Cognate Site Identifier analysis reveals novel binding properties of the Sex Inducer homeodomain proteins of Cryptococcus neoformans

Brynne C. Stanton,1 Steven S. Giles,1 Emilia K. Kruzel,1 Christopher L. Warren,2 Aseem Z. Ansari2,3 and Christina M. Hull1,4*
Departments of 1Biomolecular Chemistry and 4Medical Microbiology and Immunology, School of Medicine and Public Health, 2Department of Biochemistry and 3the Genome Center, University of Wisconsin, Madison, Madison, WI 53706, USA.

Summary
Homeodomain proteins function in fungi to specify cell types and control sexual development. In the meningoencephalitis-causing fungal pathogen Cryptococcus neoformans, sexual development leads to the production of spores (suspected infectious particles). Sexual development is controlled by the homeodomain transcription factors Sxi1α and Sxi2α, but the mechanism by which they act is unknown. To understand how the Sxi proteins regulate development, we characterized their binding properties in vitro, showing that Sxi2α does not require a partner to bind DNA with high affinity. We then utilized a novel approach, Cognate Site Identifier (CSI) arrays, to define a comprehensive DNA-binding profile for Sxi2α, revealing a consensus sequence distinct from those of other fungal homeodomain proteins. Finally, we show that the homeodomains of both Sxi proteins are required for sexual development, a departure from related fungi. Our findings support a model in which Sxi1α and Sxi2α control sexual development in a homeodomain-dependent manner by binding to DNA sequences that differ from those defined in previously established fungal paradigms.

Introduction
During development in eukaryotes, many processes, including cell growth, differentiation and morphogenesis, must be co-ordinated. Specific transcription factors are responsible for directing many of these events, and are often referred to as ‘master control genes’ (Gehring, 1998). As a result of the actions of these master transcriptional regulators, development is efficiently managed in a co-ordinated fashion.

Master control genes often encode sequence-specific DNA-binding proteins containing a homeodomain (Gehring, 1987; Levine and Hoey, 1988; Scott et al., 1989). The homeodomain is an ~60-amino-acid DNA-binding motif consisting of a three-helix bundle that has been characterized both structurally and functionally in eukaryotes ranging from yeast to humans (McGinnis et al., 1984; Qian et al., 1989; Papageorgiou, 2007). Studies in many organisms, including Drosophila melanogaster and Saccharomyces cerevisiae, have revealed that homeodomain proteins often interact with other binding partners to adopt novel DNA-binding specificities (Goutte and Johnson, 1988; Keleher et al., 1988; Mann and Chan, 1996).

Currently, two paradigms exist in fungi to explain how homeodomain proteins interact with one another to control development. The first paradigm was developed from findings in S. cerevisiae in which the a1 and α2 proteins heterodimerize and control mating type (Strathern et al., 1981; Goutte and Johnson, 1988). In this system, the homeodomain region of a1 mediates sequence-specific recognition, and the analogous region of α2 is less important for heterodimer function (Vershon et al., 1995; Jin et al., 1999). A similar situation is found in the mushroom Coprinopsis cinerea where the homeodomain of only the HD2 protein is required for heterodimeric function in vivo, while that of its partner, HD1, is completely dispensable (Kues et al., 1994; Asante-Owusu et al., 1996). The second paradigm is defined by the corn smut Ustilago maydis, in which the homeodomain regions of both of the bE and bW proteins are required for sexual development (Schlesinger et al., 1997). Outside of these systems, little else is known about the mechanism of homeodomain protein function in fungi. This is also true of the human fungal pathogen Cryptococcus neoformans, which causes meningoencephalitis in immunocompromised individuals (Casadevall and Perfect, 1998). Although sexual development of C. neoformans is not directly associated with meningoencephalitis, this process results in the production of spores,
Results

Sxi1α and Sxi2α bind to DNA with high specificity

Previous work had shown that the Sxi proteins contain homeodomains, and that fragments of these proteins are capable of binding DNA in vitro (Hull et al., 2002; 2005; Ekena et al., 2008). While it was known that these proteins were capable of binding to a 177 bp fragment of DNA found upstream of the C. neoformans CLP1 gene (Ekena et al., 2008), the precise sequences bound by Sxi1α and Sxi2α within this fragment were unknown.

To identify the specific site(s) bound by Sxi1α and Sxi2α, the original fragment of the CLP1 upstream region (to which binding was detected) was divided into overlapping fragments (Ekena et al., 2008). Successively smaller fragments of the bound upstream region were used to narrow down the minimal site required for binding by the Sxi proteins in EMSAs (data not shown). Although we attempted to produce full-length Sxi proteins in several expression systems, including Escherichia coli, Pichia pastoris, and cell-free extracts, active protein was not recovered. As a result, we carried out binding studies with fragments of each protein purified from E. coli. An amino-terminal fragment of Sxi1α containing amino acids 1–205, including the homeodomain (Sxi1α-N), and a 59-amino-acid fragment of Sxi2α (residues 143–202) comprised of only the homeodomain (Sxi2α-HD), were purified using histidine affinity tags (Fig. 1A).

