Article

Mutations to transcription factor MAX allosterically increase DNA selectivity by altering folding and binding pathways

Renee Hastings[1], Arjun K. Aditham [2,3], Nicole DelRosso[1], Peter H. Suzuki[2] & Polly M. Fordyce [1,2,3,4,5] ✉

Understanding how proteins discriminate between preferred and non-preferred ligands ('selectivity') is essential for predicting biological function and a central goal of protein engineering efforts, yet the biophysical mechanisms underpinning selectivity remain poorly understood. Towards this end, we study how variants of the promiscuous transcription factor (TF) MAX (*H. sapiens*) alter DNA specificity and selectivity, yielding >1700 K_ds and >500 rate constants in complex with multiple DNA sequences. Twenty-two of the 240 assayed MAX point mutations enhance selectivity, yet none of these mutations occur at residues that contact nucleotides in published structures. By applying thermodynamic and kinetic models to these results and previous observations for the highly similar yet far more selective TF Pho4 (*S. cerevisiae*), we find that these mutations enhance selectivity by altering partitioning between or affinity within conformations with different intrinsic selectivity, providing a mechanistic basis for allosteric modulation of ligand selectivity. These results highlight the importance of conformational heterogeneity in determining sequence selectivity and can guide future efforts to engineer selective proteins.

The ability of proteins to accurately recognize cognate binding partners amidst a landscape of chemically similar ligands is essential for nearly all biological processes. Fully describing a binding landscape requires considering not only specificity (which ligand is preferred) and affinity (the absolute strength of the interaction), but also selectivity (the quantitative difference in affinity between preferred and non-preferred ligands). Highly selective proteins bind a limited set of targets well above levels of background binding, while less selective proteins promiscuously recognize a wider variety of ligands. An enhanced understanding of how protein amino acid sequence encodes selectivity is essential for deciphering signaling networks and ultimately engineering protein interactions without off-target effects. However, this remains challenging as the quantitative, systematic data required for these investigations are scarce.

Transcription factor (TF) proteins provide an ideal model system for investigating binding selectivity. TFs naturally bind specific regulatory DNA sequences[1] across a range of affinities to control gene expression, the absolute interaction affinities[2] of which have been dictated and tuned by evolutionary pressures on both protein and DNA sequence. Differences in binding selectivity therefore have biological consequences: TF paralogs with identical sequence specificity but different absolute affinities for that sequence play non-interchangeable roles in development[3], and many closely related TFs bind the same high-affinity motifs but bind alternate low-affinity sites[4–8]. Underscoring the importance of interactions spanning the entire landscape, low-affinity DNA sites are conserved and important for gene regulation[7,9,10], and many enhancers are evolved for suboptimal affinities such that increasing affinity is pathogenic or disrupts specific and properly timed gene expression[11–15].

[1]Biophysics Program, Stanford University, Stanford, CA, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA, USA. [3]Sarafan ChEM-H, Stanford University, Stanford, CA, USA. [4]Department of Genetics, Stanford University, Stanford, CA, USA. [5]Chan Zuckerberg Biohub, San Francisco, CA, USA. ✉e-mail: pfordyce@stanford.edu

While previous attempts to understand the protein sequence determinants of TF binding landscapes have focused on analyzing structures of bound TF-DNA complexes[16,17], this approach is complicated by the structural plasticity of TFs. Some of the most abundant structural classes of human TFs are enriched in intrinsically disordered sequences[18] that can drive promiscuous binding[19], only fold when bound to DNA[20–22], and adopt different conformations to facilitate binding to different DNA sequences[19,23,24]. Therefore, understanding selectivity requires data that systematically vary both protein and DNA sequence across a range of affinities to map the full DNA binding landscape.

Towards this goal, we investigate how TF sequence shapes DNA binding landscapes by comparing binding landscapes for two basic helix-loop-helix (bHLH) TFs: *H. sapiens* Myc-associated factor X (MAX) and *S. cerevisiae* Pho4. While both TFs recognize the same CACGTG E-box motif with nearly identical bound structures and are largely unstructured in the absence of DNA[20,21], Pho4 is highly selective for this motif while MAX is more promiscuous[25]. To understand the origin of this difference, we apply a recently developed microfluidic technique (STAMMP, for Simultaneous Transcription Factor Affinity Measurements via Microfluidic Protein arrays)[26] to interrogate how 240 single amino acid substitutions impact MAX DNA binding affinities to 7 motif-variant DNA sequences (>1700 K_ds). In parallel, we develop and apply a method for measuring kinetics (k_{off}s) for hundreds of protein variants (k-STAMMP), yielding >500 rate constants. Overall, ~10% of measured MAX mutations increase binding selectivity (the difference by which MAX prefers CACGTG over mutated motifs). Surprisingly, none of these residue positions make base-specific DNA contacts in published structures, instead contacting the phosphate backbone or facing the solvent. Thermodynamic and kinetic models capable of explaining observed measurements suggest that while Pho4 traverses a single binding pathway (a classic folding-and-binding transition into a single helical state), MAX can bind DNA through more than one pathway, one of which is selective for CACGTG binding while the other binds DNA promiscuously. Thus, consideration of multiple binding pathways with different intrinsic binding selectivities for MAX—as well as the diverse mechanisms by which mutations can perturb these states and pathways—enables a mechanistic understanding of how non-base-contacting mutations can allosterically increase binding selectivity. Together, these results establish that systematic and high-throughput thermodynamic and kinetic measurements of combinatorial protein/ligand mutations can provide information difficult to obtain via standard structural approaches and highlight a need to consider heterogeneous binding modes when understanding and engineering selective binders.

Results

MAX and Pho4 are model systems for investigating selectivity
The DNA-binding domains (DBDs) of MAX and Pho4 are disordered in solution, with their DNA-contacting regions folding only upon recognition of a DNA binding site to assume highly similar structural conformations (RMSD = 1.519 Å, Fig. 1a)[27,28]. Both TFs possess similar domain architectures comprised of a DNA-contacting basic region followed by two helices separated by a flexible loop (Fig. 1a, b), make identical base contacts via identical nucleotide-contacting residues (Fig. 1c), and preferentially bind the same cognate CACGTG E-box site as dimers (Fig. 1d). Despite these similarities, prior measurements of WT Pho4 and MAX affinities for a library of 256 DNA sequences containing mutations within an E-box half site revealed differences in their binding energy landscapes (Fig. 1e)[25]. Single-nucleotide mutations to the cognate sequence led to >9-fold higher reductions in binding affinity for Pho4 than for MAX (85-fold versus 9-fold), with Pho4 binding CACGTG more tightly and mutant motifs more weakly[25] (Methods). This difference in selectivity can be conceptualized using energy landscape diagrams[29] in which Pho4 binds its cognate E-box

with a deeper energetic well (Fig. 1e, f). These differences in binding landscapes despite nearly identical DNA-binding interfaces motivated us to probe how variation in non-contacting residues (Supplementary Fig. 1) can shape binding landscapes.

STAMMP measures DNA binding for hundreds of MAX mutations
Using the STAMMP microfluidic platform (Supplementary Figs. 2–3), we previously quantified impacts of 210 single amino acid substitutions within Pho4 on binding to 9 oligonucleotides[26]. Here, we applied STAMMP to investigate how 240 single amino acid substitutions in and around the MAX DBD (Supplementary Fig. 4) impact DNA binding affinity and specificity. This mutant library included 156 alanine and valine scanning mutants to probe protein sequence determinants of DNA motif recognition, 10 mutations substituting orthologous amino acids present in other bHLHs[30] to probe how evolutionary differences alter landscapes, 30 mutants hypothesized to alter helicity and charge to probe biophysical mechanisms contributing to recognition[31,32], and 38 mutations from human allelic variants that were cataloged as pathogenic mutations or variants of unknown significance (VUS)[33,34] (Fig. 1g, h, Supplementary Table 1). C-terminally meGFP-tagged MAX variants were expressed on-chip via in vitro transcription/translation and recruited to anti-eGFP patterned device surfaces for purification and subsequent measurements of concentration-dependent DNA binding via STAMMP. Relative affinity measurements here for both WT and mutated MAX agree well with prior work for constructs lacking a meGFP tag (Supplementary Fig. 5)[25,35], suggesting the presence of a fluorescent protein tag does not alter measured binding.

Substitutions throughout the MAX DBD alter DNA binding
To identify MAX residues involved in DNA recognition, we first measured impacts of each mutation on binding to the cognate 5′-CACGTG-3′ motif. Of 240 mutants, 237 expressed consistently (Supplementary Table 2, Supplementary Fig. 6). Measured K_ds fit from processed data (Methods, Supplementary Figs. 7–8, Supplementary Table 3) were highly consistent across device replicates (RMSE < 0.3 kcal/mol over a 3.5 kcal/mol dynamic range) (Fig. 2a, Supplementary Fig. 9) and independent of immobilized TF-eGFP concentrations (as expected for TF concentrations well below measured K_ds) (Supplementary Fig. 10). Subsequent analyses aggregated measurements (Supplementary Fig. 11) for each variant with 3 or more replicates to report median K_d or $\Delta\Delta G$ value. Overall, 106 mutants significantly altered the affinity for the consensus motif relative to WT (Bonferroni-corrected $p < 0.05$). Many mutations strongly reduced DNA binding, including substitutions at conserved nucleotide-contacting residues H28, E32, R35, and R36[28] (Fig. 2b). Many pathogenic variants (18/31) and some (I71N and R47W) VUS significantly reduced binding as well (Supplementary Fig. 12). Allelic variants yielded some of the largest changes in DNA binding affinity in the library, with mutations at the dimerization interface either significantly decreasing (e.g., L46, A67, and I71) or enhancing binding (e.g., M74L[36]) (Supplementary Fig. 12). Mutations to crystal structure-predicted phosphoryl oxygen backbone-contacting residues increased or decreased affinity depending on the substituted residue (Fig. 2b). Of the 29 significant mutations that enhanced binding, 12 occurred at non-contacting, solvent-facing residues in the basic region, confirming that putative non-DNA contacting residues can have substantial impacts on binding.

Comparable substitutions decrease affinity less in MAX
Just as the magnitude of changes in affinity upon DNA mutation was smaller for MAX than for Pho4, comparable mutations at corresponding residue positions also had smaller impacts in MAX (Wilcoxon signed-rank test $p = 10^{-9}$) (Figs. 2c and 1e). Substitutions altering charged contacts with the DNA phosphate backbone yielded substantially larger reductions in binding affinity for Pho4 (Pho4 R252Q/
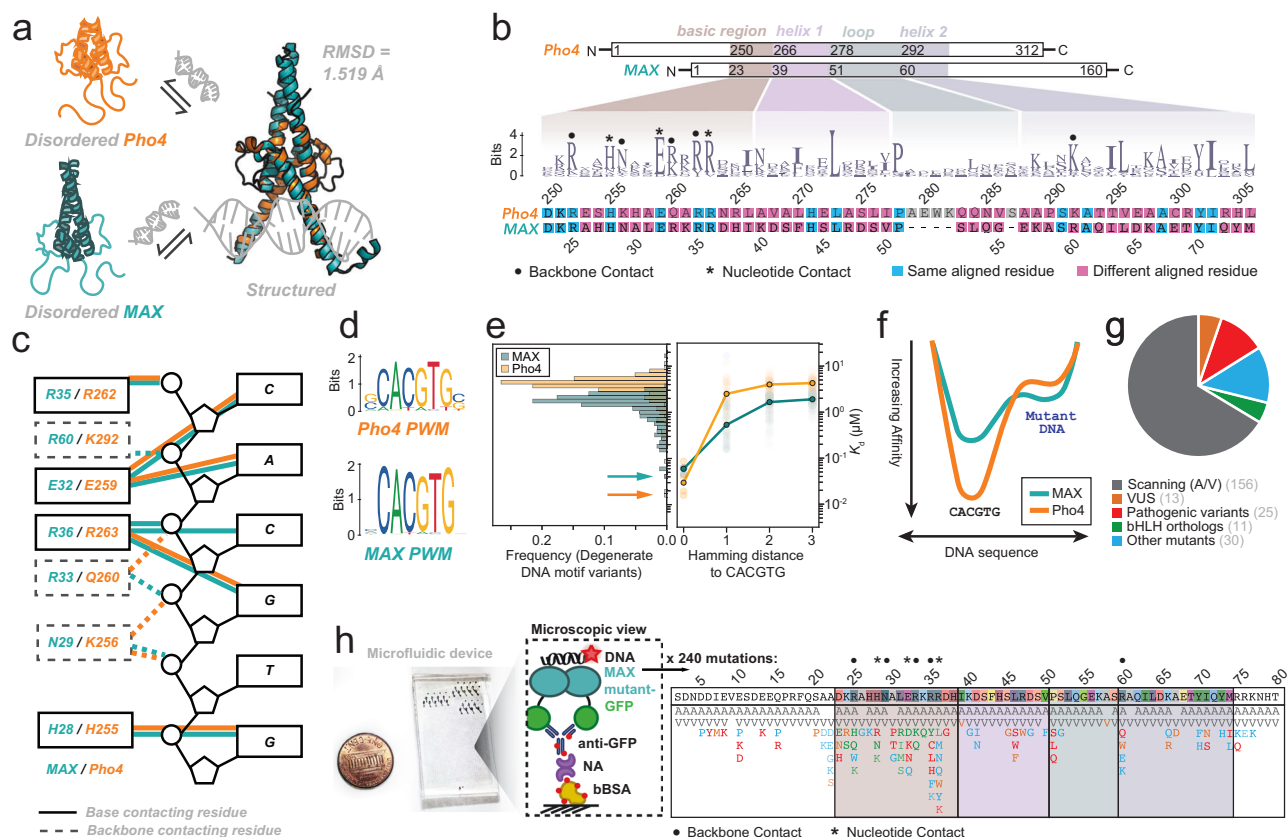
Fig. 1 | Amino acid sequence, structure, and DNA specificity of MAX and Pho4. **a** Schematic of folding and binding pathway and structural alignment for Pho4 (orange, 1A0A) and MAX (teal, 1HLO). **b** Domain architectures and sequence alignment for MAX and Pho4 DNA binding domains alongside conservation across bHLH TFs. **c** Crystallographic contacts between the CACGTG consensus E-box and MAX (teal) and Pho4 (orange). **d** PWMs for MAX (JASPAR MA0058.3) and Pho4 (JASPAR MA0357.1). **e** Distribution of binding affinities for all degenerate E-box motif variants[25] with most tightly bound sequences annotated (left); median affinity as a function of Hamming distance away from the CACGTG cognate. **f** Cartoon illustrating differential selectivity. **g** Classification of MAX mutations in this study. **h** Microfluidic device and schematic of TFs immobilized on a surface with biotinylated BSA (bBSA), neutravidin (NA), and anti-GFP antibody (left) along with location and identity of MAX mutations studied here (right).

