Basic leucine zipper transcription factor Hac1 binds DNA in two distinct modes as revealed by microfluidic analyses

Polly M. Fordyce[a,b], David Pincus[a], Philipp Kimmig[a], Christopher S. Nelson[a], Hana El-Samad[a,c], Peter Walter[a,b], and Joseph L. DeRisi[a,b,1]

[a]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158; [b]Howard Hughes Medical Institute, Chevy Chase, MD 20815; and [c]California Institute for Quantitative Biosciences, University of California, San Francisco, CA 94143

Edited by Kevin Struhl, Harvard Medical School, Boston, MA, and approved September 5, 2012 (received for review July 24, 2012)

A quantitative understanding of how transcription factors interact with genomic target sites is crucial for reconstructing transcriptional networks in vivo. Here, we use Hac1, a well-characterized basic leucine zipper (bZIP) transcription factor involved in the unfolded protein response (UPR) as a model to investigate interactions between bZIP transcription factors and their target sites. During the UPR, the accumulation of unfolded proteins leads to unconventional splicing and subsequent translation of *HAC1* mRNA, followed by transcription of UPR target genes. Initial candidate-based approaches identified a canonical *cis*-acting unfolded protein response element (UPRE-1) within target gene promoters; however, subsequent studies identified a large set of Hac1 target genes lacking this UPRE-1 and containing a different motif (UPRE-2). Using a combination of unbiased and directed microfluidic DNA binding assays, we established that Hac1 binds in two distinct modes: (*i*) to short (6–7 bp) UPRE-2-like motifs and (*ii*) to significantly longer (11–13 bp) extended UPRE-1-like motifs. Using a genetic screen, we demonstrate that a region of extended homology N-terminal to the basic DNA binding domain is required for this dual site recognition. These results establish Hac1 as the first bZIP transcription factor known to adopt more than one binding mode and unify previously conflicting and discrepant observations of Hac1 function into a cohesive model of UPR target gene activation. Our results also suggest that even structurally simple transcription factors can recognize multiple divergent target sites of very different lengths, potentially enriching their downstream target repertoire.

DNA specificity | microfluidics

The basic leucine zipper (bZIP) proteins form one of the largest families of eukaryotic transcription factors and play roles in a wide variety of biological phenomena, from responding to endoplasmic reticulum (ER) dysfunction to regulating immune responses and oncogenesis (1). Members of this superfamily contain a positively charged DNA binding region composed of basic residues linked to a leucine zipper sequence and homo- or hetero-dimerize via this leucine zipper. Invariant arginine and asparagine residues within the basic DNA binding region (NXAAXXCR) make direct contact with DNA bases within the major groove and drive binding specificity to palindromic or semipalindromic target sites (2, 3). Although considered to be the simplest known protein-DNA recognition motif, crystal structures of bZIP domains bound to DNA have revealed functional variability in how these conserved residues contact DNA (2, 4), and no universal code linking basic region sequence with target DNA preferences has been developed.

Here, we investigate the mechanisms that drive bZIP target site recognition using Hac1, a *Saccharomyces cerevisiae* transcription factor involved in the highly conserved unfolded protein response (UPR). During the UPR, cells sense an accumulation of unfolded proteins within the endoplasmic reticulum (ER) and trigger a transcriptional upregulation of genes encoding ER-resident chaperones and protein modifying enzymes, components of ER-associated protein degradation (ERAD), and enzymes for phospholipid biosynthesis (5). In *S. cerevisiae*, two main proteins are responsible for enacting the UPR: Ire1, a transmembrane kinase/endonuclease, and Hac1. Unfolded proteins bind to the Ire1 domain facing the ER lumen, triggering its oligomerization and activation of its cytoplasmic endonuclease domain. Once activated, Ire1 cleaves *HAC1* mRNA at two sites and tRNA ligase rejoins the severed exons via an unconventional spliceosome-independent mechanism (5). This splicing removes an intron to produce a new transcript (denoted $HAC1^i$ mRNA; "i" for "induced"), thereby relieving translational inhibition exerted by the intron. Following translation of the spliced mRNA, $Hac1^i$ is translocated to the nucleus, where it regulates a large set of UPR-responsive genes (6). Despite the central role played by $Hac1^i$ in activating the UPR, the rules by which $Hac1^i$ recognizes UPR target genes remain unclear.

Initial studies took a candidate-based approach to identify potential $Hac1^i$ binding sites within the promoters of known UPR target genes. Analysis of the promoter of *KAR2*, encoding the major Hsp70-type ER-resident chaperone Kar2 (or BiP), revealed a 22-bp *cis*-acting unfolded protein response element required for induction of UPR-dependent *KAR2* transcription (here referred to as UPRE-1) (7, 8). Subsequent transcriptional activity assays identified a core 7-bp consensus (5′-*CAGNGTG*-3′; here referred to as cUPRE-1), in which point mutations of palindromic half sites (6 conserved bp) or changes in the half-site spacing severely reduced activity (9). Gel shift assays demonstrated direct binding of $Hac1^i$ to the 22-bp UPRE-1, and reporter gene assays confirmed that this element was sufficient to confer UPR-responsive transcriptional activity in an otherwise silent promoter (9, 10). UPRE-1-like motifs were also found in the promoters of four additional known UPR target genes (*PDI1*, *EUG1*, *FKB2*, and *LHS1*), lending support to its proposed role (11–13).

This central role for UPRE-1 in upregulating target gene transcription was subsequently called into question by a study employing genome-wide microarray expression profiling to identify all candidate UPR target genes (14). This work identified 381 candidate target genes, representing nearly 5% of all open reading frames in the *S. cerevisiae* genome and encoding numerous

Author contributions: P.M.F., D.P., P.K, C.S.N., H. E.-S., P.W., and J.L.D. designed research; P.M.F., D.P., P.K., and C.S.N. performed research; P.M.F. analyzed data; and P.M.F., D.P., P.K., C.S.N., H. E.-S., P.W., and J.L.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: joe@derisilab.ucsf.edu.

See Author Summary on page 18253 (volume 109, number 45).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212457109/-/DCSupplemental.

proteins required in the ER, the Golgi apparatus, and throughout the secretory pathway. Bioinformatic analysis of the promoter regions of these genes revealed that although most lacked the canonical UPRE-1, many contained one or more of two alternate motifs (UPRE-2, 5′-*TACGTG*-3′; UPRE-3, 5′-*AGGACAAC*-3′) capable of driving Hac1^i-mediated transcription in reporter assays. Surprisingly, this analysis failed to recover the known UPRE-1 site (15). To account for the target site variety, it was proposed that Hac1^i bound to these alternate sites via heterodimerization with Gcn4. Further complicating the picture, a study using protein binding microarrays (PBMs) to probe Hac1^i binding preferences among all possible 8-bp nucleotide sequences revealed binding only to UPRE-2 (16).

In vivo studies of Hac1^i are complicated by both the very short half-life of the Hac1^i isoform derived from the spliced mRNA and the tendency of bZIP transcription factors to homo- and heterodimerize. Therefore by necessity, in vitro approaches provide a particularly valuable tool for accurately defining binding preferences. Here, we probe how Hac1^i regulates expression of target genes using microfluidic affinity analysis [MITOMI (17) and MITOMI 2.0 (18)] to identify and characterize Hac1^i target sites. In addition, we analyze expression of reporter genes driven by a variety of Hac1^i mutants to identify the protein residues required for target site recognition. Our results resolve the previous conundrum regarding Hac1^i binding behavior to provide an integrated model of UPR target gene regulation and shed new light on the basic mechanisms by which bZIP transcription factors recognize their target genes.

Results

Experimental Setup for Measuring Hac1^i Binding. To obtain an unbiased assessment of Hac1^i binding preferences, we used a microfluidic platform, MITOMI 2.0 (18), to measure relative binding affinities (ΔΔG) between Hac1^i and 70-bp double-stranded oligonucleotides containing overlapping instances of all possible 8 bp combinations (Fig. 1*A*). In previous work, we validated this platform using a panel of 28 *S. cerevisiae* transcription factors and demonstrated the ability to quantitatively measure relative binding affinities to each oligonucleotide and recover known binding preferences (18).

