BMC Molecular Biology

BioMed Central

Flexible promoter architecture requirements for coactivator recruitment

Derek Y Chiang (dchiang@ocf.berkeley.edu)
David A Nix (nix@bioroot.org)
Ryan K Shultzaberger (shultzab@hotmail.com)
Audrey P Gasch (agasch@wisc.edu)
Michael B Eisen (mbeisen@lbl.gov)

Flexible promoter architecture requirements for coactivator recruitment

Derek Y Chiang[1, ¶], David A Nix[2,†], Ryan K Shultzaberger[1], Audrey P Gasch[3], Michael B Eisen[1,2§]

[1] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

[2] Department of Genome Sciences, Life Sciences Division, Ernest Orlando Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

[3] Department of Genetics, University of Wisconsin, Madison, WI 53706, USA

[¶] Current address: Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA
[†] Current address: Affymetrix, Santa Clara, CA 95051, USA

[§]Corresponding author

Email addresses:
DYC: dchiang@ocf.berkeley.edu
DAN: nix@bioroot.org
RKS: shultzab@hotmail.com
APG: agasch@wisc.edu
MBE: mbeisen@lbl.gov

Abstract

Background

The spatial organization of transcription factor binding sites in regulatory DNA, and the composition of intersite sequences, influences the assembly of the multiprotein complexes that regulate RNA polymerase recruitment and thereby affects transcription. We have developed a genetic approach to investigate how reporter gene transcription is affected by varying the spacing between transcription factor binding sites. We characterized the components of promoter architecture that govern the yeast transcription factors Cbf1 and Met31/32, which bind independently, but collaboratively recruit the coactivator Met4.

Results

A Cbf1 binding site was required upstream of a Met31/32 binding site for full reporter gene expression. Distance constraints on coactivator recruitment were more flexible than those for cooperatively binding transcription factors. Distances from 18 to 50 bp between binding sites support efficient recruitment of Met4, with only slight modulation by helical phasing. Intriguingly, we found that certain sequences located between the binding sites abolished gene expression.

Conclusions

These results yield insight to the influence of both binding site architecture and local DNA flexibility on gene expression, and can be used to refine computational predictions of gene expression from promoter sequences. In addition, our approach can be applied to survey promoter architecture requirements for arbitrary combinations of transcription factor binding sites.

Background

In most eukaryotes, the sequences that regulate transcription integrate multiple signals, through the binding of different transcription factors, to modulate levels of gene expression. When bound to DNA, transcription factors anchor the assembly of multiprotein complexes that influence the recruitment of RNA polymerase. Efficient assembly depends on optimally spaced protein-protein interactions among transcription factors and auxiliary proteins [1-4]. Since transcription factors recognize specific sites on DNA, the distance between these binding sites can influence how transcription factors interact with each other and other proteins. For example, overlapping sites may prevent two transcription factors from binding simultaneously, while sites too distant from each other may hinder bound transcription factors from recruiting necessary cofactors. Furthermore, some distantly spaced sites can only properly interact when the DNA between them is looped, a process influenced by the composition of the looped DNA.

Computational approaches take into account the multifactorial nature of transcriptional regulation when discovering transcription factor binding sequence motifs. Several methods search for pairs of enriched sequences, while allowing either fixed or variable spacing between them [5-8]. Other approaches start with binding specificities for pairs of known transcription factors, then assess whether the predicted occurrences of their binding sites are closely spaced [9-11]. Notably, most of these methods assess the spacing between binding sites with arbitrary criteria, such as

minimum and maximum distance thresholds between binding sites. However, the biological relevance and suitability of these thresholds has seldom been verified experimentally.

Several mechanisms that govern transcription factor interactions have been previously described. Transcription factors may bind cooperatively to adjacent sites in DNA, thus increasing the stability of the ternary DNA-protein complex. Since this effect is mediated by direct protein-protein interactions, sites for cooperatively binding transcription factors are usually spaced within 20 bp (*e.g.*, [12-15]). Slight alterations in spacing between the binding sites can drastically reduce gene expression unless helical phasing is preserved. Computational analyses suggest that proper helical phasing between predicted binding sites may be a common property of transcriptional control regions [16, 17].

Alternatively, transcription factors may bind to DNA independently and cooperatively recruit a coactivator protein. Co-recruitment of such activators is analogous to an "AND gate" in logic. Coincident binding of two proteins increases the fidelity and specificity of signal detection [2] [18] [19]. The network of transcription factors that regulates sulfur gene derepression in yeast provides a model system to dissect the promoter architecture requirements for coactivator recruitment. Among these transcription factors, only the coactivator Met4 contains a strong activation domain. However, Met4 does not bind to DNA directly, but is recruited in sulfur limitation conditions by Cbf1 and Met28 to the *MET16* promoter, as well as by Met28 and Met31/32 on regions from the *MET3* and *MET28* promoters [20, 21]. In addition, yeast two-hybrid assays with Met4 truncation mutants revealed distinct regions that mediate interaction with Cbf1 and Met31/32 [21, 22]. Taken together, these studies suggest a model by which the coactivator Met4 is coordinately recruited by the transcription factors Cbf1, Met28 and Met31/32 to the promoters of sulfur utilization genes (reviewed in [23]). Nevertheless, the effects of distance constraints and sequence context between Cbf1 and Met31/32 binding sites have not been characterized.

We use the term promoter architecture to refer to distance constraints and sequence context effects that govern interactions among transcription factor binding sites. Our goal is to understand how promoter architectures differ for transcription factors that participate in coactivator recruitment, versus those that bind cooperatively. In this work, we developed a synthetic promoter assay to characterize how various distances between Cbf1 and Met31/32 binding sites influenced gene expression in response to methionine starvation. The relative order of binding sites affected reporter gene expression. We discovered that distance constraints on coactivator recruitment were more flexible than those for cooperatively binding transcription factors. Distances from 18 to 50 bp between binding sites could support efficient recruitment of Met4, with only slight modulation by helical phasing. Intriguingly, we found that certain sequence contexts between the binding sites abolished gene expression. Finally, we noted that the probability of coactivator recruitment could be affected by the bendability of the spacer sequence between transcription factor binding sites.

Results

Annotated promoters of sulfur-regulated genes contain closely spaced binding sites

We first surveyed the promoter architectures of 19 annotated genes in *S. cerevisiae* that were co-expressed under sulfur-limiting conditions [24]. All of these promoters contained Cbf1 and Met31/32 binding sites with perfectly identical sequences among at least 3 of 4 closely-related yeast species (Figure 1). We assessed binding site conservation based on identity within a multiple sequence alignment, but not the distances between the sites. Due to the small sample size, we did not distinguish between the possible relative orientations of the binding sites. Notably, each sulfur-regulated promoter included a Cbf1 binding site upstream of a Met31/32 binding site. These conserved binding sites could occur in either forward or reverse orientation.

In each promoter, we searched for the Cbf1 binding site upstream of the Met31/32 binding site that yielded the smallest distance between them in *S. cerevisiae*. A histogram of these smallest distances showed a peak between 10 and 30 bp (Figure 2A). This peak suggested an optimal distance between the transcription factors was necessary for efficient Met4 recruitment. When investigating whether the distances between the closest pairs of binding site were helically phased, we could not detect an enrichment of distances on a certain face of DNA (Figure 2B). Finally, the vast majority of annotated promoters contained Met31/32 binding sites within 100 to 350 bp upstream of the translation start site (Figure 2C).

Cbf1 binding sites are required upstream of Met31/32 binding sites

A larger collection of sulfur-regulated promoters would provide more statistical power to define key components of promoter architecture. To explore sequence space more diverse than that found in the yeast genome, we developed a synthetic genetic approach to select for sulfur-regulated promoters from a plasmid library. We engineered a single-copy plasmid that fused a minimal promoter upstream of the *HIS3* reporter gene (Figure 3). To test their effects on transcription levels, different promoter architectures were embedded in the context of the minimal promoter from the *S. cerevisiae MEL1* gene [25]. This promoter was chosen for its low background expression, compared to promoters derived from the *S. cerevisiae CYC1* gene. Promoter architectures with combinations of regulatory sequences that supported sufficient expression of the *HIS3* reporter gene enabled the parental yeast strain BY4742 to grow in media lacking histidine. In addition, semiquantitative measurements of *HIS3* expression can be assayed by titration with 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of the *HIS3* gene product [26]. Faster growth rates in the presence of higher concentrations of 3-AT correspond to higher expression levels of the *HIS3* gene.

We sought to define the minimal regulatory information that was sufficient to induce reporter gene expression in response to sulfur limitation. As a negative control, none of the synthetic promoters were able to induce enough *HIS3* expression in a repressive concentration of methionine to support growth on a low level (5 mM) of 3-AT (Figure 4A). In addition, neither the minimal promoter alone nor a single Met31/32 binding site could support growth in the absence of methionine with 10 mM 3-AT. A single Cbf1 binding site supported weak growth on 10 mM 3-AT. In the wild-type *MET14* promoter, a Cbf1 binding site was found 35 bp upstream from a Met31/32 binding site, as measured by center-to-center distance. Two Cbf1 binding

sites placed at the same distance showed moderate *HIS3* expression. However, two Met31/32 binding sites were unable to support growth. A promoter with a Cbf1 binding site upstream of a Met31/32 binding site showed the highest level of *HIS3* expression. In contrast, a promoter with the Met31/32 binding site found upstream of the Cbf1 binding site was unable to support growth on 10 mM 3-AT.

We compared these results obtained from a minimal promoter system to the average effects of promoter architecture in endogenous yeast genes. For each gene, we estimated the average number of mRNA copies per cell by multiplying the basal transcript levels in rich growth media and the average ratio of gene expression change from published microarray studies of sulfur limitation conditions [27-29]. We then averaged these inferred absolute transcript levels for sets of genes whose promoters shared combinations of Cbf1 or Met31/32 binding sites (Figure 4B). The presence of Cbf1 binding sites or Met31/32 binding sites was associated with a significant increase in inferred transcript levels compared to the rest of the genome. However, two Cbf1 or two Met31/32 binding sites did not contribute to higher transcript levels. Genes whose promoters shared a Cbf1 binding site upstream of a Met31/32 binding site showed significantly higher transcript levels compared to genes with only a single Cbf1 binding site. In contrast, the opposite order of binding sites was not associated with a significant increase in transcript levels. Thus, the constraint on the order of binding sites was consistent with computational predictions of the regulatory effects for various promoter architectures.