MAX R25Q, Pho4 K292A/MAX R60A (Fig. 2c, Supplementary Fig. 13). This trend held for mutations to residues that do not contact bound DNA in published structures (Pho4 K251/MAX K24), where some analogous substitutions even increased binding affinity for MAX (L31V) while reducing it for Pho4 (A258V) (Fig. 2c, Supplementary Fig. 13). These differences in 'mutational sensitivity' support prior observations[37] that the strength of otherwise similar molecular contacts is contextual and that only certain protein homologs can support rheostat positions[38].

MAX and Pho4 differ in folding and binding transitions
For TFs that are unstructured in the absence of DNA, altering helical propensity[39] can commensurately change binding affinity by modulating folding entropic penalties[16,17,26,40,41]. Thus, the observation that mutations to comparable solvent-facing residues have different impacts on measured binding affinity for CACGTG in Pho4 and MAX could indicate differing folding and binding transitions. To better understand how changes in a coupled folding and binding pathway impact measured K_ds, we modeled equilibrium bHLH TF/DNA binding in which: (1) unbound TFs can be either unfolded or helical, (2) only the helical form can bind DNA, and (3) TF mutations alter the folding equilibrium (Fig. 2d, Supplementary Fig. 14, Methods). The observed DNA concentration at which half of the TF population is DNA-bound ($K_{d,apparent}$) can then be expressed as a function of a K_d for the interaction between folded TF and free DNA, a folding equilibrium constant K_{fold} (describing the partitioning between folded and unfolded

conformations of the unbound TF), and the total concentration of free DNA and protein. Simulated binding isotherms revealed that the magnitude by which mutations shift $K_{d,apparent}$ depends on the WT value of K_{fold} (Supplementary Figs. 14–15) and provided a method to infer $K_{d,WT}$ and $K_{fold,WT}$ from measured affinities of non-contacting mutations.

Consistent with this simple folding and binding model, Pho4 mutations that altered helical propensity concomitantly shifted binding free energies relative to WT ($\Delta\Delta G_{CACGTG}$) (Pearson $r^2 = 0.55$) and were well-fit by a model in which WT Pho4's basic region is significantly disordered (~81%) in the unbound state (Methods, Fig. 2e, Supplementary Fig. 16), in agreement with NMR data[20]. By contrast, mutations that alter the predicted helicity of the MAX basic region were less correlated with changes in binding free energies (Pearson $r^2 = 0.20$). For this observation to be consistent with a 3-state model, MAX would have to be primarily helical in solution (~1% disordered), at odds with documented structural disorder in this region[21] and the observation that many solvent-facing mutations strongly modulate binding affinity (Methods, Fig. 2e, Supplementary Figs. 15–16). These results could instead be consistent with a more complex model in which MAX can adopt multiple conformations with different intrinsic affinities for DNA. In this model, rather than altering the propensity to fold into a *single* conformation, non-contacting mutations impact measured affinity by altering partitioning between *multiple* conformations. This model is consistent with previously identified bound state conformational heterogeneity within the MAX basic region[42].
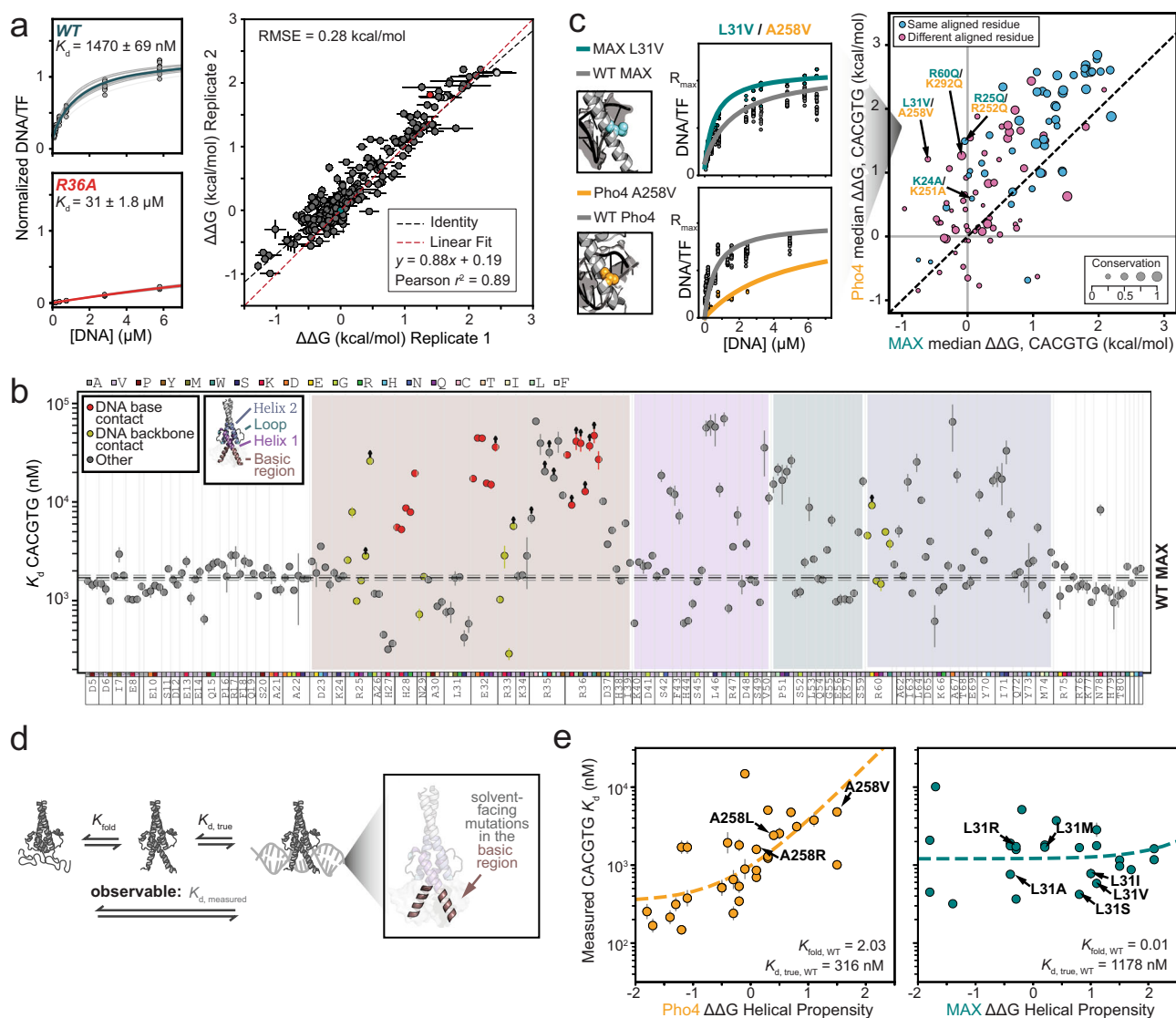
Fig. 2 | MAX and Pho4 differ in folding-and-binding to CACGTG. a Binding isotherms for WT (teal, $n = 20$) and R36A (red, $n = 3$) MAX variants binding to cognate DNA (left). Reproducibility of $\Delta\Delta G$ measurements across two microfluidic devices (right, median ± SEM per device). Light gray markers indicate mutants unresolvable from background binding for which reported K_ds represent a lower limit. **b** Affinities for MAX mutants binding CACGTG (median ± SEM). Red markers denote mutations to DNA-contacting residues, gray markers with red outlines denote mutations to phosphate backbone-contacting residues, and arrows denote K_d limits. **c** Binding isotherms for WT MAX ($n = 51$), MAX L31V ($n = 10$), WT Pho4, and Pho4 A258V (left) and comparison of $\Delta\Delta G$ measurements for aligned substitutions to MAX and Pho4 (right); marker size indicates residue conservation across the bHLH family (Methods). **d** Thermodynamic model for a three-state system such that observable K_d depends on the folding equilibrium (K_{fold}) and true binding affinity ($K_{d,\,true}$). **e** Measured change in cognate affinity (median ± SEM) versus changes in helical propensity[39] for mutations to non-DNA contacting basic region residues in MAX (teal) and Pho4 (orange); dashed line indicates fitted thermodynamic model with indicated fitted values of K_{fold} and K_d. Replicate counts for all K_d measurements are contained in Supplementary Data 1. Source data are provided as a Source Data file.

Distal MAX mutations modulate affinity independent of motif

To understand how MAX mutations alter low-affinity binding and therefore sequence specificity and selectivity, we also measured binding to 5 low-affinity sequences containing mutations within core nucleotide positions in the E-box consensus (**A**ACGTG, C**G**CGTG, CA**T**GTG, CACG**C**G, and CACGT**T**) (Fig. 3a, b, Supplementary Table 4). Across all replicates (Supplementary Data 1), expression (Supplementary Figs. 17-21) and K_d measurements remained highly reproducible (Supplementary Figs. 22–26) and affinities were independent of expression (Supplementary Fig. 27). Over 8–12 K_d replicate measurements per MAX variant (Supplementary Fig. 28), single nucleotide substitutions within the consensus E-box motif reduced WT MAX binding by 2 to 5-fold, with the **A**ACGTG mutation being most deleterious (Fig. 3b, Supplementary Fig. 29).

