Differential binding of the related transcription factors Pho4 and Cbf1 can tune the sensitivity of promoters to different levels of an induction signal

Jonathan S. Z. Aow, Xiaowei Xue, Jin-Quan Run, Geoffrey F. S. Lim, Wee Siong Goh and Neil D. Clarke*

Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis St., Singapore 138672, Singapore

ABSTRACT

Transcription factors that belong to the same family typically have similar, but not identical, binding specificities. As such, they can be expected to compete differentially for binding to different variants of their binding sites. Pho4 is a yeast factor whose nuclear concentration is up-regulated in low phosphate, while the related factor, Cbf1, is constitutively expressed. We constructed 16 GFP-reporter genes containing all palindromic variants of the motif NNCACGTGNN, and determined their activities at a range of phosphate concentrations. Pho4 affinity did not explain expression data well except under fully induced conditions. However, reporter activity was quantitatively well explained under all conditions by a model in which Cbf1 itself has modest activating activity, and Pho4 and Cbf1 compete with one another. Chromatin immunoprecipitation and computational analyses of natural Pho4 target genes, along with the activities of the reporter constructs, indicates that genes differ in their sensitivity to intermediate induction signals in part because of differences in their affinity for Cbf1. The induction sensitivity of both natural Pho4 target genes and reporter genes was well explained only by a model that assumes a role for Cbf1 in remodeling chromatin. Our analyses highlight the importance of taking into account the activities of related transcription factors in explaining system-wide gene expression data.

INTRODUCTION

Most transcription factors belong to a relatively small number of families that are distinguished by the structures and sequences of their DNA-binding domains. In most of these families, members bind to similar DNA sequences (1,2). Because their DNA-binding specificities are similar, these proteins can be expected to compete with one another for binding to their cognate binding sites. Competition is avoided in many cases by the expression of family members in different cell types or under different conditions, or by member-specific interaction with other DNA-binding proteins that result in complexes of different binding specificity. Nevertheless, competition can occur, and it can be an important part of regulatory switches, a classic example being the lysis/lysogeny decision of phage lambda, mediated by the Repressor/Cro pair of helix–turn–helix proteins (3). The control of meiosis in yeast by Ndt80 and Sum1 is another well studied example, albeit involving factors from different families (4). Computational modeling of the Ndt80/Sum1 system has highlighted how competition enhances the specificity of target gene selection beyond the specificity intrinsic to either one of the factors by itself (5,6). One of the most widespread families of transcription factors is the basic helix–loop–helix (bHLH) family, a family typically found with multiple variants in animals, plants and fungi (7). The binding specificities of yeast bHLH proteins have been determined in vitro (1,2,8) and through motif enrichment analysis at in vivo binding sites defined by chromatin immunoprecipitation (ChIP) (9). These studies all show that bHLH proteins bind to short palindromic, or nearly palindromic, motifs, typically of the form CANNTG. Members of the family differ from one another, often quite subtly, through additional specificity determinants within the core hexamer and at flanking positions (8).

One member of the yeast bHLH family, Pho4, has a well-defined role in phosphate metabolism. In low phosphate, Pho4 becomes nuclear localized, binds to its cognate binding sites and activates transcription of a set of genes relevant to phosphate physiology (10–12). In contrast, another yeast bHLH protein, Cbf1, lacks a single well-defined function. It is a much more abundant protein than Pho4 (13), and there is no evidence for regulation of its expression or activity. The name CBF1 stands

*To whom correspondence should be addressed. Tel: +65 6808 8005; Fax: +65 6808 8291; Email: clarken@gis.a-star.edu.sg

© The Author(s) 2013. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
for ‘Centromere binding factor 1’ and mutants defective in Cbf1 grow poorly, apparently because of defects in chromosome segregation (14). However, Cbf1 also appears to have a direct role in regulating sulfur metabolism, and cbf1 mutants are methionine auxotrophs (14,15). Despite their differences in function, the methionine auxotrophy of a cbf1 mutant can be complemented by overexpression of Pho4, implying that Pho4 can bind and activate genes at sites normally bound by Cbf1 (16). Furthermore, cells that grow poorly on low phosphate due to a defect in phosphate transport can be suppressed by a cbf1 null mutation (16). This suggests that Cbf1 might ordinarily compete for Pho4 binding, so that in the absence of Cbf1, Pho4 better activates its target genes, mitigating the effects of low phosphate.