In brief, each MITOMI 2.0 device contained 4,160 chambers composed of two compartments ("DNA" and "protein") controlled by three valves ("neck," "sandwich," and "button") (Fig. 1*B*). Experiments took place in six main steps (Fig. 1*C*): (*i*) DNA compartments were programmed with specific Cy5-labeled double-stranded DNA sequences by aligning devices to a spotted DNA microarray; (*ii*) BODIPY-FL-labeled His-tagged Hac1^i was flowed across the protein compartments and recruited to

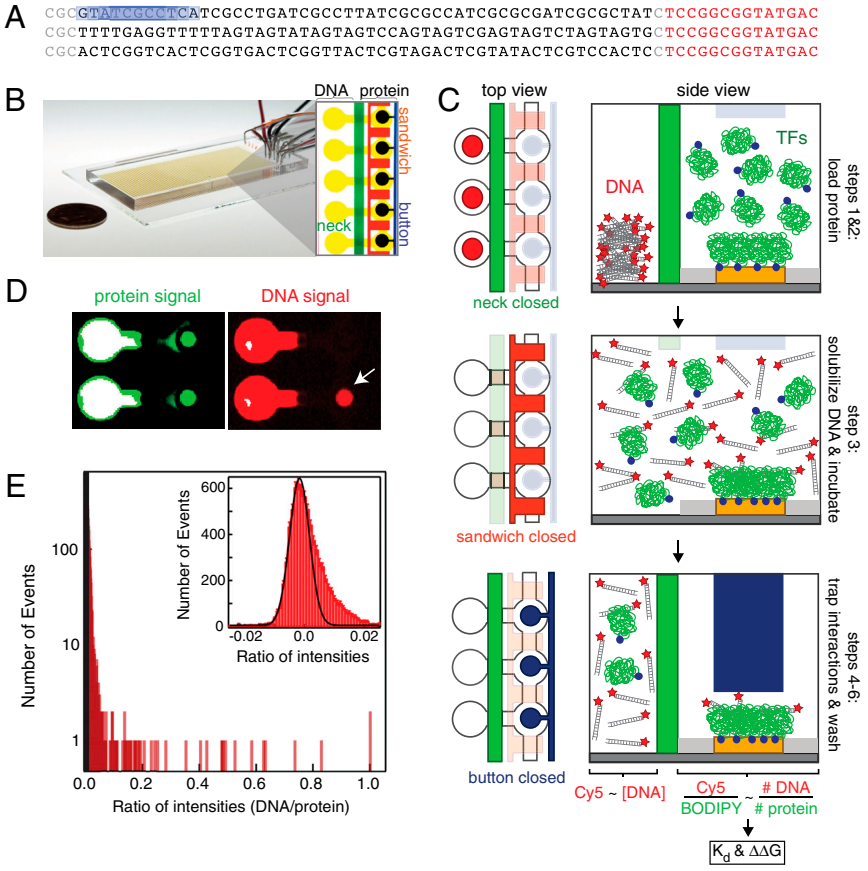

Fig. 1. MITOMI 2.0 experimental geometry. (*A*) Three example oligonucleotide library sequences illustrating sequence structure. All sequences contain a "*CGC*" clamp (gray text), a central variable region composed of overlapping 8-nt candidate binding sites (black text), a "*C*" spacer (gray text), and an identical 14-bp sequence (red text) for hybridization and extension of a universal Cy5-labeled oligonucleotide to create dsDNA. Transparent blue boxes show four potential 8-mer binding sites. (*B*) Photograph of 4,000 unit cell device with a penny for scale (left); close-up view of five individual unit cells (right) showing protein and DNA chambers (yellow), "neck" valve (green), "sandwich" valve (orange), and "button" valve (blue). (*C*) Schematic showing top and side views of experimental chambers at different points during the experiment. (*D*) Fluorescence scans showing final Cy5 (DNA, red) and BODIPY-FL (protein, green) intensities in DNA and protein chambers; white arrow highlights DNA recruited by surface-immobilized Hac1^i beneath the button valve. (*E*) Histogram of measured fluorescence intensity ratios (Cy5/BODIPY-FL) on a log-linear scale to highlight outliers; the thick black vertical bar near the *y*-axis denotes four standard deviations above the background mean. Inset: Zoomed view of background events on a linear scale with a Gaussian fit (black) to the background distribution ($\chi^2 = 1.19$, $p = 1.0$).

surfaces beneath button valves that were coated with anti-His antibodies; (*iii*) protein solution was pushed into DNA compartments, solubilizing spotted DNA and allowing Hac1^i and DNA sequences to interact; (*iv*) binding interactions were mechanically trapped at equilibrium by pressurizing button valves to squeeze out unbound material; (*v*) neck valves were closed to isolate the compartments and allow washing away of unbound material in the protein compartment while preserving equilibrium concentrations of both binding partners in the DNA compartment; and (*vi*) devices were read using a fluorescence scanner. Final Cy5 intensities in each DNA chamber were previously shown to be proportional to the soluble DNA concentration available for binding (17, 18), and the ratio of Cy5 (DNA) to BODIPY-FL (Hac1^i) intensities beneath the button valve reports the protein fractional occupancy, allowing calculation of interaction K_d and ΔΔG (Fig. 1*C*).

We measured Hac1^i binding in two independent experiments. In both cases, Hac1^i showed strong preferences for particular sequences (Fig. 1 *D* arrow and *E* and *SI Appendix*, Fig. S1*A*), with Z scores of approximately 85 for the highest affinity sequences (Fig. 1*E* and *SI Appendix*, Fig. S1*A*). Measurements were fairly reproducible both between replicates within a given experiment (Pearson $r^2 = 0.73$ and $r^2 = 0.77$; *SI Appendix*, Fig. S1*B*) and between experiments performed on different days (Pearson $r^2 = 0.51$, *SI Appendix*, Fig. S1*C*); therefore, we pooled results from both experiments for further analysis.

MITOMI 2.0 Analysis Predicts Hac1^i Binding Primarily to UPRE-2. Each 70-bp oligonucleotide contained multiple potential overlapping Hac1^i binding sites (Fig. 1*A*); consequently, additional analysis was required to deconvolve results and identify the target sites responsible for Hac1^i binding. First, we used fREDUCE (19) to search for 6-, 7-, and 8-bp motifs whose appearance within oligonucleotides correlated most strongly with their measured intensity ratios. Surprisingly, all searches exclusively returned variants of UPRE-2, with strong correlations between the appearance of this motif and observed intensity values (Fig. 2*A* and *SI Appendix*, Fig. S2 and Table S1). We then assessed the effects of single nucleotide substitutions in this consensus site on ΔΔG by using MatrixREDUCE (20, 21) to generate a position-specific affinity matrix (PSAM). Importantly, PSAMs can be used to predict binding to any sequence quantitatively, and comparisons between predicted binding profiles and measured binding profiles yield additional information: In particular, oligonucleotides bound more strongly than predicted would indicate binding to additional motifs, while oligonucleotides bound more weakly would indicate motifs that repel binding. In our data, comparisons between predicted and measured binding showed strong agreement, suggesting that Hac1^i bound nonpromiscuously to UPRE-2 in vitro and displaying no evidence for binding to additional sequences present in the oligonucleotide library (Fig. 2*A* and *SI Appendix*, Fig. S2).

Hac1^i Binds the UPRE-2 but Not the cUPRE-1. It poses a paradox that our microfluidic affinity assay data and previous PBM experiments have failed to uncover evidence of UPRE-1 binding, which was well validated in previous studies (7, 9, 10). This failure could be explained because either Hac1^i does not bind to the cUPRE-1 but requires a longer sequence that is not represented in our library or by an insufficient sensitivity of the MITOMI 2.0 assay to pick up low-affinity cUPRE-1 binding.

