The EMBO Journal vol.13 no.6 pp.1434-1442, 1994

Recognition of a DNA operator by a dimer composed of two different homeodomain proteins

Caroline Goutte[1] and Alexander D.Johnson[1,2]

[1]Department of Biochemistry and Biophysics and [2]Department of Microbiology and Immunology, University of California, San Francisco, CA 94143-0502, USA

Communicated by R.Kahmann

The yeast homeodomain proteins a1 and α2 interact to form a heterodimer that binds DNA with high specificity. The DNA recognition element consists of two similar half-sites, arranged with dyad symmetry and separated by a fixed number of base pairs. We demonstrate that in the a1α2–DNA complex, one of these half-sites is bound by a1 while the other is bound by α2. These assignments allow a comparison of the chemical and nuclease protection patterns produced by both proteins when bound together to the *hsg* operator. Contrary to simple expectations, we propose that the a1 and α2 homeodomains are arranged on the DNA in tandem, despite the fact that the recognition sequence is dyad symmetric.
Key words: combinatorial control/DNA–protein interactions/heterodimer/mating type/*S.cerevisiae*

Introduction

The yeast **a**1 and α2 proteins belong to the super-family of homeodomain proteins, members of which have been identified in many eukaryotic species (for reviews, see Gehring, 1987; Scott *et al.*, 1989). The homeodomain is a stretch of 60 amino acids that adopts a characteristic structure of three tightly packed α-helices that can recognize DNA in a sequence-specific manner. NMR spectroscopy and X-ray diffraction data available for three DNA-bound homeodomains (the *Drosophila* Engrailed, *Drosophila* Antennapedia and the yeast α2) have revealed not only that these domains share a common structure, but also that they interact with DNA in a very similar manner (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Phillips *et al.*, 1991; Wolberger *et al.*, 1991).

a1 and α2 are also related to many of the other homeodomain proteins in a functional sense since they are cell-type-specific proteins that act as regulators of cell fate. For the diploid cell type of *Saccharomyces cerevisiae* (the **a**/α cell) to differentiate appropriately, genes that are specifically required in the haploid cell types (**a** and α cells) must be repressed (for reviews, see Herskowitz, 1989; Sprague, 1990; Dolan and Fields, 1991; Johnson, 1992). These genes are collectively called the haploid-specific genes (*hsg*), and each is tagged with one or more copies of an operator called the haploid-specific gene (or *hsg*) operator (Siliciano and Tatchell, 1984, 1986; Miller *et al.*, 1985; Goutte and Johnson, 1988; Dranginis, 1990). The role of **a**1 and α2 proteins is to recognize these operators, marking the linked genes for transcriptional repression. The repression itself is carried out by additional proteins, including TUP1 and SSN6 (Mukai *et al.*, 1991; Keleher *et al.*, 1992). The key regulatory feature of haploid-specific gene repression is the requirement for both **a**1 and α2 proteins in the recognition of the *hsg* operator. Thus, neither purified **a**1 protein nor purified α2 protein alone can efficiently recognize the *hsg* operator; however, the two purified proteins together (the situation mimicking that in the **a**/α cell) give very tight and specific binding to the *hsg* operator (Goutte and Johnson, 1988; Dranginis, 1990; Goutte and Johnson, 1993). **a**1 and α2 form a weak heterodimer in solution (A.Mak and A.D.Johnson, submitted) which binds tightly to the *hsg* operator.

In addition to acting in combination with **a**1, α2 also acts in combination with the MCM1 protein, a non-homeodomain protein that is related to the human serum-response factor (Norman *et al.*, 1988). α2 and MCM1 bind as a hetero-tetramer to a DNA sequence known as the *asg* (**a**-specific gene) operator (Keleher *et al.*, 1988; Ammerer, 1990; Vershon and Johnson, 1993). As illustrated in Figure 1, the *hsg* and *asg* operators are related in that both consist of a pair of similar half-sites arranged symmetrically (Johnson and Herskowitz, 1985; Miller *et al.*, 1985; Siliciano and Tatchell, 1986).

In this paper, we demonstrate, both *in vivo* and *in vitro*, that the spacing of the operator half-sites is the primary feature through which **a**1α2 distinguishes the *hsg* operator from the *asg* operator. From a series of chemical and nuclease probe experiments, we deduce the positioning of **a**1 and α2 on the *hsg* operator, and suggest that the homeodomains are bound to DNA in a tandem arrangement that does not reflect the dyad symmetry of the recognition site.

Results

The spacing of half-sites in the hsg operator is critical for recognition by a1α2

The symmetrically arrayed half-sites of *hsg* operators (recognized by **a**1α2) are similar to those of *asg* operators (recognized by α2–MCM1). However, the spacing between

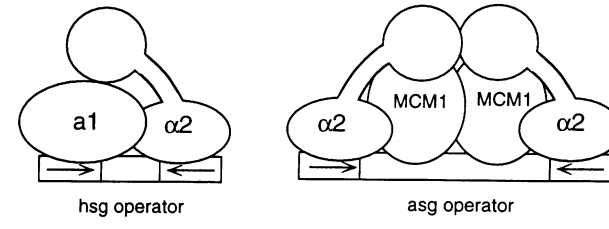

Fig. 1. The haploid-specific gene operator (*hsg*) is recognized by an **a**1α2 heterodimer and the **a**-specific gene operator (*asg*) is recognized by an α2MCM1 heterotetramer. The symmetric half-sites of the *hsg* operator (indicated by arrows) are similar to those of the *asg* operator.

A

Operator	Promoter Activity a	Promoter Activity α	Promoter Activity a/α	Fold Repression (a/a/α)	Estimated Operator Occupancy in a/α
NO OPERATOR	481	520	390	1x	–
hsg	574	538	10.6	50x	98%
hsg +13	28	30	27	1x	0%
asg*	108	64	76	1x	0%
asgΔ13	313	287	103	3x	67%