To identify TF mutations that modulate DNA-binding *affinity* (i.e. alter binding to all DNA sequences equally by shifting the binding energy landscape by a consistent free energy difference), we performed bootstrapped equivalence testing to identify MAX mutations that introduced similar changes in binding free energy across many measured DNA sequences (Methods). The 14 identified "affinity-altering" mutations (Fig. 3c) were generally located within the loop region or dimerization interface (Fig. 3d) and all weakened binding relative to WT MAX. These affinity-altering mutations include cancer-associated variants predicted to be either pathogenic (D23H and P51L) or VUS (R47W) (Supplementary Table 5). P51L uniformly decreased binding affinity by ≥0.92 kcal/mol across DNA sequences (Fig. 3e), likely by disrupting a key proline necessary for proper positioning of helix 1[27]. Leucine zipper substitutions also altered affinity equally across all DNA
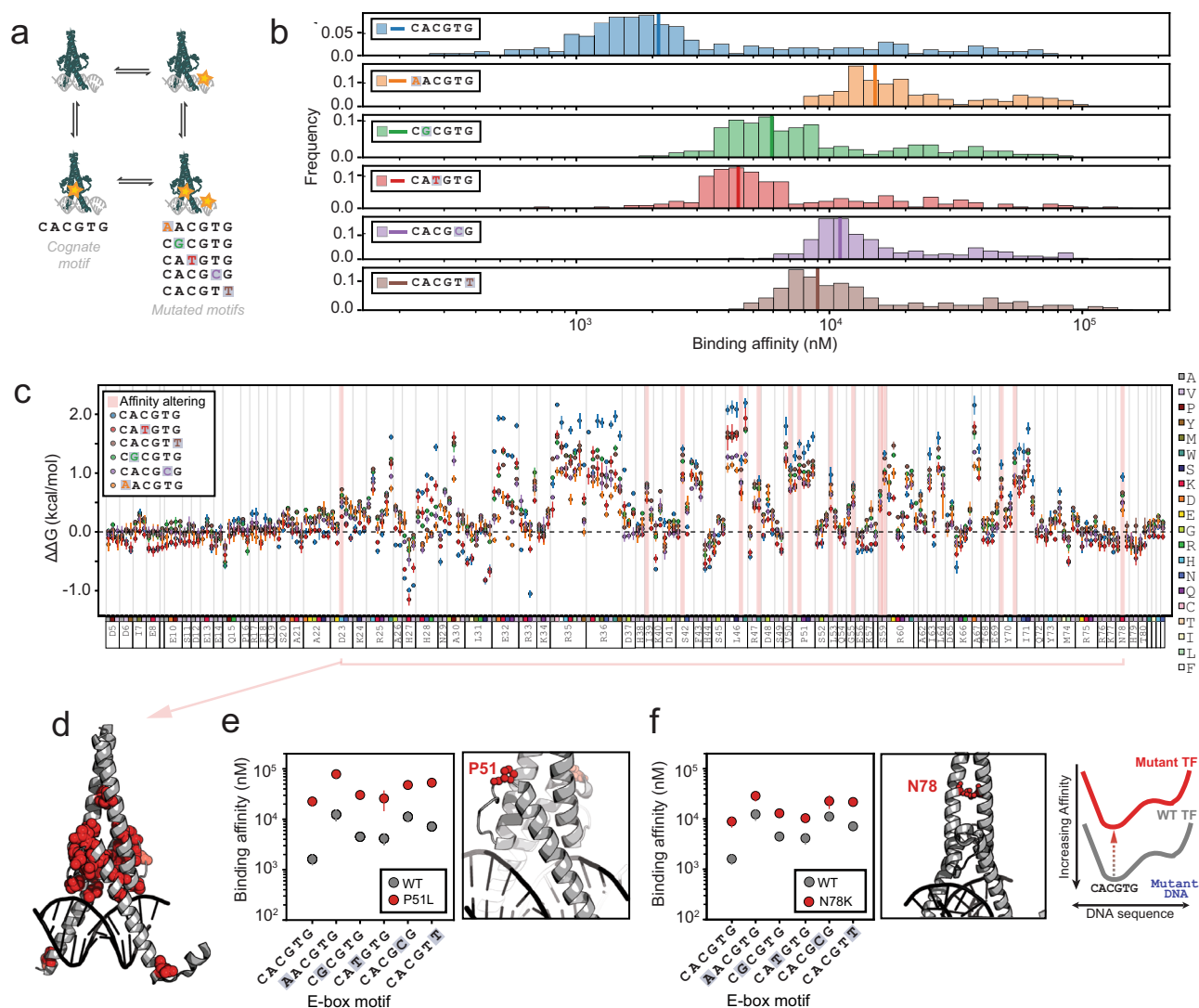
Fig. 3 | Dimer and loop mutations modulate MAX binding affinity. a Cartoon illustrating double mutant cycles across the TF/DNA interface. **b** Histograms of K_ds for all MAX mutants against each measured E-box; vertical lines denote WT. **c** Measured ΔΔG relative to WT across all E-boxes (median ± SEM). Red highlights indicate affinity altering mutations (Methods). **d** Affinity-altering mutations projected on MAX structure (1HLO). K_ds (median ± SEM) for all E-box sequences (left), location of residues of interest (red) within MAX structure (1HLO, right) for MAX P51L (**e**) and N78K (**f**), with cartoon illustrating impact of mutations on DNA-binding landscape (right). Replicate counts for all K_d measurements are contained in Supplementary Data 1. Source data are provided as a Source Data file.

sequences, likely by altering dimerization. For example, substituting a homotypically-stabilizing asparagine in the a′ position of the MAX leucine zipper at position 78 for lysine uniformly disrupted binding across all DNA sequences (median ΔΔG = 0.56 kcal/mol), consistent with prior studies of leucine zipper coupling energies[43] (Fig. 3f).
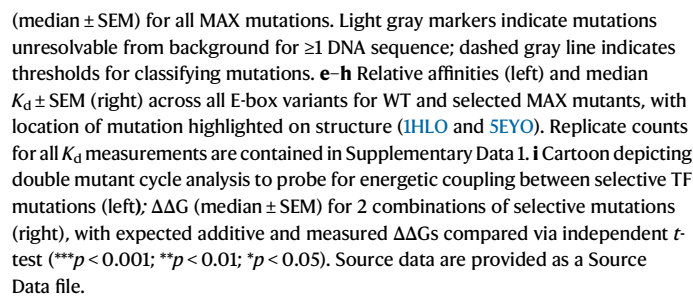
Double mutant cycles reveal drivers of specificity

To systematically identify specificity- and selectivity-altering substitutions, we performed biochemical double-mutant cycles[44] comparing energetic impacts on binding from TF mutations and mutations to cognate DNA (Fig. 4a, Supplementary Figs. 30–33). When visualized as a scatter plot, additive mutations to non-interacting residues and nucleotides lie along an additive line, while non-additive mutations to interacting and/or epistatic residue and nucleotide pairs fall off-diagonal (Fig. 4a). While most (150) mutations additively alter binding to a low-affinity motif known to be preferentially bound by Myc/MAX heterodimers in vivo (CACG**C**G)[45], 34 of 81 mutants bind this motif with a reduced energetic penalty relative to the additive expectation (Fig. 4a, Supplementary Table 6). These include mutations to structurally-predicted nucleotide contacts (e.g. E32, which contacts

the outer two nucleotide bases in the E-box) and residues that make stabilizing salt bridges with these contacts (e.g. R35, which contacts E32)[46,47]. Some epistatic mutations, such as E32A, even alter the intrinsic sequence specificity of MAX to yield significantly ($p < 0.05$ via independent t-test) tighter binding to CACG**C**G. 10 additional mutations also bind most tightly to non-CACGTG E-box motifs, all of which disrupt annotated DNA-contacting interactions (Supplementary Fig. 34). Thus, this double-mutant cycle analysis recapitulates known direct contacts within the crystal structure. Intriguingly, other mutations non-additively bound CACG**C**G without an obvious structural rationale, (e.g. solvent-facing D37 or H28, which canonically contacts the 5′ guanine[28]) (Fig. 4a, b).

Selective mutations favor cognate or disfavor mutated motifs

Selectivity-altering substitutions change the energetic difference between cognate site binding and binding to *many* or *all* non-cognate sites. To identify such mutations, we computed residuals for each pairwise comparison between a mutated E-box and the cognate sequence and defined a "selectivity score" as the median of these residual Z-scores (Fig. 4c, Supplementary Fig. 35). Mutants with negative

Fig. 4 | Mutations increase selectivity of MAX by enhancing cognate or weakening mutant DNA binding. a Cartoon depicting additive and epistatic energetic impacts for double mutants (top); pairwise comparisons between measured ΔΔG (median ± SEM) relative to WT for all MAX mutants interacting with low-affinity 5′-C CACG**C**G A−3′ versus consensus 5′-C CACGTG A-3′. Light gray markers indicate mutations unresolvable from background for ≥1 DNA sequence; black markers indicate known crystallographic contacts to mutated nucleotide bases; red marker edges indicate non-additive binding; dashed black line indicates linear regression; red dashed line indicates 1:1 identity line. **b** Mutants with epistatic energetic impacts projected onto MAX structure (1HLO). **c** Schematic illustrating calculation of selectivity scores from double mutant cycles. **d** Selectivity scores *vs.* ΔΔG$_{CACGTG}$

(median ± SEM) for all MAX mutations. Light gray markers indicate mutations unresolvable from background for ≥1 DNA sequence; dashed gray line indicates thresholds for classifying mutations. **e–h** Relative affinities (left) and median K_d ± SEM (right) across all E-box variants for WT and selected MAX mutants, with location of mutation highlighted on structure (1HLO and 5EYO). Replicate counts for all K_d measurements are contained in Supplementary Data 1. **i** Cartoon depicting double mutant cycle analysis to probe for energetic coupling between selective TF mutations (left); ΔΔG (median ± SEM) for 2 combinations of selective mutations (right), with expected additive and measured ΔΔGs compared via independent *t*-test (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$). Source data are provided as a Source Data file.

selectivity scores *decrease* the energetic penalty for binding to *mutant* motifs relative to the cognate (decreasing selectivity). As expected, many mutations to nucleotide base-contacting residues reduce selectivity in this way (e.g. H28, E32) (Fig. 4d). Mutants with positive values putatively render MAX more selective. Strikingly, 22 MAX mutations increase selectivity for CACGTG (Fig. 4d, Supplementary Fig. 36,

Methods), with nearly all selectivity-increasing mutations lying in the DNA-contacting basic region or helix 1 (Supplementary Fig. 37). These mutations are enriched in backbone-contacting residues ($p = 1 \times 10^{-3}$ via Chi-squared test) and solvent-facing basic region residues ($p = 8 \times 10^{-7}$ via Chi-squared test), at odds with the expected result in which base contacts determine selectivity through direct readout.

Mutations that *increase* selectivity can do so by either enhancing binding to the cognate motif and/or by decreasing binding to non-cognate motifs. Comparing selectivity scores versus $\Delta\Delta G_{CACGTG}$ (Fig. 4d) reveals that some mutations *increase* affinity for the cognate motif without altering affinity for mutated sequences (i.e. "deepening the energetic well"; upper-left quadrant, Fig. 4e, f). For example, all measured mutations at H27 increase affinity for the cognate sequence (Supplementary Fig. 38) and to a lesser degree CA**T**GTG (the only mutated motif for which H28 mutations are additive) (Supplementary Fig. 36, Supplementary Table 6), consistent with a model in which substitutions at residue 27 better position or reduce competition for protonation of H28 to promote selective E-box recognition. Similarly, K40A does not directly contact DNA or DNA-contacting residues, but mutations at this position may disrupt a polar contact with D37 visible in the crystal structure[28]. Our double-mutant cycle analysis suggests D37 is epistatic with the outer 2 bases in the E-box, so disrupting the K40/D37 interaction may increase affinity for all DNA sequences without mutations to these bases (Fig. 4f, Supplementary Fig. 36). The aligned position in Pho4 contains an alanine (Fig. 1b), possibly contributing to Pho4's selectivity via the same mechanism. Notably, repeating this selectivity score analysis for Pho4 reveals that, unlike for MAX, no Pho4 mutations increase selectivity for CACGTG (Supplementary Fig. 39). Instead, comparable mutations that increase selectivity in MAX by increasing affinity for CACGTG alone like H27A also increase affinity for mutated E-box sequences in Pho4 and are therefore affinity-altering. This Pho4 behavior is again consistent with a simple model of folding and binding where the same bound conformation recognizes all E-box variants and mutations therefore proportionally change binding to all sequences.

MAX mutations in the upper middle of Fig. 4d increase selectivity by decreasing binding to non-CACGTG sequences (i.e. "raising the energetic wall" for non-cognate binding). For example, N29V (which ablates a structurally predicted polar contact with the DNA phosphate backbone[48]) does not alter affinity for CACGTG but *decreases* affinity for all measured mutated E-box sequences (Fig. 4g). Similarly, mutating residue A30 (which faces the solvent on DNA-bound structures and is not adjacent to any other nucleotide base-contacting residues) to glycine leaves CACGTG binding unaltered but decreases affinities for all mutated E-boxes (Fig. 4h). Mutations that distinguish between cognate and mutated motifs without any apparent sequence preference among mutated sequences suggest that the alternative conformations previously hypothesized for MAX may be selective and promiscuous conformations, as explored below.

Double mutants of selective MAX variants are non-additive

To test if MAX mutants that allosterically enhance selectivity do so by altering partitioning between selective and promiscuous conformations (as implicated by "wall-raising" selective mutations), we examined mutant cycles between two protein residue pairs[49] binding to many motifs (Fig. 4i, Supplementary Fig. 40). To exclude convolved effects of nearby mutations, we restricted interpretation to TF mutant pairs >15 Å apart in published structures (Supplementary Fig. 41). In this analysis, additive changes to binding regardless of motif imply independent perturbations, such as rearrangement of local contacts. However, if MAX double mutants yield additive changes in binding to some motifs but non-additive changes to others (such that any resulting changes to overall binding selectivity are non-additive), this provides corroborating evidence for a model in which MAX can bind DNA with multiple distinct conformations or binding pathways with different intrinsic selectivity.

Consistent with these expectations, the MAX H27V/K40A double mutant yielded additive energetic impacts relative to the single TF mutants for all measured E-box sequences (Fig. 4i), with both mutations increasing selectivity by putatively rearranging local contacts to independently enhance cognate binding. By contrast, the K40A/A30G

double mutant yielded additive impacts for all motifs except for CACGTG, where impacts were less than additive, like A30G alone (Fig. 4i). This non-additivity suggests that some mutations, like A30G, enhance motif selectivity by changing partitioning between multiple conformations with distinct sequence preferences: one that selectively binds CACGTG and another that promiscuously binds many mutated E-box sequences.