Both proteins exhibited binding to a 19 bp fragment of DNA from the CLP1 upstream region (hereafter referred to as clp19). Increasing concentrations of Sxi1α-N produced a species that migrated slower than the probe alone (Fig. 1B, lanes 2–4), representing the Sxi1α-N/clp19 complex. The addition of an antibody that recognizes the 6x-histidine tag of Sxi1α-N led to a supershifting of the Sxi1α-N/clp19 complex, confirming the presence of Sxi1α-N (Fig. 1B, lane 5). A similar pattern of binding by the Sxi2α-HD protein was also observed (Fig. 1B, lanes 6–9). Surprisingly, a shorter fragment of Sxi1α containing only the homeodomain showed no detectable DNA binding under the conditions tested (data not shown).

To assess the specificity of Sxi1α-N binding, we carried out EMSAs using two different DNA sequences as probes. Sxi1α-N bound to a 177 bp fragment of the CLP1 upstream region (Fig. 1C, lanes 1 and 2), but did not exhibit any binding to a control DNA fragment in which the clp19 sequence had been deleted (Fig. 1C, lanes 3 and 4). EMSAs with clp177 and clp177Δ19 as probes were also carried out with Sxi2α-HD and revealed a similar pattern of binding (data not shown). In addition, competition assays with Sxi2α-HD were carried out using both specific (clp19) and non-specific (random 19-mer), unlabelled competitor double-stranded oligonucleotides (Table S1). Binding of Sxi2α-HD to labelled DNA was decreased in the presence of increasing concentrations of specific competitor (Fig. 1D, lanes 3–6). However, non-specific competitor had no effect on the amount of Sxi2α-HD binding to the labelled DNA (Fig. 1D, lanes 8–11), indicating that under these solution conditions, binding was specific. Although the binding of both Sxi proteins to clp19 is specific, only Sxi2α-HD exhibited strong relative binding, making it amenable to further characterization in vitro.

To assess the affinity of the specific interaction between purified Sxi2α-HD and the CLP1 site, an apparent dissociation constant (Kd) for this binding was determined. In this assay, the concentration of DNA probe was held constant at 1 nM, and the Sxi2α-HD protein was varied from...
0 to 200 nM to identify the concentration of protein at which half of the DNA probe was bound. The apparent \( K_d \) was determined to be between 25 and 50 nM (Fig. 1E). This apparent \( K_d \) for Sxi2a-HD is consistent with values that have been determined for other transcription factors in vitro, which are generally in the nanomolar range (Affolter et al., 1990).

The Sxi2a homeodomain binds to clp19 in a one-hybrid assay

While it was clear that the Sxi2a-HD protein bound to clp19 with high affinity in vitro, it was unknown whether this site was recognized by the Sxi2a-HD protein in vivo. To determine if Sxi2a-HD was capable of binding clp19 in...
The Sxi2 expression is observed in the presence of Sxi2. An asterisk indicates sample. Average values and standard error are reported and an CPRG assays were performed in triplicate for an n = 6 for each sample. Average values and standard error are reported and an asterisk indicates P < 0.0001. An increase of 40% and 36% in lacZ expression is observed in the presence of Sxi2-HD alone and in combination with Sxi1α-FL respectively.

**Fig. 2.** The Sxi2a-HD binds clp19 in vivo.

A. Schematic representation of the reporter plasmid pLGΔ-178 + clp19. The CYC1 promoter drives lacZ expression.

B. The Sxi2a-HD activates lacZ expression in the clp19 reporter. CPRG assays were performed in triplicate for an n = 6 for each sample. Average values and standard error are reported and an asterisk indicates P < 0.0001. An increase of 40% and 36% in lacZ expression is observed in the presence of Sxi2a-HD alone and in combination with Sxi1α-FL respectively.

In an in vivo context, one-hybrid assays were carried out in *S. cerevisiae*. In this assay, two lacZ reporter plasmids were used: one containing the CYC1 promoter driving lacZ expression (pLGΔ-178), and another where the clp19 sequence had been cloned into the promoter region of this vector (Fig. 2A). Each reporter plasmid was individually transformed and assayed for lacZ expression in the presence and absence of Sxi2a-HD.

Cells transformed with a control plasmid lacking clp19 did not show lacZ induction over background levels in the presence of the Sxi proteins (data not shown). However, for those cells that contained the reporter plasmid harbouring clp19, an increase of 40% in lacZ expression was detected in the presence of Sxi2a-HD (Fig. 2B). Addition of the SXI1α-FL expression vector to these cells did not result in an increase in lacZ expression. These results indicate that the Sxi2a-HD protein is capable of binding to the clp19 site in vivo.

The Sxi2a homeodomain binds a unique DNA sequence

It was clear from binding experiments that the Sxi2a homeodomain exhibits high-affinity sequence-specific binding in vitro and binds to clp19 in an in vivo context. We next sought to define the spectrum of sequences to which Sxi2a-HD binds. Although Sxi2a-HD bound specifically to a 19 bp sequence, the core sequence being recognized was unclear. To generate a comprehensive DNA-binding profile for the Sxi2a protein, CSI analysis was carried out (Warren et al., 2006). For the purposes of this study, our efforts were focused on only the Sxi2a-HD protein, because the Sxi1α-N protein did not exhibit high-affinity DNA binding.