To distinguish between these possibilities, we directly measured concentration-dependent binding of Hac1^i to a series of oligonucleotides containing either the cUPRE-1 or the UPRE-2 embedded within random sequence (Fig. 2*B*). In three separate experiments, we observed high-affinity binding to the oligonu-

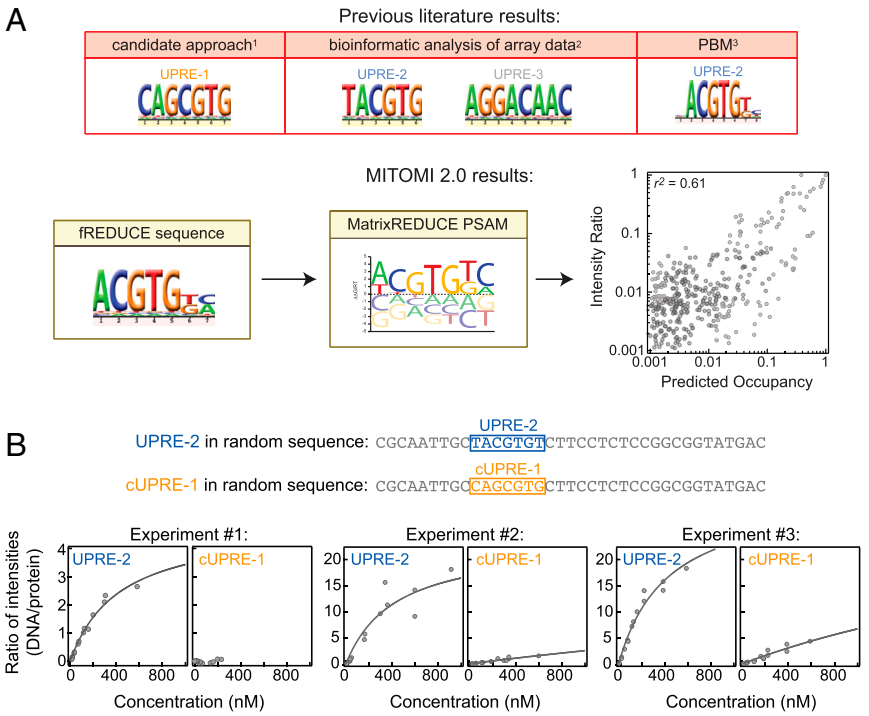

Fig. 2. Hac1^i target binding sites revealed by MITOMI 2.0 microfluidic affinity analysis using an 8-mer oligonucleotide library. (*A*) Top row: Previously published motifs determined via a candidate-based approach (UPRE-1) (9); bioinformatic analysis of promoters associated with genes upregulated during the UPR (UPRE-2 and UPRE-3) (15); and in vitro protein binding microarray experiments (UPRE-2) (16). Bottom row, left: 7-bp sequence whose appearance within oligonucleotide sequences correlates most strongly with measured intensity ratios, as determined using fREDUCE (19). Bottom row, middle: PSAM for this sequence determined using MatrixREDUCE (21). Bottom row, right: Comparison between measured and predicted binding. (*B*) Measured fluorescence intensity ratios (gray circles) as a function of soluble DNA concentration for UPRE-2 (blue box) and cUPRE-1 (orange box) embedded within standard MITOMI library sequence (gray text) in three separate experiments.

cleotide containing the UPRE-2, with no measurable binding above background levels to the oligonucleotide containing the cUPRE-1 (Fig. 2*B*). Fits to the UPRE-2 binding data yielded a K_d of 427 ± 37 nM, similar to values we previously obtained for other bZIP transcription factors (18).

Reporter assays have suggested that the central *C* within the cUPRE-1 can be replaced by alternate nucleotides with only a slight reduction in activity (9). We therefore assessed binding to these variants to see if any of the variations restore binding. For all variants, binding remained at the level of random sequence over multiple experimental replicates (*SI Appendix*, Fig. S3).

Hac1^i Binding to UPRE-1 Requires an Extended Target Site. Previous work showed that the cUPRE-1 is necessary for transcriptional activity (9). However, it was never shown to be sufficient, and phylogenetic comparisons suggest that cUPRE-1 flanking sequences are important for Hac1^i binding. UPRE-1 sites from the promoters of multiple known Hac1^i targets (*KAR2*, *EUG1*, *PDI1*, *FKB2*, and *LHS1*) show conservation of several nucleotides both upstream and downstream from the 7-bp core, even as the core is imperfectly conserved (11–13) (Fig. 3*A*). The same pattern is also seen for UPRE-1 sites within the promoters of *KAR2* orthologs from distant species (Fig. 3*B*).

To test whether flanking sequences are critical for Hac1^i binding, we measured concentration-dependent binding for the cUPRE-1 embedded within either a fragment of the *KAR2* promoter or within the *ERO1* promoter, which typically contains a UPRE-2-like motif (Fig. 3*C*). Consistent with the notion that

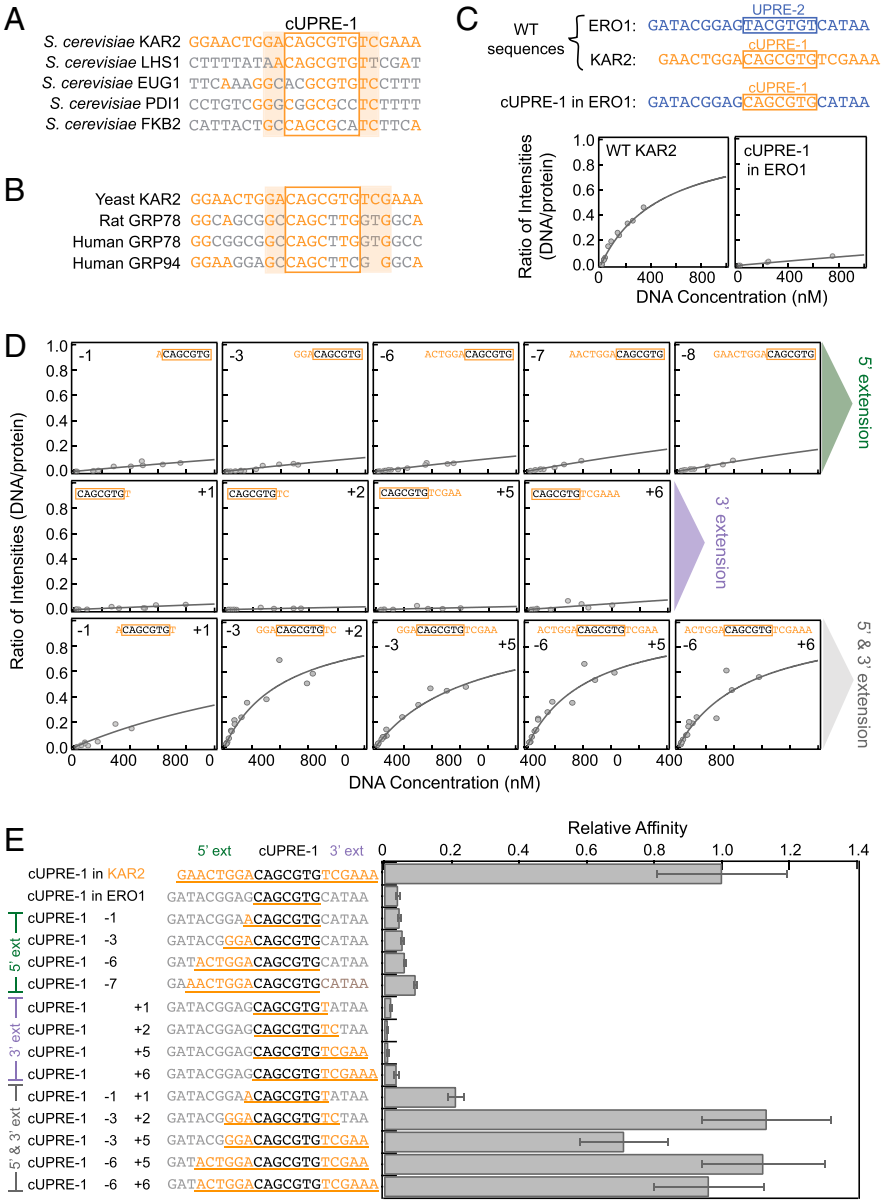

Fig. 3. Efficient binding of Hac1^i to cUPRE-1 requires an additional 2–3 nucleotides both upstream and downstream from the 7-bp core. (*A*) Alignments of known Hac1^i target promoters containing UPRE-1 variants. Conserved 5′ and 3′ flanking nucleotides are indicated by orange text and shading; imperfectly conserved nucleotides within the cUPRE-1 are indicated by light gray text. (*B*) Alignments of *KAR2* ortholog promoters. Conserved 5′ and 3 flanking nucleotides are indicated by orange text and shading. (*C*) Fluorescence intensity ratios as a function of DNA concentration for Hac1^i binding to the cUPRE-1 (orange box) in the context of either cUPRE-1-associated *KAR2* promoter sequence (orange text) or orthologous UPRE-2-associated *ERO1* (blue text) promoter sequence. (*D*) Fluorescence intensity ratios as a function of DNA concentration for DNA constructs including increasing portions of 5′ *KAR2* promoter flanking sequence (top row), 3′ *KAR2* promoter flanking sequence (middle row), and both 5′ and 3′ *KAR2* promoter flanking sequence (bottom row). (*E*) Bar chart showing relative binding affinities for different constructs; error bars represent errors from global fits to a single-site binding model.