Kinetic measurements provide insights into binding mechanism

Next, we asked if selective mutations caused changes in the MAX bound conformational ensemble, unbound ensemble, or both. As equilibrium binding measurements cannot resolve at which stages of the folding and binding reaction selective mutations alter microscopic transitions[50], we developed a kinetic version of the STAMMP microfluidic binding assay (k-STAMMP, derived from k-MITOMI[51]) (Supplementary Fig. 42). Specifically, we measured macroscopic dissociation rates ($k_{off,macroscopic}$, hereon referred to as k_{off}) (Fig. 5a) and inferred apparent on-rates (inferred $k_{on,} = k_{off}/K_d$, calculated assuming a macroscopic 2-state model in which TFs are either bound or not bound to DNA)[52] for Pho4 and MAX variants interacting with 7 motif-variant DNA sequences[52], totaling 561 total measured rate constants (Supplementary Figs. 43–48). Off-rates were well-fit by a single exponential for both Pho4 and MAX (Fig. 5a), with measured rates typically varying by ~2-fold between devices.

Mutations to both Pho4 and MAX yielded larger changes to inferred on-rates than measured off-rates (6.2/5.6-fold change difference between fastest and slowest off-rates and 28.9/36.7-fold for on-rates for Pho4 and MAX respectively) (Fig. 5b, Supplementary Data 2), consistent with recent work suggesting TF affinity and specificity is primarily governed by variation in association rates[6,53]. The subset of mutations that significantly changed dissociation rates tended to ablate or disrupt nucleotide (such as E32N), backbone, or dimer contacts (Fig. 5c), likely due to destabilization of the bound state(s).

A reduced affinity conformation is more stable in MAX

For a folding and binding reaction with a single binding conformation, preformation of structure should increase k_{on} and decrease k_{off}[54]. Consistent with this model, increasing helical propensity in Pho4 increased $k_{on,apparent}$ and decreased k_{off} for cognate E-box binding (Fig. 5d). In contrast, increasing helical propensity in MAX slightly decreased $k_{on,apparent}$ and had little impact on k_{off} for cognate E-box binding (Fig. 5d). This observation is again consistent with the existence of multiple unbound binding-competent states in MAX such that the energetic impact of a mutation on cognate binding becomes uncoupled with intrinsic changes to helicity and instead alters conformational partitioning. Moreover, the observation that putatively helix-breaking mutations such as A30G speed up on-rate suggests that a weaker binding conformation for CACGTG may be more stable. This is also consistent with stopped-flow kinetics data that suggest a conformational change is the rate-limiting step for binding in MAX[55].

Selective mutations act via different microscopic mechanisms

By the Hammond postulate[56], on-rates are more impacted by changes in the unbound state and off-rates by changes in the bound state. Investigating the inverse relationship between k_{on} and k_{off} across multiple DNA sequences (relative to WT) can therefore provide insight into which microscopic transition(s) are impacted[53]. MAX mutations to phosphate backbone-contacting residues that enhanced selectivity by weakening binding to non-cognate motifs (e.g. N29V, R60V; Fig. 4g) primarily altered k_{off} with little to no changes in inferred $k_{on,apparent}$ (Fig. 6a, Supplementary Fig. 49, Supplementary Data 3). Other mutations to non-DNA-contacting, solvent-facing residues that enhanced selectivity by increasing affinity for the consensus motif (e.g. H27V, K40A) (Fig. 4e, f) primarily increased $k_{on,apparent}$ for only the cognate motif (Fig. 6b, Supplementary Fig. 49) with little changes to off-rate
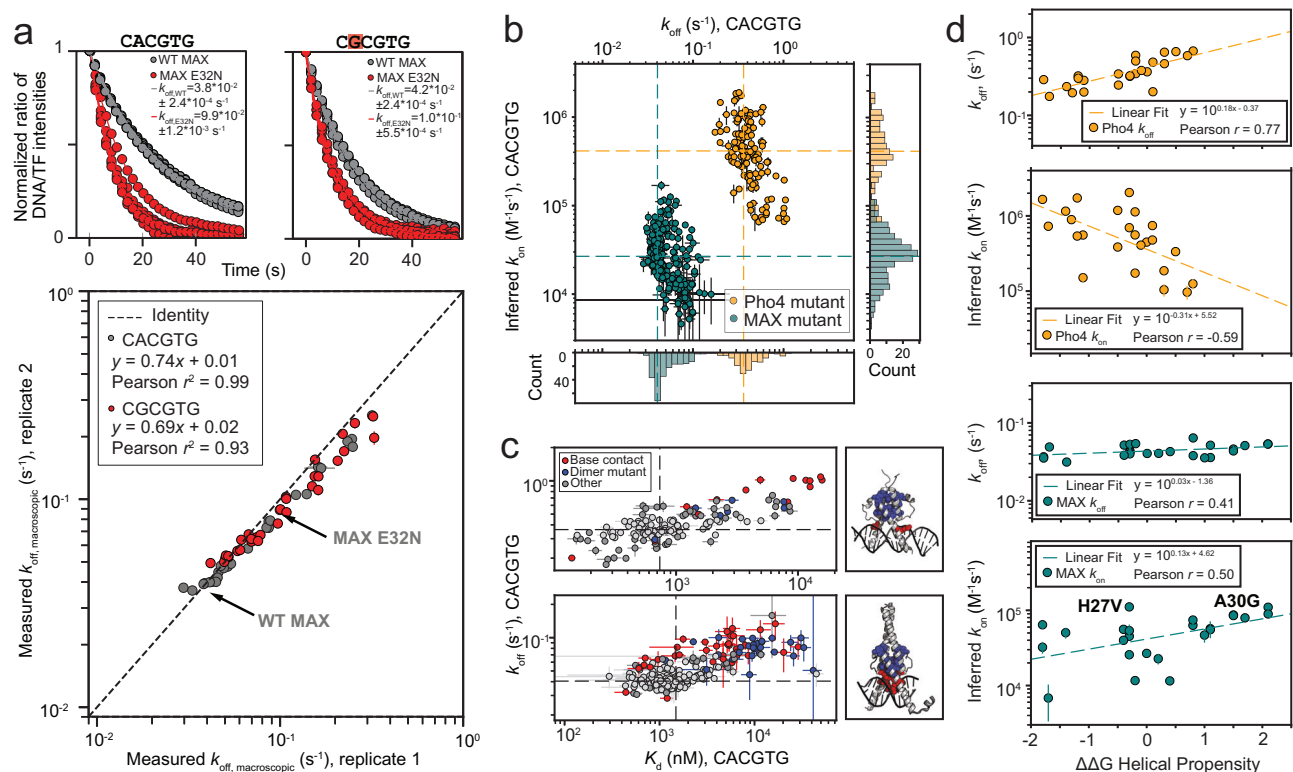
Fig. 5 | Binding kinetics suggest MAX and Pho4 differ in folding and binding transition. a Example dissociation traces for WT (gray) and E32N (red) variants of MAX (top, $n = 5$ for all) and comparison of k_{off} measurements (median ± SEM) across two devices and two DNA sequences (bottom). Dashed black line indicates 1:1 relationship. **b** Measured k_{off} versus inferred k_{on} for all Pho4 (orange) and MAX (teal) mutants interacting with cognate E-box (median ± SEM); dashed lines denote WT values. **c** Measured k_{off} versus inferred k_{on} for all Pho4 (top) and MAX (bottom) mutants interacting with cognate E-box (median ± SEM); dashed lines denote WT values. Marker color indicates mutations to known crystallographic DNA base

contacts (red) or dimerization interface contacts (blue), also shown imposed on Pho4 (1AOA) and MAX (1HLO) crystal structures. WT and variant k_{off}s were compared via independent t-test, and WT-like k_{off} values ($p \geq 0.05$) are indicated by light gray markers. **d** Measured k_{off} and inferred k_{on} (both median ± SEM) versus change in helical propensity[39] for non-DNA contacting basic region substitutions in MAX (teal) and Pho4 (orange); dashed line indicates linear regression. Replicate counts for all K_d and k_{off} measurements are contained in Supplementary Data 2. Source data are provided as a Source Data file.

across many measured motifs (Fig. 6b). This again implies that these mutations may change the unbound ensemble to increase the rate of initial MAX DNA association. Finally, some selective mutations to solvent-facing residues (e.g. A30G) (Fig. 4h) altered both k_{off} and $k_{on,apparent}$ (Fig. 6c, Supplementary Fig. 49), suggesting changes to both bound and unbound states or to the transition state itself.

Modeling selective and promiscuous states reconstitutes data

To test our proposed multi-state model of MAX binding and the microscopic origins of selective binding, we employed Gillespie simulations to model binding for a single TF and DNA molecule via multiple reaction schemes. For each reaction scheme, we sought to identify which, if any, changes in microscopic rate constants altered binding selectivity through similar kinetic and affinity pathways to measured selective MAX mutations.

First, we examined a 3-state model in which TFs are either unbound to DNA ("free"), non-specifically bound and "testing" to see if a site underfoot represents the target site, or specifically "bound" (Fig. 6d), identical to a previously used TF binding scheme[53]. In this model, the rate constant for transitioning between the free and testing states is given by $k_{on,max}$, a theoretical upper bound for the on-rate when all non-specific TF-DNA interactions result in specific binding. The rate constant for transitioning from the bound state to the testing state is given by $k_{off,\mu}$, and the probability of transitioning to the bound state depends on the likelihood of binding a given sequence (f_{motif}) and the rate at which TFs transition from testing back to the free state

($f_{motif} \times k_{off,M}$). Similar to observations for Pho4, systematically varying or co-varying rate constants globally shifted the binding landscape without changing selectivity (Fig. 6e, Supplementary Fig. 50). Explicitly modeling folding-and-binding transitions also did not change selectivity ($\Delta\Delta G_{f\_motif=0.99 - f\_motif=0.01}$) (Supplementary Figs. 51–52). These results are consistent with previous experimental observations that Pho4 mutations that alter helicity globally tune affinity (Supplementary Fig. 39) as well as our hypothesis that the mechanism by which MAX mutations enhance selectivity can only be understood by invoking an additional state.

Consistent with MAX observations, changes to microscopic rate constants within a 5-state model with 2 binding-competent conformations with different intrinsic selectivities (Fig. 6f) yielded a variety of distinct affinity and selectivity effects (Fig. 6g, Supplementary Figs. 53–57). In this model, proteins can transition to 2 different helical[20] testing states that either bind DNA specifically or promiscuously. Transitions to and from the selective and promiscuous testing states are described by $k_{on,max,s}$, $k_{off,M,s}$, $k_{on,max,p}$ and $k_{off,M,p}$. Transitions to and from the selective and promiscuous bound states are described by the microscopic rate constants $f_{motif} \times k_{off,M,s}$, $k_{off,\mu,s}$, $k_{on,p}$, and $k_{off,\mu,p}$. Increasing $k_{off,\mu,p}$ simultaneously increases observable macroscopic k_{off} and decreases observable macroscopic k_{on}, reconstituting measured changes in binding affinities and kinetics for selective mutations hypothesized to disrupt phosphate backbone contacts (e.g. N29V and R60V, Figs. 4e and 6g). Thus, we suggest these mutations may increase selectivity by preferentially destabilizing
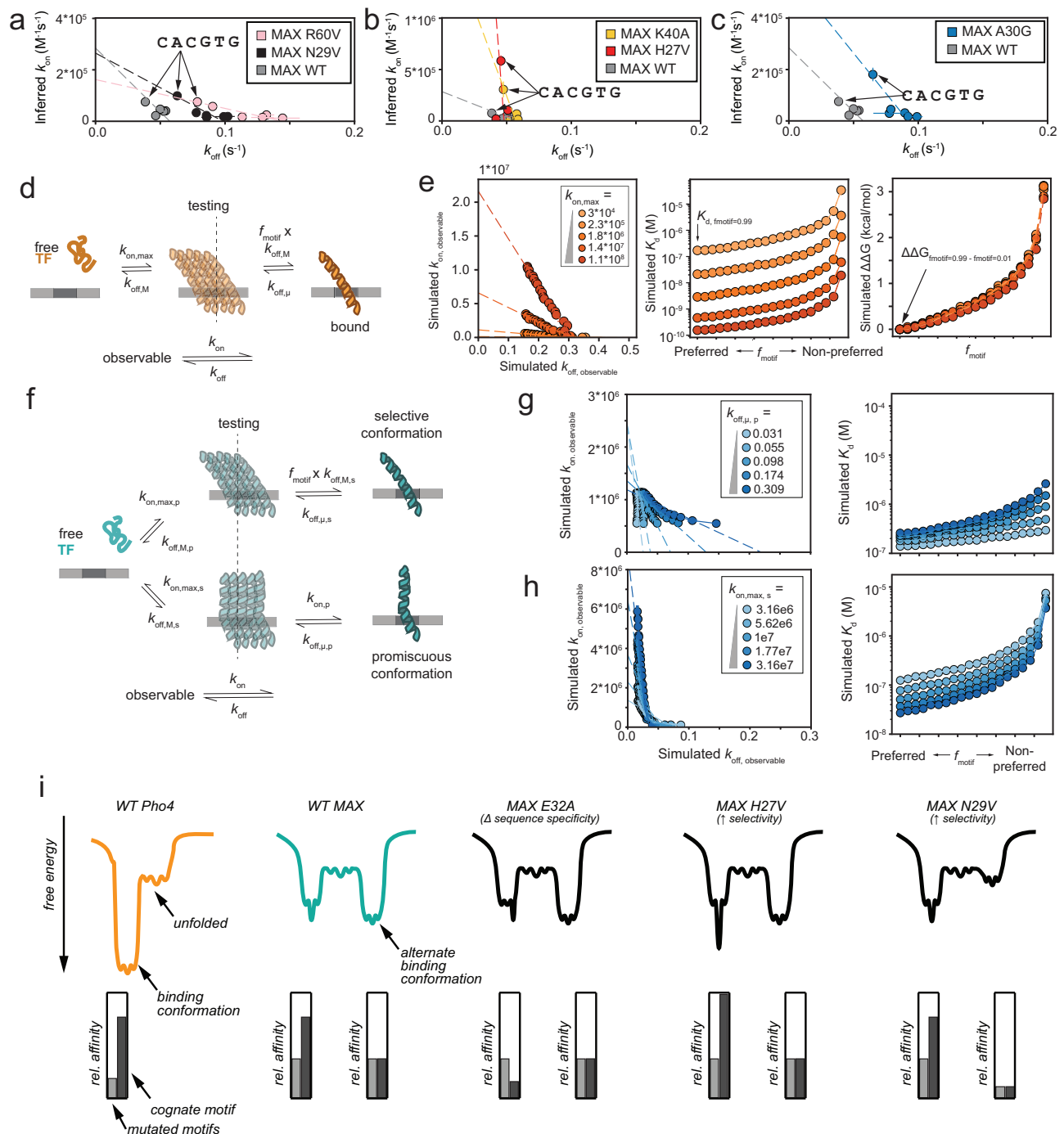
Fig. 6 | Selective mutations change the binding landscape through different microscopic mechanisms. a–**c** Inferred k_{on} versus measured k_{off} (both median ± SEM) for WT MAX and selective mutations across many E-box variants; dashed line indicates linear fit minimizing error in x- and y-dimensions. Replicate counts for all K_d and k_{off} measurements are contained in Supplementary Data 3. Source data are provided as a Source Data file. **d** Three-state model and associated microscopic rate constants for TF binding with a single bound conformation. **e** Simulated rate constants (left) and affinities (middle and right) for many f_{motif} values with binding model illustrated in (**d**). **f** Five-state model and associated microscopic rate constants for TF binding with multiple unbound and bound conformations with different intrinsic selectivities. **g**, **h** Simulated rate constants (left) and affinities (middle and right) for many f_{motif} values with binding model illustrated in (**f**). **i** Cartoon model illustrating idealized energetic landscapes for Pho4 (orange), MAX (teal), and three MAX mutations.