On the CSI arrays, every permutation of a double-stranded DNA (10 bp in size) was represented on a microfabricated array. Sequences (1 049 088 total) were represented by 39-mer oligonucleotides, which were induced to form hairpins that mimic B-form DNA (Fig. 3A) (Warren et al., 2006). Hairpin arrays were probed with fluorescently labelled Sxi2a-HD (see Experimental procedures for labelling details) under conditions conducive to DNA binding. The resulting fluorescence intensity values were translated into measurements of relative binding values (Z-scores) of the protein for each specific 10-mer duplex DNA. After statistical analysis, normalized Z-scores ranged from 19.7 to −4.1. From these values, a profile of relative binding affinity of Sxi2a-HD for all of the represented sequences was generated.

At a concentration of 50 nM, the Sxi2a-HD produced a clear pattern of sequence-specific binding. The resulting intensity values were subjected to statistical analyses, and sequences with significant, normalized intensity values were evaluated using the motif discovery algorithm MEME. The 500 sequences with the highest intensity values were input into MEME to produce motifs 1, 2 and 3 (Fig. 3B), each containing an invariant 5′-CAATC-3′ (5′-GATTG-3′) core. Each of these motifs was represented by 285, 86 and 70 individual sequences respectively. Evaluating sequences having Z-score values greater than the 90th percentile using MEME reliably generated the 5′-CAATC-3′ motif in all five possible positions of the 10-mer (4591 sequences out of the top 51 736 contain this motif). In addition, 8-mer duplex DNA arrays were also probed with 50 nM Sxi2a-HD, and the MEME algorithm produced an identical 5′-CAATC-3′ core sequence (data not shown).

To validate the data generated by CSI and MEME analysis, EMSAs were carried out using hairpin DNAs representing several Z-score sequences (Fig. 3C). Hairpins representative of Z-scores of > 14, 7–8, 4–5, 0 and < 0 were assayed for binding by 50 nM Sxi2a-HD. A general trend of binding was observed in which hairpins with higher Z-scores showed a greater portion of the probe shifted by Sxi2a-HD. These data indicate a positive correlation between binding to the hairpin microarray and binding to hairpin oligonucleotides in solution, showing that our solution studies mimic the data trend rendered by the CSI arrays.

To further refine the sequences to which Sxi2a-HD binds in clp19, we identified all of the 10-mers represented in clp19 in the CSI 10-mer data set and evaluated...
their Z-scores (Fig. 4A). We then carried out EMSAs with Sxi2a-HD, and various 10-mers within clp19, using a surrounding sequence context that mimics the sequence context of the CSI experiments (Fig. 4A). Based on the Z-score analysis, we predicted that Sxi2a-HD would bind well to the 10-mer containing the sequence 5′-TTTCATTCAA-3′. This was, in fact, the case. DNA probes with the highest Z-scores were bound more efficiently than those with lower scores, and the probe showing the most robust binding contained the sequence 5′-CATTC-3′, a sequence that is very similar to the CSI-identified core (5′-CAATC-3′, Fig. 4A).

To ensure that the Sxi2a-HD bound specifically to the 5′-CATTC-3′ core, the specificity of binding by Sxi2a-HD to the clp19 10-mer D was examined (Fig. 4A, panel D). Using a competition experiment similar to the one shown in Fig. 1E, we observed that binding to fragment D was specific (data not shown), indicating that the 5′-CATTC-3′ sequence of clp19 is the most likely target of binding by Sxi2a-HD by EMSA. Odd-numbered lanes contain only the hairpin DNA, while even-numbered lanes include 50 nM Sxi2a-HD. The corresponding Z-score bins are labelled above each individual set of shifts.
Sxi2α-HD binds to both 10-mer D (lane 2) and motif 1 (lane 4) with similar affinities (Fig. 4B). Therefore, the CSI-identified motif and the in vivo identified 10-mer D are similar not only in sequence content, but also in binding affinity by Sxi2α-HD.

Bioinformatic analysis of target sequences does not reveal a Sxi1α binding site

Previous studies using two-hybrid analysis indicate that the Sxi proteins interact with one another, suggesting that, like a1 and α2 of S. cerevisiae, Sxi1α and Sxi2α function as a heterodimer (Goutte and Johnson, 1988; Hull et al., 2005). Having established that Sxi2α binds to the sequence 5′-CAATC-3′, we sought to determine a binding sequence for Sxi1α. However, due to the limited binding activity exhibited by Sxi1α-N in vitro, CSI array analysis could not be carried out. While chromatin immunoprecipitation (ChIP) assays with the Sxi proteins would be an ideal approach to identify binding sequences, because of technical challenges our attempts with these assays have not been successful. Using a wide variety of approaches and methods, it has thus far been impossible to generate extracts from crosses amenable to either immunoprecipitation or Western blot analyses. Instead, we undertook a bioinformatic approach utilizing microarray expression data, in conjunction with the Sxi2α binding site, to analyse sequences near the CSI consensus for conserved sequences to which Sxi1α might bind.

Genes regulated by the Sxi proteins were determined by comparing the expression profiles of crosses between sxiΔ mutants to crosses between wild-type cells using whole-genome expression microarrays. Wild-type (α × α) and mutant (sxi2α × sxi1α) crosses were cultured on V8 medium for 24 h prior to the isolation of RNA. This RNA was used to generate fluorescently labelled cDNA for each cross. The resulting pools were mixed and hybridized to an oligo-based expression array representing each predicted C. neoformans gene. The resulting data were analysed, and we found that of the genes...
meeting a significance threshold of \( P < 0.05 \), 245 showed differential expression greater than twofold.