cUPRE-1 flanking sequences are required, addition of the *KAR2* flanking sequences to cUPRE-1 restored high-affinity binding (Fig. 3*C*). Insertion of the cUPRE-1 into a heterologous flanking context (the *ERO1* promoter) did not restore Hac1^{i} binding (Fig. 3*C*), establishing that UPRE-1-specific flanking sequences are required.

To identify the precise boundaries of flanking sequences required for Hac1^{i} binding to the cUPRE-1, we started with the cUPRE-1 embedded within heterologous *ERO1* flanking sequences and systematically substituted these sequences with increasing portions of *KAR2* sequences in the upstream and/or downstream direction (Fig. 3*D*). Restoration of upstream or downstream flanking sequences alone did not significantly increase Hac1^{i} binding affinity to cUPRE-1 (Fig. 3*D*, top and middle rows, respectively). By contrast, simultaneous addition of both upstream and downstream *KAR2* flanking sequences had a strong effect on binding (Fig. 3*D*, *Bottom* row).

Addition of one nucleotide on either side of the cUPRE-1 increased affinity fivefold, and addition of two nucleotides on either side of the core restored affinity to that measured for the cUPRE-1 in its native *KAR2* context (Fig. 3*E*). Inclusion of additional *KAR2* flanking sequence did not significantly alter binding affinities, suggesting that the 12-bp sequence 5′-*GGACAGCGTGTC*-3′ (hereafter termed the extended core UPRE-1, or xcUPRE-1) is sufficient for Hac1^{i} binding. Further corroboration of the importance of these nucleotides comes from the observation that single point mutations 2-bp upstream and downstream from the cUPRE-1 in the *KAR2* promoter previously caused a reduction in transcriptional activity (7, 9).

Systematic Mutation of xcUPRE-1 and UPRE-2 Target Motifs Reveals Two Distinct Target Motifs. Understanding precisely how xcUPRE-1 and UPRE-2 differ requires a comprehensive map of individual nucleotide preferences for each binding mode. To create such a map, we measured concentration-dependent binding curves for systematic substitutions of all possible nucleotides at each position within both targets (Fig. 4*A* and *SI Appendix*, Table S2). In each case, we performed 3–4 experimental replicates (*SI Appendix*, Figs. S4–S7) and computed the average relative binding affinity for each substitution (Fig. 4*A*). From these relative affinities, we then computed an average PSAM (Fig. 4*B* and *SI Appendix*, Tables S3 and S4 and Fig. S8) for each motif.

The relative nucleotide preferences for UPRE-2 derived from these measurements agree well with those from our MITOMI 2.0 analysis (compare Fig. 4*B* and Fig. 2*A*). Taken together, these results establish that the complete UPRE-2 is short, subtending 6–7 nucleotides, with little degeneracy tolerated at most positions. The UPRE-2 appears to be an imperfect palindrome: attempts to create a more fully symmetric site by either adding a 5′ C (Fig. 4*A*) or by altering multiple nucleotides (*SI Appendix*, Fig. S9) do not lead to statistically significant increases in binding affinity. By contrast, mutations at nearly all positions within

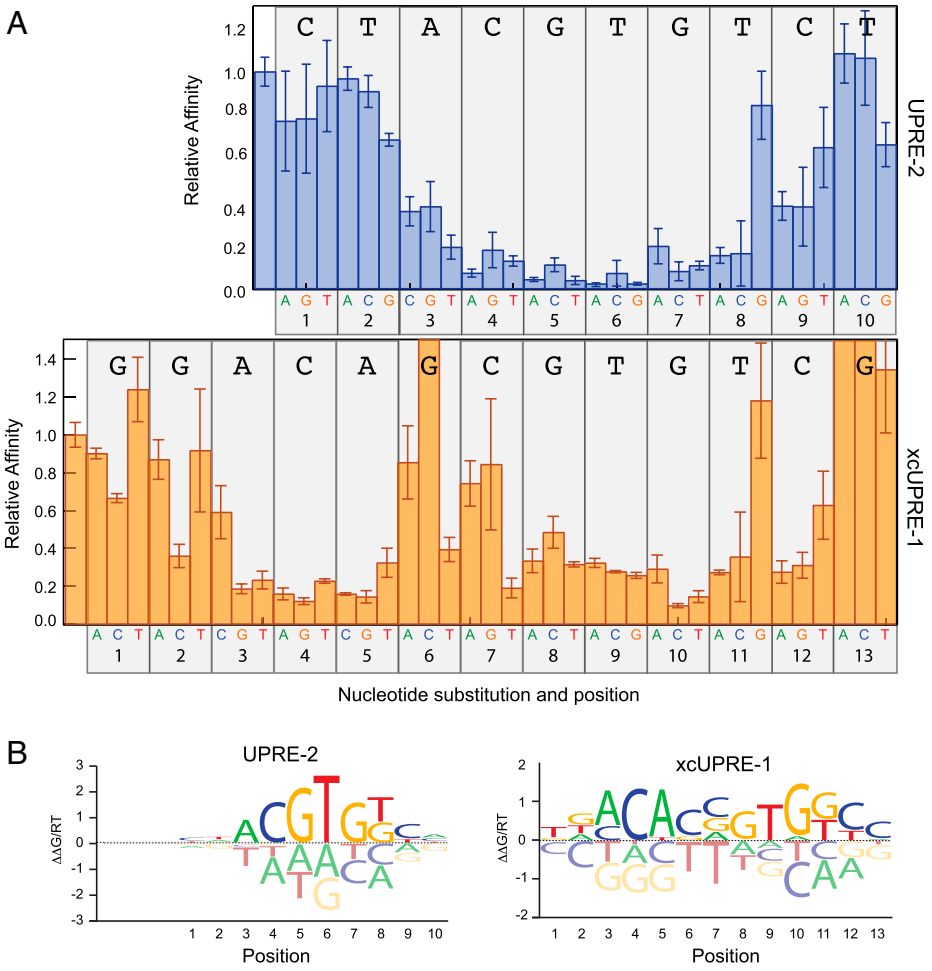

Fig. 4. Maps of nucleotide binding preferences at each position within xcUPRE-1 and UPRE-2. (*A*) Measured relative binding affinities for all possible single nucleotide substitutions at each position within both UPRE-2 (top, blue bars) and xcUPRE-1 (bottom, orange bars) sites. Values represent the average affinity for each substitution relative to the previously reported canonical sequence (shown at top) measured over multiple replicates; errors reported are the standard error on the mean. Canonical UPRE-2 and xcUPRE-1 sequences are from the *ERO1* and *KAR2* promoters, respectively. (*B*) Affinity Logos[21] for UPRE-2 (left) and xcUPRE-1 (right) PSAMs derived from relative affinities.

xcUPRE-1 have significant effects on affinity, further confirming that xcUPRE-1 subtends on the order of 11–12 bp (Fig. 4*A* and *B*). In addition, the overall composition of the motif is different: xcUPRE-1 appears to be composed of two palindromic dyad repeats (5′-*G[A/C]CAC*-3′) separated by a central degenerate nucleotide.