High cooperativity between Cbf1 and Met31/32 binding sites spaced at least 18 bp apart

We predicted that efficient recruitment of Met4 to the promoters of sulfur utilization genes should depend on the spacing between Cbf1 and Met31/32 binding sites. To investigate the effect of varied spacing on reporter gene activation, we constructed a set of promoter libraries that differed by 2-bp increments from 6 bp to 34 bp, as well as 5-bp increments from 40 bp to 50 bp. Each promoter library had a fixed size but degenerate nucleotide sequences between the Cbf1 and Met31/32 binding sites. The binding sites were flanked by 10 bp of sequence from the *MET16* promoter of *S. bayanus*, which lacks an adjacent Gcn4 site.

By pooling hundreds of yeast transformants for each library, we reasoned that the contribution of nucleotide composition on Met4 recruitment and subsequent gene activation would be averaged out. Growth rates for each promoter library thus represent the aggregate effect of a certain distance on reporter gene expression. At several steps in the procedure, we took care to reduce the potential of selecting only the fastest-growing strains from each pool. First, we picked transformants of similar colony size. Second, we amplified these pools overnight in dropout leucine media, which ensured that the plasmids were retained but did not select for levels of promoter expression. Third, we measured growth rates for 7 hours after promoter induction and selection with 3-AT, which corresponded to fewer than 2.5 doubling times. Finally, we verified the complexity of the library pools by isolating single colonies on non-selective growth plates and sequencing of the promoters in twenty different colonies.

Pooled measurements of growth rates in sulfur limiting conditions determined that a minimum distance between Cbf1 and Met31/32 binding sites was required for the highest levels of gene expression (Figure 5). Yeast harboring promoter libraries of varying sizes grew at similar rates in the absence of 3-AT, indicating low levels of leaky transcription from the reporter construct. Expression levels of the *HIS3* reporter gene were titrated with the addition of 1 mM 3-AT; similar results were obtained with

different concentrations of 3-AT (data not shown). Binding sites whose centers were spaced fewer than 14 bp apart promoted weak reporter gene expression. At these close distances, Cbf1 and Met31/32 may be sterically constrained from assembling a complex with Met4. Reporter gene expression increases sharply as the distance between binding sites is increased from 14 bp to 18 bp. The highest levels of gene expression were observed for promoter libraries with binding sites spaced from 18 bp to 50 bp apart, whereas helical phasing modulated the average growth rate by less than 20%.

Sequence context between binding sites can inhibit gene activation

In addition to characterizing the aggregate effects of binding site spacing, we also examined the effects of different spacer sequences on reporter gene expression. We assayed the growth rates of individual yeast transformants on solid media containing 10 mM or 25 mM 3-AT. Each of the 70 to 72 transformants tested for a certain distance harbored a promoter with a different, random sequence between the Cbf1 and Met31/32 binding sites. We observed reproducible variability in growth rates among transformants with the same distance, but different spacer sequences, between Cbf1 and Met31/32 binding sites (Figure 6).

At each distance surveyed, a certain proportion of intervening sequences was compatible with reporter gene expression. Since the pooled growth rates in liquid media were qualitatively similar over this distance range, we interpret these proportions as the probability that a random intervening sequence would support gene expression at a given distance. At a distance of 12 bp between sites, less than 30% of the sequences supported reporter gene expression. At distances between 16 and 50 bp, the proportion of transformants that showed moderate to high levels of growth on 25 mM 3-AT varied from 38% to 60%. We observed a modest dependence of this proportion on helical phasing in the distance between binding sites.

To investigate what features of spacer sequences correlated with gene activation, we sequenced a sample of 28 promoters with distances of 12 bp, as well as 41 promoters with distances of 20 bp, between the Cbf1 and Met31/32 binding sites (Table 2). Promoters that supported gene expression (positives) were similar in nucleotide composition to promoters that inhibited gene expression (negatives). Since no trimers or tetramers were enriched in the positive or negative promoter sets, additional sequence-specific transcription factors probably did not contribute to gene expression. The most discriminating feature of negative promoters was a shared G or T immediately 5' to the Met31/32 binding site in 15 of 17 examples of distance 12, as well as in all 13 examples of distance 20. However, about half of the positive examples contained a G or T at that position, as expected.

We searched for additional residues that could discriminate among sequences that shared a G or T at the most 3' position of the spacer region using WebLogo [30, 31]. We compared sequence logos between the positive and negative promoters to calculate whether any nucleotides were enriched at particular positions in the spacer sequences (Figure 7A). By focusing on the three most informative positions, we derived nucleotide combinations that predicted negative promoters with an overall sensitivity of 80% and a specificity of 89% (Table 3).

To test whether the A_{11}-T_{17} nucleotide combination was sufficient to inhibit gene expression in spacer sequences of length 20, we identified five promoters with a B_{11}-T_{17} sequence combination and converted the nucleotide at position 11 to an adenine by site-directed mutagenesis. Similar levels of reporter gene expression were driven by the original and mutant promoters, as assayed by serial dilutions on media

containing 10 mM or 25 mM 3-AT (Figure 7B). Thus, the effects of sequence context are not encoded by specific positions within the primary nucleotide sequence.

Discussion

Promoter architecture features of yeast sulfur utilization genes

We have developed a synthetic promoter assay to test how various features of promoter architecture affected *HIS3* reporter gene expression in the context of a common minimal promoter. Although this reporter gene assay is indirect, it has been successfully used to obtain semi-quantitative measurements of transcript levels [26]. We applied this system to characterize the collaborative recruitment of the coactivator Met4 by the transcription factors Cbf1 and Met31/32 in response to methionine starvation. We found that the relative order of binding sites was crucial, since a Cbf1 binding site was required upstream of a Met31/32 binding site for full gene expression. The influence of Cbf1 and Met31/32 binding site order on reporter gene expression implies that the spatial orientation of the Met4 activation domain is required for the recruitment of downstream targets. Two Cbf1 binding sites could moderately increase reporter gene expression, yet the mechanism for this enhanced activation is unclear. Synergistic activation of reporter gene expression occurred when Cbf1 and Met31/32 binding sites were spaced at least 18 bp apart. Notably, the allowed distances for coactivator recruitment extend beyond the maximal range for cooperatively binding transcription factors. Finally, we discovered that different sequence contexts between binding sites produced considerable heterogeneity of reporter gene expression, whereas helical phasing showed comparatively little effect.

Although the transcription factors Cbf1 and Met31/32 lack canonical activation domains, they can serve as activators via collaborative recruitment of the coactivator Met4 when they are jointly bound to the promoters of sulfur utilization genes (reviewed in [23]). Our genome-wide computational survey found that genes with single Cbf1 or Met31/32 binding sites in their promoters were associated with significantly higher transcript levels, on average, when compared to the rest of the genome (Figure 4B). However, the presence of two binding sites for the same transcription factor was not associated with a further increase in transcript levels on average. In contrast, we found that two Cbf1 binding sites separated by 35 bp in a minimal promoter conferred increased reporter gene expression (Figure 4A). This discrepancy could be explained by distinct distance constraints between two binding sites for the same transcription factor.

Our data could not rule out the possibility that Met31/32 may serve as transcriptional repressors when bound to other promoters that lack Cbf1 binding sites. We found that reporter gene expression in methionine starvation was lower for minimal promoters with one or two Met31/32 binding sites, compared to a minimal promoter with a Cbf1 binding site alone (Figure 4A). Several models could explain how the binding of Cbf1 could convert Met31/32 from a repressor to an activator. Since the recruitment of the coactivator Met4 requires interactions with both Cbf1 and Met31/32, Met4 could displace a corepressor that may be constitutively bound to Met31/32. In contrast, the binding of Cbf1 could recruit enzymes that confer posttranslational modifications on Met31/32. These modifications could induce a conformational change that relieves repressive activity of Met31/32. Kinetic analyses of transcription factor binding and subsequent recruitment of multiprotein regulatory complexes by chromatin immunoprecipitation could help distinguish between these models.

Distinct promoter architecture requirements for different transcription factor combinations

The promoter architecture requirements for Met4 coactivator recruitment differ considerably from previously characterized yeast promoters. For instance, the transcription factor Rap1 can efficiently recruit Gcr1/2 only when their binding sites are found 13 or 23 bp apart [13]. Notably, a distance of 18 bp that altered the helical phasing between these factors abolished gene activation. Similarly, helical phasing between Pho2 and Swi5 binding sites modulates cooperative binding by almost three-fold [14]. As an extreme case, the insertion of a single base pair between the a1 and α2 binding sites abolishes cooperative binding [15]. In contrast with the above transcription factor pairs that bind cooperatively, levels of reporter gene activation were fairly consistent when Cbf1 and Met31/32 binding sites were spaced between 18 bp and 50 bp apart. The tolerance of Met4 coactivator recruitment on a wide distance range contradicts the model that transcription factor interactions are predominantly determined by the precise spacing between their binding sites. Intriguingly, the recruitment of Met4 to a common minimal promoter seems to depend more on the sequence context between Cbf1 and Met31/32 binding sites than on the distance between them, provided that the minimum distance requirements were met. In light of these results, previous studies that varied distances between transcription factor binding sites should be reassessed, since they usually considered only a single sequence context for each distance.

The rather flexible distance constraints between Cbf1 and Met31/32 binding sites suggest that Met4 recruitment may not require rigid, simultaneous protein-protein interactions among the bound transcription factors. Taken together, these experiments suggest that the process of Met4 recruitment differs considerably from the lock-and-key arrangements of bound transcription factors that govern the mammalian interferon beta enhanceosome [1, 2]. Instead, an intrinsic property of the intervening sequence context, such as DNA bendability, may facilitate an induced fit between the bound transcription factors and Met4. Whereas the distance between binding sites plays a diminished role in bridging bound transcription factors, intervening sequences with low intrinsic bendability could impair coactivator recruitment. Thus, the key requirements of promoter architecture may rely heavily on the molecular mechanism of transcription factor interactions at a particular set of co-regulated promoters.