Based on the positioning of homeodomain regulatory sites in other fungi (Johnson and Herskowitz, 1985; Goutte and Johnson, 1988; Romeis et al., 2000; Brachmann et al., 2001; Galgoczy et al., 2004), we evaluated regions up to 500 bp upstream of each regulated gene for the presence of the Sxi2a motif (5′-CAATC-3′). One hundred and two out of the 245 genes contained the motif within 500 bp of the predicted gene start site; however, because of the small size of the core CSI motif (5 bp) and the high frequency with which it occurs in random sequence space, we could not identify a statistically significant pattern of enrichment in the predicted promoter regions of regulated genes or conservation of nearby sequences (data not shown).

To narrow our pool of genes to those most likely to be regulated directly by the Sxi proteins, we carried out EMSAs with the Sxi2a-HD protein on a subset of highly regulated genes from the microarray. In this analysis, 100 bp probes were made from sequences from 27 regulated genes containing the motif, five regulated genes without the motif and five unregulated genes (Tables S2 and S3). High-affinity binding by Sxi2a-HD was observed for four out of the 37 probes examined; each of the four bound probes contained the 5′-CAATC-3′ motif, and little or no binding by the Sxi2a homeodomain to the remaining 33 probes was observed (Fig. 5). We analysed sequences surrounding the 5′-CAATC-3′ motif in the four bound promoters (100 bp upstream and downstream) for a conserved sequence to which Sxi1α might bind. However, no clear conserved motifs were detected (data not shown). There are many possible explanations for the absence of an identifiable Sxi1α binding site, including the possibilities that (i) the site bound by Sxi2a does not represent a heterodimer half site, but instead represents a binding sequence to which Sxi2a binds in the absence of Sxi1α in vivo, (ii) the set of sequences we examined are not representative of the sequences bound by the Sxi proteins in vivo, (iii) the sequences recognized by Sxi1α are not as well conserved as those that are recognized by Sxi2a, or (iv) regulation by the Sxi proteins does not require binding to DNA by Sxi1α, akin to the situations with HD1 and HD2 in C. cinerea and a1 and a2 in S. cerevisiae, where both proteins are required for sexual development, but direct binding via the homeodomain regions of both proteins is not (Kues et al., 1994; Vershon et al., 1995; Asante-Owusu et al., 1996; Jin et al., 1999). To test this last possibility and determine whether or not the homeodomains of both Sxi proteins are required during sexual development, we examined their roles in vivo.

**Conserved homeodomain DNA-interacting residues are required for sexual development**

Functional and structural studies of many homeodomain proteins have identified several residues that are generally required for binding DNA (Pabo and Sauer, 1984; Gehring et al., 1994). Many of these DNA-interacting residues, located primarily in the third helix, are conserved across diverse homeodomain proteins, including Sxi1α and Sxi2a (Fig. 6A). To examine the roles of the homeodomains of Sxi1α and Sxi2a during sexual development, three highly conserved residues, N51, R53 and R55, were individually mutated to alanine in the context of the full-length proteins. The mutant and wild-type alleles of Sxi1α and Sxi2a were transformed into the respective deletion strains (sxi1αΔ or sxi2aΔ) and tested for the ability to undergo sexual development (Figs 6B and 5C respectively). If DNA binding is required during the regulation of sexual development, then crosses with mutant homeodomain proteins would be defective in this process, because they presumably could not exert transcriptional regulation.

Crossing an a strain expressing wild-type Sxi1α by an a strain results in robust filament formation, indicating sexual development (Fig. 6B, panel 1), whereas crosses with a sxi1αΔ strain, having no source of Sxi1α protein, show no filamentation (panel 2). When the strains containing the mutant sxi1α alleles were tested (panels 3–5), none was capable of restoring sexual development in a cross. The same pattern of sexual development was observed for the Sxi2a constructs; only limited, aberrant filamentation was detected in the absence of Sxi2a (Fig. 6C, panel 2) and for the mutant sxi2a alleles (panels 3–5). Therefore, none of the Sxi1α or Sxi2a third-helix mutants was capable of promoting sexual development in a cross. These results suggest that the DNA-binding ability of the homeodomains of both Sxi1α and Sxi2a is required during sexual development.
Conserved homeodomain DNA-interacting residues are required for DNA binding by Sxi2-HD

To test whether the defects in sexual development observed with substitutions in residues 51, 53 and 55 of the Sxi1a and Sxi2a homeodomains resulted from defects in DNA binding, we examined the ability of the mutant proteins to bind DNA in vitro. Mutant proteins were expressed recombinantly and purified from E. coli and then assayed for the ability to bind to clp19. Because only Sxi2a-HD (and not Sxi1a-N) exhibits high-affinity DNA binding in vitro, we focused on evaluating the binding properties of only Sxi2a-HD in this experiment.

As expected, the wild-type Sxi2a homeodomain exhibits high-affinity binding to clp19 in vitro (Fig. 7A, lanes 2–4), whereas binding by all of the mutant proteins was poor. Two of the mutants (N51A and R53A) showed no detectable DNA binding (Fig. 7A, lanes 5–10), and the third mutant (R55A) did not bind to clp19 as well as wild-type Sxi2a-HD (Fig. 7A, lanes 11–13). While the...
mutations in the DNA binding region were likely to disrupt binding due to alterations of specific DNA-contacting residues, the possibility that they disrupted protein folding could not be ruled out.