promiscuously bound conformations. Similarly, increasing $k_{on,max,s}$ reconstitutes measured changes in binding affinities and kinetics for selective mutations like H27V and K40A, increasing observable k_{on} while leaving observable k_{off} relatively unchanged (Figs. 4f and 6h). Consequently, these mutations may increase selectivity by changing the unbound ensemble to "preconfigure" certain conformations with side-chains positioned for specific recognition. Combinations of

changes to microscopic rate constants can also recapitulate more complex behaviors, such as those observed for A30G (Supplementary Fig. 58). While we find that models with two differentially selective states (and not models with a single binding conformation) are consistent with selective MAX mutant data, this toy model likely approximates two "macrostates" that are each the sum of some large number of microstates in which residues at the DNA interface are differentially

positioned within the folding landscape (Fig. 6i). We conclude that consideration of multiple states with different intrinsic selectivity for the same set of sequences is necessary to explain kinetic and thermodynamic data for MAX.

Discussion

Understanding the protein sequence determinants of selectivity—the quantitative difference in binding energy between preferred and non-preferred ligands—remains an unsolved biophysical challenge, with applications to binding generally beyond TF-nucleic acid interactions. Many bioengineering[57] and gene therapy[58] objectives require highly selective binding, yet mitigating off-target events remains challenging[59]. Thus, an enhanced understanding of the protein binding landscapes will also enhance the ability to engineer therapeutically useful protein-DNA interactions[57,60].

Here, we investigated how mutations to MAX, a promiscuous bHLH TF, alter binding to motif-variant DNA sequences in comparison to the structurally similar yet highly selective TF, Pho4. These measurements revealed putative non-contacting mutations in MAX that increased selectivity for the cognate motif via diverse molecular mechanisms. While some mutations likely stabilize selective microstates prior to binding (similar to mechanisms thought to drive antibody affinity maturation[61]), others change partitioning between different macroscopic conformations possessing different selectivities (Fig. 6i). Pho4, in contrast, lacks evidence of appreciable alternate binding states, suggesting highly selective binding may be achieved with narrow folding funnels (lacking ability to access alternate conformations or rearrangements) (Fig. 6i). Overall, our results demonstrate that high-throughput measurements of mutational impacts on binding affinities and kinetics can reveal important properties about conformational ensembles difficult to resolve via other methods, and that these properties can dramatically impact the selectivity of otherwise highly similar proteins.

The observed selectivity differences between Pho4 and MAX might represent distinct evolutionary pressures that stem from their different biological roles and speed/specificity tradeoffs within different genome sizes[62]. Pho4 initiates transcription in response to phosphate stress[63], while MAX acts as a heterodimerization node to control cell proliferation in concert with other TFs[64]. Previous observations—both that Pho4 binding is enriched at fewer genomic locations as a function of Hamming distance away from the cognate sequence when compared to other closely related yeast TFs[65,66] and that MAX dimers are enriched in non-canonical E-box binding[4,67]—suggest that the biochemical differences in selectivity investigated here may have in vivo relevance. Moreover, the observed decreased mutational sensitivity of MAX compared to Pho4 (Fig. 2c) may result from a need to preserve a wide variety of existing functions and reflect the fact that mutations in promiscuous binders may be more likely to yield functional binding proteins[68–71].

This work is aligned with many other investigations linking conformational ensembles to TF specificity, from bispecific binding to divergent motifs[19,23,72–74] to structural characterization of selective and promiscuous complexes[75–78]. These selective and promiscuous conformations are not just static bound states; TFs undergo conformational rearrangements between these complexes with varying degrees of selectivity as part of the binding pathway[24,79,80]. Moreover, the ability to access different conformations—and therefore bind increasingly diverse sites—can originate from decreased global fold stability[71] or differing degrees of frustration[81]. Currently, many structure-based binding algorithms cannot capture this information. We predict that incorporating conformational heterogeneity will be essential for properly predicting and engineering molecular specificity and selectivity. Our work suggests that prediction and design of selective binders (beyond TF-DNA interactions) will necessitate consideration of energy landscapes that govern both folding and recognition, fueled by systematic measurements of protein/ligand affinity, specificity, and selectivity.

Methods

Fabrication of microfluidic molds and devices

Flow and control molds were fabricated as described previously[65,82] and all design files are available on the Fordyce Lab website (http://www.fordycelab.com/microfluidic-design-files).

We fabricated two-layer MITOMI devices from these molds using polydimethylsiloxane (PDMS) polymer (RS Hughes, RTV615) in the Stanford Microfluidics Foundry. To fabricate the control layer, we combined 55 g of PDMS (1:5 ratio of cross-linker to base), mixed and degassed the components within a centrifugal mixer at 2000 (THINKY ARE-250, 11.2 g) and/or 2200 rpm (THINKY ARE-310, 17.g) for 3 min. We then poured the mixture onto the molds, degassed them in a vacuum chamber for 45 min under house vacuum, and baked them in an 80 °C convection oven for 40 min. We then cut control layers for individual devices from the cast PDMS and punched fluid inlet lines using a drill press (Technical Innovations) with a mounted catheter hole punch (SYNEO, CR0350255N20R4).

To fabricate the flow layer, we combined PDMS at a 1:20 ratio (cross-linker to base) and mixed and degassed the components within a centrifugal mixer at 2000 (THINKY ARE-250, 11.2 g) and/or 2200 rpm (THINKY ARE-310, 17.g) for 3 min. We then spin-cast the PDMS onto molds for 10 s at 500 rpm followed by 1850 rpm for 75 s. Spin-cast layers were allowed to relax on a flat surface at room temperature for 5 min before baking at 80 °C for 40 min. We then manually aligned individual control layers to the partially cured flow layer and baked the aligned devices for an additional 40 min at 80 °C. Bonded two-layer devices were cut from the flow mold with a scalpel and the flow-layer fluid inlet lines were punched as described above.

QuikChange mutagenesis for MAX mutant library

We generated a MAX plasmid carrying the full sequence of the MAX transcription factor with a c-terminal monomeric eGFP tag[83] separated from the MAX coding sequence via a gly-ser linker (GGSGGGGS). We used Gibson assembly to clone the MAX-eGFP fusion into a purified, linearized PURExpress vector with ampicillin resistance. The construct was sequenced validated using Sanger sequencing prior to generating mutants.

Primers encoding mutants were generated as described previously[26,84] using a custom-made program, available at (https://github.com/FordyceLab/designQuikChangePrimers). Briefly, the program takes as input the DNA sequence encoding the MAX ORF sequence and a list of desired mutants (e.g., "A67D" for Ala 67 to Asp mutation), generates a set of candidate primers for each mutant, and returns suggested mutagenic primer pairs scored according to criteria previously published in the QuikChange manual. Primers were ordered in a 96-well plate format from IDT (Integrated DNA Technologies) at the 10nmol synthesis scale with standard desalting purification; the forward and reverse primers for each mutant were premixed in each well. For library design, pathogenic mutations and VUS were curated from clinical sequencing databases as of March 2021.

Mutagenesis reactions were performed as previously reported[26] in a 96-well plate format. Each well contained its own mutagenesis reaction. Reactions were performed using the QuikChange protocol as directed by the manufacturer (Agilent Technologies, New England Biolabs). Upon completion of mutagenesis, we digested any remaining methylated wildtype plasmid using Dpn1 (New England Biolabs, R0176L) for 1 h at 37 °C. We then transformed 1 μL of each reaction into 5 μL of competent E. coli DH5alpha cells (New England Biolabs, C2987I). Transformants were grown to saturation in 5–8 mL of LB media supplemented with ampicillin (100 μg/mL) and miniprepped (Qiagen) for Sanger sequencing. To validate successful mutagenesis, we aligned each sequence to the template ORF and ensured that only the intended

mutation was present in the plasmid. We re-picked colonies in the event of errant mutations elsewhere in construct (eg. indels, additional mutations in plasmid), or poor sequencing quality.

Plasmid array printing

Prior to printing plasmids, we transferred mini-prepped plasmid into 96-well plates. To standardize volumes of plasmids, the wells were evaporated to dryness. We resuspended each plasmid with 50 μL of Milli-Q water. Plasmids were transferred from 96-well plates into 384-well plates using a Biomek FX Automated Workstation (Beckman Coulter, model A31843). Each plasmid was pipetted into 4 consecutive wells within the 384-well plate, and each well of the 384-well plate contained 10 μL of plasmid. We recorded positions to keep track of empty wells for adding subsequent mutants manually.

We evaporated 384-well plates to dryness at room temperature and resuspended dried wells in print buffer formulated as below:

- 1% (10 mg/mL) Bovine Serum Albumin (Sigma Life Science, B4287-25G)
- 200 mM (11.65 mg/mL) NaCl (Sigma Life Science, 71376-1KG)
- 12 mg/mL trehalose dihydrate (Sigma Life Science, T9531-25G)

All reagents were combined in Milli-Q water and mixed to dissolution at room temperature and sterile filtered to remove aggregates. To each well in the 384-well plate, we added 12–15 μL of print buffer for arrayer printing. When not in use, we sealed plasmid plates with foil covers and stored them at −20 °C. Prior to printing, plates were defrosted overnight at 4 °C and centrifuged at 2000 RPM/773 g for 5–10 min. Over the course of subsequent prints, we added ~3–5 μL of Milli-Q water (or additional print buffer) as needed to ensure sufficient volumes of sample in plates for printing.

We printed plasmids using a SciFlex Arrayer (SCIENION AG) using either the PDC50 or PDC70 nozzle (Type 1 coating). We generated a "field file" to map each well on a 384-well plate to positions within the printed plasmid array. To prevent cross-contamination between plasmids, the glass nozzle was washed with room temperature Milli-Q water in between spotting different plasmid samples. We printed plasmid arrays on epoxysilane-coated glass slides (ArrayIt SME2, SuperChip C50-5588-M20, or self-coated as previously described[85]). After drying arrays overnight at room temperature, we aligned microfluidic devices to "program" each chamber with its own plasmid spot. Prior to alignment, we pre-baked microfluidic devices at 80 °C for 20–25 min using a hotplate (Torrey Pines Scientific) and allowed them to cool to room temperature. These devices were then baked for 4–4.5 h at 95 °C on a hotplate.