Pho4 and Cbf1 bind to many promoters in common (9) (Supplementary Figure S1), and recent ChIP-seq experiments demonstrate that the two proteins bind to some of the very same binding sites (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17). Many CACGTG motifs in the genome are bound more by Pho4 in a cbf1 mutant (17).

Pho4 and Cbf1 differ subtly in their binding specificities. These differences have been studied most thoroughly in vitro using an assay that rapidly traps protein–DNA complexes in a microfluidic device (8). This technique was used to measure the affinities of Pho4 and Cbf1 to each of the 16 possible dinucleotides flanking one side of the CACGTG core. Assuming that each half of the binding site contributes independently to affinity, these values allow the calculation of relative affinities for all 256 variants of the NNCACGTGNN motif (Figure 1). The richness and presumptive accuracy of these data offer a unique opportunity to ask how well in vitro binding affinity data can explain expression in vivo. We show that they do, in fact, do a remarkably good job explaining gene expression data for 16 reporter constructs over a range of phosphate concentrations. However, to explain the data at phosphate concentrations higher than ~80 μM, we find that a modest, but significant, contribution to transcriptional activation is required from Cbf1; also required to explain the data is competition for binding between Pho4 and Cbf1. Finally, we show that differences in the affinity for Cbf1 can explain much of the differences among natural Pho4 target genes in terms of their sensitivity to induction at moderate phosphate concentrations.

**MATERIALS AND METHODS**

**Strains**

Diploid strains containing a single vtc4::GFP reporter variant, with HIS4- and LEU2-tagged PCR products integrated by homologous recombination. For each of the vtc4::GFP variants, bridging PCR was used to produce a single PCR product from two input PCR products, one that was common to all the constructs and one that was unique to each variant. The one in common contains yEGFP and HIS4, flanked at the 5’ end by VTC4 5’ UTR sequences adjacent to the initiation codon and at the 3’ end by VTC4 3’ UTR sequences. The unique input consisted of the desired Pho4/Cbf1-binding site motifs, with sequences at the 3’ end overlapping the yEGFP-HIS4 product. The two inputs were then amplified with primers that extended the flanking regions of identity to the genome. The vtc4::GFP haploid strains were mated to BY4742 PHO4-myc producing the diploids that were used in the FACs expression analysis. For ChIP analyses, two epitope-tagged variants of strain W303 were constructed in which the epitope tags for Pho4 and Cbf1 were switched: W303 PHO4-myc::TRP1 CBF1-3HA::LEU2 and W303 PHO4-3HA::LEU2 CBF1-myc::TRP1. W303 PHO4-myc::TRP1 was used in the gene expression microarray analyses, demonstrating that a C-terminal epitope tag does not destroy PHO4’s regulatory activity.

**Flow cytometry**

vtc4::GFP bearing diploid strains were grown in minimal media supplemented with high phosphate (10 mM). Overnight cultures were spun down, washed twice in water, resuspended again in water and then diluted into 3 ml of minimal media supplemented with 0 μM, 20 μM, 50 μM, 80 μM, 100 μM, 1 mM or 10 mM phosphate. Because yeast grows faster in high phosphate, the cells were diluted to ~0.3 K cells/ml for the two highest concentrations of phosphate, ~1–2 K cells/ml for the 50–100 μM phosphate samples, ~5 K/ml for the 20 μM sample and ~25 K/ml for the 0 μM sample. Following overnight rocking at 30°, cell fluorescence was analyzed on a FacsCalibur flow cytometer (Becton-Dickinson). At least three biological replicates, each consisting of at least three technical replicates, were performed for each combination of reporter and phosphate concentration. For each technical replicate, the median fluorescence signal from 5000–10 000 cells was used for further analysis. Values reported here are the average of these median values, taken over the set of biological replicates. The reasonable precision of these data is evident from the way the averaged median values vary smoothly and monotonically across phosphate concentrations. The values and their standard deviations, as well as the means and standard deviations for the first and third quartiles of the fluorescence signals, are provided in Supplemental Dataset S1.
Expression microarrays
W303 PHO4-myc::TRP1 was grown to mid-log phase in Yeast extract, bacto Peptone, Dextrose (YPD), washed in water, resuspended in minimal media lacking phosphate and incubated at 30°C with shaking for 3h. The culture was split and phosphate added to final concentrations of 10 mM, 1 mM, 100 μM, 10 μM, or 0 μM. The cultures were incubated a further 80 min. Biotinylated cRNA were prepared according to the Affymetrix protocol. After fragmentation, 10 μg of cRNA were hybridized for 16 h at 45°C on the S98 Yeast Genome Array. GeneChips were washed and stained in the Affymetrix Fluidics Station 450. Data were analyzed with Microarray Suite version 5.0 using Affymetrix default settings and global scaling as normalization method. Data are available at GEO: GSE26770.