The absolute affinity for UPRE-2 in these experiments was slightly higher than the absolute affinity for xcUPRE-1 (UPRE-2 $K_d = 497 \pm 60$ μM; xcUPRE-1 $K_d = 720 \pm 80$ μM) (*SI Appendix*, Table S5). However, the range of affinities measured for all UPRE-2 and xcUPRE-1 variants largely agree, with the strongest binding measured to be 220 ± 30 μM and 360 ± 40 μM, respectively (*SI Appendix*, Table S5).

Prediction of Potential Genomic Targets Using xcUPRE-1 and UPRE-2 PSAMs. An advantage of PSAMs over position weight matrices (PWMs) is that they allow de novo prediction of protein binding affinities to arbitrary sequences. To test the performance of our xcUPRE-1 and UPRE-2 PSAMs, we compared measured and predicted Hac1^i binding affinities for a variety of genomic UPREs, including those present within the *ERO1*, *KAR2*, *EUG1*, *LHS1*, *FKB2*, *SEC66*, and *PDI1* promoters (*SI Appendix*, Fig. S10). Measured and predicted affinities showed relatively strong agreement ($r^2 = 0.64$, $p = 0.03$), confirming our ability to accurately predict binding to novel sequences. Next, we calculated predicted binding affinities to all annotated promoters within the yeast genome as well as to known UPR target genes (14) (*SI Appendix*). Promoters predicted to be bound with high affinity via UPRE-2-like binding were more likely to be present within the original UPR-induced data set (14) (*SI Appendix*, Table S6), with the top 20 hits including multiple known UPR targets (*ULI1*, *TRA1*, *SFB3*, *MCD4*, *SNF11*, *HNT1*, and *KIC1*). Known UPR target promoters included 10, 19, and 20 genes with predicted affinities within the top 10% of the distribution for the xcUPRE-1 PSAM alone, the UPRE-2 PSAM alone, or both PSAMs, respectively.

To test the ability of binding affinities measured in vitro to predict in vivo transcriptional response, we compared levels of expression of green fluorescent protein (GFP) in two *S. cerevisiae* strains following addition of dithiothreitol, which impairs the formation of disulfide bonds and leads to induction of the UPR. In one strain, GFP expression was driven by a synthetic promoter containing four repeats of the full *KAR2* UPRE-1 (14); in the second strain, cUPRE-1 sequences were replaced by UPRE-2 sequences, resulting in higher measured in vitro affinities (*SI Appendix*, Fig. S11). In both strains, basal GFP expression was low and identical. Following induction of the UPR, GFP levels in the strain containing the UPRE-2 substitutions were approximately twofold higher (*SI Appendix*, Fig. S11), establishing that changes in affinity measured here predict target promoter activity in vivo.

A Region Of Extended Homology N-Terminal to Basic Region Is Required for Dual Site Recognition. Given that xcUPRE-1 and UPRE-2 differ significantly both in their overall length and relative nucleotide compositions (Fig. 4*A*), recognition of each motif must accommodate distinct arrangements of contacts between Hac1^i and target site nucleotides. If this is the case, it should be possible to create Hac1^i mutants that disrupt binding to one site while largely preserving binding to the other. In the Maf subfamily of bZIP transcription factors, a region of extended homology positioned N-terminal to the basic DNA binding domain is critical for recognition of extended (13–14 nucleotide) target sites (4). Phylogenetic alignment of Hac1 orthologs across ascomycetes reveals a similar region of extended homology (*SI Appendix*, Fig. S12), suggesting that these residues may be important for DNA specificity.

To identify mutants with altered binding preferences, we used a genetic screen to assess levels of binding via each mode. To do so, we used error-prone PCR to generate a library of Hac1^i constructs containing random mutations to the protein between the N terminus and the first heptad repeat of the leucine zipper (Fig. 5*A* and *SI Appendix*, Fig. S13). We transformed this library into a yeast strain containing two synthetic promoters controlling the expression of two fluorescent proteins. The first promoter consists of four repeats of the *KAR2* UPRE-1 motif driving mApple expression, and the second consists of four repeats of the UPRE-2 driving GFP expression (Fig. 5*B*). To generate the 4x-UPRE-2 promoter, we mutated two nucleotides within the *KAR2* UPRE-1 to create a UPRE-2 target site (Fig. 5*B*). Importantly, the PSAMs derived here (Fig. 4*B* and *SI Appendix*, Tables S3 and S4) predict that these mutations are sufficient to switch Hac1^i binding toward the UPRE-2 recognition mode (*SI Appendix*, Fig. S14). To ensure that differences in fluorescence intensity were due to changes in Hac1^i binding and not indirect effects from other UPR components, we ectopically expressed both wild-type Hac1^i and this mutant library using an estradiol-inducible system (22) (*SI Appendix*, Fig. S15).

Using this approach, we identified multiple Hac1^i mutants with altered levels of binding to either one or both target promoters relative to wild-type constructs (Fig. 5*C* and *SI Appendix*, Figs. S16 and S17). Reassuringly, constructs sharing a given mutation displayed the same fluorescence phenotype (*SI Appendix*, Fig. S17 and Tables S7 and S8). Most constructs with altered binding retained the ability to recognize xcUPRE-1 even as UPRE-2 recognition was impaired; this tendency could reflect the fact the xcUPRE-1 recognition appears to take place via both binding modes (*SI Appendix*, Fig. S14) or could simply be due to the increased length and tolerance of degeneracy within xcUPRE-1 (Fig. 4 *A* and *B*). Mutations in positively charged arginines or lysines within the extended homology region or near the N terminus of the basic DNA binding region preferentially reduced UPRE-2 binding while maintaining xcUPRE-1 binding (Fig. 5*D*). Interestingly, a single arginine within the basic region plays a crucial role in xcUPRE-1 recognition (Fig. 5*D*). The diversity of these binding phenotypes and their emergence from individual mutations strongly argues that Hac1^i binds DNA via distinct binding modes, with individual protein residues playing different roles within each interaction.

An N-Terminal Truncation Mutant Lacking Extended Homology Regions Binds UPRE-2-Like Sequences with Reduced Affinity. To further probe this idea and test the notion that residues within the Hac1^i extended homology region are required for UPRE-2 recognition, we created Hac1^i constructs with truncations at different locations within the extended homology region and mapped their xcUPRE-1 and UPRE-2 binding preferences using microfluidic affinity analysis. One truncation mutant (N25) retained three residues identified as being important for UPRE-2 binding, while the other truncation mutant (N35) lost these residues (Fig. 6*A*). Although comparisons between relative binding affinities for nearly full-length (N10) Hac1^i and the N25 truncation mutant showed strong agreement ($r^2 = 0.90$; Fig. S18*A*), similar comparisons between nearly full-length (N10) Hac1^i and the N35 truncation mutant showed much weaker agreement ($r^2 = 0.40$; *SI Appendix*, Fig. S18*B*), suggesting a change in binding preferences. Calculation of the difference in binding preference relative to average xcUPRE-1 and UPRE-2 behaviors for each construct reveals that although all constructs show similar binding preferences for oligonucleotides containing xcUPRE-1 and single-site substitutions (Fig. 6*B*), the N35 construct shows dramatically reduced binding for UPRE-2 and single-site substitutions (Fig. 6*C*). In particular, the N35 UPRE-2 PSAM shows a decreased tolerance for nucleotide substitutions at the 5′ end of