Preparation of DNA for fluorescence-based binding assays

We designed all DNA sequences for binding assays with a 3′ region complementary to a AlexaFluor-647 dye-conjugated primer (anneal temperature: 37 °C) (Supplementary Table 4).

We ordered all DNA sequences as single-stranded oligonucleotides from Integrated DNA Technologies (IDT) with standard desalting purification and shipped in 'LabReady' formulation (100 μM in IDTE buffer, pH 8.0). We then duplexed these single-stranded DNA sequences by (1) annealing the universal AlexaFluor-647-labeled primer to the 3′ region of the oligonucleotide and (2) extension using the primer as a template using Klenow fragment, exo⁻, polymerase. Both steps (1) and (2) were performed as previously described[26].

After the Klenow extension, we filtered the DNA reactions using a 0.45 μm filter spin column. We subsequently equilibrated duplexed DNA in the final assay buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM DTT, pH 7.5; aliquoted and filtered using 0.45 mM Steriflip vacuum (Millipore, SE1M179M6)) using 10 K filter spin concentrator columns (Amicon Ultra, UFC501096). We added ~100 μL of the duplexed DNA to the filter spin columns, added 200 μL of assay buffer, mixed by pipetting, and concentrated the reaction to 100 μL by centrifugation

(9000 RPM/7.8 g for 8 min). We repeated this process 5 times, and subsequently eluted the equilibrated DNA via manufacturer's instructions for the 10 K filter spin concentrator column.

We serially diluted equilibrated DNA in final assay buffer as previously described[26]. For this dilution and the subsequent assay, the assay buffer was supplemented with 50 μg/mL of UltraPure BSA (ThermoFisher, AM2618). To calibrate each step of the binding assay with a DNA concentration, we measured the highest concentration of DNA using a DeNovix to measure absorbance at 260 nm.

For all experiments involving a mutation within the core-site, we also performed this procedure for the consensus DNA sequence 5′-C CACGTG A-3′. For these oligonucleotides, we measured binding isotherms for 5 DNA concentrations. For the sixth measurement, we introduced the duplexed and labeled reference DNA sequence at a high concentration (~7–9 μM) so that we could accurately quantify the saturation ratio with which to fit all binding isotherms.

Microscopy and instrumentation

We made measurements as previously described[26,84] using a Nikon Ti-S microscope. Devices were controlled using a pneumatics manifold[86]. Custom scripting and automation enabled integrated control of both the microscope and the pneumatics manifold (https://github.com/FordyceLab/RunPack)[26].

Measuring K_d values on-chip via STAMMP

Measuring K_d values on-chip have 3 major steps: (1) On-chip expression and purification of MAX mutants, (2) titration of fluorescently labeled DNA, (3) image analysis and calculation of K_d values.

On-chip expression & purification of MAX mutants

At assay start, "control" valve lines (Supplementary Fig. 2) were dead-end filled with 550 mM NaCl solution to prevent premature solubilization of arrayed DNA by osmotic balancing. All other control lines were dead-end filled with Milli-Q water. All pneumatic valves on the device were pressurized at 32–35 psi and reagents were introduced at pressures of 3–4 psi.

On-chip expression and purification require the immobilization of expressed proteins for subsequent assay steps. To accomplish this, we performed a series of passivation steps as described previously[26,84] to immobilize biotinylated anti-GFP (Abcam ab6658, loaded at 1:20 dilution in PBS) antibodies selectively underneath the 'button valves' of the STAMMP microfluidic device. As all TF variants are fused to a GFP, these antibodies will trap recruited TF variants for subsequent assays. Briefly, the device flow layer was first dead-end filled with biotinylated BSA (2 mg/mL, ThermoFisher Pierce, 29130). After all air bubbles were expelled, we opened the outlet valve to allow reagent flow, and introduced the biotinylated BSA for an additional 5 min with the "button" valves pressurized and then for 30 min with the "button" valves open. Next, we flushed the device with phosphate buffered saline (10X stock, Corning, 46-013-CM; diluted to 1X in Milli-Q water) for 10 min. We then introduced neutravidin (1 mg/mL, Thermo Scientific, 31000) for 30 min with "button" valves opened, followed by another PBS wash for 10 min and an additional biotinylated BSA wash with the "button" valves pressurized again for 30 min. After an additional 10 min PBS wash, we introduced biotinylated anti-GFP antibody (100 μg/mL, Abcam, ab6658) into the device for 2 min with the "button" valves pressurized and then opened "button" valves and flowed for an additional 11 min and 20 s. Finally, we washed the device with PBS for 10 min.

To express all TF variants simultaneously, we used PURExpress (NEB E6800L). Briefly, we equilibrated Parts A and B of PURExpress on ice until defrosted. For one device using 25 μL total of PURExpress, we first incubated 10 μL of Part A with 7.5 μL of Part B on ice for 45 min. Then, we added 1.5 μL of recombinant RNAsin (Promega N2515) and 6 μL of nuclease-free water (Promega P1193) and mixed by pipette until

no phase separation was visible. We introduced PURExpress onto the device as previously described[26,84] Devices were then placed on a pre-heated hotplate at 37 °C for 45 minutes to express all proteins. We then placed devices on the scope and allowed the GFP to fold over the course of 45-60 minutes with the button valves on the device closed. After this was completed, we opened the button valves and recruited GFP-tagged protein to the antibodies for 20–30 min. We then closed the buttons to shield trapped TFs while we washed the device with PBS and TrypLE (ThermoFisher 12604-013) to remove nonspecifically bound TFs from the device walls. After this, we equilibrated the device with assay buffer to remove trace amounts of TrypLE, composed of the following unless otherwise specified:

- 20 mM Tris-HCl pH 7.5 (from 100 mM stock)
- 100 mM NaCl (from 100 mM stock)
- 1 mM DTT (from 1 M stock) (Sigma-Aldrich, D9779)
- 50 μg/mL ultrapure BSA (ThermoFisher, AM2618)

DNA binding measurements

To perform binding affinity measurements, we introduced fluorescent DNA (prepared as described above) at 6 concentrations ranging between ~60 nM to ~6 μM on the device. At each concentration step, DNA was introduced to the device by:

(1) closing "neck" and "button" valves and opening "sandwich" and inlet and outlet valves

(2) flowing labeled DNA across the device for 10 min

(3) closing "sandwich", inlet, and outlet valves

(4) opening "button" valves and incubating for 20 min to allow reactions to come to equilibrium

(5) imaging all chambers within the device in the Cy5 (DNA) channel

(6) closing "button" valves (to trap TF-bound DNA)

(8) washing with assay buffer for 10 min

(9) imaging all chambers in both the GFP (TF) and Cy5 (DNA) channels to quantify the relative intensities of trapped species in each chamber.

For binding measurements with DNA sequences containing mutations within the core binding site, only five concentration points were measured. For the sixth and final concentration point, we measured DNA binding for the reference DNA sequence 5′-C CACGTG A-3′ at a high concentration to determine DNA to protein fluorescence intensity ratio denoting saturation of all binding sites for global fitting K_d values. For prewash Cy5 images, we imaged the device at multiple exposure times, ranging from 30 ms to 100 ms. We imaged postwash GFP images using an exposure time of 500 ms. For postwash Cy5 images, we used exposure times of either 1200 ms or 3000 ms to ensure we did not collect measurements at a saturating intensity.

Quantification of bound DNA fluorescence intensities

Images were stitched using in-house Python packages ImageStitcher (https://github.com/FordyceLab/ImageStitcher)[26]. These images were then analyzed using the ProcessingPack package (https://github.com/FordyceLab/ProcessingPack)[26]. To quantify affinities for each TF mutant binding to a given DNA sequence, we acquired per-chamber calibration curves relating observed AlexaFluor-647 fluorescence to spectroscopically measured dsDNA concentrations (Supplementary Fig. 7), converted intensities to DNA concentrations based on orthogonal measurements using a DeNovix instrument, and then fit concentration-dependent binding curves as described below.

To identify TF mutants with DNA binding statistically indistinguishable from background, we compared Cy5 intensities from TF-containing chambers with those from blank chambers by repeated measures ANOVA (providing a conservative estimate of mutants with detectable binding); we report measured K_ds for these variants as a lower limit (Supplementary Fig. 8, Supplementary Table 3).

Fitting K_d values

To fit dissociation constants, we first measured the amount of DNA bound to surface-immobilized TF mutants over multiple concentrations and converted these to ratios of bound DNA intensities (Cy5 channel) over immobilized TF (eGFP channel). We then applied a global fit to the measured DNA/TF ratios and fit data from each individual chamber to single-site binding models[25,82].

$$R([\text{DNA}]) = \frac{R_{\max} \cdot [\text{DNA}]}{K_d + [\text{DNA}]} \quad (1)$$

$$R([\text{DNA}]) = \frac{(R_{\max} - C) \cdot [\text{DNA}]}{K_d + [\text{DNA}]} + C \quad (2)$$

Here, R is the intensity of DNA/TF as a function of DNA concentration within the chamber, $R_{\max}$ is the constant shared across all chambers corresponding to the value at which all binding curves saturate (assuming an identical molecular stoichiometry), [DNA] corresponds to the concentration of free DNA within the chamber, and K_d is the dissociation constant for a particular chamber. We determined the $R_{\max}$ value by taking the median of the top 10% of DNA binding MAX mutants at the highest DNA concentration point in an experiment for the reference DNA sequence; for experiments with a mutated DNA sequence, we measured the highest DNA concentration point using a reference DNA sequence to prevent underestimation of $R_{\max}$.

In addition to fitting to a Langmuir isotherm, we fit our data to a modified single-site binding model with an offset value, C, to correct for variations in background intensities between experiments that can affect ratio values. The fitting method that minimized per-chamber RMSE of fits for each technical replicate was used for final determination and export of K_d values.

Measuring k_{off} values on-chip via k-STAMMP

At the end of a STAMMP binding assay, k_{off} values can also be optionally obtained. Measuring k_{off} values on-chip adds two additional steps to a STAMMP assay: (1) titration of unlabeled DNA and (2) image analysis and calculation of kinetic constants.

Dissociation measurements

Dissociation rate data was acquired after equilibrium binding procedures largely as previously described[6,51]. We first flushed each device with non-fluorescent (dark) competitor dsDNA oligonucleotides containing an E-box motif at a concentration of ~0.9 μM diluted in assay buffer for 10 min with button valves closed after the acquisition of the "post-wash" image. These oligos were prepared with Klenow polymerase as described above but with unlabeled primers. The inclusion of non-fluorescent (dark) competitor at high concentrations during dissociation is critical to prevent rebinding of labeled material, which leads to systematic underestimation of dissociation rates[6].

Next, after stopping flow of unlabeled competitor dsDNA and closing sandwich valves, we then opened the buttons for 2.0 s to allow dissociation of bound fluorescent DNA from surface-immobilized TF. Finally, we closed buttons, flushed the device, and imaged in both the Cy5 and eGFP channels to quantify loss of DNA binding and surface-bound TF, respectively. For each experiment, we iterated this process for 40 button duty cycle iterations.

Fitting of kinetic constants

After acquiring and processing images as described for STAMMP assays, kinetic constants (k_{off}) were determined by first calculating the ratio (R) of "post-wash" DNA fluorescence (Alexa 647) to "post-wash" GFP fluorescence per chamber at each time point. This ratio was then

used to fit a single exponential value:

$$R(t) = R_{t=0} \cdot e^{-k_{off} \cdot t} + C \quad (3)$$

where $R(t)$ is the fluorescence ratio as a function of time, k is the dissociation constant, and C is a constant term which accounts for background fluorescence or non-specific sticking of DNA. From these fitted k_{off} values, we can infer k_{on} through the definition of the dissociation constant from measured K_d and measured k_{off} as previously described[6,52]:

$$k_{on} = \frac{k_{off}}{K_d} \quad (4)$$

Fold-reduction in binding from MITOMI measurements

To calculate fold-reduction in binding for MAX and Pho4 from previous measurements[25], we collected the measured affinities for all sequences with a Hamming distance of one away from the consensus motif (Fig. 1e), compared these binding affinities to the median of all consensus motif measurements (with variable flank nucleotides), and calculated and reported the median for fold-reduction in binding.

Calculation of conservation

To calculate conservation of individual positions in the bHLH DBD, we used curated multiple sequence alignment (MSA) of bHLH DBDs (PF00010 [https://www.ebi.ac.uk/interpro/entry/pfam/PF00010/entry_alignments/]). We culled this MSA to only include non-gapped positions and positions aligned to MAX. Finally, we re-aligned this filtered MSA with the skbio.alignment.TabularMSA function, and calculated conservation from this using the skbio.alignment.TabularMSA.conservation function (Fig. 2c).

Thermodynamic modeling of folding and binding equilibria

Thermodynamic model fitting and binding simulations were defined by the following variables:

(1) H, the percentage of TF that is folded (helical) in solution.

(2) C, the percentage of TF that is unfolded (coil) in solution.

(3) D, the concentration of free DNA.

(4) Co (for Complex), the bound TF-DNA complex.