Possible effects of sequence context between transcription factor binding sites

Sequence context could alter Met4 recruitment in several ways. First, residues adjacent to binding sites could reduce the binding affinity of Cbf1 or Met31/32. Accordingly, we found that all spacer sequences that were incompatible with reporter gene expression contained a guanine or thymine immediately 5' to the Met31/32 binding site. Secondly, the DNA bendability of the spacer sequence could alter the conformation of Cbf1, which bends DNA by approximately 68° [32]. Conformational changes in Cbf1 could affect its protein-protein interactions with Met28 or Met4, thus reducing Met4 recruitment. A requirement for DNA bendability on protein-protein interactions has been recently shown for the transcription factor Mcm1, which bends DNA by 66°, comparable to the bend angle induced by Cbf1 [33]. A point mutant of Mcm1 with a DNA bending angle of 46° had a lower affinity for cooperative binding with Fkh2 than a mutant with a DNA bending angle of 49°, suggesting that a certain threshold of DNA bending was required for ternary complex formation *in vitro* [33]. Circular permutation assays on promoters with different sequence contexts could test

whether the extent of bendability correlates with reporter gene activation. In addition, chromatin immunoprecipitation studies could identify the transcription factors whose binding *in vivo* is affected by sequence context.

Whereas the influence of sequence context on gene activation has been widely reported *e.g.,* [34-36], the key determinants of sequence context have been poorly defined. Except for the residue adjacent to the Met31/32 binding site, we could not identify features of the primary nucleotide sequence that correlated with gene activation. Previous studies have reported that protein-DNA interactions can be affected by physicochemical properties of DNA, such as twist [37]. Although we assessed several dinucleotide parameters, we could not find any significant correlation between the average parameter value of a spacer sequence and reporter gene activation (data not shown).

Epigenetic effects could account for some of the observed variability in gene activation among promoters with different sequence contexts. By examining multiple independent serial dilutions for several promoter sequences (Supplementary Figure 1), we believe that this variability is reproducible and not due to stochastic effects on individual clones. In order to sample a large number of promoter architectures, we assayed reporter gene expression from a single-copy plasmid, which yields over 10,000-fold higher transformation efficiency than chromosomal integration. We have not explored how the flanking sequence composition of wild-type promoters may affect the basal or Met4-induced nucleosomal accessibility of Cbf1 and Met31/32 binding sites in the genome. Cbf1 can also modulate nucleosome positioning and recruit the Isw1 chromatin remodeling complex [38, 39]. Thus, additional determinants of local sequence context that affect the binding or DNA bending of Cbf1 may influence Met4 recruitment and gene activation in a chromosomal context.

Implications for computational predictions of transcription factor interactions

The development of computational methods to predict the transcriptional output of an arbitrary regulatory sequence has attracted considerable interest, as reviewed in [40, 41]. Most computational approaches assess the enrichment of predicted binding sites within a large sequence region, while ignoring the spatial arrangement of the binding sites. Moreover, only a handful of methods explicitly consider whether binding sites are more closely spaced than expected [9, 10, 42-44]. These methods typically specify minimum and maximum distance thresholds between which transcription factors are predicted to interact. Whereas the use of thresholds roughly approximates the range of transcription factor interactions, our above experiments suggest two major improvements for more accurate predictions.

First, different mechanisms of transcription factor interactions may impose distinct distance constraints between their binding sites. We found that the minimum spacing between Cbf1 and Met31 binding sites was the key distance constraint on reporter gene activation. We interpret this minimum distance to be a consequence of the coactivator's role in bridging the bound transcription factors. Various coactivators likely have different minimum distance requirements, based on their size and the relative locations of their interaction surfaces with DNA-bound transcription factors. Experimental studies on the promoter architecture requirements for other common coactivators should provide empirical distance thresholds that could improve the prediction accuracy of their regulated target genes.

In addition, the pronounced effects of sequence context on reporter gene activation suggest that highly accurate predictions of target gene regulation may not be easily extrapolated from targeted experimental studies. Further investigations of

promoter architecture may benefit from a framework that formalizes how enthalpy gains from protein-protein interactions are offset by the entropy loss of multiprotein complex formation. Thermodynamic measurements on promoter variants with different spacing and sequence contexts between transcription factor binding sites could then be associated with changes in gene activation. Such a theory on the energetics of multiprotein complex formation could provide the quantitative precision needed to predict how a particular transcriptional control region adopts a conformation that enables transcriptional activation.

Conclusions

The main goal of this work was to characterize the influence of various components of promoter architecture on transcription factor interactions. We found that the requirements for Met4 coactivator recruitment were considerably more flexible than those for cooperatively binding transcription factors. The characteristic requirements of Met4 recruitment included the precise order of Cbf1 and Met31/32 binding sites, a large distance range between the binding sites that was insensitive to helical phasing, and the pronounced inhibitory effects of sequence context. Given the modular design of our synthetic promoter system, our approach can be readily used to characterize the promoter architecture constraints between arbitrary combinations of yeast transcription factors.

Methods

Plasmid construction

Plasmid pDC204 was constructed in five steps. 1) The *HIS3* coding region was PCR amplified from *S. cerevisiae* genomic DNA using the primers HIS3_F_BamHI and HIS3_R (Table 1) and cloned downstream of the *MEL1* minimal promoter (P_{MEL1}) by ligating into the BamHI + EcoRV-cleaved plasmid YIpMELβ2 from EUROSCARF [25]. Two changes were then made to the *MEL1* minimal promoter. 2) An NcoI site was introduced into P_{MEL1} 31 bp upstream of the existing XhoI site by site-directed mutagenesis (oligos MEL1_NcoI_W and MEL1_NcoI_C). 3) An out-of-frame ATG codon located 17 bp upstream of the *HIS3* coding region was removed by site-directed mutagenesis (oligos ATG_W and ATG_C). 4) The P_{MEL1}-*HIS3* fusion construct was PCR amplified (primers pMH14-F_ApaI & pMH14-R_AscI-SacII) and cloned into the ApaI + SacII-cleaved plasmid pRS314 [45]. 5) The *Kluyveromyces lactis LEU2* gene was PCR amplified from pUG73 (primers pUG73_F and pUG73_R) [46] and cloned into the AscI site of the above plasmid. Restriction digests confirmed the same-strand orientation of the HIS3 and LEU2 coding regions, and sequencing verified the promoter and coding regions.

Promoter library construction

Degenerate oligonucleotides were designed with a Cbf1 binding site at a fixed distance upstream of a Met31/32 binding site (Operon) (Table 1). Ten bp of flanking sequence upstream of the Cbf1 binding site and downstream of the Met31/32 binding site were included from the wild-type *MET16* promoter. Double-stranded DNA was synthesized by Bio-X-Act polymerase (Bioline) from the primer MET16_reverse (Table 1), digested with NcoI and XhoI and ligated into pDC204.

Yeast strains and media

Strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was obtained from Invitrogen. Growth media were prepared by mixing yeast nitrogen base (Bio101), 2% dextrose and amino acid supplements lacking leucine or lacking histidine, leucine and methionine (BD Biosciences). Histidine or 3-amino-1,2,4-triazole (3-AT) (Sigma) were supplemented to the indicated concentrations.

Pooled growth rates for promoter libraries

Plasmids containing promoter libraries with the indicated spacings between binding sites were introduced into the BY4742 parental strain by lithium acetate transformation [47]. Transformants that harbored these single-copy plasmids were selected by growth on dropout medium lacking leucine. For each growth rate experiment, over 100 yeast colonies from a separate transformation were pooled and amplified by culturing overnight in dropout media lacking leucine.

To induce reporter gene expression, the pooled yeast cultures were diluted to early log phase (OD_{600} ~ 0.04) in 20 mL of dropout media lacking leucine, histidine and methionine and grown at 30°C with shaking at 250 rpm for 3 hours (OD_{600} ~ 0.1). Each culture was then split in half and 3-AT was added to one half, to a final concentration of 1 mM. To acclimatize the yeast cultures to 3-AT, the cultures were grown for a further 2.5 hours. Subsequently, we measured the OD_{600} of each pooled culture every 45 minutes until 7 hours after 3-AT addition. These measurements were transformed to log (base 2) values and a linear regression was calculated in Excel. The doublings per hour corresponds to the slope of the linear regression for a single growth curve.

Computational association of promoter architectures with gene expression

The regulatory information associated with a particular sequence – such as a transcription factor binding site – can be quantified as the average change in gene expression for all genes that contain that sequence in its transcriptional control region. We followed the standard practice of defining yeast transcriptional control regions as the 500 bp upstream of each coding region, as obtained from the Saccharomyces Genome Database [48]. We searched for exact matches to the core recognition sequences for Cbf1 (TCACGTG) or Met31/32 (TGTGGC) on either strand of these upstream regions.

We first computed the relative ratio of transcript levels between sulfur starvation and complete media for each gene. The relative ratio for each gene was obtained by averaging the log base 2 expression ratios for the first four timepoints of an amino acid starvation microarray experiment, as well as four replicates of a 1 mM cadmium treatment [28, 29]. To convert these average relative ratios to absolute mRNA levels, we multiplied these relative ratios by the average basal transcript levels in rich growth media as reported by [27]. We reported the average of these mRNA levels for all genes that shared each promoter architecture under consideration.

Authors' contributions

DYC, DAN, APG and MBE conceived and designed the experiments; DYC performed the experiments; DYC, DAN and RKS analyzed the data; DYC, DAN, RKS, APG and MBE wrote and approved the manuscript.