To confirm that the purified DNA-binding mutants were folded properly, circular dichroism (CD) was carried out. Mutant and wild-type fragments of the Sxi2a homeodomain were expressed recombinantly and purified from E. coli, and all preparations were subjected to CD over a range of concentrations (0.3–0.6 mM). The spectra produced by wild-type and mutant proteins clearly represented a characteristic α-helical spectrum (Fig. 7B), as indicated by the high molar ellipticity at 191 nm and low molar ellipticity at 208 and 222 nm. Based on these findings, it appears that the mutations do not significantly alter folding of the homeodomain in vitro. We concluded therefore that the mutations examined interfere with DNA binding, indicating that the mutated residues do indeed facilitate DNA binding by Sxi2a-HD and that DNA binding is required for sexual development.

**Discussion**

In this study, we discovered that the developmental regulators Sxi1α and Sxi2a challenge previously established paradigms of fungal homeodomain function. First, we show that the Sxi2a-HD protein alone achieves high-affinity binding in the absence of a partner protein. Previous studies show that homeodomain proteins achieve sequence specific binding only when in complex with a partner protein (Goutte and Johnson, 1988; Stark et al., 1999). Second, we identified a binding site for the Sxi2a-HD protein that is different from the sites bound by fungal homeodomain proteins in other systems that is bound by the Sxi2a-HD both in vitro and in vivo. Lastly, we show that the specific DNA contact residues in the homeo-

**Fig. 7.** Sxi2a-HD mutants exhibit little or no DNA binding but are generally folded in solution.

A. Lanes 1–13 contain 75 nM clp19 probe. Lanes 2–4, 5–7, 8–10 and 11–13 contain 5, 10 or 20 nM purified Sxi2a-HD wild type, N51A, R53A and R55A respectively.

B. The CD spectra of purified Sxi2a-HD wild type (black continuous line), N51A (dotted line), R53A (long dashed line) and R55A (short dashed line) are plotted. The x-axis refers to the wavelength at which data were collected (in nm), and the y-axis refers to the molar ellipticity (in degree cm$^2$ decimol$^{-1}$).
domains of both of the Sxi proteins are required during sexual development, a finding that differentiates their mechanism of developmental control from that of \textit{S. cerevisiae} and \textit{C. cinerea}.

\textbf{Homeodomain control of development paradigms in fungi}

Homeodomain proteins in fungi are known to control sexual development and have been studied in detail in \textit{S. cerevisiae} and a few ascomycete relatives, as well as in the more distantly related basidiomycetes \textit{U. maydis} and \textit{C. cinerea}. From these studies, two paradigms have emerged that detail how homeodomain proteins control development. In the first paradigm, modelled after \textit{S. cerevisiae} and \textit{C. cinerea}, two homeodomain proteins together regulate sexual development, but sequence-specific DNA recognition is mediated largely by only one of these proteins; the other homeodomain is less important for function and is, in some cases, completely dispensable (Kues et al., 1994; Vershon et al., 1995; Asante-Owuwu et al., 1996; Jin et al., 1999). In the second paradigm, established in \textit{U. maydis}, the homeodomains and specifically the DNA contact residues of both proteins are necessary for proper developmental control (Schlesinger et al., 1997). While our data reveal that, as in \textit{U. maydis}, both of the homeodomains of the Sxi proteins are required to promote sexual development, other features of this binding are distinct.

Because homeodomain-mediated developmental control in fungi requires two compatible homeodomain proteins heterodimerizing to bind DNA and effect changes in development (Kues and Casselton, 1992), the sites bound by these proteins are of vital importance. In the case of \textit{U. maydis}, a be–w complex binds a \textit{b}-binding sequence (bbs) that contains the core sequences bound by the \textit{S. cerevisiae} \textit{a}1 and \textit{a}2 proteins, 5′-ACA-3′ and 5′-GATG-3′, respectively, indicating substantial sequence conservation across phyla (Romeis et al., 2000). Although we had predicted that Sxi1α–Sxi2α would likely bind a sequence similar to the \textit{S. cerevisiae} \textit{a}1–\textit{a}2 site and \textit{U. maydis} bbs (5′-GATG\textit{N}3ACA-3′), we found instead that the sequence bound by Sxi2α (5′-CAATC-3′) is distinct from the binding sites identified previously for fungal homeodomain proteins. Because this sequence is different from the 5′-ACA-3′ core bound by \textit{a}1 (a homologue of Sxi2α), our findings suggest two possibilities: (i) the Sxi1α–Sxi2α complex recognizes a core sequence that is distinct from its fungal counterparts, or (ii) the Sxi1α–Sxi2α complex binds a ‘traditional’ fungal homeodomain binding site, while Sxi2α alone has evolved to have binding properties disparate from those of other fungal proteins that allow it to function in the absence of a binding partner. In either case, we have discovered a less conserved and more diverse binding pattern for fungal homeodomains than previously realized.

