and then select for the variant that would bind the optimal ZFHD1 site

with highest affinity. Variants could also be selected for binding to a site of the form used in Chapter 5 for C1 complex formation. Any demonstration that the linker could influence specificity would demonstrate a general potential that could be exploited in other proteins.

For targeted gene regulation, it may not be necessary to have the capacity to bind to any sequence in the genome. Because transcription factors can function in a distance-independent and orientation-independent manner, one might only need to inspect the DNA sequence within 1000 base-pairs or so 5' or 3' to the initiation site for a gene-specific sequence which could serve as the target in a design scheme. For example, one might identify several occurrences of a TAAT homeodomain core binding site in the vicinity of the initiation site. Sequences adjacent to one of the TAAT subsites might resemble the subsite for a different domain which could be fused to the homeodomain in a sterically compatible arrangement. Alternatively, the adjacent sequences might be used as the target in phage display experiments for the selection of a zinc finger that could be fused to the homeodomain in the appropriate register. Regulation by designed factors would not necessarily require any knowledge of the normal mechanisms that regulate the gene in question.

A designed regulatable transcription factor might be used to affect the expression of a particular gene in an inducible, reversible manner, either in a cell line or a transgenic organism. An advantage of using designed transcription factors for gene regulation is that regulation could be conferred without irreversibly altering portions of the coding region of the gene under study, as is done in embryonic stem cell gene targeting (Bronson and Smithies, 1994). Some knockout methods allow temporal or spatial control of the deletion event (Gu et al., 1994), but they are not reversible. If made regulatable, a designed factor might be useful, for example, for the study of genes during particular temporal windows in the lifetime of an organism in situations where presence of the gene product is desired both before and after the window.

Until it is demonstrated that a factor can be designed to regulate an endogenous gene, it is difficult to invoke applications with confidence. For this experiment, any endogenous gene which can be assayed easily and quantitatively could be targeted. For example, the human growth hormone (hGH) gene, which is normally only expressed in the pituitary, might be used. The DNA sequence in the vicinity of the transcription initiation site is known (DeNoto et al., 1981). Expression vectors encoding hGH are often used as efficiency controls in transient transfection assays, and hGH is secreted into the medium and measured easily with readily available RIA kits (Selden et al., 1986). Any number of genes could be used for this experiment, and it would be practical to demonstrate the function of a designed factor in a tissue culture setting before introducing it into an organism. It would be important to consider how one might detect the inadvertent activity of the designed factor at genomic loci other than the one targeted. Apart from evaluating the gross toxicity of the factor, the best one might be able to do is to assay the activity of control elements that contain sequences that are related to the target but should not bind the factor.

Therapeutic applications of designed transcription factors might become possible, either for the activation of expression of a particular desired gene product, or for the inhibition of expression of a disease-causing gene. Such applications would depend upon systems that would ensure the safe and efficient delivery of factors, and progress is being made in these areas for other gene therapy approaches (Mulligan, 1993). In the meantime, there exists the opportunity to focus on the possibilities for manipulating protein-DNA interactions to achieve desired specificities.

REFERENCES

- Bronson, S.K., and O. Smithies. 1994. Altering mice by homologous recombination using embryonic stem cells. J. Biol. Chem. **269**: 27155-27158.
- Choo, Y. and A. Klug. 1994. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc. Natl. Acad. Sci. USA **91**: 11168-11172.
- Choo, Y. and A. Klug. 1994. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. USA **91**: 11163-11167.
- DeNoto, F.M., D.D. Moore, and H.M. Goodman. 1981. Human growth hormone DNA sequence and mRNA structure: possible alternative splicing. Nucl. Acids Res. **9**: 3719-3730.
- Gu, H., J.D. Marth, P.C. Orban, H. Mossmann, K. Rajewsky. 1994. Deletion of a DNA Polymerase β gene segment in T cells using cell type-specific gene targeting. Science **265**: 103-106.
- Jamieson, A. C., S.-H. Kim, and J.A. Wells. 1994. In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry **33**:5689-5695.
- Mulligan, R.C. The basic science of gene therapy. 1993. Science **260**: 926-932.
- Rebar, E. and C.O. Pabo. 1994. Zinc finger phage: Affinity selection of fingers with new DNA-binding specificities. Science **263**:671-673.
- Selden, R.F., K. Burke-Howie, M.E. Rowe, H.M. Goodman, and D.D. Moore. 1986. Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Mol. Cell. Biol. **6**: 3173-3179.