(5) pT (for total protein), the total amount of TF available in the reaction.

(6) dT (for total DNA), the total amount of DNA available in the reaction.

These variables were used to construct the following equations and define equilibrium constants:

(1) Mass balance equation for protein species, defined as:

$$pT = H + C + Co \quad (5)$$

(2) Mass balance equation for DNA, defined as:

$$dT = D + Co \quad (6)$$

(3) Equilibrium constant defining partitioning between folded and unfolded states in the unbound state:

$$K_{fold} = \frac{H}{C} \quad (7)$$

(4) True binding equilibrium constant, where only the folded (helical) form can complex with DNA:

$$K_d = \frac{H * D}{Co} \quad (8)$$

In a STAMMP experiment, only the total amount of free DNA, total amount of immobilized TF, and fractional occupancy of bound TF-DNA complex is known; the distribution of folded/unfolded unbound states and true values underlying equilibrium constants is not known. Therefore, first we used the preceding 4 equations and sympy.solve to define: Co(pT, dT; K_{fold}, K_d), the concentration of bound TF-DNA complexes as a function of total protein and DNA given the folding and binding equilibrium constants, as follows:

$$Co(pT, dT; K_{fold}, K_d) = \frac{1}{2}[dT + K_d K_{fold} + K_d + pT$$
$$- \sqrt{dT^2 + 2dT K_d K_{fold} + 2dT K_d - 2dTpT + K_d^2 K_{fold}^2 + 2K_d^2 K_{fold} + K_d^2 + 2K_d K_{fold}pT + 2K_dpT + pT^2}] \quad (9)$$

For all subsequent calculations, pT was defined as 50 nM, based on previous estimates for the concentration of immobilized protein on MITOMI microfluidic devices[84]. Apparent (measured) DNA-binding affinities were then obtained by: (1) calculating equilibrium occupancy of Co at dT spanning from 0 to 10 μM, analogous to the procedure for measuring binding affinities in STAMMP experiments, and then (2) defining the dT resulting in $\frac{Co}{pT} = 0.5$ as the apparent DNA-binding $K_{d, apparent}$.

TF mutations that alter the propensity to fold or unfold in the unbound state can also alter the observed DNA binding affinity. We assumed that all surveyed TF mutations only change the free energy difference between the folded and unfolded state, changing K_{fold} but not K_d. The amount by which a TF mutation changes K_{fold} can then be defined as the change in helical propensity relative to WT TF, which changes the folding equilibrium as follows:

$$K_{fold, mutant} = \frac{K_{fold}}{e^{\frac{\Delta\Delta G_{HP}}{-1.987*10^{-3}*298}}} \quad (10)$$

where $\Delta\Delta G_{HP}$ is defined as the change in helical propensity, which defines the free energy difference for partitioning between unfolded and folded, helical states[39].

Fitting STAMMP data to derived thermodynamic model

To first develop intuition for how $K_{d,true}$, K_{fold}, and $\Delta\Delta G_{HP}$ alter the range of $K_{d,measured}$ for both WT and mutant TFs, we determined how the expected linear free energy relationships between helical propensity-altering TF mutants and apparent DNA binding affinity impact $K_{d,measured}$ for TF mutations with intrinsic changes in helical propensity spanning −2.0 to 2.5 kcal/mol (Supplementary Figs. 14–15).

Next, we fit $K_{d,measured}$ for MAX and Pho4 to the thermodynamic model defined for Co(K_d, K_{fold}, dT, pT) to extract $K_{d,true}$ and K_{fold} for WT TF (Supplementary Fig. 16). To accomplish this, we first restricted our analysis to TF mutations in the basic region that do not make crystal contacts with DNA or at the dimerization interfaces, as these mutants presumably alter $K_{d,measured}$ through mechanisms other than changes to K_{fold}. Each of these TF mutations was then defined to have a $\Delta\Delta G_{HP}$ in accordance with previously measured changes in Gibbs free energy for helix formation[39]. Next, for both MAX and Pho4, we calculated the RMSE of $\log_{10}$ thermodynamic model-predicted $K_{d,measured}$ for all measured TF mutations to the STAMMP-derived $\log_{10} K_{d,measured}$ for values of $K_{d,true,WT}$ ranging from 1 to 103.5 nM and values of $K_{fold,WT}$ where the fraction of unfolded TF in the unbound state ranged from 99 to 1%. The fitted values of $K_{d,true,WT}$ and $K_{fold,WT}$ were defined as those that minimized RMSE for the mutations and $K_{d,measured}$ measured via STAMMP relative to the thermodynamic model of folding-and-binding. Code to reproduce all simulations and fitting procedures is available at https://osf.io/jmz8t.

Identification of affinity altering mutations

To identify affinity-altering mutations, we first performed boot-strapped equivalence testing for all MAX mutations, rejecting the null hypothesis that a given mutation does not equally alter affinity for all sequences ($||zi = \bar{z}|| > \delta$) if $p < 0.05$. For this procedure, we used a noise tolerance threshold δ set by the noise in WT MAX measurements (the median standard deviation in measured K_ds for across all motifs). Next, we exclude mutations that are WT-like ($\Delta\Delta G \approx 0$) by filtering mutations that do not significantly alter affinities relative to WT MAX ($p < 0.05$ via independent t-test) in any measured E-box sequences. Finally, we excluded mutations unresolvable from background binding to any DNA sequence (Supplementary Fig. 8, Supplementary Table 3) from our definition of affinity-altering.

Identification of non-additive TF/DNA mutation pairs

Identifying epistasis across the TF-DNA interface requires 4 affinity measurements: (1) WT TF binding a 'reference' DNA sequence, (2) mutant TF binding a 'reference' DNA sequence, (3) WT TF binding a 'mutant' DNA sequence, and (4) mutant TF binding a 'mutant' DNA sequence. We then determined if each pair was statistically significantly non-additive in K_d space, largely as previously reported[26].

Briefly, to estimate the affinity that would have been expected if the energetic effects of TF and oligonucleotide mutations were purely additive, we first calculated an expected 'additive' K_d value using the median reference K_d value (for WT TF interacting with the 'reference' oligonucleotide), the median K_d resulting from the relevant oligonucleotide mutation alone, and the median K_d resulting from the TF mutation alone as follows:

$$K_d^{TF_{mut}, DNA_{mut}} = \frac{K_d^{TF_{mut}, DNA_{CACGTG}} * K_d^{TF_{WT}, DNA_{mut}}}{K_d^{TF_{WT}, DNA_{CACGTG}}} \quad (11)$$

To determine whether the candidate TF mutant appeared epistatic with the DNA nucleotide mutation, we used measurements of (1) WT MAX and CACGTG, (2) WT MAX and mutant DNA, and (3) mutant MAX and WT DNA to generate a distribution of additive K_d measurements ($n = 500$ simulated additive measurements). We then performed an independent t-test comparing the distribution of additive affinities with the experimentally measured affinities for the double-mutant and used a Bonferroni-corrected p-value cut-off of 0.05 to define TF mutants that are epistatic with DNA mutants.

Identification of selective MAX mutations

To identify mutations that differentially increase selectivity, we computed residuals for each pairwise comparison between a mutated E-box motif and the CACGTG cognate, calculated Z-scores for each residual (to account for the fact that residual distributions vary with absolute affinity), and defined a score as the median of all Z-scores across each double mutant cycle comparison (Fig. 4c, Supplementary Fig. 35), with selective mutations exceeding a threshold defined by the standard deviation of the Gaussian fit to the residuals (Supplementary Fig. 37). Mutations which were unresolvable from background in the cognate motif measurement (for which reported K_ds are underestimated) were excluded from the list of reported selective mutations, and candidate selective mutations were inspected and culled by eye.

Gillespie model of TF binding kinetics

Gillespie algorithms are stochastic simulations based on reaction rates that use discrete molecule counts and variable time steps[87]. Here, we simulated TF selectivity using Gillespie models of different binding pathways depicted in Fig. 6d, f. At each time step, we compute: (1) How long until the next reaction occurs? and (2) Which reaction happens?

First, we calculated reaction propensities (a) from reaction probabilities (c) and the number of reactants available for each

reaction. Reaction probabilities can be derived from the kinetic rate constants as previously described[6]. For all simulations, microscopic rate parameters previously estimated from CTMC modeling were used as a starting estimation[6]. In this model, we initialized with 1 molecule of MAX and DNA and set the volume to 1.66×10^{-12} pL (chosen for simplicity so simulated s^{-1} values equal M s^{-1} on-rate constants).

Observed off-rates (macroscopic k_{off}) were calculated as the median value across 3 replicates of the inverse of the average time it takes MAX to become fully dissociated once specifically bound; observed on-rates (macroscopic k_{on}) were calculated as the median value across 3 replicates of the inverse of the average time it takes MAX to become specifically bound once dissociated in solution. The observed K_d is calculated as the ratio between macroscopic on- and off-rates.

To calculate selectivity, we calculated k_{off}, k_{on}, and K_d for a range of "motifs", ranging from strongly to weakly bound sequences. Motif "strength" is defined in all models by f_{motif}, an implicit parameterization of the probability of binding such that an increase in f_{motif} (tightly bound motifs) causes an increase in association rate and a decrease in off-rate, as previously described[6,53]. Given that all observed specificity-increasing mutations do not occur at conserved nucleotide contacting residues, we assumed that TF mutations do not change the intrinsic probability of recognizing a motif (f_{motif}) but instead only alter microscopic rate parameters. Selectivity was defined as the free energy difference between the strongest ($f_{motif} = 0.99$) and weakest ($f_{motif} = 0.01$) motif surveyed. Code to reproduce all simulations is available at https://osf.io/jmz8t.

Sensitivity analysis for 3-state model

Reaction likelihoods were defined according to the 3-state model consisting of unbound, testing, and bound states shown in Fig. 6d. Simulated k_{on}, k_{off}, and K_d values across 20 "motifs" of strengths ranging from $f_{motif} = 0.99$ to 0.01 were obtained by coarsely varying 3 free microscopic rate constants $k_{on, max}$, $k_{off, max}$, and $k_{off, u}$ across 4 orders of magnitude each in 10 step increments. For each combination of free parameters, we simulated 3 independent trajectories with 10^4 reaction steps. The resulting free energy difference between the tightest and weakest surveyed motifs was calculated.

Sensitivity analysis for 4-state model

Reaction likelihoods were defined according to the 4-state model consisting of an unbound binding-incompetent conformation, an unbound binding-competent conformation, testing states, and bound states shown in Supplementary Fig. 55. Simulated k_{on}, k_{off}, and K_d values across 20 "motifs" of various strengths (from $f_{motif} = 0.99$ to 0.01) were obtained by coarsely varying 3 free microscopic rate constants $k_{on, max}$, $k_{off, max}$, and $k_{off, u}$ across 4 orders of magnitude each in 10 step increments, and an additional equilibrium folding constant K_{fold} (defined as $k_{on, fold}/(1- k_{on, fold})$) over 5 increments (spanning percent folded in the unbound state from 1 to 99%). For each combination of free parameters, we simulated 3 independent trajectories with 10^4 reaction steps. The resulting free energy difference between the tightest and weakest surveyed motifs was calculated to report selectivity.

Sensitivity analysis for model with multiple conformations

Reaction likelihoods were defined according to a 5-state model of consisting of unbound TF, a selective testing and bound state, and a promiscuous testing and bound state (Fig. 6f). Simulated k_{on}, k_{off}, and K_d values across 10 "motifs" of various strengths (from $f_{motif} = 0.99$ to 0.01) were obtained by coarsely varying 6 free microscopic rate constants $k_{on, max, s}$, $k_{off, max, s}$, $k_{off, u, s}$, $k_{on, max, p}$, $k_{off, max, p}$, and $k_{off, u, p}$ across 4 orders of magnitude each in 5 step increments. For each combination of free parameters, we simulated 3 independent trajectories with 3×10^3 reaction steps. The resulting free energy difference

between the tightest and weakest surveyed motifs was calculated to report on selectivity. For relevant parameter spaces, trajectories were re-simulated with 10^4 steps across 20 "motifs" with 5 independent replicates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The processed affinity and kinetic measurements used in main text Figures are reproduced as Source Data; all other data (including quantified imaging data, fitted affinity and dissociation measurements, and simulated data) used to generate Supplemental Figures, Tables, and analyses are deposited as an Open Science Framework under accession code jmz8t (https://osf.io/jmz8t/) in the Data and Simulations subfolders, with a README.md file delineating which data relate to which figure. Due to the large size of the imaging data, raw microscopy images are available upon request. Figures using protein structures were generated using PDB accession no. 1HLO or 5EYO (for MAX), as well as 1A0A (for Pho4). Source data are provided with this paper.

Code availability

All source code generated and used in this study for simulations and figure generation is available as an Open Science Framework in the Figures and Simulations3 subfolders under accession code jmz8t (https://osf.io/jmz8t/) as well as Zenodo[88] (https://doi.org/10.5281/zenodo.14218355).