Chromatin immunoprecipitation-qPCR
For ChIP-qPCR analysis, a pair of haploid strains was used that contained either myc-tagged Pho4 and 3HA-tagged Cbf1, or myc-tagged Cbf1 and 3HA-tagged Pho4 (see section on Strains for details). Each strain was grown in minimal media containing 0 or 10 mM phosphate as described in the section on flow cytometry, and ChIP was performed essentially as previously described (18). Two biological replicates were performed for each of the eight combinations of transcription factor (Pho4/Cbf1), epitope tag (myc/3HA) and phosphate concentration (0/10 mM), and each biological replicate was analyzed by qPCR twice, each time with three technical replicates on the plate. Along with the eight Pho4-target genes, two genomic regions lacking Pho4/Cbf1 motifs were used as controls, along with the input DNA. Ct values for the triplicate on-plate replicates were first averaged, and then used to calculate the ΔCt values for each locus relative to the input DNA. These data are provided as Supplemental Dataset S2. Differences in the efficiency of crosslinking or immunoprecipitation ΔCt can lead to systematic variation in ΔCt values across biological replicates, so we first median-normalized all of the sets of ΔCt values, then averaged these values across the four replicates (two biological × two technical) and, finally, subtracted from each normalized averaged ΔCt value the lowest of the ΔCt values for the two control sequences. These ΔΔCt values were converted into fold-enrichment values. There are eight sets of fold-enrichment values for each gene in the myc-tagged dataset, εgene,m myc according to $E_{scaled} = 1 + S(E_{gene,myc} - 1)$ where the scale factor $S$ was defined as $S = \frac{\sum E_{gene,3HA} - 1}{\sum E_{gene,myc} - 1}$.

Computational modeling of binding and expression
For each of the 16 motif variants analyzed in this work, we calculated relative equilibrium constants based on the ΔΔG values for flanking dinucleotides affinities determined in vitro (8) under the assumption that the two flanking regions contribute additively to the free energy of binding (Supplementary Table S1). The affinity values are expressed relative to the mean $K_0$ of all 256 variants. Binding site occupancies, $\theta$, were determined by standard binding isotherms: the fractional occupancy for factor A is defined as $\theta_A = \frac{K_A^A}{1 + K_A^A [A]}$ where $[A]$ is the concentration of factor and $K_A^A$ is the association constant of factor A for the site. In this context, the association constant is the predicted affinity for the NNCACGTGNN variant found at the site of interest, normalized to the average for all NNCACGTGNN variants. This equation was used to infer the effective concentration of the factor in the cell by fitting gene expression values to binding site affinity, under the assumption that gene expression is directly proportional to transcription factor occupancy. The protein concentration in turn can be used to calculate occupancy, yielding a linear relationship between expression and occupancy. To accommodate competition from a second factor, $B$, the binding isotherm formula is modified to:

$$\theta_A = \frac{K_A^A [A]}{1 + K_A^A [A] + K_B^B [B]}$$

where $[B]$ is the concentration of the competing factor and $K_B^B$ is its binding affinity. Reporter gene expression levels, $E$, in arbitrary units, were predicted by summing the weighted contributions of Cbf1 and Pho4 occupancies: $E \propto \theta_{Cbf1} + 5.6 \theta_{Pho4}$. The weighting factor of 5.6 for Pho4-bound sites was determined from the ratio of slopes for the expression versus occupancy fits for Pho4 and Cbf1 under conditions when each dominates the expression differences (i.e., 0 mM phosphate for Pho4 and 10 mM phosphate for Cbf1). A fixed concentration of $[Cbf1] = 1$ was used to fit expression data at all phosphate concentrations; [Pho4] was varied systematically to maximize the fit between experimental and predicted
expression levels. For calculations involving reporter constructs, which differ at a single site, these calculations were performed in Excel. Excel was also used for calculating expression sensitivity values for the natural promoters and for modeling Cbf1 as a chromatin remodeling activity. For this purpose, if the promoter had more than one CACGTG motifs, calculations were performed independently for each motif, and the total expression activity was based on the probability that one (or more) Pho4 was bound and the probability that one (or more) Cbf1 was bound. For other purposes, occupancies of the natural promoter were predicted with the program GOMER, which uses a Position Weight Matrix (PWM) and an assumed protein concentration to score a genomic region for the probability of being bound at least one location (5). To generate PWMs for Pho4 and Cbf1, we started with a shared CACGTG core. The weights for this core were obtained by averaging the relative free energy terms from several PWMs for Pho4 and Cbf1, and then symmetricizing the PWM around the palindromic center (9,19,20). We then added weights for the flanking dinucleotides based on experimental binding affinities (21). To derive the PWM weights from the 16 experimental values for dinucleotide-binding affinity, a linear regression was performed to estimate the six independent parameters required (three at each position; the weight for the fourth base is determined by the other three). The free energy differences predicted by the PWMs are well correlated with the experimental free energy differences ($R = 0.95$ for Pho4 and $R = 0.85$ for Cbf1). The PWMs are available as Supplementary Tables S2 and S3.

RESULTS

GFP reporter assays

The gene VTC4 is a direct regulatory target of Pho4 as defined by ChiP experiments and by a joint statistical analysis of sequence motifs and the effect of constitutive Pho4 expression on gene expression (Methods; Supplementary Figure S2) (5,22,23). We used the VTC4 promoter as a backbone for construction of 16 GFP fusions that differ at a single Pho4/Cbf1 binding site. The wild-type VTC4 promoter contains two perfect CACGTG motifs but, based on data from in vitro affinity measurements (8), one of these sites (caCAGT Gaa) is predicted to bind Pho4 ~30-fold less well than the other (cgCAGTGGcc). Nevertheless, to avoid complications in interpretation, we knocked out the weaker site in all of our reporter constructs, using several substitutions in the core motif. The remaining site was replaced by all 16 possible palindromic variants of the sequence nnCACGT Gnn. (The promoter sequences are provided as Supplementary Figure S3). The 16 reporter constructs were separately integrated by homologous recombination into a diploid yeast strain, replacing one of the VTC4 alleles with the gene fusion (Methods). Each of the 16 reporter strains was then assayed for GFP expression at eight different phosphate concentrations ranging from 0 to 10 mM (Figure 2).

For most of the promoter variants, there is a substantial increase in expression as the phosphate concentration is lowered, as is expected for a gene under the control of Pho4. Furthermore, there is a clear qualitative association between expression level and the affinity of the motif for Pho4 (as indicated by the brightness of the yellow color in the Pho4 column of the key) and expression under maximally induced conditions (0 mM phosphate) (Figure 2). However, for the variants that have the highest affinity for Cbf1 (brightness of blue color in the Cbf1 column of the key), the phosphate-dependent induction is more modest. In fact, in the extreme case, the two variants with the highest Cbf1 affinity (GT and AT in the 5’ flanking region), there is no increase in expression whatever in low phosphate. Interestingly, the expression level for each of these two promoters is actually higher than that of all others under non-inducing conditions (10 mM phosphate), when Pho4 activity is at its lowest level. As these two promoters are expected to bind Cbf1 most strongly, this result suggests that Cbf1 itself drives gene expression in the absence of strong competition from Pho4.
maximum expression level and the concentration of Pho4, we find a good fit at a Pho4 concentration four times the mean $K_d$ of the binding sites. Using this value of $[\text{Pho4}] = 4$, we can transform affinities into predicted fractional occupancies, yielding a linear relationship between expression and predicted occupancy ($R = 0.94$; Figure 3B). This is a remarkably strong correlation considering that the occupancy of binding sites in vivo was predicted on the basis of subtle differences in affinity in vitro; we conclude that the specificity in vivo must be very similar to the specificity in vitro (8). Furthermore, the experimental value that was correlated to predicted binding is the expression level of reporter genes. Expression levels need not have been related in some simple fashion to binding, but the strength and linearity of the correlation suggests that median single-cell GFP fluorescence values are, in fact, an accurate proxy for transcription factor binding in this system. The effective Pho4 concentration is also likely to be reasonably well estimated because both the linearity of the fit and the value of the correlation coefficient fall off as Pho4 concentration values vary from the optimum (Supplementary Figure S4). We have less confidence in the absolute Pho4 occupancy values because these values are based on the assumption that expression asymptotically approaches its limit entirely because Pho4 binding asymptotically approaches saturation. In reality, there may be additional factors, such as the saturation of RNA polymerase activity, that affect maximal expression level. However, these effects do not materially affect our interpretation, as the excellent fit that we find between in vitro affinity and expression requires none of the assumptions that are required for the estimation of occupancy.