B

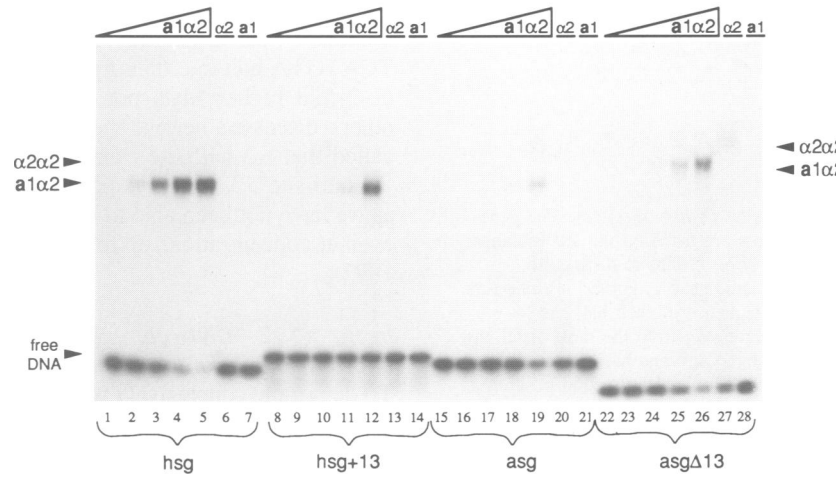

Fig. 2. Behavior of operators with altered half-site spacing. (**A**) Repression measured *in vivo*. The indicated operators (see Materials and methods for sequences) were each inserted into a test promoter that drives the expression of the *lacZ* gene; promoter activity could thus be measured in terms of β-galactosidase activity. Promoter activity was tested in the three different yeast cell types for each operator. Numbers are units of β-galactosidase activity and represent an average of at least six separate experiments. The fold repression is calculated as a ratio of promoter activity in **a** cells to that in **a**/α cells. Operator occupancy is calculated as the percentage of operators that are occupied by **a**1α2 ($100-100/x$, where x is the fold repression); this calculation is based on the assumption that the extent of repression directly correlates with operator occupancy. The operators diagrammed in white are derivatives of a naturally occurring *hsg* operator, while those indicated by shading are derived from a naturally occurring *asg* operator (see Materials and methods for details). The striped region represents a replacement of the MCM1 binding site with an unrelated sequence. (**B**) Relative binding affinities measured *in vitro*. Four different operators (indicated at the bottom) were used in an electrophoretic mobility shift experiment. Increasing amounts of purified **a**1 and α2 protein were added as follows: lanes 2, 9, 16 and 23, 4×10^{-9} M **a**1 and 2×10^{-9} M α2; lanes 3, 10, 17 and 24, 10^{-8} M **a**1 and 7×10^{-9} M α2; lanes 4, 11, 18 and 25, 3×10^{-8} M **a**1 and 2×10^{-8} M α2; lanes 5, 12, 19 and 26, 10^{-7} M **a**1 and 6×10^{-8} M α2; lanes 6, 13, 20 and 27, 6×10^{-8} M α2; lanes 7, 14, 21 and 28, 10^{-7} M **a**1; lanes 1, 8, 22, no protein. **a**1α2–operator complexes and α2α2–operator complexes are indicated.

these sites is significantly different, as depicted in Figure 1. In every known *hsg* operator, the half-sites are separated by 6 bp, while in every known *asg* operator the half-sites are separated by 18 or 19 bp (Johnson and Herskowitz, 1985; Miller *et al.*, 1985). The spacing of half-sites in *asg* operators has been shown to be critical for the co-operative binding of α2 and MCM1 (Smith and Johnson, 1992). Based on these considerations, it seemed likely that the spacing of the two half-sites is the key determinant that distinguishes an operator that is to be recognized efficiently by the **a**1α2 heterodimer from one that is to be recognized by α2MCM1. To test this idea, we created two artificial constructs: an *asg* operator from which 13 bp were deleted from the center (referred to as *asgΔ13*) so that the half-sites of this operator are brought into register with those found in naturally occurring *hsg* operators (6 bp apart), and an *hsg* operator to which 13 bp were added between the two half-sites such that the half-sites were brought into register with those found in *asg* operators (19 bp apart; referred to as *hsg+13*). The operators are diagrammed in Figure 2.

Each operator was inserted into a yeast CYC1 promoter that directs expression of *Escherichia coli* β-galactosidase and then tested for its ability to bring about cell-type-specific repression *in vivo*. While a wild-type *hsg* operator produces a 50-fold repression of promoter activity in **a**/α cells (but not in **a** or α cells), it loses this cell-type-specific effect if its half-sites are separated by an extra 13 bp (*hsg+13*; Figure 2A). In the reciprocal experiment, the half-sites of an *asg* operator are brought closer by 13 bp (*asgΔ13*); this operator is unable to function as an *asg* operator, but shows a reproducible 3-fold repression of promoter activity in **a**/α cells. [Assuming a direct relationship between operator occupancy and promoter repression, the wild-type *hsg* operator (50-fold repression) is calculated to have an occupancy of 98% in the **a**/α cell, and the *asgΔ13* operator (3-fold repression) is calculated to have an occupancy of 67% in **a**/α cells; see Figure 2.] This experiment shows that the spacing of the half-sites is a key feature that distinguishes an *hsg* operator from an *asg* operator. In addition to the correct spacing, the *asg* operator must also contain a

PROMOTER	a1 HALF	α2 HALF
MATα1:	TGATGTACTT	CAATGTAGAA
RME(1):	TGATGTAATC	AGATGTCACA
STE5(1):	TGATGTGTAA	GCTTGTTAAT
STE5(2):	TGATGCAGAA	TCATGTACTT
HO(1):	TGATGAAGCG	GCGTTTAGAA
HO(2):	TGATGTTAAT	TCATGTCCAC
HO(3):	TGATGTAAAT	TCATGTTATT
HO(4):	TGATGTAACT	TCGTGTATTT
HO(5):	TGATGCAGTT	ACATGTCTTC
HO(6):	TGATGTGAAT	TCATGTATTC
HO(7):	CGATGTGCTT	TAGAGTGAAA
HO(8):	TGATGTATCT	GCCTGCGATG
HO(9):	GGATGTAACT	TTATGTTAAA
HO(10):	TGATGTAGGT	CCGCGTTAAA

position: 1 2 3 4 5 6 7

Consensus a1 half (out of 14): $T_{12}G_{14}A_{14}T_{14}G_{14}T_{11}A_{10}A_7A_5T_{10}$

Consensus α2 half (out of 14): $T_7\ C_{10}A_8\ T_{12}G_{13}T_{13}A_5\ A_7A_7A_6$

Consensus half site of asg operators: $C_9A_7T_{10}G_{10}T_{10}A_{10}A_8T_6T_{10}$ (out of 10)

Fig. 3. Comparison of the two half-sites of *hsg* operators. The genes whose promoters contain *hsg* operators are listed on the left [numbers in parentheses are used as in Miller *et al.* (1985) to distinguish different *hsg* operators found in the same gene promoter]. Because the operators have dyad symmetry, the sequence of each half-site is presented in a 5′ to 3′ direction, progressing from the most distal base inwards toward the center of symmetry. Only one half-site in each operator starts with TG (or at least with a G in the second position); we used this criterion to group all such half-sites into one group (**a**1 half) and the corresponding half-sites were put into the other group (α2 half). Sequences were obtained from Miller *et al.* (1985), and references therein, and Covitz *et al.* (1991). Consensus sequences for the two half-sites are derived from the list presented here. Numbers next to each base correspond to the number of incidences (out of 14) of that base. The consensus half-site sequence of *asg* operators is shown for comparison [see Johnson and Herskowitz (1985), and references therein, for additional sequence information].

recognition sequence for MCM1, as indicated by the null behavior of an *asg* operator whose MCM1 recognition sequence has been replaced by an unrelated sequence (*asg* in Figure 2; see Keleher *et al.*, 1988).

We next compared the affinity of purified **a**1α2 for the different operators *in vitro* using an electrophoretic mobility shift assay (Figure 2B). The affinity of **a**1α2 for the *hsg* operator is reduced ~10-fold when its half-sites are spread from 6 bp apart to 19 bp apart (compare lanes 1–5 with 8–12). In the reciprocal experiment, the affinity of **a**1α2 for the *asg* operator is enhanced ~3-fold when its half-sites are brought closer together by 13 bp (compare lanes 15–19 with 22–26). These results correlate with the effects of the operators *in vivo*, demonstrating that the **a**1α2 heterodimer has a preference for operators with a half-site spacing of 6 bp (as found in the naturally occurring *hsg* operators) rather than 19 bp (as found in the naturally occurring *asg* operators).