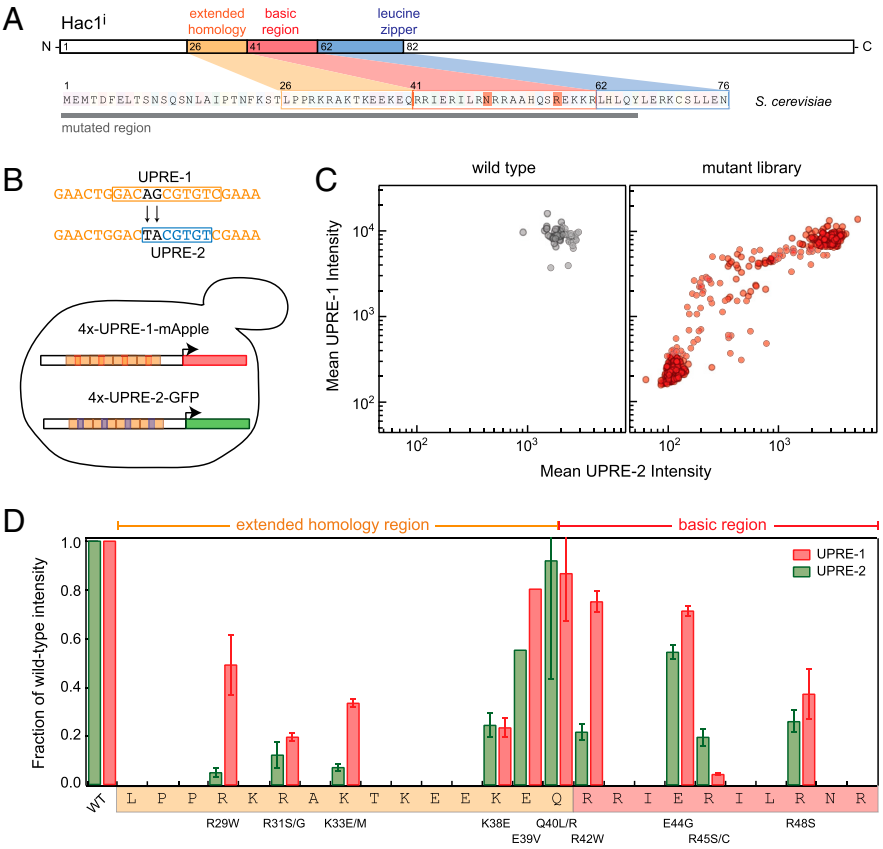

Fig. 5. Mutations within the Hac1^i DNA binding domain and an N-terminal region of extended homology can disrupt two-mode binding. (*A*) N-terminal region of Hac1^i protein sequence, including proposed extended homology region (orange), basic DNA binding region (pink), beginning of leucine zipper region (blue), and region mutated via error-prone PCR (gray bar). (*B*) Top: Details of mutation of *KAR2* UPRE-1 site to generate a UPRE-2 target site. Bottom: Yeast strains used in flow cytometry assays contained (*i*) a reporter with mApple expression and driven by four repeats of the *KAR2* UPRE-1, (*ii*) a reporter with GFP expression driven by four repeats of UPRE-2 within *KAR2* flanking sequences, and (*iii*) either wild-type *HAC1*^{i} or mutant *hac1*^{i} under the control of the Gal1 promoter within an estradiol-inducible system. (*C*) Measured mean UPRE-1-driven (mApple) and UPRE-2-driven (GFP) intensities for wild-type Hac1^i (left) and mutant Hac1^i (right) 4 h after addition of estradiol. Each individual point represents the mean fluorescence intensity in each channel for a population of yeast cells grown from a single clone. (*D*) UPRE-1-driven (mApple) and UPRE-2 driven (GFP) intensities as a fraction of wild-type intensity for sequenced Hac1^i populations sharing a given mutation; bars show average values for all populations with a specific mutation. Error bars display standard error on the mean; measurements without error bars were derived from a single population.

the motif, suggesting a shift towards a more extended binding site (Fig. 6*C* and *SI Appendix*, Fig. S19).

The N25 and N35 truncation mutants showed twofold and 10-fold decreases in overall binding affinities, respectively (*SI Appendix*, Fig. S20). As a result, mapping N35 binding preferences required that experiments be performed at fourfold higher DNA concentrations to accurately measure affinities. Comparisons between relative affinities measured for the N10 construct at both concentrations showed good agreement ($r^2 = 0.76$; Fig. S18*C*), signifying that changes in binding preferences do not result merely from changes in experimental conditions. These results lend additional support to the idea that residues within the extended homology region are required for dual-mode binding of Hac1^i to target sites.

Discussion

Here, we show that Hac1^i binds two divergent DNA binding sites, a compact 6- or 7-bp UPRE-2 site and a significantly longer 11- or 12-bp xcUPRE-1 site. While the compact UPRE-2 appears to be a slightly asymmetric half-site, the xcUPRE-consists of two palindromic 5′-*GJA/CJCAC*-3′ dyad repeats separated by a central bp that is relatively degenerate (Fig. 4*A*), with mutations at this position having little effect on transcriptional activity (11). These differences in both site length and nucleotide composition suggest that Hac1^i must contact each site via distinct modes of binding. In support of this hypothesis, mutational analysis reveals that a region of extended homology N-terminal to the basic DNA binding domain is required for Hac1^i dual site recognition, and microfluidic affinity analysis confirms the importance of these residues for UPRE-2 recognition. Based on these conclusions, Hac1^i emerges as the first natural bZIP transcription factor shown to operate in at least two different modes.

The idea that the xcUPRE-1 subtends 11–12 bp is supported by multiple previous observations. Although necessary for UPR-responsive transcriptional activation, the 7-bp cUPRE-1 was not sufficient for activation, and mutations in flanking nucleotides outside of this core motif caused severe reductions in reporter assay activity (7, 9). Such a long recognition sequence may also explain the prior failure of short word-based bioinformatic analysis of promoters to recover this motif from known UPR target genes (15). Our results are therefore consistent with previous observations and clarify our understanding of Hac1^i function.

Several arguments suggest that the binding observed here reflects the behavior of Hac1^i alone and not of Hac1^i heterodimers, as previously proposed (15). In our experiments, 6x-His tagged Hac1^i was produced in an in vitro translation system that was then flowed over a surface coated with anti-His antibodies, effectively concentrating and purifying Hac1^i on-chip prior to affinity measurements. We consider it likely that Hac1^i produced in this manner exists as an equilibrium of monomeric and homodimeric species. In addition, the shapes of the concentration-dependent binding curves suggest that both xcUPRE-1 and

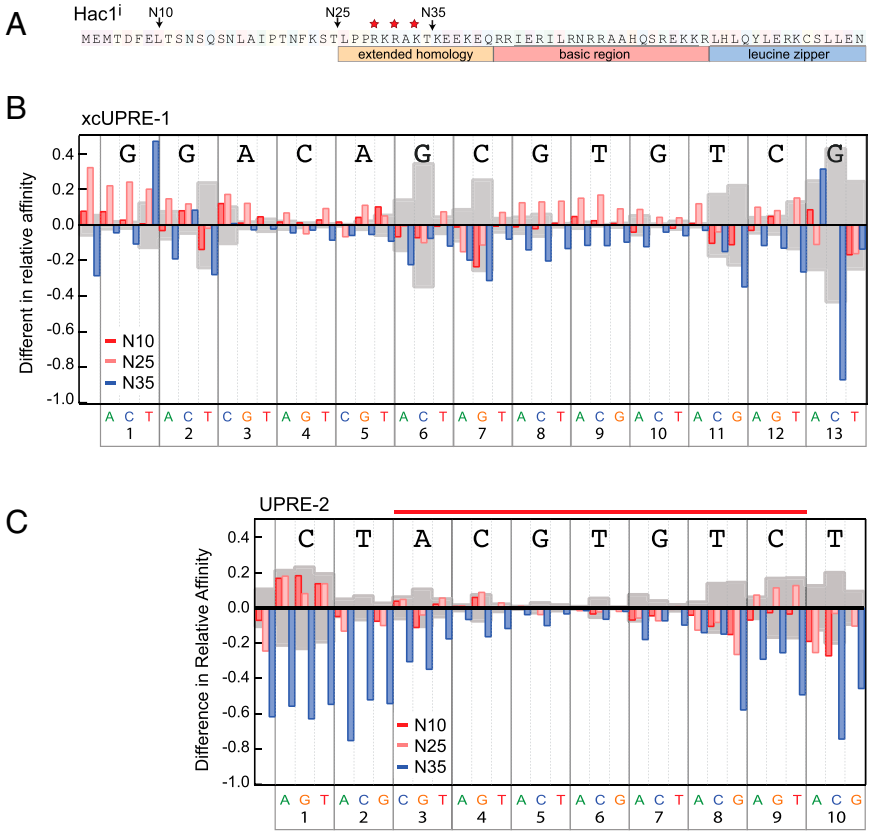

Fig. 6. Microfluidic affinity analysis of Hac1^i N-terminal truncation mutants. (*A*) Schematic showing truncation points for wild-type Hac1^i (N10) and mutants either retaining (N25) or lacking (N35) three residues identified as being important for UPRE-2 binding in the genetic screen (red stars). (*B*) Differences in measured relative binding preferences for N10 (red bars), N25 (pink bars), and N35 (blue bars) Hac1^i constructs as compared with average xcUPRE-1 binding preferences (Fig. 4*A*, bottom, and *SI Appendix*, Table S3*A*) for *KAR2* xcUPRE-1 and single-nucleotide substitutions. Shaded gray area shows standard deviation from the mean for each oligonucleotide over three experimental replicates. (*C*) Differences in measured relative binding preferences for N10 (red bars), N25 (pink bars), and N35 (blue bars) Hac1^i constructs as compared with average UPRE-2 binding preferences (Fig. 4*A*, top and *SI Appendix*, Table S4*A*) for *ERO1* UPRE-2 and single-nucleotide substitutions. Red bar indicates core UPRE-2 target site (as defined in Figs. 2 and 4); shaded gray area shows standard deviation from the mean for each oligonucleotide over four experimental replicates.