Cbf1-mediated activation and competition with Pho4 are required to explain expression under moderate-phosphate conditions

Under non-inducing conditions (10 mM phosphate), the differences in expression among the 16 promoter variants are much smaller than under inducing conditions, and the differences are uncorrelated with Pho4 affinity (Figure 4A). Surprisingly, though, they are well correlated with Cbf1 affinity ($R = 0.90$ for the linear fit to predicted occupancy $[\text{Cbf1}] = 1$). The strength of this correlation owes much to the GT- and AT- variants, which have unusually high Cbf1 affinities and are expressed at relatively high levels. However, there remains a substantial correlation even without these two promoters ($R = 0.58$). In contrast, there is no positive correlation whatsoever between Cbf1 affinity and expression under inducing conditions, just as there is no correlation between Pho4 affinity and expression under non-inducing conditions (Figure 4A). If anything, there is a slight negative correlation to each of these relationships. Thus Pho4 binding appears to dominate differences in gene expression when Pho4 concentrations are low (0 mM phosphate) while differences in Cbf1 binding dominate the (smaller) differences in expression when Pho4 levels are low (10 mM phosphate). The greater range of expression at 0 mM phosphate, where Pho4 dominates, than at 10 mM, PhD...
where Cbf1 dominates, implies that Pho4 may be a more potent transcriptional activator. Based on the ratio of the correlation slopes at low and high phosphate, it appears that gene expression is 5.6 times more responsive to differences in Pho4 occupancy than it is to differences in Cbf1 occupancy. For purposes of modeling gene expression, then, we interpret this to mean that Pho4 is 5.6 times better as a transcriptional activator.

As Pho4 specificity is unable to explain expression at high phosphate concentrations, and Cbf1 specificity is unable to explain expression at low phosphate concentrations, it is not surprising that there are intermediate phosphate concentrations at which neither explains the data very well (Figure 4B). However, excellent correlations were achieved between observed and predicted expression across the full range of phosphate concentrations using a very simple model (Figure 4B; Methods). In brief, (i) Pho4 and Cbf1 compete for binding, (ii) Cbf1 concentration is constant while Pho4 increases with decreasing phosphate and (iii) Pho4 and Cbf1 each activate expression, albeit with different efficiencies. The efficiency of Pho4 as a transcriptional activator was fixed at a value 5.6× greater than Cbf1 for reasons described above. Cbf1 concentration was fixed at a value of 1 relative to the mean Cbf1 affinity of the 16 sites. The concentration of Pho4 was parameterized at each phosphate concentration to maximize the correlation between predicted occupancy and expression (Figure 4B and C). Interestingly, the inclusion of Cbf1 in the model improved the correlation to expression even under fully induced conditions, where Pho4 dominates, increasing from  \( R = 0.94 \), for Pho4 only, to a value of 0.98 for Pho4 and Cbf1 together, and at the same concentration of Pho4 (Supplementary Figure S5). Both the hypothesized activation activity of Cbf1 and the increase in [Pho4] with decreasing phosphate are required to achieve a high correlation between predicted and observed expression at each phosphate concentration (Figure 4B).

Not surprisingly, the expression profiles for individual reporters show a similarly good improvement (Supplementary Figure S6). A particularly interesting example is the behavior of the CCCACGTGGG and GGCACGTCGGG reporters. These motifs are nearly identical in Pho4 affinity, so a model that uses only Pho4 predicts the concentration dependence of their expression to be nearly identical (Supplementary Figure S7). Experimentally, however, the GG-flanked motif is more active than the CC-flanked motif in high phosphate, and less active in moderate phosphate. The model that incorporates Cbf1 correctly reproduces this behavior because the higher Cbf1 affinity of the GG site makes a modest positive contributor to expression in high phosphate (low [Pho4]), but it inhibits the binding of Pho4, the more potent activator, at moderate phosphate concentrations (Supplementary Figure S7).