Acknowledgements

References

1.   Carey M: **The enhanceosome and transcriptional synergy**. 1998, *Cell* **92**:5-8.
2.   Merika M, Thanos D: **Enhanceosomes**. *Curr Opin Genet Dev* 2001, **11**:205-208.
3.   Ogata K, Sato K, Tahirov TH: **Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar**. *Curr Opin Struct Biol* 2003, **13**:40-48.
4.   Remenyi A, Scholer HR, Wimanns M: **Combinatorial control of gene expression**. *Nat Struct Mol Biol* 2004, **11**:812-817.
5.   van Helden J, Rios AF, Collado-Vides J: **Discovering regulatory elements in non-coding sequences by analysis of spaced dyads**. *Nucleic Acids Res* 2000, **28**:1808-1818.
6.   GuhaThakurta D, Stormo GD: **Identifying target sites for cooperatively binding factors**. *Bioinformatics* 2001, **17**:608-621.
7.   Liu X, Brutlag DL, Liu JS: **BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes**. *Pac Symp Biocomput* 2001, **6**:127-138.
8.   Robin S, Daudin JJ, Richard H, Sagot MF, Schbath S: **Occurrence probability of structured motifs in random sequences**. *J Comput Biol* 2002, **9**:761-773.
9.   Klingenhoff A, Frech K, Quandt K, Werner T: **Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity**. *Bioinformatics* 1999, **15**:180-186.
10.   Pilpel Y, Sudarsanam P, Church GM: **Identifying regulatory networks by combinatorial analysis of promoter elements**. *Nat Genet* 2001, **29**:153-159.
11.   Bulyk MJ, McGuire AM, Masuda N, Church GM: **A motif co-occurrence approach for genome-wide prediction of transcription-factor-binding sites in *Escherichia coli***. *Genome Res* 2004, **14**:201-208.

12. Amin J, Fernandez M, Ananthan J, Lis JT, Voellmy R: **Cooperative binding of heat shock transcription factor to the Hsp70 promoter in vivo and in vitro**. *J Biol Chem* 1994, **269**:4804-4811.

13. Drazinic CM, Smerage JB, Lopez MC, Baker HV: **Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p)**. *Mol Cell Biol* 1996, **16**:3187-3196.

14. Brazas RM, Bhoite LT, Murphy MD, Yu Y, Chen Y, Neklason DW, Stillman DJ: **Determining the requirements for cooperative DNA binding by Swi5p and Pho2p (Grf10p/Bas2p) at the *HO* promoter**. *J Biol Chem* 1995, **270**:29151-29161.

15. Jin Y, Mead J, Li T, Wolberger C, Vershon AK: **Altered DNA recognition and bending by insertions in the α2 tail of the yeast a1/α2 homeodomain heterodimer**. *Science* 1995, **270**:290-292.

16. Ioshikhes I, Trifonov EN, Zhang MQ: **Periodical distribution of transcription factor sites in promoter regions and connection with chromatin structure**. *Proc Natl Acad Sci USA* 1999, **96**:2891-2895.

17. Makeev VJ, Lifanov AP, Nazina AG, Papatsenko DA: **Distance preferences in the arrangement of binding motifs and hierarchical levels in organization of transcription regulatory information**. *Nucleic Acids Res* 2003, **31**:6016-6026.

18. Naar AM, Lemon BD, Tjian R: **Transcriptional coactivator complexes**. *Annu Rev Biochem* 2001, **70**:475-501.

19. Spiegelman BM, Heinrich R: **Biological control through regulated transcriptional coactivators**. *Cell* 2004, **119**:157-167.

20. Kuras L, Barbey R, Thomas D: **Assembly of a bZIP-bHLH transcription activation complex: Formation of the yeast Cbf1-Met4-Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding**. *EMBO J* 1997, **16**:2441-2451.

21. Blaiseau PL, Thomas D: **Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA**. *EMBO J* 1998, **17**:6327-6336.

22. Kuras L, Cherest H, Surdin-Kerjan Y, Thomas D: **A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism**. *EMBO J* 1996, **15**:2519-2529.

23. Thomas D, Surdin-Kerjan Y: **Metabolism of sulfur amino acids in *Saccharomyces cerevisiae***. *Microbiol Mol Biol Rev* 1997, **61**:503-532.

24. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**:241-254.

25. Melcher K, Sharma B, Ding WV, Nolden M: **Zero background yeast reporter plasmids**. *Gene* 2000, **247**:53-61.

26. Horecka J, Sprague GFJ: **Use of imadazoleglyercerolphosphate dehydratase (His3) as a biological reporter in yeast**. *Methods Enzymol* 2000, **326**:107-119.

27. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA: **Dissecting the regulatory circuitry of a eukaryotic genome**. *Cell* 1998, **95**:717-728.

28. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes**. *Mol Biol Cell* 2000, **11**:4241-4257.

29. Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, Marguerie G, Sentenac A, Werner M, Labarre J: **Sulfur sparing in the yeast proteome in response to sulfur demand**. *Mol Cell* 2002, **9**:713-723.

30. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: A sequence logo generator**. *Genome Res* 2004, **14**:1188-1190.

31. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences**. *Nucleic Acids Res* 1990, **18**:6097-6100.

32. Niedenthal RK, Sen-Gupta M, Wilmen A, Hegemann JH: **Cpf1 protein induced bending of yeast centromere DNA element I**. *Nucleic Acids Res* 1993, **21**:4726-4733.

33. Lim FL, Hayes A, West AG, Pic-Taylor A, Darieva Z, Morgan BA, Oliver SG, Sharrocks AD: **Mcm1p-induced DNA bending regulates the formation of ternary transcription factor complexes**. *Mol Cell Biol* 2003, **23**:450-461.

34. Elledge SJ, Davis RW: **Position and density effects on repression by stationary and mobile DNA-binding proteins**. *Genes Dev* 1989, **3**:185-197.

35. Mai X, Chou S, Struhl K: **Preferential accessibility of the yeast *HIS3* promoter is determined by a general property of the DNA sequence, not by specific elements**. *Mol Cell Biol* 2000, **20**:6668-6676.

36. Tornoe J, Kusk P, Johansen TE, Jensen PR: **Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites**. *Gene* 2002, **297**:21-32.

37. Olson WK, Gorin AA, Lu X-J, Hock LM, Zhurkin VB: **DNA sequence-dependent deformability deduced from protein-DNA crystal complexes**. *Proc Natl Acad Sci USA* 1998, **95**:11163-11168.

38. Moreau JL, Lee M, Mahachi N, Vary J, Mellor J, Tsukiyama T, Goding CR: **Regulated displacement of TBP from the PHO8 promoter in vivo requires Cbf1 and the Isw1 chromatin remodeling complex**. 2003, **11**(6):1609-1620.

39. Kent NA, Eibert SM, Mellor J: **Cbf1p is required for chromatin remodeling at promoter-proximal CACGTG motifs in yeast**. *J Biol Chem* 2004, **279**(26):27116-27123.

40. Futcher B: **Transcriptional regulatory networks and the yeast cell cycle**. *Curr Opin Genet Dev* 2002, **14**:676-683.

41. Istrail S, Davidson EH: **Logic functions of the genomic cis-regulatory code**. *Proc Natl Acad Sci USA* 2005, **102**:4954-4959.

42. Wagner A: **Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes**. *Bioinformatics* 1999, **15**:776-784.

43. Chiang DY, Moses AM, Kellis M, Lander ES, Eisen MB: **Phylogenetically and spatially conserved word pairs associated with gene-expression changes in yeasts**. *Genome Biol* 2003, **4**:R43.

44. Beer MA, Tavazoie S: **Predicting gene expression from sequence**. *Cell* 2004, **117**:185-198.

45. Sikorski RS, Hieter P: **A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae***. *Genetics* 1989, **122**:19-27.

46. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast**. *Nucleic Acids Res* 2002, **30**:e23.

47.   Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method**. *Methods Enzymol* 2002, **350**:87-96.
48.   Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hong EL, Nash R, Oughtred R, Skrzypek M, Theesfeld CL, Binkley G, Dong Q, Lane C, Sethuraman A, Weng S, Botstein D, Cherry JM: **Fungal BLAST and model organism BLASTP best hits: new comparison resources at the Saccharomyces Genome Database (SGD)**. *Nucleic Acids Res* 2005, **33**:D374-D377.
49.   Cliften PF, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston RH, Cohen BA, Johnston M: **Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting**. *Science* 2003, **301**:71-76.

Figures

Figure 1 - Promoter architectures of annotated sulfur-regulated genes

Conserved binding sites for Cbf1 (blue rectangles, TCACGTG), Met31 or Met32 (green circles, TGTGGC), Gcn4 (orangle triangles, TGA[C|G]TCA) and TBP (TATAA) are drawn to scale in the indicated intergenic regions. A binding site was considered conserved if invariant copies were aligned in at least 3 out of 4 closely-related *Saccharomyces* species in a multiple sequence alignment [24, 49].

Figure 2 - Minimum distances between conserved Cbf1 and Met31/32 binding sites in annotated sulfur-regulated promoters

(A) Histogram of minimum distances between a Cbf1 binding site (TCACGTG) and a Met31/32 (TGTGGC) binding site. Distances were calculated from the center of each binding site, as indicated by the arrows between the consensus sequences. (B) Helical wheel projection of minimum distances. Cbf1 binding sites were aligned at the top of the helical wheel (position 0). Each green dot represents the remainder of a minimum distance from (A) divided by 10.5 bp. Since the helical pitch of DNA is 10.4 bp, each dot approximates the position of the Met31/32 binding site relative to the Cbf1 binding site. (C) Histogram of distances between the Met31/32 binding sites from (A) and the translation start site.

Figure 3 - Synthetic promoter system

A minimal promoter from the *MEL1* gene (P_{MEL1}) was fused upstream of a *HIS3* reporter gene on a single-copy plasmid. Selected restriction enzyme sites are labeled with their coordinates. Various combinations of Cbf1 and Met31/32 binding sites were inserted between the NcoI and XhoI restriction enzymes sites in the minimal promoter.

Figure 4 - Reporter gene expression driven by various combinations of Cbf1 and Met31/32 binding sites

(A) Yeast strains were grown under either repressive (+Met) or activating (–Met) growth conditions, in addition to varying concentrations of the inhibitor, 3-AT. Each column represents a 10-fold serial dilution of a yeast strain containing a reporter plasmid with a different binding site combination, labeled as follows. V: vector alone, C: Cbf1 binding site, M: Met31/32 binding site, C2: Two Cbf1 binding sites spaced by 35 bp, M2: Two Met31/32 binding sites spaced by 35 bp, CM: Cbf1 binding site placed 35 bp upstream of a Met31/32 binding site, MC: Met31/32 binding site placed 35 bp upstream of a Cbf1 binding site. Yeast strains were grown on the indicated media for 5 days at 30°C. Serial dilutions were performed in triplicate, and a representative dilution is displayed for each growth condition. (B) Average inferred transcript levels in sulfur-limitation conditions associated with various promoter architectures. For each gene, the gene expression log ratio between sulfur limitation and complete media conditions was calculated as the average log base 2 expression ratio from previously published gene expression studies [28, 29]. To infer the average number of mRNA transcripts per cell, the gene expression ratio for each gene was multiplied by the basal transcript level as measured during growth in rich media [29]. Inferred transcript levels were averaged over sets of genes that shared the indicated binding site combinations in the 500 bp upstream of their translation start sites; the

first row indicates the average for all genes. Error bars indicate the standard error of the mean.