\textbf{\textit{C. neoformans} homeodomain binding}

Although the CSI-identified binding sequence for Sxi2α does not resemble sites bound by fungal homeodomain proteins, the same site (5′-CAATC-3′) is recognized by the \textit{Homo sapiens} pre-B-cell leukaemia homeobox 1 (Pbx1) protein, and the \textit{Coturnix coturnix japonica} (Japanese quail) homeodomain protein Hox-A7 (Zhu et al., 2005). The Hox-A7 protein has been identified in humans, mice and other metazoans, and is known to be involved in differentiation and development (Banerjee-Basu et al., 2003; Papageorgiou, 2007). Interestingly, the homeodomain of the Hox-A7 protein shares a great deal of sequence identity (85%) with that of the \textit{D. melanogaster} protein Antennapedia (Antp). The binding site of Antp (5′-TAAT-3′) also shares some features with the Hox-A7 and Sxi2α binding sites (5′-AAT-3′), suggesting a similar pattern of recognition among the Antp, Hox-A7 and Sxi2α proteins (Muller et al., 1988). In fact, many of the predicted DNA contact residues for these three proteins are identical (Q50, N51, Q44, W48, T13) and could explain the similar sequence recognition pattern shared among them (Fraenkel and Pabo, 1998). Other homologues of these proteins, such as \textit{a}1, share fewer identical DNA contact residues with the Antp protein; such discrepancies could account for the differing sequence specificities exhibited by these proteins.

The common binding motif among Sxi2α, Antp and Hox-A7 indicates that Sxi2α may have a binding profile that resembles its metazoan counterparts more so than those found in fungi. This idea is exemplified by the high-affinity, sequence-specific DNA binding exhibited by Sxi2α-HD alone. This is in sharp contrast to \textit{S. cerevisiae} where the \textit{a}1 protein becomes a high-affinity DNA-binding protein only in the presence of the tail region of \textit{a}2 (Goutte and Johnson, 1993). The behaviour of Sxi2α is more aligned with other non-fungal homeodomain proteins, such as the Antp. Antp binds DNA on its own with a \textit{K}_d of \textasciitilde10^{-9} \text{M} (Affolter et al., 1990), and the \textit{K}_d for Sxi2α-HD is in the same range. This similarity contributes to the many parallels that exist between Sxi2α and metazoan homeodomain proteins. One implication of strong Sxi2α binding is that, like Antp, Sxi2α could bind DNA on its own \textit{in vivo} and regulate gene expression in the absence of a partner protein. This possibility leads to myriad regulation models, all distinct from the paradigms established in fungi.

Because we were limited in our studies to using only the homeodomain portion of Sxi2α, we cannot rule out that other portions of the full-length protein will modulate DNA binding. However, we can assert that several previous
studies with homeodomain proteins have found full-length proteins to bind the same sequences as fragments that include the DNA-binding domain alone. For example, in the case of the α1 and α2 proteins, minimal fragments including the homeodomain only (and a 21-amino-acid tail of α2) bind an identical sequence as compared with the full-length proteins, but do so with a 10-fold reduction in affinity (Phillips et al., 1994; Li et al., 1995).

Mapping Sxi protein transcriptional control

Fully understanding how the Sxi proteins control gene regulation in *C. neoformans* requires identification of not only the sequences to which the proteins bind *in vitro*, but also the direct targets of these proteins (and the sites to which they bind *in vivo*). ChiPs to identify Sxi binding sites *in vivo* would certainly facilitate this mapping; however, ChiP assays using extracts from *C. neoformans* crosses are currently not possible due to technical difficulties. Unexpected limitations of the system have also thwarted efforts to assess the role of the clp19 site *in vivo*. For reasons that remain unclear, constructs in which the CLP1 upstream region was used to drive reporter gene expression showed highly variable levels of expression, independent of mutations in the clp19 sequence. Parallel experiments to assess CLP1 expression at its genomic locus gave similar results; in each case, any molecular manipulation of the CLP1 gene resulted in considerable variation in transcript levels among individual transformants, making it impossible to assess the effects of any specific sequence alterations. While another approach, *in vivo* footprinting, would provide evidence that the Sxi proteins bind this site *in vivo*, this assay has not been used successfully in *C. neoformans* under any condition (Zhang et al., 1999). The one-hybrid data demonstrate that the consensus sequence functions in an *in vivo* context, but it remains to be determined whether the sites identified here are bound by the Sxi proteins in *C. neoformans* to regulate sexual development.

Overall, identification and characterization of a Sxi protein binding site *in vitro* provides evidence for unique interactions of Sxi2a with its binding sequence. Our CSI array results demonstrate that this powerful *in vitro* approach is capable of identifying motifs similar to those identified *in vivo*. For example, it was particularly striking that the clp19 binding site originally identified in the promoter of a candidate target gene contained a sequence with only a single-base-pair change from the CSI motif (5′-CAATC-3′ versus 5′-CATTCC-3′). Further analysis of the CSI data has revealed many motifs with single-base changes from the CSI motif that also exhibit strong relative binding by Sxi2a (B.C. Stanton and C.M. Hull, unpubl. data).

Additional CSI investigations will focus not only on determination of the comprehensive binding profile for the Sxi1α protein, but also on the detection of any changes that may occur in the binding profile of Sxi2a when complexed with Sxi1α. The results of these experiments can then be used to carry out additional bioinformatic analyses of the *C. neoformans* genome to map the transcriptional circuitry of this eukaryotic developmental pathway.

Summary

The homeodomain proteins of *C. neoformans*, Sxi1α and Sxi2a, comprise a system of homeodomain-mediated developmental control where the mechanism of action is unknown. Delineating the role of the Sxi proteins during sexual development will further our understanding of how *C. neoformans* produces spores, potential infectious particles. Utilizing *in vitro* and *in vivo* binding assays, CSI arrays, bioinformatics and *in vivo* complementation experiments, we have identified features of the Sxi proteins as well as the DNA sites to which they bind that distinguish them from the fungal homeodomain protein paradigms and indicate similarities to their metazoan counterparts. The data presented here advance the understanding of eukaryotic regulation of transcriptional networks by identifying potential regulatory sites and defining key aspects of sexual developmental regulation in an important human fungal pathogen.