BIOGRAPHICAL NOTE

Joel L. Pomerantz

EDUCATION

Massachusetts Institute of Technology

Ph.D., Biology, 1995

September 1991-present
Department of Biology
Thesis Advisor: Phillip A. Sharp, Ph.D.

Harvard Medical School, Boston, MA

M.D. candidate

September 1989-present
Harvard-M.I.T. Division of Health Sciences and Technology
M.D.-Ph.D. Program, Medical Scientist Training Program
Sterling Winthrop Research Fellow in Health Sciences and Technology, 1992-1993

Brandeis University, Waltham, MA

B.A., 1989

September 1985-May 1989
summa cum laude, with Highest Honors in Biochemistry
Phi Beta Kappa, elected 1988
Julian J. and Helen R. Behr Scholarship Prize, 1989
Elihu A. Silver Prize, 1988
Nathan and Bertha Richter Award, 1988
CRC Press Freshman Chemistry Achievement Award, 1986
Justice Brandeis Scholar, 1985-1989
Undergraduate Thesis Advisor: Ranjan Sen, Ph.D., Dept. of Biology

POSITIONS

June 1990-August 1991

Research Assistant
Massachusetts Institute of Technology, Cambridge, MA
Laboratory of Phillip A. Sharp, Ph.D.

May 1987-September 1987

Summer Student/Technician
Brigham and Women's Hospital,
Harvard Medical School, Boston, MA
Laboratory of K. Frank Austen, M.D.

TEACHING EXPERIENCE

7.60 Cell Biology of the Nucleus, Spring 1993, MIT

Graduate Teaching Assistant

7.02 Introductory Biology Laboratory, Spring 1994, MIT

Graduate Teaching Assistant

PUBLICATIONS

Pomerantz, J.L., C.O. Pabo, and P.A. Sharp. 1995. Analysis of homeodomain function by structure-based design of a transcription factor. Proc. Natl. Acad. Sci.USA, in press.

Pomerantz, J.L., P.A. Sharp, and C.O. Pabo. 1995. Structure-based design of transcription factors. Science, 267, 93-96.

Kristie, T.M., J.L. Pomerantz, T.C. Twomey, S.A. Parent, And P.A. Sharp. 1995. The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. Journal of Biological Chemistry, 270, 4387-4394.

Pomerantz, J.L., and P.A. Sharp. 1994. Homeodomain determinants of major groove recognition. Biochemistry, 33, 10851-10858.

Pomerantz, J.L., T.M. Kristie, and P.A. Sharp. 1992. Recognition of the surface of a homeo domain protein. Genes & Development 6, 2047-2057.

Pomerantz, J.L., F. Mauxion, M. Yoshida, W.C. Greene, and R. Sen. 1989. A second sequence element located 3' to the NF-kB binding site regulates IL-2 receptor-alpha gene induction. Journal of Immunology 143, 4275-4281.

Rothenberg, M.E., J.L. Pomerantz, W.F. Owen, Jr., S. Avraham, R.J. Soberman, K.F. Austen, and R.L. Stevens. 1988. Characterization of a human eosinophil proteoglycan, and augmentation of its biosynthesis and size by interleukin 3, interleukin 5, and granulocyte/macrophage colony stimulating factor. Journal of Biological Chemistry 263, 13901-13908.

INVENTIONS AND PATENTS

Chimeric Transcription Factors

Inventors: Joel L. Pomerantz, Phillip A. Sharp, Carl O. Pabo

Patent Application filed on December 29, 1994