The two half-sites of hsg operators are not functionally equivalent

The observation that a naturally occurring *hsg* operator functions better both *in vivo* and *in vitro* than the *asgΔ13* operator (Figure 2A and B) suggests that in naturally occurring *hsg* operators there exists more information than simply the presence of two appropriately spaced half-sites. A comparison of the sequences of all known *hsg* operators reveals that a given *hsg* operator is composed of two different types of half-sites [Figure 3 and Miller *et al.* (1985)] and two slightly different consensus sequences can be derived for the two half-sites. One half, and only one half, of all *hsg* operators has the consensus sequence 5′-TGATGTA-3′. (We will refer to the positions in a half-site as positions #1–#7, reading in the 5′ to 3′ direction.) The other half-site of an *hsg* operator usually has a C in place of the G at position #2, and positions #1 and #7 are less conserved; the consensus for this half-site is 5′-(T)CATGT(A)-3′. Since both **a**1 and α2 are known to contact the *hsg* operator (Goutte and Johnson, 1993), it seemed likely that one class of half-sites would be specifically recognized by **a**1 and the other by α2. We tested the binding preference of purified α2 to isolated half-sites of each class and found that an α2 monomer has a slightly higher affinity for a single TCATGTA half-site than for a single TGATGTA half-site (~2-fold higher; data not shown). For this reason, and others discussed below, the TCATGTA half-site will be called the 'α2 half-site', and the TGATGTA half-site the '**a**1 half-site'. A similar experiment for **a**1 was not possible as we have not been able to observe **a**1 binding on its own, even at concentrations up to 10^{-5} M (Goutte and Johnson, 1993).

In the a1α2–DNA complex, a1 contacts the 'a1 half-site' while α2 contacts the 'α2 half-site'

To determine conclusively which protein is bound to which half-site, we performed a protein–DNA cross-linking experiment in which a single half-site was tagged with bromodeoxyuridine (BrdU). Upon exposure to UV light, BrdU is covalently cross-linked to protein more readily than are other nucleotides [see, for example, Chodosh (1988)]. The increased sensitivity of BrdU allowed us to obtain experimental conditions in which **a**1 and α2 could be cross-linked to a labeled DNA operator only if the protein was contacting the DNA in proximity to a BrdU substitution. We prepared two operators (Figure 4A): one in which thymidine-to-BrdU substitutions were made in the **a**1 half (B1) and one in which they were made in the α2 half (B2). An operator with the same sequence, but lacking BrdU substitutions (WT), was used as a control for background levels of non-BrdU-mediated cross-linking events. **a**1α2–operator complexes were exposed to UV light, denatured and immunoprecipitated with antibody against **a**1 or α2. These immunoprecipitates were electrophoresed through an SDS gel and visualized by autoradiography to determine whether the immunoprecipitated protein had become ^{32}P labeled by virtue of being cross-linked to the [^{32}P]BrdU DNA (Figure 4A).

The results of this experiment, interpreted in Figure 4B, demonstrate that both **a**1 and α2 can be cross-linked to the *hsg* operator, but only if BrdU substitutions have been made in the appropriate half-site. When **a**1 and α2 were mixed with ^{32}P-labeled B1 operator, **a**1 but not α2 was covalently cross-linked to the operator, and hence acquired a label (Figure 4A, lanes 4–6). In contrast, when ^{32}P-labeled B2 operator was used, α2 but not **a**1 was labeled (lanes 7–9). The observed protein–DNA cross-linking is dependent on the presence of both **a**1 and α2, as no cross-linking is observed if either of the proteins is omitted (lanes 1–3 and 13–15).

The entire **a**1α2 complex could be precipitated with

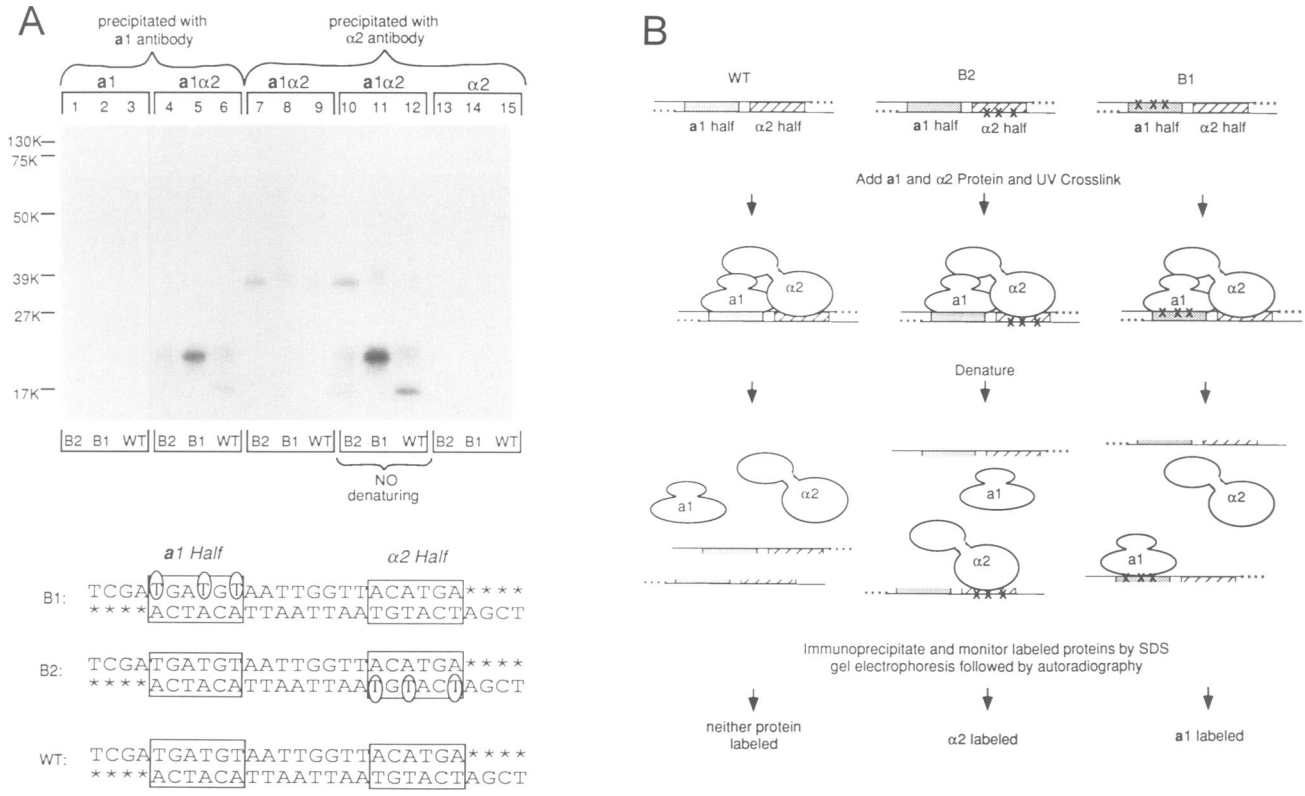

Fig. 4. Binding of **a**1 and α2 to specific half-sites of the *hsg* operator. Each protein−DNA cross-linking reaction contained one of three different operators which are depicted at the bottom of **panel A**. Asterisks indicate ^{32}P-labeled bases and circled Ts indicate the positions of the three BrdU substitutions made in each operator (positions #1, #4 and #6). The only difference between the **a**1 half and the α2 half of the operators is the base pair at position #2 of each half-site. Operators were incubated with 7×10^{-8} M **a**1 (lanes 1−3), or 4×10^{-8} M α2 (lanes 13−15), or both 7×10^{-8} M **a**1 and 4×10^{-8} M α2 (lanes 4−12). Following UV exposure, all reactions, except those of lanes 10−12, were denatured. All reactions were immunoprecipitated with the indicated polyclonal antibodies and precipitated products were denatured prior to electrophoresis through a 19% gel. Prestained molecular weight markers (BRL) were used as size standards. (**B**) Schematic representation of the results of the experiment. The **a**1 half of the operator is filled in and the α2 half is hatched. BrdU substitutions are represented by Xs.