Figure 5 - Average growth rates for sequence libraries with defined distances between Cbf1 and Met31/32 binding sites

We generated libraries of pooled transformants containing single-copy plasmids with the indicated distances between Cbf1 and Met31/32 binding sites. We induced reporter gene expression by transferring these pooled cultures into media lacking histidine, leucine and methionine, plus the indicated concentration of 3-AT (see Methods for details). The growth rate, expressed in doublings per hour, for each pooled culture was obtained from regular measurements of OD_{600}. The average growth rate and standard error of the mean are plotted for three independent trials.

Figure 6 - Different sequences between Cbf1 and Met31/32 binding sites show a range of reporter gene activation

(A) Serial dilutions of yeast containing reporter plasmids with the same distance (20 bp) between binding sites, but different spacer sequences. Yeast strains were grown on the indicated media for 5 days at 30°C. (B) Proportions of transformants that displayed moderate to high levels of growth on solid media with 10 mM or 25 mM 3-AT. For each distance between binding sites, growth rates of 72 transformants with different spacer sequences were assayed with serial dilutions. Average growth rates in liquid media with 1 mM 3-AT, as in Figure 4, is also shown for comparison purposes.

Figure 7 - Sequence context effects were not determined by individual nucleotides

(A) Sequence logos for a sample of promoters with a center-to-center distance of 20 bp between Cbf1 and Met31/32 binding sites. At each position, the height of the nucleotide corresponds to the information content at that position, which weights its frequency in the sequenced sample compared with its expected frequency [31]. The sample was divided into sequences that supported or inhibited gene activation, as well as whether a guanine or thymine was found adjacent to the Met31/32 binding site. A separate logo was generated by WebLogo for each sub-sample [30]. Since three invariant nucleotides on both the 5' and 3' ends of the spacer sequence represented the Cbf1 and Met31/32 binding sites, respectively, only the central 14 bp are displayed. (B) Mutation of single nucleotides in positive promoters failed to inhibit reporter gene activation *in vivo*. Each column corresponds to a yeast strain containing a different spacer sequence with a distance of 20 bp between the binding sites. The clone number and nucleotide at position 11 are indicated above each panel. Ten-fold serial dilutions were performed in triplicate, and a representative dilution is shown.

Table 1) List of oligonucleotides used in this study

Plasmid construction

Oligo name	Sequence (5'→3') [restriction sites underlined]
HIS3_F_BamHI	CG<u>GGATCC</u>CGAAGATGACAGAGCAGAAAGC
HIS3_R	CCTCGTTCAGAATGACACG
MEL1_NcoI_W	CCCTGAAAGGTTTTT<u>CCATGG</u>AATAGTCAGGACGC
MEL1_NcoI_C	GCGTCCTGACTATT<u>CCATGG</u>AAAAACCTTTCAGGG
ATG_W	GTAATAAAAGCAACGACGTTGAACGGATCCCGAAAG
ATG_C	CTTTCGGGATCCGTTCAACGTCGTTGCTTTTATTAC
pMH14-F_ApaI	ATA<u>GGGCCC</u>GGAAATTTGTGTAAAACCCCC
pMH14-R_AscI-SacII	AACAA<u>CCGCGG</u>ATAAT<u>GGCGCGCC</u>CTCGTTCAGAATGACAC(
pUG73_F	AA<u>GGCGCGCC</u>GCATAGGCCACTAGTGGATCTG
pUG73_R	AGTAA<u>GGCGCGCC</u>CAGCTGAAGCTTCGTACGC

Promoter library construction

Oligo name	Sequence (5'→3') [restriction sites underlined]
MET16_reverse	CCGCTCGAGTTACTGAAGTTG
Cbf1_6_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_8_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_10_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_12_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_14_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_16_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_18_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_20_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_22_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_24_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_26_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_28_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_30_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAG</u>CGG
Cbf1_32_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGAGC</u>
Cbf1_34_Met31	CATG<u>CCATGG</u>TATCATCATTTCACGTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNCCACAACTTCAGTAA<u>CTCGA</u>

Table 2) Promoter sequences associated with reporter gene activation

Clone	Growth rate	Intervening sequence
5	-	TGGGTT
8	-	GAGGCG
20	-	GAGCAT
22	-	TGGATG
30	-	GTGAGT
32	-	AAAGAG
33	-	GTGACT
35	-	TGGTGT
36	-	AGAATG
44	-	TGCACT
47	-	AATTGG
48	-	AAACTC
3	+	TAAGAG
6	+	CATAGT
19	+	CGGTCC
25	+	GTTAAT
42	+	CGTCGT
2	++	CGCGTT
4	++	AACCGC
7	++	TGAGGC
27	++	CAAACG
43	++	CCATGG
46	++	GGTTGT
18	+++	ATTGGC
23	+++	AGGCAA
29	+++	ATATAT
31	+++	AAATGA
34	+++	TTGTGA

Clone	Growth rate	Intervening sequence
2	-	GAGTCTGATGGTCT
7	-	TGGGTTGTCAACGG
23	-	GGGCAATCGCGATG
28	-	CGTGGGGTGCTTAG
31	-	TATAAGGCGTTGGG
47	-	CGAGGGAAAACAG
48	-	TGAGGAGATGAAGT
68	-	GAAGTGAGGAGCGG
69	-	AAGAATTACCCGGT
3	+	CCTGATGCCTACAG
16	+	CAAGGCTAGGAGCG
24	+	GCGCAGGATCGGCT
67	+	GGGTGTGAAGGGCT
9	++	TGAGCTCTTGACAT

14	++	GGTTCAACGTTACT
30	++	GCAAGGAGCGAGGG
32	++	AGGGGAACGGAGAG
33	++	TAGTGGGATTTGCG
34	++	GGTGACTAGGCCTC
46	++	GAAGTGGATTGCGT
5	+++	GGACGTAATTTCAA
8	+++	TTTACAAACTAGGG
10	+++	CGATGTACTGCCAA
11	+++	GTTTGTTGGGATGG
12	+++	GGCATTTATGGGAA
13	+++	CCCTTCCTGTGGGC
15	+++	GGTGGTTCATGGGA
18	+++	CGCGCGGGCGTCTT
25	+++	TCAGGGTTCAGCCA
26	+++	CGCGCCGAACGGGC
27	+++	TAGTGTCGGGGGCG
29	+++	GTGGTAGACGCTGC
35	+++	TTATGGTACCACCA
36	+++	TCATGCGTCGTACG
37	+++	TTGCTGGCAAGGAT
38	+++	AAAAGAGGAGATTC
39	+++	ATGTCGTCATGTGT
40	+++	AATGGCATGCTGCG
41	+++	AGAGGCAGTATCAA
43	+++	GTTTGGGTCCGGGC
72	+++	GGCATTTATGGGAA

Table 3) Nucleotide combinations that correlate with lack of reporter gene activation

Size	Sequence combination	Sensitivity	Specificity
12	R_5-G_9	6 / 6 (100%)	6 / 7 (86%)
12	K_6-T_9	8 / 9 (89%)	8 / 9 (89%)
20	W_8-G_{17}	5 / 8 (62%)	5 / 5 (100%)
20	A_{11}-T_{17}	5 / 5 (100%)	5 / 6 (83%)

IUPAC symbols: R = A or G; K = A or C; W = A or T.