UPRE-2 motifs are bound by Hac1^i complexes with identical stoichiometries: All curves in a given experiment asymptote to an identical fluorescence intensity ratio, establishing that the number of DNA molecules bound per labeled protein molecule remains constant. The notion that Hac1^i binds as a homodimer is further supported by detection of UPRE-2 binding in PBM experiments employing Hac1^i proteins expressed in an *Escherichia coli* system that does not contain potential orthologous binding partners (16, 23). Moreover, gel shift assays performed in yeast extracts showed indistinguishable shifts for Hac1^i bound to oligonucleotides containing either xcUPRE-1 or UPRE-2 motifs (15). Finally, the palindromic structure of the xcUPRE-1 target site is consistent with expectations of homodimeric binding.

How does Hac1^i bind both long xcUPRE-1 and compact UPRE-2 DNA target sites? Although most bZIP transcription factors are thought to bind relatively compact binding sites (*SI Appendix*, Fig. S21), Maf subfamily transcription factors recognize unusually long motifs (13–14 bp) via an unconventional conformation of the invariant arginine and asparagine residues within the basic region of all bZIP proteins (4). Similarly, a crystal structure of Pap1, a *S. pombe* bZIP transcription factor, complexed with DNA demonstrated that Pap1 target site specificity was also due to alternate positioning of these two residues (2). In our genetic screen, none of the constructs with altered binding affinities were found to have mutations in these invariant residues (Fig. 5*D* and *SI Appendix*, Fig. S17 and Table S8), although such mutations were found in colonies lacking fluorescence in either channel (*SI Appendix*, Table S7). We therefore suggest that these invariant residues could be required for recognition of both target sites, with changes in their conformation leading to recognition of one site or the other. Notably, the ability to recognize two closely related sites via conformational shifts has previously been proposed for a minimal bZIP construct (24). In a manner analogous to the Maf proteins, we propose that the extended homology region could stabilize invariant bZIP residues in the conformation required for UPRE-2 recognition. With the exception of MafG, most bZIP crystal structures have been based on constructs truncated to include only 1–9 nucleotides N-terminal to the basic DNA binding region (4, 25–28) (*SI Appendix*, Fig. S22). It remains to be seen whether N-terminal regions of extended homology facilitate binding of alternate sites by other bZIP proteins.

Several recent studies have noted plasticity in bZIP binding preferences, although to date, this plasticity has been confined to tolerance for binding multiple related sites of the same or very similar lengths. A synthetic bZIP protein composed of the Gcn4 basic region fused to the C/EBP leucine zipper was shown to bind with high affinity to both cognate and alternate sites, indicating that protein architecture beyond the basic region can affect binding preferences (24). In addition, a recent study employing PBMs to characterize the DNA binding specificities of multiple bZIP TFs noted that several proteins (Yap1, Yap3, and Sko1) possessed the ability to bind closely related dyad repeat sites with variable length (1–2 bp) spacers or extensions at either end (23). This study also noted that the DNA binding domain for Hac1^i shares multiple residues with the basic regions of bHLH proteins,

possibly explaining the similarity between UPRE-2 and bHLH E-box target sites.

Our results have broad implications for experimental approaches to characterize transcription factor binding preferences. Many current methods [e.g., PBMs (16, 29–31) and HiTS-FLIPS (32)] entail measurements of binding interactions, which are interpreted to yield relative preferences for particular nMer (usually 8-mer) sequences. In these analyses, the strength of binding for a given oligonucleotide is assumed to result solely from the presence or absence of a particular nMer sequence, without consideration for sequence context. Here, we show that even transcription factors thought to be simple in their DNA binding properties can bind multiple target sites of very different lengths, complicating attempts to identify target sites via ranked nMer preferences. The MITOMI 2.0 technique presented here attempts to avoid these pitfalls by applying a statistical mechanical model to extract binding preferences. It is important to note, however, that the size of the library that can be accommodated within the devices is still too small for simultaneous detection of binding to both Hac1^{i} motifs. The experiments presented here were guided by our previous knowledge of Hac1^{i} regulatory targets, underscoring the importance of integrating both biophysical and biological data to understand transcriptional regulation.

To what purpose does Hac1^{i} recognize multiple distinct sites? For the glucocorticoid receptor, DNA sequences can act as allosteric ligands, inducing conformational changes to preferentially recruit specific cellular co-factors with functional consequences for transcriptional activation (33). A similar scenario may apply to Hac1^{i}, and perhaps to other bZIP family members, although additional studies will be required to determine whether changes in protein conformation within the DNA binding domain can propagate elsewhere within the protein. Alternatively, dual site recognition could represent a snapshot in evolutionary time of a transcriptional network rewiring event in progress. According to this notion, it may have been advantageous to place an additional set of target genes under Hac1^{i} control, perhaps as a handoff of some other transcriptional program. In this light, it is interesting to note that the Hac1^{i}-driven transcription program in *S. cerevisiae* has been split into multiple transcriptional branches in metazoans, indicating evolutionary network plasticity.

Materials and Methods

DNA Library and Hac1^{i} Production. DNA libraries for MITOMI 2.0 experiments were synthesized as described previously (18). Briefly, all possible 65,536 8-bp DNA sequences were assembled into a compact DNA library spread over 1,457 oligonucleotides, each of which contained an identical 3 nt 5′ ‘*CGC*’ clamp and an identical 14 nt 3′ universal sequence allowing hybridization of a single Cy5-labeled primer (Fig. 1*A*). Following hybridization, all sequences were extended using Klenow exo- (New England Biolabs). Prior to printing, 8-mer libraries were dried down and resuspended to a final concentration of 1.25 μM in 3X SSC containing polyethylene glycol (PEG) (Fluka) and D-(+)-trehalose dihydrate (Fluka) to improve spot visibility and solubilization. Libraries for measuring concentration-dependent binding behavior were synthesized largely as described previously (18) with final preprinting concentrations of 10 μM, 6.7 μM, 4.4 μM, 3.0 μM, 2.0 μM, 1.3 μM, 0.9 μM, and 0.4 μM. For truncation mutant experiments, final pre-printing concentrations were 16 μM, 10.7 μM, 7.1 μM, 4.7 μM, 3.2 μM, 2.1 μM, 1.4 μM, and 0.94 μM; each spot was printed twice to increase concentrations even further. Libraries were printed on custom 2″ × 3″ Scientific SuperChip Epoxysilane (Thermofisher) slides.

Linear expression templates for Hac1 proteins were created via a two-step PCR amplification reaction, as described previously (18). In the first PCR reaction, gene-specific primers were used to amplify gene templates and add an upstream Kozak sequence and a C-terminal 6xHis tag. In the second PCR reaction, universal primers were used to add an upstream T7 promoter, 3′ poly-A tail, and downstream T7 terminator. PCR-generated templates were added to TNT T7 Quick Coupled in vitro transcription translation kits (rabbit reticulocyte lysates, Promega) according to the manufacturer's instructions in the presence of BODIPY-labeled lysine tRNA (Fluorotect Green, Promega) to produce fluorescently labeled protein.