**Reporters with high Cbf1:Pho4 affinity ratios are unusually sensitive to low levels of induction signal**

Thus far we have shown that the expression level for 16 reporters can be fit by a model in which Cbf1 has modest activation activity and competes with Pho4 for binding.
readily in the sub-population of cells in which the both Pho4 and Cbf1 itself to access binding sites more from ‘closed’ chromatin to ‘open’ in a way that permits dissociates. Thus, Cbf1 binding could shift the equilibrium can persist even after the protein that induced that state by a sequence-specific DNA-binding protein, like Cbf1, histone modifications, chromatin states that are induced enzymes involved in chromatin remodeling (27,28). Cbf1 has been reported to have a role in recruiting chromatin to ‘open’

Cbf1 binding promotes conversion of ‘closed’
chromatin to ‘open’

Cbf1 has been reported to have a role in recruiting enzymes involved in chromatin remodeling (27,28). Thanks to positive feedback mechanisms involving histone modifications, chromatin states that are induced by a sequence-specific DNA-binding protein, like Cbf1, can persist even after the protein that induced that state dissociates. Thus, Cbf1 binding could shift the equilibrium from ‘closed’ chromatin to ‘open’ in a way that permits both Pho4 and Cbf1 itself to access binding sites more readily in the sub-population of cells in which the chromatin had been opened by a previous Cbf1-binding event. In this way, Cbf1, though competing with Pho4 for binding to the same sites, could, under certain conditions, function as if it were binding cooperatively. We modeled this process by assuming that Pho4 and Cbf1 each bind to sites in the open state with 10× higher affinity than in the closed state, that the equilibrium constant for chromatin favors the closed state by a factor of 10 and that sites occupied by Cbf1 are more likely to shift into the open state. The distribution between closed and open states was modeled by recursively calculating the fraction of each state at progressively lower concentrations of Pho4. In every other way, expression levels and the sensitivity to induction were calculated as already described. As shown in Figure 5, this modeling of Cbf1-induced chromatin opening dramatically improves the correlation between the predicted induction sensitivity and the observed. Although this does not, of course, prove that Cbf1-induced chromatin remodeling is responsible for differences in sensitivity, it does suggest that it is a plausible explanation.

Cbf1 binding appears to modulate the sensitivity of expression of natural promoters

It is well established that there are differences in the sensitivity to induction among Pho4-target genes. Lam et al. (24), using GFP fusions for seven Pho4 target genes, classified the genes into two groups on this basis. We validated this observation using a microarray gene expression data obtained at five phosphate concentrations (Figure 6A). Although the microarray and GFP reporter assays are measuring rather different things, the baseline expression

A quantitative model for high induction sensitivity: Cbf1 binding promotes conversion of ‘closed’ chromatin to ‘open’

Sensitivity to induction is of special interest because natural Pho4-regulated promoters differ in this regard, and these differences have been rationalized as being due to differences in chromatin structure (24–26). As our reporters are identical in sequence, except for the small differences flanking the CACGTG core motif, it is unlikely that will differ substantially in their intrinsic chromatin structure.

For a family of promoters with a single binding site, and which are otherwise identical, we can expect a monotonic relationship between the affinity for the binding site and the sensitivity to induction (Supplementary Figure S8). While the precise shape of this relationship depends on the concentrations of the transcription factor with respect to the affinities, the overall trend is robust to these differences. We therefore calculated sensitivity values from the experimental GFP expression values for the 16 reporter constructs and compared these values with the in vitro affinities of the binding sites (Methods). For 12 of the 16 promoters, those with sites flanked by [N][ACG], we find a strong linear correlation between Pho4 affinity and induction sensitivity ($R = 0.85$). The correlation is improved to $R = 0.92$ if, instead of using Pho4 affinity, we use predicted induction sensitivity (Figure 5). This can be done using the predicted expression levels obtained previously from the model, with no additional parameterization.

Although most of the reporters adhere well to our expectations, the four sites with T adjacent to the core have experimental sensitivity values far above those predicted from the model (Figure 5). Two of these sites, those flanked by GT and AT, have exceptionally high Cbf1 affinities. That is not the case, however, for CT and AT, which rank 11th and 13th, respectively, in Cbf1 affinity. What does distinguish all four of these sites from the other twelve is that they have exceptionally high ratios of Cbf1:Pho4 affinity. This suggests that there is some aspect of Cbf1 binding, and in particular Cbf1/Pho4 competition, that is contributing to induction sensitivity but has not yet been captured by the model.