Additional Files

Additional File 1 (PDF) - Reproducibility of growth assays

Two different sequence contexts were tested between Cbf1 and Met31/32 binding sites with a center-to-center distance of 20 bp. The sequences of these clones were confirmed by isolating plasmids and sequencing. Clones 8A and 8B had the sequence <u>TCACGTG</u>TTTACAAACTAGGGG<u>GCCACA</u>; clones 12A and 12B had the sequence <u>TCACGTG</u>GGCATTTATGGGAA<u>GCCACA</u>. These plasmids were transformed independently into yeast strains. Serial dilutions of separate isolates were plated on the indicated growth media.
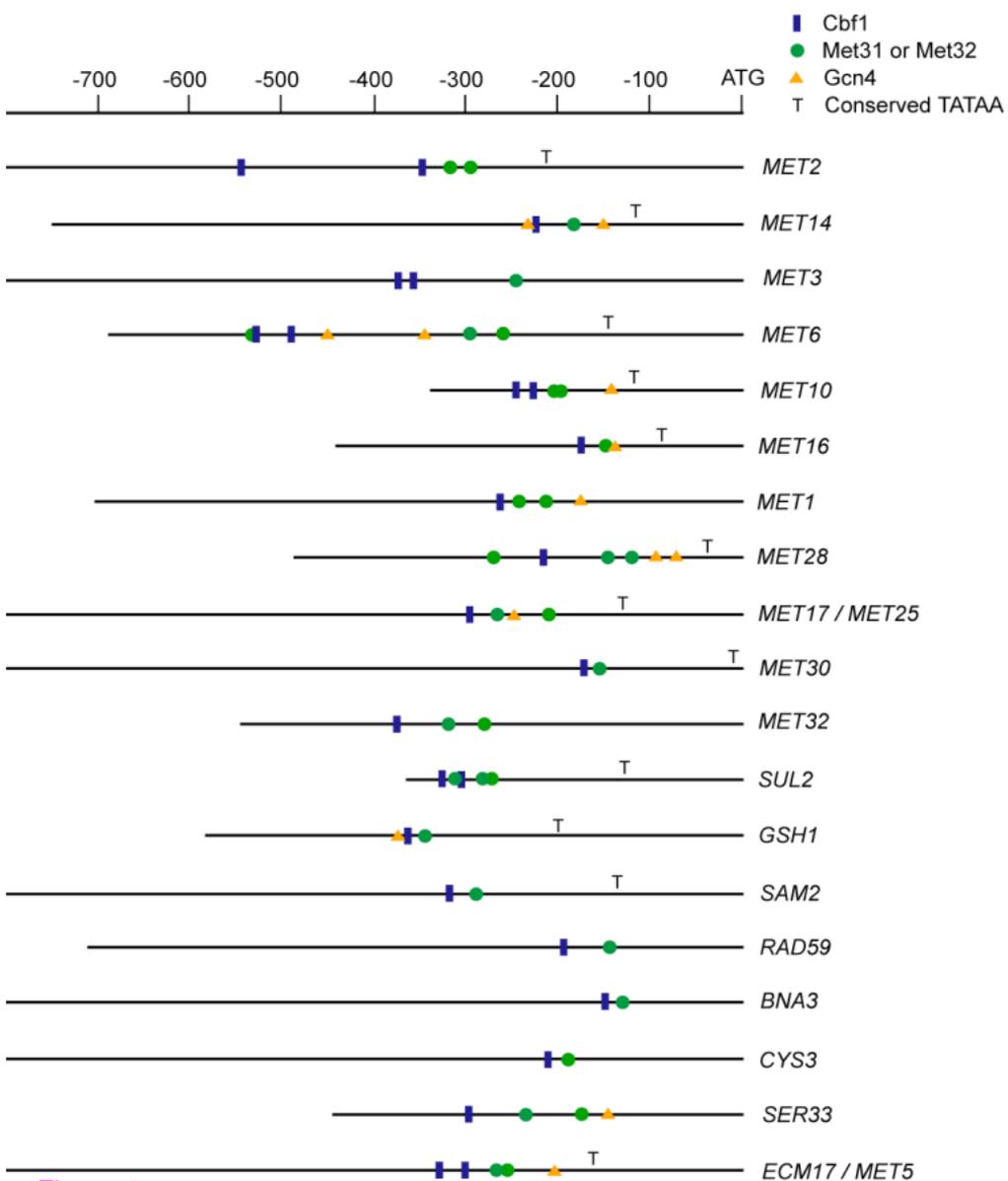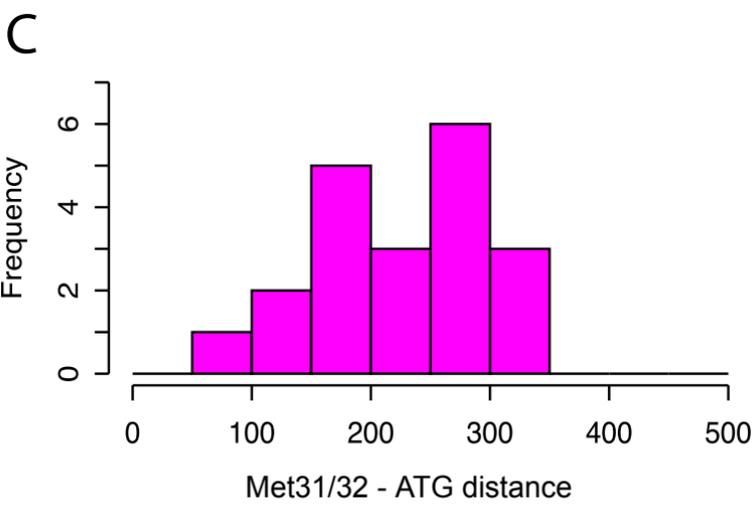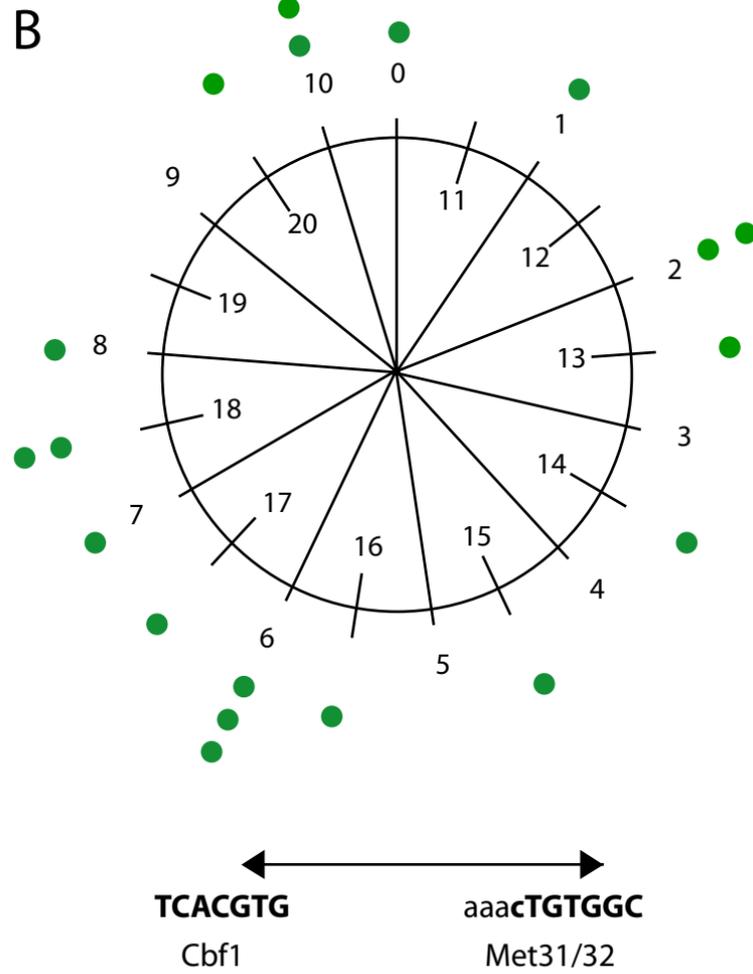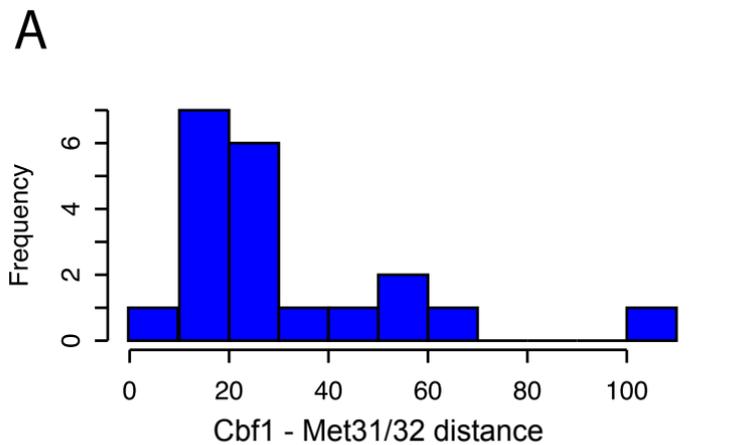
Figure 1
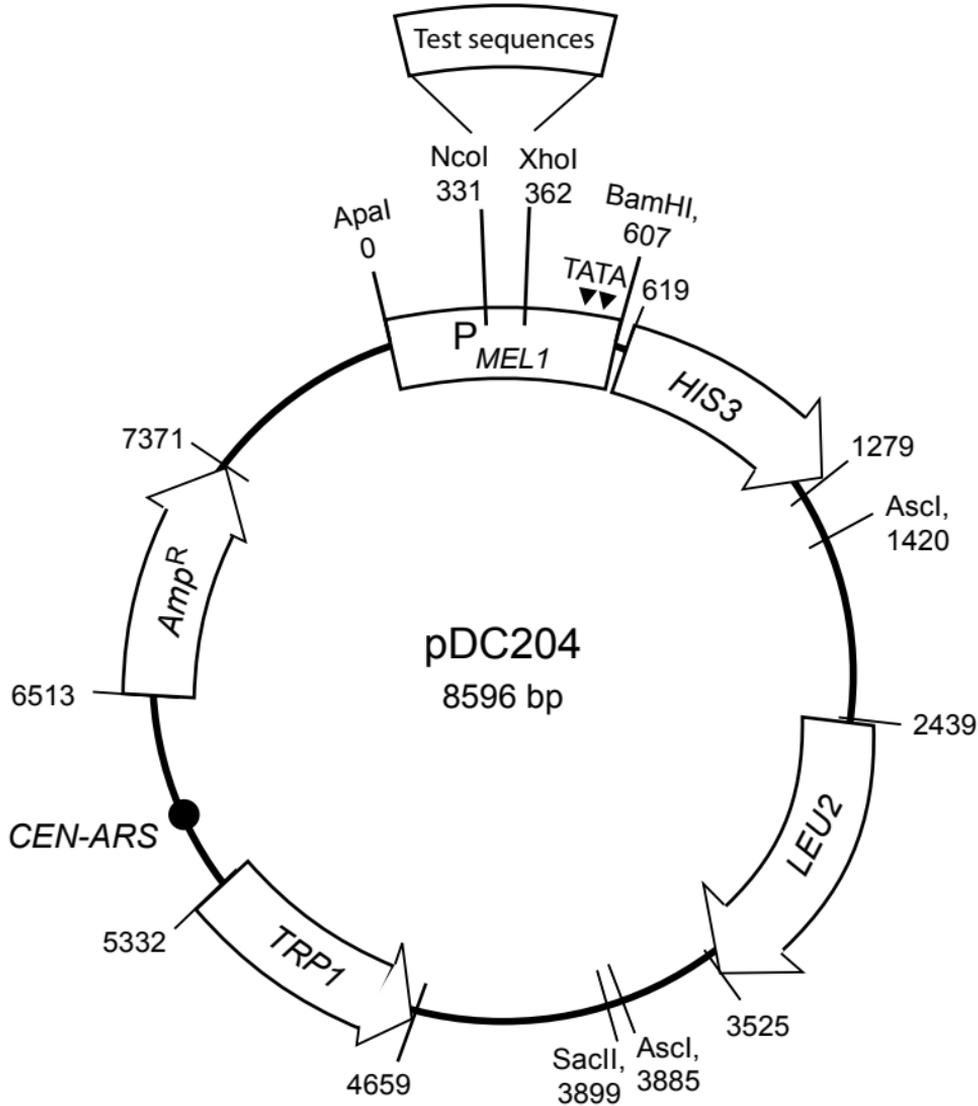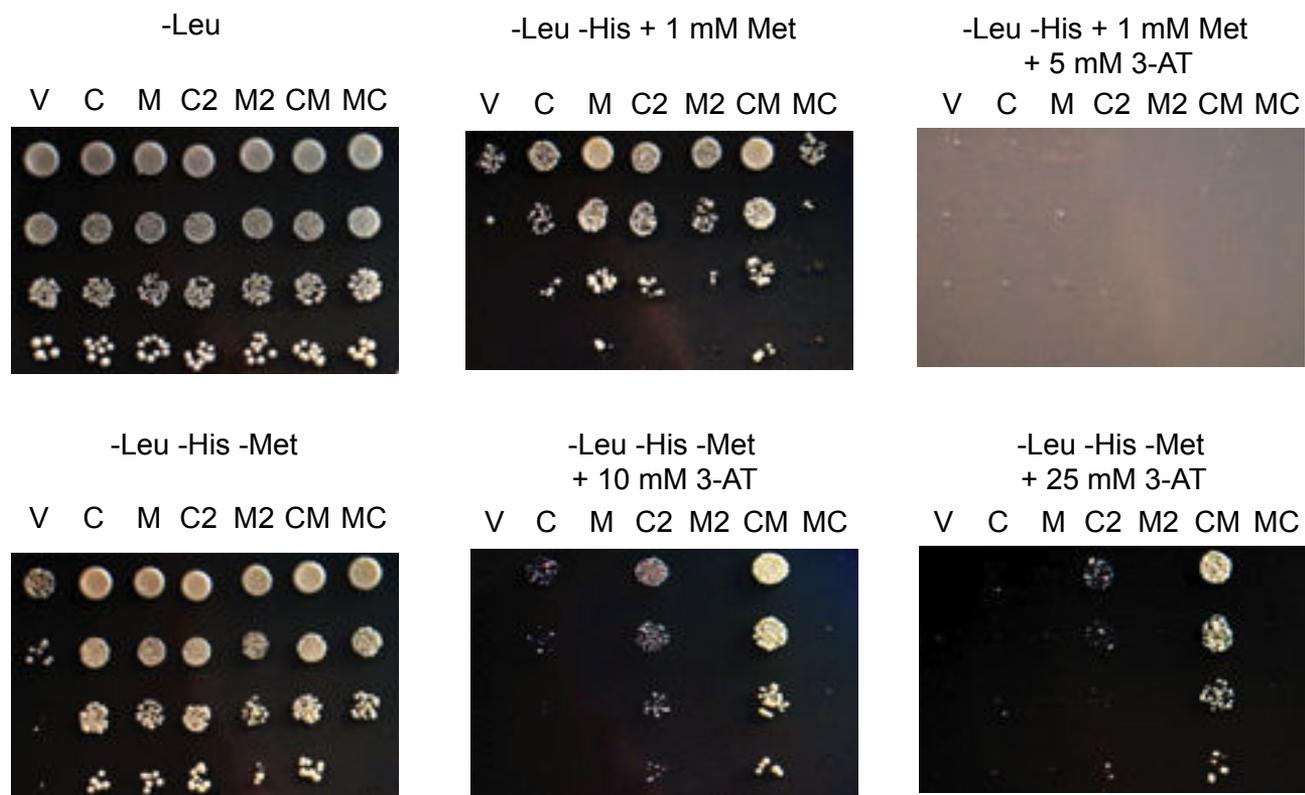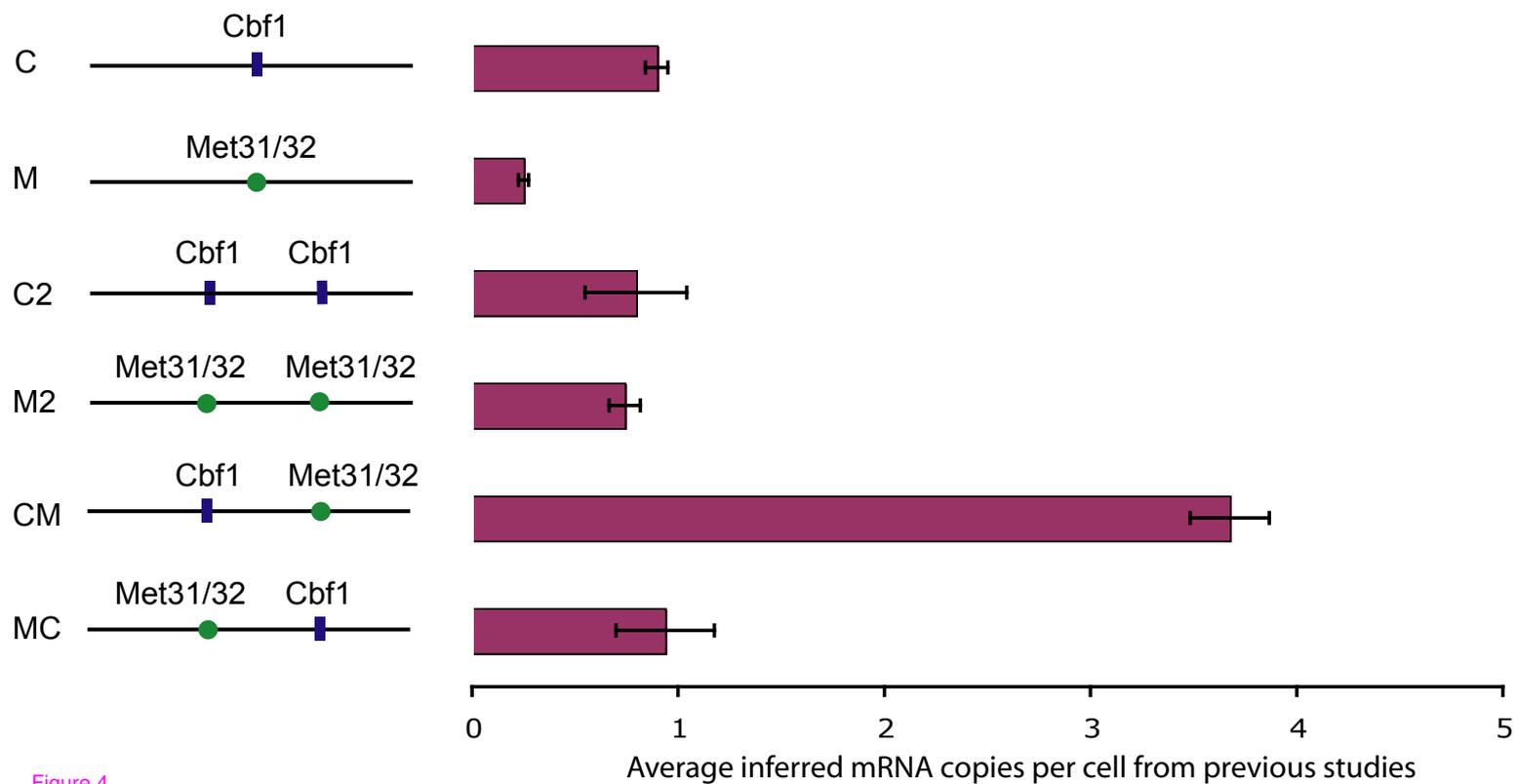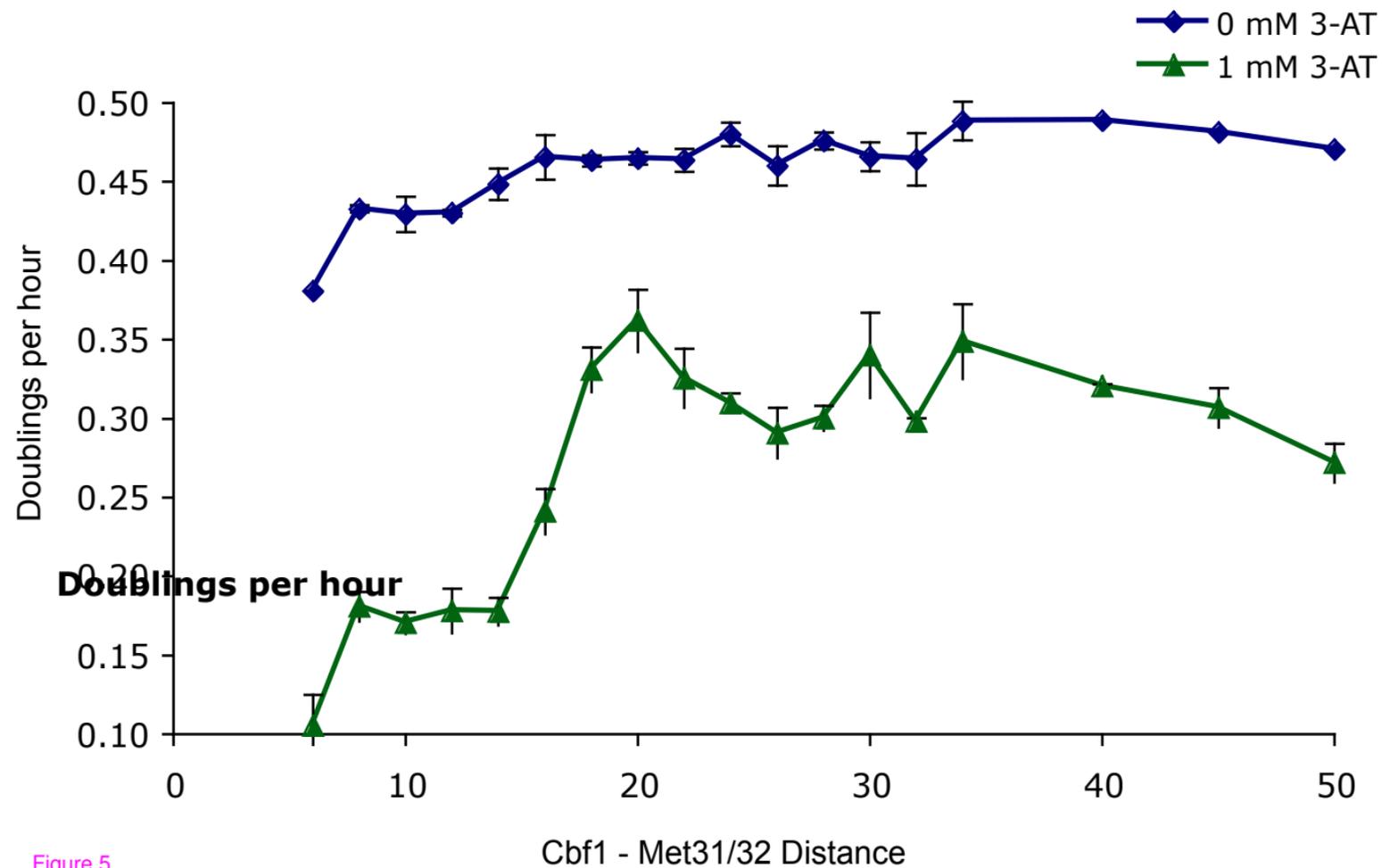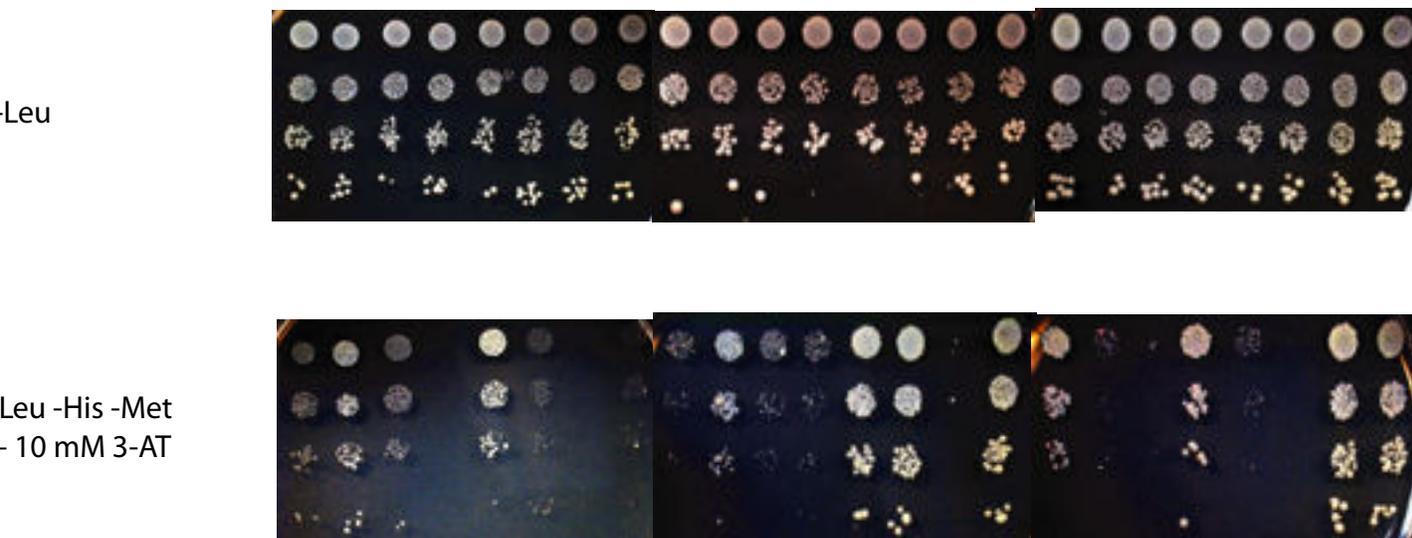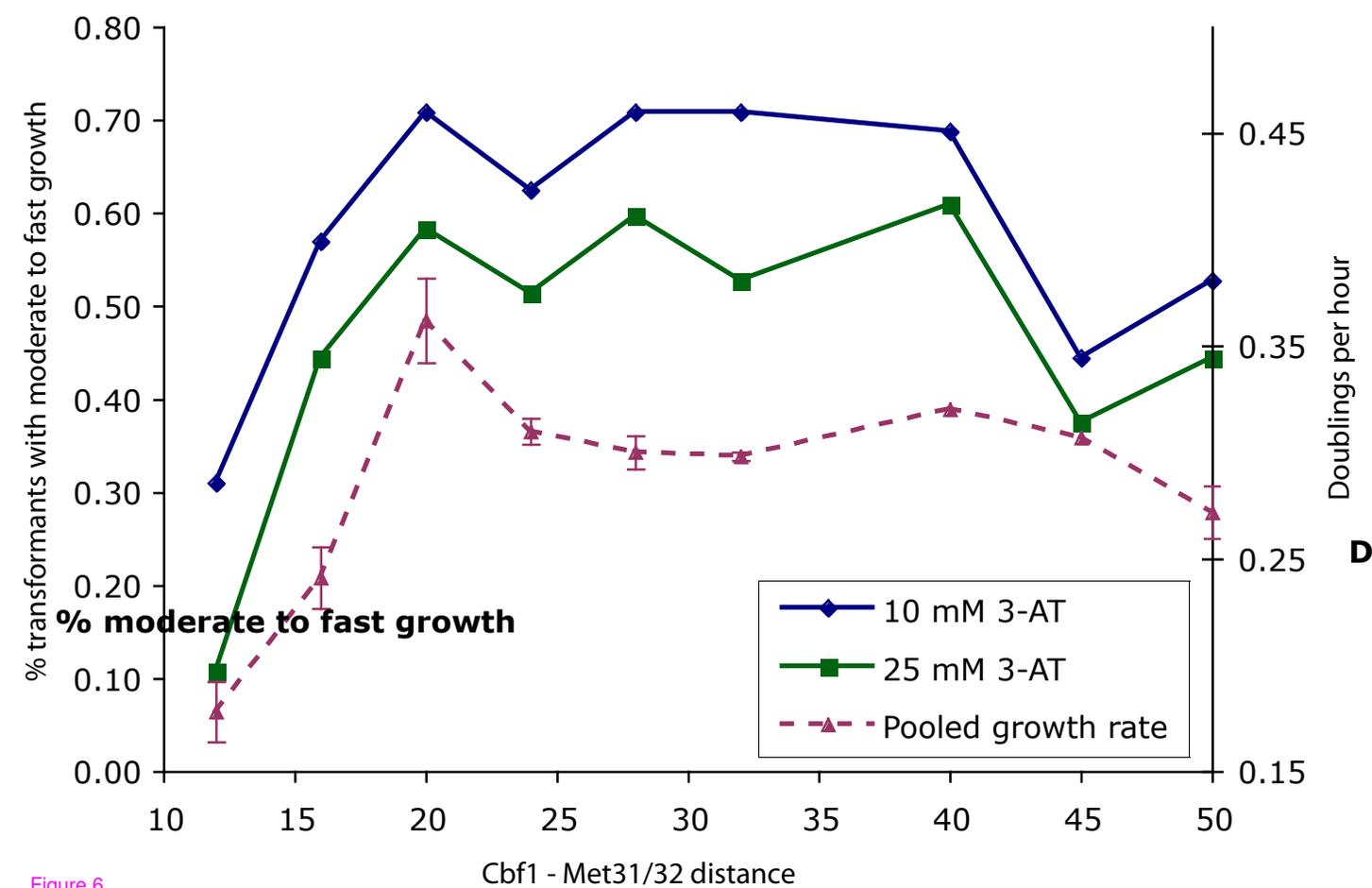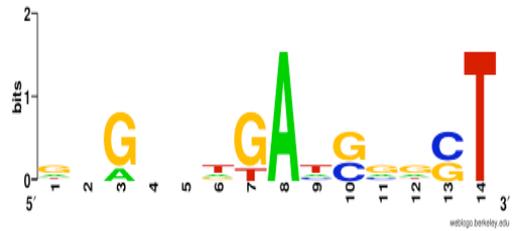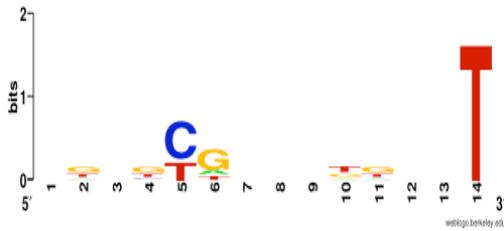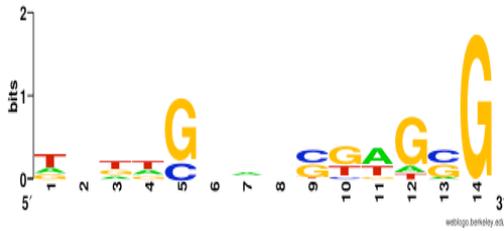
A