antibodies directed against α2 if the denaturation step was omitted (Figure 4A, lanes 10−12). Again, under these conditions, **a**1 was labeled only with the B1 operator and α2 only with the B2 operator. The results of these experiments prove that the 'α2 half-site' of *hsg* operators is indeed bound by the α2 protein, while the '**a**1 half-site' is bound by the **a**1 protein of the **a**1α2 heterodimer.

a1α2 contacts its operator in an asymmetric fashion

When the **a**1α2 heterodimer is bound to the *hsg* operator, it protects a continuous stretch of 26 bp, including both half-sites of the operator, from cleavage by DNase I (Figure 5A and summarized in Figure 7). Within this 26 bp region, **a**1α2 protects several regions of the DNA backbone from attack by hydroxyl radical (Figure 5B). The protected backbone positions all map to the same face of the DNA helix, indicating that **a**1α2 interacts with its operator on one side of the DNA helix (see Figure 7).

Although the *hsg* operator is roughly symmetric in sequence, the DNA protection patterns produced by the binding of **a**1α2 are noticeably asymmetric. For example, an enhanced DNase I cleavage site is observed at position #6 (T) of the **a**1 half-site (filled arrowhead, Figure 5A, and also see Figure 7), but not on the α2 half-site (Figures 5A and 7). More strikingly, the pattern of DNA backbone contacts revealed by hydroxyl radical protection experiments is not 2-fold symmetric with respect to the center of the operator. For example, the protection pattern on the **a**1 half shows an extra stretch of four backbone contacts that are not observed on the α2 half (Figure 5B and see comparison in Figure 7). Dimethyl sulfate (DMS) protection experiments also reveal an asymmetric pattern: on the **a**1 half-site, the guanine at position #2 is protected from methylation while the methylation of the guanine at position #5 is enhanced; on the α2 half-site, the pattern is reversed (Figure 6B, lanes 2 and 7, and Figure 7). These results indicate that although each half-site of the *hsg* operator is contacted by a homeodomain protein, the way in which these two proteins interact with the similar sequences is detectably different.

Comparison of DNA contacts made by the a1α2 heterodimer and the α2α2 homodimer

To learn more about the contributions of **a**1 and α2 to the asymmetric **a**1α2 protection patterns, this pattern was compared with that observed for an α2 homodimer bound to the same sequence. Although binding of the α2 homodimer is considerably weaker than that of the **a**1α2 heterodimer, it can be forced to occupy the *hsg* operator at high protein concentration (Goutte and Johnson, 1993). Both the homodimer and the heterodimer protect the same 26 bp of the *hsg* operator from cleavage with DNase I (Figure 6A). However, in contrast to the asymmetric protection patterns produced by the binding of **a**1α2, the α2 homodimer produces protection patterns that are 2-fold symmetric around the center of the operator, indicating that both monomers make similar protein−DNA contacts on the two half-sites.

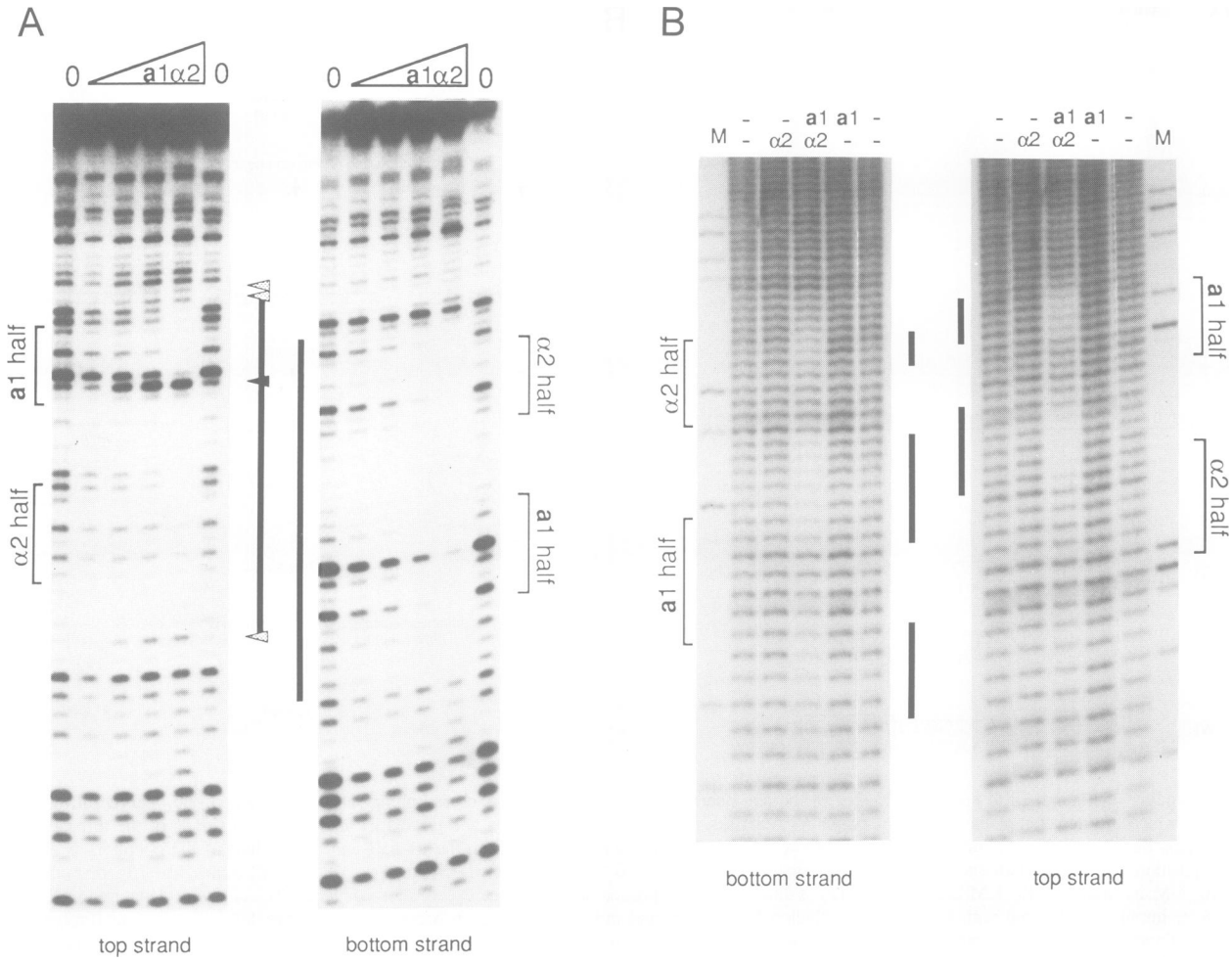

Fig. 5. DNA protection patterns produced by the binding of **a**1α2 to the *hsg* operator. Protection patterns are shown for both DNA strands of the *hsg* operator (see Materials and methods for sequences). (**A**) Protection from DNase I. (**B**) Protection from hydroxyl radical. The **a**1 and α2 halves of the operator are indicated by brackets. Thick lines indicate the regions of protection. Arrowheads indicate enhanced DNase I cleavage sites (gray arrowheads for weak enhancements and black arrowheads for strong enhancements). In panel A, the outer two lanes have no protein, all other lanes have 2×10^{-8} M α2 and the concentration of **a**1 increases from left to right as follows: 2×10^{-9} M; 6×10^{-9} M; 1.8×10^{-8} M; 5×10^{-7} M. In panel B, the outer lanes have no protein; other lanes contain 7×10^{-8} M **a**1 or 8×10^{-8} M α2, or both, as indicated above each lane. Marker lanes contain DMS reactions performed on the operator as sequence markers.