![Figure 5. Reporters with high Cbf1: Pho4 affinity ratios show a sensitivity to induction that is anomalously high based on the simple Pho4 + Cbf1 model (X’s), but are explained well a model that assumes a role for bound Cbf1 in shifting promoter chromatin structures toward a more accessible state (circles). Sensitivities were calculated as described in the text. Colors indicate the NA (green), NC (blue), NG (yellow) and NT (red) reporters as in other figures.](https://academic.oup.com/nar/article-abstract/41/9/4877/2409613)
in high phosphate for the seven genes that were assayed in common are well correlated \((R = 0.95)\), as is the extent of their induction upon shifting to 0 mM phosphate \((R = 0.95)\) (Supplementary Figure S9). We prefer to use the term expression sensitivity rather than threshold because induction is not an all or none phenomenon, but the concepts are related, and our data can readily be interpreted to produce the same two groups defined by GFP assays (Figure 6A). In addition to the seven genes studied previously, we included in our analysis VTC4, a Pho4 target gene that shows an induction sensitivity intermediate between the two classes defined earlier (Figure 6A).
In contrast to the reporter constructs, for which the 12 [N][ACG] sites showed the expected correlation between Pho4 affinity and induction sensitivity, the natural promoters show no hint of a link between Pho4 affinity and sensitivity (Supplementary Figure S10). This lack of a correlation is consistent with the suggestion that differences in sensitivity may be dominated by differences in nucleosome occupancy at Pho4 sites (24–26). We also noted, though, that all four of the more sensitive genes have predicted occupancies for Cbf1 that are at least slightly higher than that of all the other genes (Supplementary Figure S10). Although the number of genes is too small for the predicted differences to be statistically significant, this is reminiscent of our observation that the reporter genes with high Cbf1:Pho4 affinity ratios have anomalously high induction sensitivities, and suggests that sensitivity of natural promoters might also be associated in some way with Cbf1 affinity.

To examine experimentally the relationship between expression sensitivity and transcription factor binding we performed ChIP-qPCR assays for both Pho4 and Cbf1 on all eight promoters in cells exposed to high and low phosphate (Figure 6B). As expected, Pho4 binds best in low phosphate while Cbf1 binds best in high phosphate because there is less Pho4 to compete with it. Although it is problematic to compare absolute enrichment values across different promoters because of possible differences in crosslinking efficiencies, fold-changes in enrichment can be more confidently compared (Figure 6C). We find that Pho4 generally increases in enrichment more than Cbf1 decreases, and that the changes are inversely correlated: genes that show a relatively large change in Pho4 enrichment tend to show a relatively small change in Cbf1 enrichment. This is an expected result because the total occupancy of the promoter increases under inducing conditions, when the sum of the concentrations of Pho4 and Cbf1 is highest. Perhaps less intuitive is the inverse relationship between Pho4 fold-changes and Cbf1 fold-changes. Nevertheless, this is also an expected result. This is because a high-affinity Pho4 site is further along the binding curve at low [Pho4] than a low-affinity site, and therefore will have a smaller fold-increase than the low-affinity site for the same change in [Pho4] (also, see Supplementary Figure S8); conversely, the higher occupancy of Pho4 at the high-affinity site means that there is less Cbf1 bound, so Cbf1 binding is effectively binding as if it were at a lower concentration, and thus has more room for increasing.

The more important result shown in Figure 6C is that there is a strong inverse correlation between the expression sensitivity of the promoters and the fold-change in Pho4 binding. Simulations of simple one-site promoters show that this is precisely the relationship that is expected between sensitivity and changes in transcription factor binding (Supplementary Figure S8). Thus, the microarray expression data and the Pho4 ChIP data are consistent with one another. Furthermore, simulations of simple promoters show that both the expression data and the ChIP data are consistent with high-sensitivity promoters (red) having higher effective affinities for Pho4 than do low sensitivity promoters (blue). However, as noted above, there is no correlation between induction sensitivity and the predicted affinity of promoters for Pho4. We reasoned, therefore, that there might be other features of the sequence, such as Cbf1 affinity, that modulate the effective affinities of the promoters for Pho4.