Frequency (y-axis), Cbf1 - Met31/32 distance (x-axis)

B

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

TCACGTG
Cbf1

aaa**cTGTGGC**
Met31/32

C

Frequency (y-axis), Met31/32 - ATG distance (x-axis)

Figure 2

A

-Leu

V C M C2 M2 CM MC

-Leu -His + 1 mM Met

V C M C2 M2 CM MC

-Leu -His + 1 mM Met
+ 5 mM 3-AT

V C M C2 M2 CM MC

-Leu -His -Met

V C M C2 M2 CM MC

-Leu -His -Met
+ 10 mM 3-AT

V C M C2 M2 CM MC

-Leu -His -Met
+ 25 mM 3-AT

V C M C2 M2 CM MC

B

C   Cbf1

M   Met31/32

C2   Cbf1   Cbf1

M2   Met31/32   Met31/32

CM   Cbf1   Met31/32

MC   Met31/32   Cbf1

Average inferred mRNA copies per cell from previous studies

0   1   2   3   4   5

Figure 4

Figure 5

A

-Leu

-Leu -His -Met
- 10 mM 3-AT

B

% moderate to fast growth

Legend:
- 10 mM 3-AT
- 25 mM 3-AT
- Pooled growth rate

x-axis: Cbf1 - Met31/32 distance

left y-axis: % transformants with moderate to fast growth

right y-axis: Doublings per hour

D
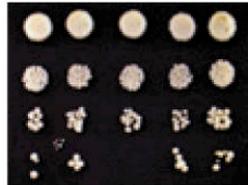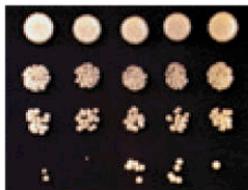
Figure 6

Figure 7

Additional files provided with this submission:

Additional file 1 : Additional file 1.pdf : 137Kb
http://www.biomedcentral.com/imedia/311167238101 8046/sup1.PDF