Experimental procedures

Protein expression and purification

The DNA sequences encoding Sxi1α-N (residues 1–205) and Sxi2a-HD (residues 145–201) were PCR amplified using primers listed in Table S1. Templates used in these reactions were pCH286 for Sxi1α-N and pCH287 for Sxi2a [pRSET-A (Invitrogen) vectors containing the full-length cDNA of each protein]. The resulting PCR products were cloned into the BamHI site of the pRSET-A vector (Invitrogen). Positive clones were transformed into BL21 DE3 pLysS cells (Stratagene), and transformants were cultured and induced in Zymo Research EB/OB medium according to manufacturer’s instructions. Proteins were purified as described previously (Ekena et al., 2008). Briefly, cells were re-suspended in histidine-binding buffer (500 mM NaCl, 20 mM Tris, pH 8.0, 5 mM imidazole, 50 mM phenylalanine, 50 mM isoleucine), lysed by sonication, and the cell extract was applied to a gravity-flow column packed with Qiagen nickel resin. Protein was eluted from the resin using histidine-binding buffer supplemented with 0.5 M imidazole.

Electrophoretic mobility shift assays

DNA probes were labelled using α-32P-dNTPs in a Klenow fill-in reaction or by PCR (see Table S1). The resulting probes were
used in binding assays containing 100 mM NaCl, 0.5 mM EDTA, 60 mM Tris-Cl pH 7.9, 5 mM MgCl₂, 5 mg ml⁻¹ bovine serum albumin, 50 μg ml⁻¹ polydeoxyinosinic-deoxycytidylic acid, 10% glycerol, and purified protein(s). DNA-binding assays were carried out as described previously (Ekena et al., 2008) wherein purified protein was incubated with a DNA probe for 15 min at 4°C. The reaction mixtures were electrophoresed on a 5% non-denaturing gel for 2 h at 200 V, except in Fig. 1B and E, where the samples were electrophoresed for 1 and 4 h respectively. Dissociation constant approximation (Fig. 1E) was performed in quadruplicate, with labelled DNA concentrations ranging from 1 to 20 nM.

**One-hybrid analysis**

A reporter construct was created by cloning the clp19 sequence (GTATGTGTATCTCA) into a CYC1-lacZ reporter (pLGΔ-178) (Guarente and Mason, 1983). Various combinations of vectors expressing the Sxi2α-HD fragment (with the GAL4 activation domain, pCH577), the Sxi1α-FL fragment (without any accessory domains, pCH568) and pLGα-178 (with and without the clp19 binding site, pCH564 and pCH848 respectively) were transformed into chemically competent EG123 cells (Siliciano and Tatchell, 1986). Transformants were selected on minimal medium lacking tryptophan, leucine and uracil (SD -trp -leu -ura).

**CPRG assays**

Transformants were grown to log-phase in minimal medium (SD -trp -leu -ura), diluted in rich medium, and incubated at 30°C until an OD₆₀₀ of ~0.5–1.0 was reached. The cultures were pelleted, washed, re-suspended and lysed. Chlorophenol red beta-D-galactopyranoside (CPRG) was added to lysed cells and optical densities (OD₆₅₀) and time were recorded to track the colorimetric change. Each sample was analysed in triplicate. Miller units were calculated using the following formula: β-galactosidase units (Miller units) = 1000 × OD₆₅₀(T × V × OD₆₀₀), T = elapsed time (in min) of CPRG incubation, V = 0.1 × concentration factor, OD₆₀₀ = OD₆₅₀ at time of harvest. Statistical analysis was performed using an unpaired Student’s t-test, and significant differences are indicated by an asterisk to denote P < 0.0001.

**Cognate Site Identifier analysis**

The Sxi2α homeodomain was assayed for DNA binding using 10- and 8-mer CSI arrays (Warren et al., 2006) with the following modifications: the arrays were blocked in 2.5% non-fat dried milk for 1 h at room temperature with gentle agitation. Binding assays were performed in 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM MgCl₂, 0.5 mM EDTA supplemented with bovine serum albumin (3 mg ml⁻¹), non-fat dried milk (0.5%), DTT (0.5 mM), anti-6x-histidine Cyanine-5-conjugated antibody (Qiagen) and purified Sxi2α homeodomain protein (50 nM). The binding assay was incubated on ice for 40 min prior to its application to the array (75 min, 4°C, gentle agitation). Fluorescence data were acquired using an Axon microarray scanner (Molecular Devices Corporation, Union City, CA). Data were analysed using the GenePix Pro software and statistical analysis was carried out as described previously (Warren et al., 2006).

**MEME analysis and motif generation**

Motifs were generated using the motif finding algorithm MEME (Bailey and Elkan, 1994). Logos were created by inputting sequences from a particular Z-score bin representing that motif into the WEBLOGO program (Crooks et al., 2004). In the case of Fig. 3B, motifs were generated by inputting representative sequences from the top 2585 sequences (top 0.5% of data).