References

1. Slattery, M. et al. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
2. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* **35**, 357–379 (2019).
3. Bruno, L. et al. Selective deployment of transcription factor paralogs with submaximal strength facilitates gene regulation in the immune system. *Nat. Immunol.* **20**, 1372–1380 (2019).
4. Allevato, M. et al. Sequence-Specific DNA Binding by MYC/MAX to Low-Affinity Non-E-Box Motifs. *PLOS ONE* **12**, e0180147 (2017).
5. Shen, N. et al. Divergence in DNA specificity among paralogous transcription factors contributes to their differential in vivo binding. *Cell Syst* **6**, 470–483.e8 (2018).
6. Horton, C. A. et al. Short Tandem Repeats Bind Transcription Factors to Tune Eukaryotic Gene Expression. *Science* **381**, eadd1250 (2023).
7. Crocker, J. et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
8. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471.e4 (2020).
9. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature* **451**, 535–540 (2008).
10. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **16**, 962–972 (2006).
11. Jindal, G. A. et al. Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. *Dev. Cell* **58**, 2206–2216 (2023).
12. Lim, F. et al. Affinity-optimizing enhancer variants disrupt development. *Nature* **626**, 15–159 (2024).
13. Scardigli, R., Bäumer, N., Gruss, P., Guillemot, F. & Le Roux, I. Direct and Concentration-Dependent Regulation of the Proneural Gene *Neurogenin2* by Pax6. *Development* **130**, 3269–3281 (2003).
14. Jiang, J. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741–752 (1993).
15. Crocker, J. Preger-Ben Noon, E. & Stern, D. L. The Soft Touch. In *Current Topics in Developmental Biology*, (Elsevier, 2016)
16. Turner, E. C., Cureton, C. H., Weston, C. J., Smart, O. S. & Allemann, R. K. Controlling the DNA binding specificity of bHLH proteins through intramolecular interactions. *Chem. Biol.* **11**, 69–77 (2004).
17. Künne, A. G. E. & Allemann, R. K. Covalently linking BHLH subunits of MASH-1 increases specificity of DNA binding. *Biochemistry* **36**, 1085–1091 (1997).
18. Liu, J. et al. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888 (2006).
19. Morgunova, E. et al. Two distinct DNA sequences recognized by transcription factors represent enthalpy and entropy optima. *eLife* **7**, e32963 (2018).
20. Cave, J. W., Wemmer, D. E. & Kremer, W. Backbone dynamics of sequence specific recognition and binding by the yeast *Pho4* bHLH domain probed by NMR. *Protein Sci.* **9**, 2354–2365 (2000).
21. Sauvé, S., Tremblay, L. & Lavigne, P. The NMR solution structure of a mutant of the max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J. Mol. Biol.* **342**, 813–832 (2004).
22. Fuxreiter, M., Simon, I. & Bondos, S. Dynamic protein–DNA recognition: beyond what can be seen. *Trends Biochem. Sci.* **36**, 415–423 (2011).
23. Rogers, J. M. et al. Bispecific forkhead transcription factor FoxN3 recognizes two distinct motifs with different DNA shapes. *Mol. Cell* **74**, 245–253.e6 (2019).
24. Ferreiro, D. U. & De Prat-Gay, G. A Protein–DNA binding mechanism proceeds through multi-state or two-state parallel pathways. *J. Mol. Biol.* **331**, 89–99 (2003).
25. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
26. Aditham, A. K., Markin, C. J., Mokhtari, D. A., DelRosso, N. & Fordyce, P. M. High-Throughput affinity measurements of transcription factor and DNA mutations reveal affinity and specificity determinants. *Cell Syst.* **12**, 112–127.e11 (2021).
27. Shimizu, T. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J.* **16**, 4689–4697 (1997).
28. Brownlie, P. et al. The crystal structure of an intact human Max–DNA complex: new insights into mechanisms of transcriptional control. *Structure* **5**, 509–520 (1997).
29. Dill, K. A. & Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Mol. Biol* **4**, 10–19 (1997).
30. De Masi, F. et al. Using a structural and logics systems approach to infer bHLH–DNA binding specificity determinants. *Nucleic Acids Res.* **39**, 4553–4563 (2011).
31. Shammas, S. L. Mechanistic roles of protein disorder within transcription. *Curr. Opin. Struct. Biol.* **42**, 155–161 (2017).
32. Vuzman, D. & Levy, Y. Intrinsically disordered regions as affinity tuners in Protein–DNA interactions. *Mol BioSyst* **8**, 47–57 (2012).
33. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**. https://doi.org/10.1126/scisignal.2004088c (2013).
34. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
35. Maerkl, S. J. & Quake, S. R. Experimental determination of the evolvability of a transcription factor. *Proc. Natl. Acad. Sci. USA* **106**, 18650–18655 (2009).
36. Nair, S. K. & Burley, S. K. X-Ray structures of Myc-Max and Mad-Max recognizing DNA. *Cell* **112**, 193–205 (2003).

37. Meinhardt, S., Manley, M. W., Parente, D. J. & Swint-Kruse, L. Rheostats and toggle switches for modulating protein function. *PLoS ONE* **8**, e83502 (2013).

38. Page, B. M. et al. Odd One out? Functional tuning of *Zymomonas mobilis* Pyruvate Kinase is narrower than its allosteric, human counterpart. *Protein Sci.* **31**, e4336 (2022).

39. O'Neil, K. T. & DeGrado, W. F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**, 646–651 (1990).

40. Afek, A. et al. DNA mismatches reveal conformational penalties in Protein–DNA recognition. *Nature* **587**, 291–296 (2020).

41. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl Acad. Sci. USA* **112**, 9614–9619 (2015).

42. Sicoli, G., Vezin, H., Ledolter, K., Kress, T. & Kurzbach, D. Conformational tuning of a DNA-Bound transcription factor. *Nucleic Acids Res.* **47**, 5429–5435 (2019).

43. Acharya, A., Rishi, V. & Vinson, C. Stability of 100 Homo and Heterotypic Coiled–Coil **a** – **a** ' Pairs for Ten Amino Acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* **45**, 11324–11332 (2006).

44. Horovitz, A., Fleisher, R. C. & Mondal, T. Double-mutant cycles: new directions and applications. *Curr. Opin. Struct. Biol.* **58**, 10–17 (2019).

45. Perna, D. et al. Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. *Oncogene* **31**, 1695–1709 (2012).

46. del Olmo Toledo, V., Puccinelli, R., Fordyce, P. M. & Pérez, J. C. Diversification of DNA binding specificities enabled SREBP transcription regulators to expand the repertoire of cellular functions that they govern in fungi. *PLOS Genet.* **14**, e1007884 (2018).

47. Parraga, A., Bellsolell, L., Ferré-D'Amaré, A. & Burley, S. K. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 å resolution. *Structure* **6**, 661–672 (1998).

48. Wang, D. et al. MAX is an epigenetic sensor of 5-Carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **45**, 2396–2407 (2017).

49. Pagano, L. et al. Double mutant cycles as a tool to address folding, binding, and allostery. *Int. J. Mol. Sci.* **22**, 828 (2021).

50. Shammas, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I. M. & Clarke, J. Insights into coupled folding and binding mechanisms from kinetic studies. *J. Biol. Chem.* **291**, 6689–6695 (2016).

51. Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci. USA* **109**, 16540–16545 (2012).

52. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and evaluate binding affinities. *eLife* **9**, e57264 (2020).

53. Marklund, E. et al. Sequence Specificity in DNA binding is mainly governed by association. *Science* **375**, 442–445 (2022).

54. Iešmantavičius, V., Dogan, J., Jemth, P., Teilum, K. & Kjaergaard, M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chem. Int. Ed.* **53**, 1548–1551 (2014).

55. Ecevit, O., Khan, M. A. & Goss, D. J. Kinetic analysis of the interaction of b/HLH/Z transcription factors Myc, Max, and Mad with Cognate DNA. *Biochemistry* **49**, 2627–2635 (2010).

56. Leffler, J. E. Parameters for the description of transition states. *Science* **117**, 340–341 (1953).

57. Ichikawa, D. M. et al. A universal deep-learning model for Zinc finger design enables transcription factor reprogramming. *Nat. Biotechnol.* **41**, 1117–1129 (2023).

58. Yee, J. Off-target effects of engineered nucleases. *FEBS J.* **283**, 3239–3248 (2016).

59. Bogdanove, A. J., Bohm, A., Miller, J. C., Morgan, R. D. & Stoddard, B. L. Engineering altered Protein–DNA recognition specificity. *Nucleic Acids Res.* **46**, 4845–4871 (2018).

60. Glasscock, C. J. et al. Computational Design of Sequence-Specific DNA-Binding Proteins; preprint (2023).

61. Schmidt, A. G. et al. Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody. *Proc. Natl. Acad. Sci. USA* **110**, 264–269 (2013).

62. Suter, D. M. Transcription factors and DNA play hide and seek. *Trends Cell Biol.* **30**, 491–500 (2020).

63. Kaffman, A., Rank, N. M. & O'Shea, E. K. Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. *Genes Dev.* **12**, 2673–2683 (1998).

64. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad Network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16**, 653–699 (2000).

65. Le, D. D. et al. Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proc. Natl. Acad. Sci. USA* **115**, E3702–E3711 (2018).

66. Zhou, X. & O'Shea, E. K. Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. *Mol. Cell* **42**, 826–836 (2011).

67. Guo, J. et al. Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biol.* **15**, 482 (2014).

68. Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11**, 20140419 (2014).

69. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).

70. Meier, S. & Özbek, S. A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution? *BioEssays* **29**, 1095–1104 (2007).

71. Koulechova, D. A., Tripp, K. W., Horner, G. & Marqusee, S. When the Scaffold cannot be ignored: the role of the hydrophobic core in ligand binding and specificity. *J. Mol. Biol.* **427**, 3316–3326 (2015).

72. Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).

73. Kophengnavong, T., Michnowicz, J. E. & Blackwell, T. K. Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformations. *Mol. Cell. Biol.* **20**, 261–272 (2000).

74. Fordyce, P. M. et al. Basic Leucine zipper transcription factor Hac1 Binds DNA in two distinct modes as revealed by microfluidic analyses. *Proc. Natl. Acad. Sci.* USA **109**, E3084-E3093 (2012)

75. Kalodimos, C. G. et al. Structure and flexibility adaptation in nonspecific and specific Protein-DNA Complexes. *Science* **305**, 386–389 (2004).

76. Aishima, J. & Wolberger, C. Insights into nonspecific binding of homeodomains from a structure of MATα2 Bound to DNA. *Proteins Struct. Funct. Bioinform.* **51**, 544–551 (2003).

77. Iwahara, J., Zweckstetter, M. & Clore, G. M. NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. USA* **103**, 15062–15067 (2006).

78. Ye, X. et al. Two distinct binding modes provide the RNA-binding protein RbFox with extraordinary sequence specificity. *Nat. Commun.* **14**, 701 (2023).

79. Sánchez, I. E., Ferreiro, D. U., Dellarole, M. & De Prat-Gay, G. Experimental snapshots of a protein-DNA Binding Landscape. *Proc. Natl. Acad. Sci. USA* **107**, 7751–7756 (2010).

80. Ferreiro, D. U., Sanchez, I. E. & de Prat Gay, G. Transition state for Protein-DNA recognition. *Proc. Natl. Acad. Sci. USA* **105**, 10797–10802 (2008).

81. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration, function and folding. *Curr. Opin. Struct. Biol.* **48**, 68–73 (2018).

82. Fordyce, P. M. et al. De Novo identification and biophysical characterization of transcription factor binding sites with microfluidic affinity analysis. *Nat. Biotechnol.* **28**, 970–975 (2010).

83. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of Lipid-Modified Monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).

84. Markin, C. J. et al. Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. *Science* **373**, eabf8761 (2021).

85. Volpetti, F., Garcia-Cordero, J. & Maerkl, S. J. A microfluidic platform for high-throughput multiplexed protein quantitation. *PLOS ONE* **10**, e0117744 (2015).

86. Brower, K. et al. An open-source, programmable pneumatic setup for operation and automated control of single- and multi-layer microfluidic devices. *HardwareX* **3**, 117–134 (2018).

87. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).

88. Hastings, R. Code: Mutations to transcription factor MAX allosterically increase DNA selectivity by altering folding and binding pathways. *Zenodo* https://doi.org/10.5281/zenodo.14218355 (2024).

Acknowledgements

Author contributions

Conceptualization, R.H., A.K.A., and P.M.F.; Formal Analysis, R.H. and A.K.A. Investigation, R.H., A.K.A., and N.D.; R.H. and A.K.A. operated microfluidic devices to collect binding and dissociation data, A.K.A. and N.D. performed preliminary experiments. Resources, P.H.S.; P.H.S. designed blocked microfluidic devices used in this study. Writing: R.H. and P.M.F. Supervision & Funding Acquisition: P.M.F. All authors read and approved this manuscript.

Competing interests

P.M.F. is a co-founder of Velocity Bio and a member of the Evozyne scientific advisory board. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55672-2.

Correspondence and requests for materials should be addressed to Polly M. Fordyce.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.