Microfluidic Affinity Assays. Photolithography molds and microfluidic devices were produced as described previously (17, 18). Microfluidic affinity assays and data analysis were also performed largely as described previously (17, 18), with one modification. Cy5 intensities of printed slides can decrease rapidly with time, rendering calibration curves linking intensities with DNA concentration inaccurate. To sidestep this issue, we measured a single calibration curve within one day of an experiment assessing concentration-dependent binding behavior for Hac1^{i} interacting with UPRE-2 variants. The K_d values for these interactions were then used to determine the appropriate conversion between intensity and DNA concentration in the experiments shown in Figs. 2 and 3. All data from Figs. 4–6 were measured and calibrated independently. Average xcUPRE-1 and UPRE-2 PSAMs were calculated by (*i*) determining binding preferences (K_a) for wild-type and all single-substitution oligonucleotides relative to the most strongly bound xcUPRE-1-like or UPRE-2-like oligonucleotide and (*ii*) computing the average relative affinity over all replicates. Differences in binding preferences for Hac1^{i} truncation mutants were calculated by subtracting measured relative binding affinities (calculated relative to the most strongly bound oligonucleotides) from the average relative binding affinities given in the xcUPRE-1 and UPRE-2 PSAMs (*SI Appendix*, Tables S3 and S4).

Error-Prone PCR. The Hac1^{i} mutant library was created using error-prone PCR (34), using a total of 48 cycles and 11 serial dilution steps (performed every four cycles). All sequence alignments and phylogenetic alignments were created using Geneious v4.8.2 (35).

Yeast Strains and Plasmids. Standard cloning and yeast techniques were used for construction, transformation, and integration of the plasmid within strain W303. The transcription reporters used here controlled expression of the fluorescent proteins mApple and GFP via a crippled *cyc1* promoter containing four repeats of a 22-bp UPR-responsive *cis* element (xcUPRE-1 for mApple or mUPRE-1 for GFP). The UPRE-2 reporter was generated by site directed mutagenesis of 4xUPRE-1 (Fig. 5*B* and *SI Appendix*, Fig. S14). The xcUPRE-1-mApple was cloned into a single integration, *HIS3*-marked vector (pNH603), while the mUPRE-1-GFP was cloned into a single integration, *LEU2*-marked vector (pNH605).

Flow Cytometry. A dual reporter strain containing xcUPRE-1-mApple (integrated in *his3*) and mUPRE-1-GFP (integrated in *leu2*) also expressed a chimeric estradiol-responsive transcriptional activator with an N-terminal activation domain derived from Msn2 and a C-terminal DNA binding domain from Gal4. This parent strain was then transformed with either wild-type *HAC1*^{i} or the mutant *hac1*^{i} library (cloned into a single integration, *TRP1*marked vector under the control of the *GAL1* promoter). Cells were cultured in 2x SDC at 30 °C in 96 well plates (2 mL) in an Innova plate shaker at 900 rpm. After induction with estradiol (100 nM), cells were sampled after 4 h using a BD LSR-II equipped with a high throughput sampler, a 488 nm 150 mW laser, 532 nM 150 mW laser, FITC and PE-Texas red emission filters, and FACS DIVA software. Flow cytometry data were analyzed using custom software written in Python. Reported mean fluorescence intensities for mutant strains were calculated via Gaussian fits to binned intensity distributions for individual cells.

ACKNOWLEDGMENTS. We thank Doron Gerber, Danh Tran, and Stephen Quake for assistance with fabrication of microfluidic devices and early microfluidic assays; Marshall Burke for photographs of microfluidic devices; and Matthew Larson and Florencia Caro for careful reading of the manuscript. P.M.F was supported by a Howard Hughes Medical Institute/Helen Hay Whitney Foundation Postdoctoral Fellowship. J.D.R. and P.W. are investigators of the Howard Hughes Medical Institute.

1. Asada R, Kanemoto S, Kondo S, Saito A, Imaizumi K (2011) The signalling from endoplasmic reticulum-resident bZIP transcription factors involved in diverse cellular physiology. *J Biochem* 149:507–518.
2. Fujii Y, Shimizu T, Toda T, Yanagida M, Hakoshima T (2000) Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat Struct Mol Biol* 7:889–893.
3. Miller M (2009) The importance of being flexible: The case of basic region leucine zipper transcriptional regulators. *Curr Protein Pept Sci* 10:244–269.
4. Kurokawa H, et al. (2009) Structural basis of alternative DNA recognition by Maf transcription factors. *Mol Cell Biol* 29:6232–6244.
5. Chapman R, Sidrauski C, Walter P (1998) Intracellular signaling from the endoplasmic reticulum to the nucleus. *Annu Rev Cell Dev Biol* 14:459–485.
6. Rüegsegger U, Leber JH, Walter P (2001) Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. *Cell* 107:103–114.

7. Mori K, et al. (1992) A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. *EMBO J* 11:2583–2593.
8. Kohno K, Normington K, Sambrook J, Gething MJ, Mori K (1993) The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum. *Mol Cell Biol* 13:877–890.
9. Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T (1996) Signalling from endoplasmic reticulum to nucleus: Transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. *Genes Cells* 1:803–817.
10. Cox JS, Walter PA (1996) Novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. *Cell* 87:391–404.
11. Mori K, Ogawa N, Kawahara T, Yanagi H, Yura T (1998) Palindrome with spacer of one nucleotide is characteristic of thecis-acting unfolded protein response element in *Saccharomyces cerevisiae*. *J Biol Chem* 273:9912–9920.
12. Partaledis JA, Berlin V (1993) The FKB2 gene of *Saccharomyces cerevisiae*, encoding the immunosuppressant-binding protein FKBP-13, is regulated in response to accumulation of unfolded proteins in the endoplasmic reticulum. *Proc Natl Acad Sci USA* 90:5450–5454.
13. Shamu CE, Cox JS, Walter P (1994) The unfolded-protein-response pathway in yeast. *Trends Cell Biol* 4:56–60.
14. Travers KJ, et al. (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* 101:249–258.
15. Patil CK, Li H, Walter P (2004) Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. *PLoS Biol* 2:e246.
16. Badis G, et al. (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* 32:878–887.
17. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315:233–237.
18. Fordyce PM, et al. (2010) De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nat Biotechnol* 28:970–975.
19. Wu R, Chaivorapol C, Zheng J, Li H, Liang S (2007) fREDUCE: Detection of degenerate regulatory elements using correlation with expression. *BMC Bioinformatics* 8:399–411.
20. Foat BC, Houshmandi SS, Olivas WM, Bussemaker HJ (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc Natl Acad Sci USA* 102:17675–17680.
21. Foat BC, Morozov AV, Bussemaker HJ (2006) Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* 22:e141–e149.
22. Stewart-Ornstein J, Weissman JS, El-Samad H (2012) Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol Cell* 45:483–493.
23. Gordân R, et al. (2011) Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biol* 12:R125.
24. Fedorova AV, Chan I-S, Shin JA (2006) The GCN4 bZIP can bind to noncognate gene regulatory sequences. *Biochim Biophys Acta* 1764:1252–1259.
25. Schumacher MA, Goodman RH, Brennan RG (2000) The structure of a CREB bZIP•somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J Biol Chem* 275:35242–35247.
26. Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein α. *J Biol Chem* 278:15178–15184.
27. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein-DNA complex. *Cell* 71:1223–1237.
28. Keller W, König P, Richmond TJ (1995) Crystal structure of a bZIP/DNA complex at 2.2 \AA: Determinants of DNA specific recognition. *J Mol Biol* 254:657–667.
29. Berger MF, et al. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* 24:1429–1435.
30. Badis G, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324:1720–1723.
31. Zhu C, et al. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res* 19:556–566.
32. Nutiu R, et al. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat Biotechnol* 29:659–664.
33. Meijsing SH, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 324:407–410.
34. McCullum EO, Williams BAR, Zhang J, Chaput JC (2010) Random mutagenesis by error-prone PCR. *Methods Mol Biol* 634:103–109.
35. Drummond AJ, et al. (2012) *Geneious v4.8.2*, available at www.geneious.com.