**RNA preparation and cDNA synthesis**

Strains were grown to stationary phase in liquid YPD broth with 2% glucose, washed and suspended in 1× PBS. Strains were mixed in equal numbers (JEC20 and JEC21 for the wild-type a × α cross, and CHY768 and CHY610 for the sxi2α × sxi1Δ cross) and plated onto V8 medium. Crosses were incubated at room temperature, in the dark, for 24 h. RNA was extracted using hot acid-phenol as described previously (Ausubel et al., 1997), and was further purified using the RNEasy Mini Kit (Qiagen, Valencia, CA). RNA quality was assessed using the Agilent 2100 BioAnalyser (Agilent Technologies, Santa Clara, CA) prior to cDNA synthesis. cDNA was synthesized using the SuperScript™ III Direct Labelling Kit and Purification Module (Invitrogen, Carlsbad, CA), according to manufacturer’s instructions.

**Microarray analysis**

Differentially labelled cDNAs were hybridized, in triplicate, to C. neoformans whole genome spotted microarrays of 70-mer oligonucleotides (Cryptococcus Community Microarray Consortium, Washington University Genome Sequencing Center). Slides were incubated in pre-hybridization buffer (5× SSC, 0.1% SDS, 1% BSA) for 1 h at 42°C. cDNAs were then applied to slides in 1× formamide buffer (25% formamide, 5× SSC, 0.1% SDS) containing 10 μg each sheared salmon sperm DNA (Eppendorf, Westbury, NY) and yeast tRNA (Sigma-Aldrich, St Louis, MO). Hybridizations were incubated for 12 h at 42°C. Arrays were scanned on a GenePix 4000B scanner and the data extracted using GenePix Pro 4.0 (Molecular Devices, Sunnyvale, CA) at the University of Wisconsin-Madison Gene Expression Center (http://www.biotech.wisc.edu/servicesresearch/gec/). Extracted data were analysed using the GeneSpring 9.0 software package (Agilent Technologies, Santa Clara CA). After initial background correction, data were normalized using the LOWESS (locally weighted scatter plot smoothing) algorithm (Quackenbush, 2002) and assessed for significance using a Student’s t-test.

**Site-directed mutagenesis**

The QuikChange Mutagenesis Kit (Stratagene) was used to selectively mutate the coding sequences of Sxi1α and Sxi2α to change specific residues to alanine. Reactions were
carried out according to the manufacturer’s instructions. Mutations were made in the context of full-length genomic constructs, and pCH274 was used as template for the Sxi1α gene and pCH269 for the Sxi2α gene. Constructs were sequenced to identify positive clones, and the mutated open reading frames (and their corresponding 3′UTRs) were cloned into a telomeric vector, pCH390 (Garcia-Hermoso et al., 2001). Plasmids were transformed biolistically into the appropriate C. neoformans strain, CHY619 (α sxi1α::NAT) for Sxi1α plasmids and CHY926 (a sxi2α::URA5 5-FOAβ) for Sxi2α plasmids (Toffaletti et al., 1993). Transformants were evaluated for complementation of sexual development.

**Strain manipulations and media**

All strains used were of the serotype D background and were handled using standard techniques and media as described previously (Sherman et al., 1986; Alspaugh et al., 1998). C. neoformans JEC34 (a ura5) and JEC43 (α ura5) strains were maintained on yeast extract peptone dextrose (YPD, 1% yeast extract, 2% peptone and 2% dextrose) agar (Moore and Edman, 1993). Transformants were maintained on minimal medium agar. Crosses were carried out on 5% V8 medium. Sxi1α transformants were crossed by JEC34, and Sxi2α transformants were crossed by JEC43. Crosses were incubated at room temperature in the dark for 2 days. Sexual development was evaluated by microscopically observing the periphery of test spots on V8 medium for the presence of aerial hyphae and spore chains. The mating control strains used were JEC20 (a) and JEC21 (α) (Kwon-Chung et al., 1992; Alspaugh et al., 1998).

**Circular dichroism**

Three independently purified preparations of each 6x-histidine-tagged protein were diluted to concentrations ranging from 0.3 to 0.6 mM in 0.1 M sodium phosphate buffer (pH 7). The CD spectra for purified proteins were obtained using an Aviv 202SF spectrometer at the University of Wisconsin Biophysics Instrumentation Facility. The signal was averaged for 5 s during wavelength scans. Igor Pro software was used to analyse data.

**Acknowledgements**

Thanks to D.R. McCaslin, C.D. Carlson and W.C. Hallows for technical assistance and to C.R. Kent, J.L. Ekena, K.G. Klein and R.K. Borchardt for laboratory support. Thanks to P.J. Kiley, D.A. Brow, J.L. Keck, R.M. Brazas, M.W. Staudt and M.R. Botts for critical reading of and comments on the manuscript. This work was supported by R01 AI064287 to C.M.H. from the NIH, a Burroughs Wellcome Fund Career Award in the Biomedical Sciences to C.M.H., and R01 G069420 to A.Z.A. from the NIH. Circular dichroism experiments were carried out at the Biophysics Instrumentation Facility (Department of Biochemistry, University of Wisconsin-Madison), which was established with funding from the NSF (BIR-9512577), the NIH (S10 RR13790) and the University of Wisconsin, Madison.

**References**


Goutte, C., and Johnson, A.D. (1993) Yeast a1 and α2 homeodomain proteins form a DNA-binding activity with

Supporting information
Additional supporting information may be found in the online version of this article.
Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.