For example, binding of an α2 homodimer causes the guanine at position #2 of both the **a**1 half and the α2 half to show enhanced methylation, while the guanines at position #5 in the two half-sites are both protected from methylation (Figure 6B, lanes 3 and 6). We also note that the protection patterns observed for α2 (with one exception—the enhanced methylation of guaninase at the ends of the operator) are the same whether it is bound as a homodimer or as a heterodimer with **a**1 (compare lanes 6 and 7 of Figure 6B, and see also Figure 7). Taken together, these results suggest that α2 contacts the DNA in a similar way whether or not it interacts with **a**1, and that **a**1 contacts the DNA in a way quite different from that of α2. Specific models that account for this difference are proposed in the Discussion.

Discussion

Relevance of the experiments performed in vitro to the situation in vivo

Several observations indicate that our conclusions established from biochemical experiments with **a**1 and α2 purified from *E.coli* apply *in vivo* as well. First, the reconstituted **a**1α2 activity behaves identically in DNA binding experiments to the **a**1α2 activity found in extracts of yeast of the appropriate cell type, the **a**/α (Goutte and Johnson, 1993). Second, the DNA binding site (the *hsg* operator) used for the chemical protection experiments of Figures 4–8 is known to respond to the **a**1α2 combination *in vivo* both in its natural setting (Siliciano and Tatchell, 1986), and when synthesized and inserted into a naive promoter (Goutte and Johnson, 1988). In these cases, **a**1 or α2 alone have no effect; in cells that express both proteins, the *hsg* operator brings about strong repression of the nearby promoter. Third, mutations of α2 that selectively disrupt its ability to act with **a**1 *in vivo* selectively disrupt its ability to bind to the *hsg* operator with **a**1 in the purified system (Mak and Johnson, 1993). Fourth, the results presented in Figure 2 show an excellent correlation between the *in vivo* and *in vitro* affinities of **a**1α2 for a series of operators. Taken together, these arguments indicate that the properties of the **a**1α2–operator deduced from biochemical experiments reflect the *in vivo* situation.

Key features of the hsg operator

Purified **a**1 and α2 homeodomain proteins bind co-operatively to a 20 bp operator (the *hsg* operator). We first

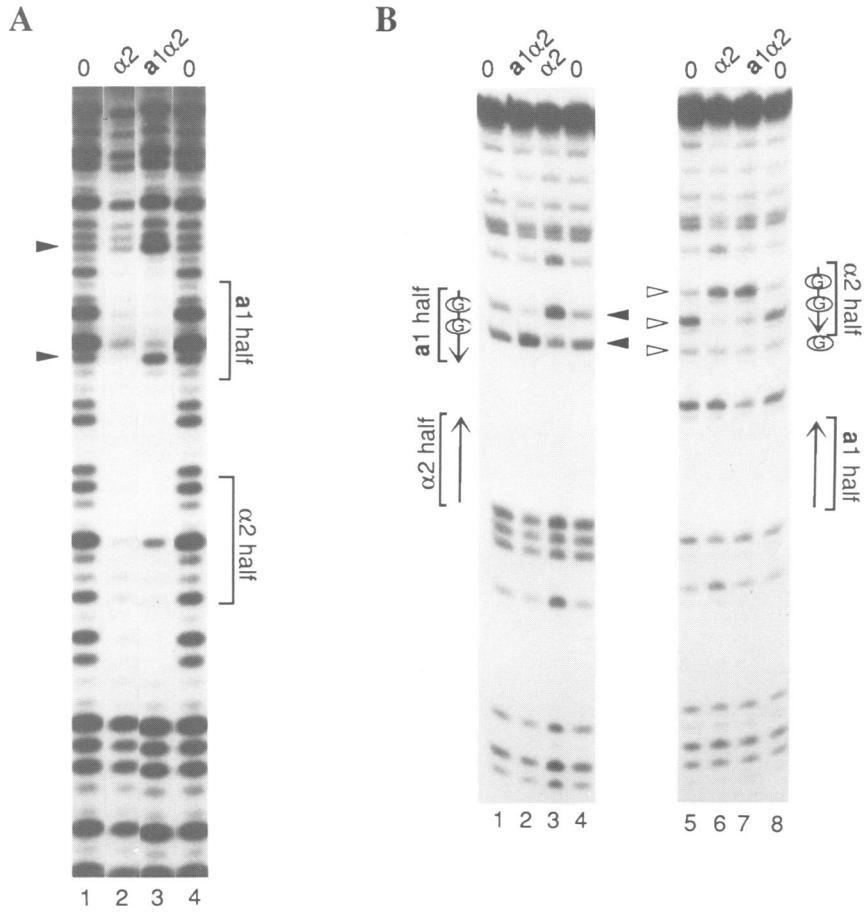

Fig. 6. Comparison of protection patterns produced by **a**1α2 with those produced by an α2 homodimer. **(A)** Protection against DNase I. The top strand of the *hsg* operator is shown with its half-sites indicated with brackets. Arrowheads point to the strongest differences between **a**1α2 and α2 alone. Protein concentrations are as follows: lanes 1 and 4, no protein; lane 2, 4×10^{-6} M α2; lane 3, 4×10^{-7} M α2 and 4×10^{-8} M **a**1. **(B)** Protection against DMS. The *hsg* operator used here has an extra guanine base (see Materials and methods) so that the positions that can be attacked by DMS are equivalent for each half-site; this modification does not alter the binding of **a**1α2 (data not shown). The half-sites are indicated with brackets for the two strands of the operator, and the relative position of Gs in each half-site is indicated. Open arrowheads indicate protections and enhancements that are the same for α2 and **a**1α2; filled arrowheads indicate patterns that differ in the two cases. Protein concentrations are as follows: lanes 1, 2, 5, 8, no protein; lanes 2 and 7, 2×10^{-8} M α2 and 6×10^{-7} M **a**1; lanes 3 and 6, 8×10^{-7} M α2.

consider the features of the *hsg* operator that are responsible for its recognition by the **a**1α2 heterodimer. The *hsg* operator consists of two half-sites which resemble the half-sites found in *asg* operators (Figure 1). Despite this similarity, the two types of operators are bound by distinct activities: the *hsg* operator by an **a**1α2 heterodimer and the *asg* operator by an α2MCM1 heterotetramer. We have found that **a**1α2 distinguishes the *hsg* operator from the *asg* operator by two criteria: the spacing of half-sites appears to be the most important difference, and the precise sequences of the half-sites serve to fine tune the strength of the interaction. For example, an *asg* operator can be transformed into an *hsg* operator (albeit a weak version) simply by reducing the spacing of its half-sites from 19 to 6 bp. Conversely, the affinity of **a**1/α2 for the *hsg* operator is significantly reduced when its half-sites are separated by an extra 13 bp, an addition which brings them into register with those of the *asg* operator (see Figure 2).

Which homeodomain is bound to which half-site?