Based on the success of the chromatin remodeling model in reconciling the predicted and experimental sensitivities of the reporter constructs, we applied the same model to the eight natural promoters. Inclusion of the Cbf1 chromatin remodeling term yields a respectable correlation between predicted and experimental sensitivity values (R = 0.75) (Figure 6D). Bootstrap resampling implies that the correlation is significant (95% confidence interval: 0.39–0.95; Figure 6E), and sensitivity analyses show that the correlation is reasonably robust to the choice of parameters (Supplementary Figure S11). Furthermore, all of the terms relevant to the modeling of chromatin restructuring (i.e., open–closed equilibrium; preferential binding of Pho4 and Cbf1 to open DNA; Cbf1-mediated chromatin opening) results in correlation coefficients near zero (Figure 6E).

**DISCUSSION**

For reporter genes that contain single Pho4-binding sites, we found the maximum gene expression levels to be remarkably well correlated with the predicted occupancy of these sites by Pho4. We are not aware of another example of this type in which the affinity of so many sites, measured in vitro, have been shown to correlate so well with expression. The strength of the correlation implies that both the in vitro affinities and expression values were determined with a fair degree of accuracy. More importantly, it implies that expression in this system is an excellent proxy for transcription factor binding and that in vivo binding affinity differences are very similar to those in vitro.

The strong correlation between Pho4 affinity and expression under induced conditions is mirrored by a similarly good correlation between Cbf1 affinity and expression under non-inducing conditions. The simplest explanation for this is that Cbf1 itself has transcriptional activation activity. Excellent correlations were also achieved at intermediate conditions by modeling contributions from both Pho4 and Cbf1 in competition with each other. The one characteristic of reporter gene expression that was not well explained by this model was the anomalously high sensitivity to induction for the four reporters that have exceptionally high ratios of Cbf1:Pho4 affinity. These reporters are not well expressed under fully induced conditions, but to the degree that they are expressed, low levels of Pho4 are sufficient to achieve a substantial fraction of that expression (which is what we mean by sensitivity). We propose that high sensitivity derives from an activity of Cbf1 that leads to a remodeling of chromatin, allowing both Cbf1 itself and Pho4 to more readily bind their cognate sites. Incorporation of this feature into the model results in an excellent correlation between predicted and experimental sensitivities of the reporter constructs.
The discovery that natural Pho4 target genes can be classified into genes with low or high induction thresholds has spawned a number of articles that seek to explain the observation in terms of nucleosome occupancy and the competition of nucleosomes with Pho4 binding sites of varying affinities (24–26). More recently, Zhou and O’Shea (17) found that a cbf1Δ strain expresses Pho4 target genes at higher than wild-type levels in high phosphate. This suggests that Cbf1 binding itself could raise the threshold for induction of Pho4 target genes. Here, we have shown that we can achieve a significant correlation between predicted gene expression sensitivity values and experimental only by using the model that allowed us to reconcile the sensitivities of the 16 reporter constructs. That is, we propose that Cbf1 binds in competition with Pho4 at CACGTG motifs, that it has some transcriptional activation activity itself and that it shifts the equilibrium for nucleosome occupancy at CACGTG sites from high to low. The chromatin remodeling activity that is implied by the model’s ability to fit to the expression data is consistent with previous experimental characterizations of Cbf1 function (27,28).

It is not yet clear how much of the differences in the responsiveness to phosphate can be attributed to Cbf1 and how much is due to intrinsic nucleosome positioning, but the experiments reported here indicate that the contribution of Cbf1 binding is substantial. We suggest that at least some of the differences in Pho4 target gene sensitivity that have been attributed to differences in nucleosome occupancy are ultimately due to differences in Cbf1/Pho4 affinities. As it is very common for transcription factor family members to have related DNA-binding specificities, our analyses suggest that differential competition for binding sites, and differences in the contributions of the proteins to gene expression, could be a common mechanism for modulating the responsiveness of promoters to regulatory signals.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–11 and Supplementary Datasets 1–2.

FUNDING

Agency for Science, Technology and Research (A*STAR, Singapore); A*STAR National Science Scholarship (to J.S.Z.A.). Funding for open access charge: Operating funds from Genome Institute of Singapore.

Conflict of interest statement. None declared.

REFERENCES