In contrast to *asg* operators, which have two virtually identical half-sites, *hsg* operators are composed of two half-sites which differ slightly in sequence (Miller *et al.*, 1985). By cross-linking **a**1α2 to synthetic *hsg* operators that are photoreactive on only one or the other half-site, we have proven that in the **a**1α2–DNA complex one half-site of the *hsg* operator is specifically bound by **a**1 and the other by α2. The half-site bound by α2 in the *hsg* operator corresponds in sequence to the half-sites of *asg* operators. The half-site bound by **a**1 differs from the α2 half-site at the second position of the half-site (see Figure 3).

The a1 and α2 homeodomains contact DNA differently

The recent structure determinations of three homeodomain proteins (α2, Antennapedia and Engrailed), each complexed with a DNA sequence, revealed that the three proteins interact with DNA in a very similar way. We had, therefore, expected that **a**1 and α2 would interact with the two similar half-sites of the *hsg* operator in virtually the same way. Contrary to this simple expectation, our results indicate that **a**1 and α2 contact the two similar half-sites of *hsg* operators quite differently. The protection pattern produced by α2, as part of the **a**1α2 heterodimer, is virtually indistinguishable from that produced by α2 as part of an α2 homodimer. The DNA protection patterns produced by the interaction

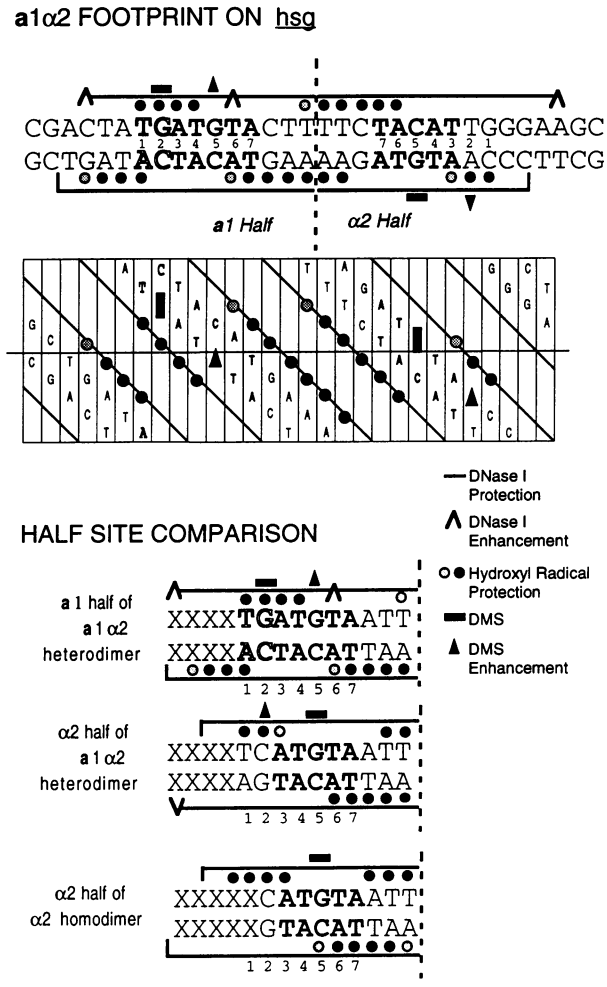

Fig. 7. Summary of all DNA protection results. (**Top half**) the *hsg* operator is shown with its axis of symmetry indicated by a dashed line and all protection patterns displayed. The operator is also shown as a split projection of B-form DNA onto which the hydroxyl radical and DMS protection patterns are mapped. (**Bottom half**) DNA protection patterns are compared for the **a**1α2 halves of the *hsg* operator and a half-site of the *asg* operator bound by an α2 homodimer. The sequences represent the consensus for each half-site with Xs representing non-conserved bases. The protections observed on the **a**1 half when **a**1α2 binds the *hsg* operator are presumed to be caused by **a**1, and those on the α2 half by α2; however, the contacts made in the center of the operator (towards the axis of symmetry of the operator) cannot be clearly ascribed to one protein or the other. The protection data for the α2 homodimer on the *asg* operator are taken from Sauer *et al.* (1988). As shown in the figure, the protection patterns of **a**1 and α2 differ in the following ways. First, **a**1 binding creates a strong and reproducible enhanced DNase I cleavage site at position #6 of the half-site. This effect is never observed as the consequence of α2 binding, either on *hsg* operators or *asg* operators. Second, **a**1 protects the guanine at position #2 of the half-site from DMS methylation and enhances the methylation rate of the guanine at position #5. α2 binding produces the reciprocal pattern. Third, **a**1 protects from hydroxyl radical attack a stretch of four sugar residues along the DNA backbone that are not protected upon α2 binding.

of **a**1 with its half-site, however, differ in a number of ways from those observed for α2 (Figure 7). The main differences are summarized in the legend to Figure 8.

Models for a1α2 configuration on the hsg operator

Wolberger *et al.* (1991) have proposed a general model for homeodomain–DNA interactions based on the known structures of the Antennapedia, Engrailed and α2 homeodomains complexed with DNA. A comparison of these structures has revealed that the positioning of the homeodomain on its binding site is very similar for the three cases, even though the DNA sequences optimally recognized by each protein are different. In particular, the protein–DNA contacts that fix the position of the recognition helix (helix 3) appropriately in the major groove of DNA are proposed to be similar for the three proteins. If this model is generally applicable, all homeodomains should produce roughly the same patterns of protection against enzymatic and chemical probes. Indeed, DNA-bound α2 and Antennapedia give roughly the same protection patterns [compare Sauer *et al.* (1988) with Affolfer (1990)]. Why then does the binding pattern of **a**1 differ from that of α2, even though both are homeodomain proteins and both are bound to similar sequences? We suggest two possible models to explain this variation.

Model 1. According to this model, the two homeodomains of the **a**1α2 heterodimer are arranged symmetrically on the *hsg* operator to match the dyad symmetry of the two half-sites (Model 1 of Figure 8). However, in order to account for the chemical and nuclease protection patterns experimentally observed for **a**1 (particularly the protection of the four sugar residues at the left-hand end of the projection diagram), the interactions between **a**1 and DNA would have to differ substantially from those observed for α2 and, by extension, those of Engrailed and Antennapedia. For example, the amino-terminal arm of the **a**1 homeodomain could be located in a different position relative to that of α2, Engrailed and Antennapedia. Alternatively, the **a**1 homeodomain structure could be similar to that of other homeodomains, but the DNA could be significantly distorted (e.g. by bending), allowing a different and more extensive set of contacts with **a**1. In this regard, we note that the DNA structure in the α2–DNA and Engrailed–DNA co-crystals does not deviate drastically from B-form. However, the binding of **a**1α2 to DNA does induce a bend (A.Desai, A.Mak and A.D.Johnson, unpublished). This model is, however, difficult to reconcile with the available knowledge of homeodomain–DNA interactions and would suppose that the **a**1–DNA interaction departs significantly from the general scheme proposed for homeodomains.

Model 2. According to this model, the **a**1 homeodomain interacts with DNA like other homeodomains, but is arranged on the DNA in tandem with α2, rather than with dyad symmetry (Model 2, Figure 8). This model goes against the common assumption that a symmetric DNA binding site is bound by symmetrically oriented DNA binding domains, but easily accounts for our data (especially the protection of the four sugar residues at the left end of the operator) in a way that is consistent with the general scheme for homeodomain–DNA interactions. According to this arrangement, the orientation of the **a**1 homeodomain relative to the half-site sequence is opposite to that of α2, i.e. α2 would read the sequence 5′-TCATGTA-3′, while **a**1 would see the sequence 5′-TACATCA-3′. Since the residues in helix 3 and the amino-terminal arm of the α2 homeodomain (responsible for the specific base contacts in the major and minor groove, respectively) differ from those of **a**1, it is plausible that very similar DNA sequences could

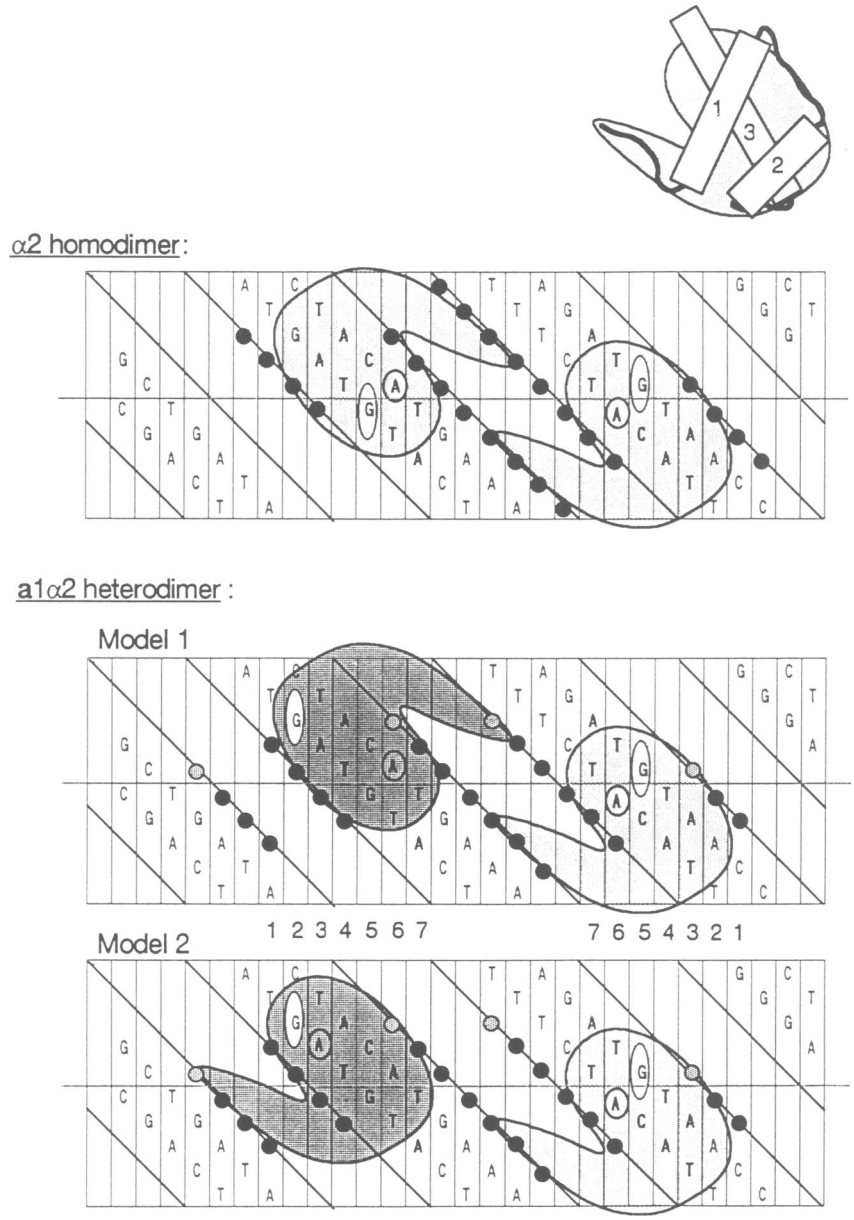

Fig. 8. Models for the interaction of **a**1α2 with the *hsg* operator. The structure of the α2 homeodomain is shown schematically in the upper right of the figure. It consists of three tightly packed α-helices and an amino-terminal arm. When bound to DNA, helix 3 lies in the major groove and the amino-terminal arm lies in the minor groove. The *hsg* operator (from the MATα1 gene promoter) is represented as a split projection of B-form DNA onto which hydroxyl radical protections (small circles) and DMS protections (circled Gs) have been mapped. The top DNA projection shows the interaction of an α2 homodimer with the operator; the middle and lower projections show two models for **a**1α2 that attempt to account for the experimentally observed protection patterns. Although the three-dimensional structure of the **a**1 homeodomain has not been determined, it is likely, based on the sequence similarity, to resemble that of α2; it is represented by darker shading. Model 1 supposes that positioning of the **a**1α2 heterodimer on DNA resembles that of the α2 homodimer, shown above it. Although attractive in theory, this model does not easily account for the four left-most DNA backbone contacts, nor does it account for the asymmetric DMS protection pattern. Model 2, invoked in order to account for these apparent discrepancies, supposes that the **a**1 homeodomain is oriented in tandem with that of α2, an arrangement that does not reflect the dyad symmetry of the operator. The general model for homeodomain–DNA interactions proposed by Wolberger *et al.* (1991) suggests that an asparagine in helix 3 (which is highly conserved among homeodomains from diverse species) contacts an A base. The A known to be contacted by this asparagine in α2 is circled, as are those proposed to be contacted by the corresponding asparagine of **a**1 in the two models. The protection data for **a**1α2 are from Figure 7; those for the α2 homodimer (top) were obtained on the related operator *asgΔ13* (C.Goutte, D.Smith and A.D.Johnson, unpublished; see Materials and methods for the sequence) and superimposed on the *hsg* operator shown here.

be recognized by the two proteins positioned in opposite orientations.

An important feature of the general scheme for homeodomain–DNA interactions proposed by Wolberger *et al.* (1991) is the interaction of an arginine in helix 3 (Asn51, which is highly conserved among homeodomains of diverse species) with an adenine. Both models of Figure 8 provide an appropriately located A (indicated by circles) for interaction with Asn51 of **a**1. The tandem model, however, accounts better for the methylation protection pattern observed for **a**1. The guanine at position #2 is protected from attack; according to the tandem model, this position would be adjacent to the Asn51–A interaction, and predicted to be covered by the homeodomain. Furthermore, the G

at position 5 is not protected; a second prediction of this model. The dyad symmetry model, in contrast, would predict the opposite pattern of methylation protection.

Although at this point we cannot definitively distinguish between these two models of **a**1α2–operator interaction, we favor Model 2, according to which the two homeodomain proteins interact to bind the *hsg* operator in tandem. Tandem arrangements of DNA-binding proteins are rare for proteins that bind as homodimers, perhaps because all protein–protein interaction surfaces would not be satisfied upon simple tandem dimer formation and higher order polymers would tend to form. However, in the case of a heterodimer, this argument does not necessarily apply and perhaps this arrangement will be common to other types of heterodimeric DNA-binding proteins.

Materials and methods

Yeast strains and plasmids

Yeast strains were provided by P.Siliciano and K.Tatchell. They are strain EG123 (*MAT***a** *trp1 leu2 ura3 his4*), the isogenic *MAT*α strain (246-1-1), and the diploid *MAT***a**/*MAT*α product of EG123 and 246-1-1 (see Siliciano and Tatchell, 1984).

Reporter plasmids carrying the *LacZ* gene under the control of the CYC1 promoter [based on pLGΔ-312S, see Guarente and Hoar (1984)] were constructed with different operator sequences inserted into the unique *SalI* site at position −184 relative to the transcription start site. The different operators were synthesized as oligonucleotide duplexes with TCGA 5′ overhangs. The sequences are as follows: *hsg*: 5′-TCGAGCTTCCCAATGTAGAAAAGTACATCATAG-3′; *hsg+13*: 5′-TCGACAATGTAGAAACCCAGATCTGAAAAGTACATCATAG-3′; *asg**: 5′-TCGACATGTAATTACCCAGATCTGAAATTTACACGC-3′; *asg*: 5′-TCGACATGTAATTACCTAATAGGGAAATTTACACGC-3′; *asgΔ13*: 5′-TCGACATGTAATTATTTACACGC-3′; 2x**a**1: 5′-TCGATGATGTAATTAATTACATCA-3′. All operator insertions were verified by sequencing.

β-Galactosidase assays were performed as previously described (Goutte and Johnson, 1988).

DNA binding assays

Electrophoretic mobility shift assays were performed using purified **a**1, α2 and truncated $\alpha 2_{128-210}$ proteins (Sauer *et al.*, 1988; C.Goutte and A.D.Johnson, 1993). Each reaction contained 5% glycerol, 10 mM Tris–HCl (pH 8.0), 0.1 mM EDTA, 100 mM $NaCl_2$, 5 mM $MgCl_2$, 0.1% NP40, 10 mg/ml bovine serum albumin (BSA; Sigma, Fraction V) and 3 μg/ml *Hae*III-cut *E.coli* DNA as non-specific DNA. Reactions were incubated for 30 min at room temperature and electrophoresed through 5% native TBE [80 mM Tris–borate (pH 8.0), 2 mM EDTA] polyacrylamide gels at 200 V for 1–2 h. Gels were dried and visualized by autoradiography. The target DNA was end labeled with ^{32}P using Klenow fragment and used at a concentration of $5-10 \times 10^{-11}$ M in each reaction. The labeled DNA fragments are either synthetic oligonucleotide duplexes of 23–36 bp (see above for sequences) or restriction fragments isolated from different pUC18 constructs which carry the different synthetic oligonucleotide sequences each cloned into the *SalI* site of the polylinker region.

All DNA protection reactions were performed in a buffer containing 10 mM Tris–HCl (pH 7.0), 0.1 mM EDTA, 50 mM $NaCl_2$, 5 mM $MgCl_2$, 2 mM $CaCl_2$, 2.5 mg/ml calf thymus DNA and 50 mg/ml BSA. **a**1 and α2 proteins were added along with end-labeled operator DNA (~0.5 nM) and incubated at room temperature for 45 min before chemical or enzymatic treatment. DNase I cleavage reactions were performed at 20°C for 10 min with 0.01 μg/ml pancreatic DNase I (Worthington Co.). Reactions were stopped and precipitated with 1.6 M NH_4Ac. Hydroxyl radical footprinting reactions were performed using the procedures of Tullius and Dombroski (1986). DMS reactions were performed according to Maxam and Gilbert (1980). All reaction products were electrophoresed through 8 or 10% denaturing TBE polyacrylamide gels. DMS reactions performed on the labeled operator in the absence of protein were used as sequence markers. ^{32}P-labeled fragments were obtained by filling in (using Klenow fragment) the *Hin*dIII end of an 80 bp restriction fragment from pCG25 or pCG60, derivatives of pUC18 into which was inserted the *hsg* operator sequence (see above) in forward and backward orientation, respectively. The operator used in the DMS reactions was different at one position: the adenine at position #2 of the α2 half-site in the *hsg* operator was replaced by a guanine so that methylations could be compared directly on both half-sites (this change does not alter the DMS protection patterns of any of the other guanines; data not shown). The operator used in Figure 6A is the 2x**a**1 operator (the same results are obtained on the *hsg* operator; data not shown).

Protein–DNA cross-linking reactions were performed as previously described (C.Goutte and A.D.Johnson, 1993), except for the use of BrdU-substituted operators (synthesized by UCSF Biomolecular Resource Center) as described in Figure 4B, and the reactions were only exposed to UV light for 40 min.

Acknowledgements

We are grateful to Drew Vershon for supplying purified α2 protein and Dana Smith for sharing unpublished results. We thank Cynthia Wolberger, Jim Jaynes, Ed Giniger, Peter Sorger, Peter Pryciak, Rick Myers and Ira Herskowitz for helpful comments on the manuscript, and Mary Jo Kelley for invaluable assistance in preparing the manuscript. This work was supported by an institutional training grant from the National Cancer Institute and the National Institutes of Health (#5 T32 CA 09270), and by grants from the NIH (GM37049) and the Pew Scholars Program.

References

Affolfer,M., Percival-Smith,A., Muller,M., Leupin,W. and Gehring,W.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4093–4097.

Ammerer,G. (1990) *Genes Dev.*, 4, 299–312.

Chodosh,L.A. (1988) In Ausubel,F.M., Brent, R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, pp. 12.5.1–12.5.8.

Covitz,P.A., Herskowitz,I. and Mitchell,A.P. (1991) *Genes Dev.*, **5**, 1982–1989.

Dolan,J.W. and Fields,S. (1991) *Biochim. Biophys. Acta*, **1088**, 155–169.

Dranginis,A.M. (1990) *Nature*, **347**, 682–685.

Gehring,W.J. (1987) *Science*, **236**, 1245.

Goutte,C. and Johnson,A.D. (1988) *Cell*, **52**, 875–882.

Goutte,C. and Johnson,A.D. (1993) *J. Mol. Biol.*, **233**, 359–371.

Guarente,L. and Hoar,E. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 7860–7864.

Herskowitz,I. (1989) *Nature*, **342**, 749–757.

Johnson,A.D. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcription*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, in press.

Johnson,A.D. and Herskowitz,I. (1985) *Cell*, **42**, 237–247.

Keleher,C.A., Goutte,C. and Johnson,A.D. (1988) *Cell*, **53**, 927–936.

Keleher,C.A., Redd,M.J. Schultz,J., Carlson,M. and Johnson,A.D. (1992) *Cell*, **68**, 709–720.

Kissinger,C.R., Liu,B., Martin-Blonco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol*, **65**, 499–560.

Miller,A.M., MacKay,V.L. and Nasmyth,K.A. (1985) *Nature*, **314**, 598–603.

Mukai,Y., Harashima,S. and Oshima,Y. (1991) *Mol. Cell. Biol.*, **11**, 3773–3779.

Norman,C., Runswick,M., Pollock,R. and Treisman,R. (1988) *Cell*, **55**, 989–1003.

Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.

Philips,C.L., Vershon,A.K., Johnson,A.D. and Dahlquist,F.W. (1991) *Genes Dev.*, **5**, 764–772.

Sauer,R.T., Smith,D.L. and Johnson,A.D. (1988) *Genes Dev.*, **2**, 807–816.

Scott,M.P., Tamkun,J.W. and Gehring,W.H. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.

Siliciano,P.G. and Tatchell,K. (1984) *Cell*, **37**, 969–978.

Siliciano,P.G. and Tatchell,K. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 2320–2334.

Smith,D.L. and Johnson,A.D. (1992) *Cell*, **68**, 133–142.

Sprague,G.F.,Jr (1990) *Adv. Genet.*, **27**, 33–62.

Tullius,T.D. and Dombroski,B.A. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 5469–5473.

Vershon,A.K. and Johnson,A.D. (1993) *Cell*, **72**, 105–112.

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

Received on August 31, 1993; revised on December 20, 1993